



Σχολή Θετικών Επιστημών & Τεχνολογίας

Μεταπτυχιακή Εξειδίκευση στα Πληροφοριακά Συστήματα  
(MSc)

Διπλωματική Εργασία

Σχεδίαση δημιουργία & διαχείριση αποθήκης δεδομένων στο νέφος

Γεώργιος Θ. Τρίφτης

Επιβλέπουσα καθηγήτρια: Γεωργία Γκαράνη

Πάτρα, Μάιος 2025

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή («συγγραφέας/δημιουργός») που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης ο συγγραφέας/δημιουργός εκχωρεί στο ΕΑΠ, μη αποκλειστική άδεια χρήσης του δικαιώματος αναπαραγωγής, προσαρμογής, δημόσιου δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσής τους διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος και για όλο το χρόνο διάρκειας των δικαιωμάτων πνευματικής ιδιοκτησίας. Η ανοικτή πρόσβαση στο πλήρες κείμενο για μελέτη και ανάγνωση δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, αποθήκευση, πώληση, εμπορική χρήση, μετάδοση, διανομή, έκδοση, εκτέλεση, «μεταφόρτωση» (downloading), «ανάρτηση» (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού. Ο συγγραφέας/δημιουργός διατηρεί το σύνολο των ηθικών και περιουσιακών του δικαιωμάτων.



Σχεδίαση, δημιουργία και διαχείριση αποθήκης δεδομένων στο  
νέφος

Γεώργιος Θ. Τρίφτης

Επιτροπή Επίβλεψης Διπλωματικής Εργασίας

Επιβλέπουσα Καθηγήτρια:  
Γεωργία Γκαράνη  
Καθηγήτρια Πανεπιστημίου Θεσσαλίας

Συν-Επιβλέπων Καθηγητής:  
Σωτήριος Κωτσιαντής  
Επίκουρος καθηγητής  
Πανεπιστημίου Πατρών

Πάτρα, Μάιος 2025

*Θα ήθελα να ευχαριστήσω την επιβλέπουσα καθηγήτριά μου κα. Γκαράνη Γεωργία για την καθοδήγηση και την βοήθεια που μου πρόσφερε καθ' όλη τη διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας.*

*Ευχαριστώ, επίσης, θερμά την οικογένειά μου για την υπομονή και τη στήριξή της στο έργο αυτό.*

## Περίληψη

Η παρούσα εργασία εξετάζει τη σημασία και τις τεχνολογικές εξελίξεις στις αποθήκες δεδομένων, τόσο σε παραδοσιακό περιβάλλον όσο και στο υπολογιστικό νέφος. Μελετά τη διαχρονική ανάπτυξη της αποθήκευσης και διαχείρισης δεδομένων, παρουσιάζοντας τις βασικές αρχές των αποθηκών δεδομένων και τη σύγκρισή τους με τις λίμνες δεδομένων. Παράλληλα, αναλύεται το υπολογιστικό νέφος, τα διαφορετικά μοντέλα ανάπτυξης και υπηρεσιών του, καθώς και οι προκλήσεις και τα ζητήματα ασφάλειας που σχετίζονται με αυτό.

Αρχικά, η εργασία εστιάζει στην παρουσίαση των αποθηκών δεδομένων, αναλύοντας την αρχιτεκτονική τους, τις βασικές λειτουργίες, και τις τεχνικές επεξεργασίας και ανάλυσης δεδομένων. Εξετάζονται επίσης τα πλεονεκτήματα και οι προκλήσεις που αντιμετωπίζουν οι οργανισμοί κατά την υιοθέτηση αποθηκών δεδομένων. Στη συνέχεια, αναλύεται η μετάβαση των αποθηκών δεδομένων στο υπολογιστικό νέφος, η οποία προσφέρει επεκτασιμότητα, ευελιξία και αυξημένη αποδοτικότητα. Η εργασία συγκρίνει τις παραδοσιακές και τις νεφοϋπολογιστικές αποθήκες δεδομένων, εξετάζοντας την απόδοσή τους, την ασφάλεια και το κόστος συντήρησης.

Η ανάλυση περιλαμβάνει επίσης πρακτική υλοποίηση και αξιολόγηση τόσο μιας παραδοσιακής αποθήκης δεδομένων όσο και μιας αποθήκης δεδομένων βασισμένης στο υπολογιστικό νέφος, χρησιμοποιώντας την πλατφόρμα Apache Druid. Παρουσιάζονται τα βήματα σχεδιασμού, η διαδικασία υλοποίησης, καθώς και η αξιολόγηση της απόδοσης των δύο προσεγγίσεων. Η εργασία συγκρίνει τις δύο υλοποιήσεις βάσει της ταχύτητας επεξεργασίας δεδομένων, της ευκολίας διαχείρισης, των απαιτήσεων αποθήκευσης και της συνολικής αποδοτικότητας.

Ένα από τα κύρια ευρήματα της μελέτης είναι ότι οι αποθήκες δεδομένων στο υπολογιστικό νέφος προσφέρουν μεγαλύτερη ευελιξία και κλιμάκωση σε σύγκριση με τις παραδοσιακές λύσεις. Παρόλα αυτά, η μετάβαση στο νέφος απαιτεί ιδιαίτερη προσοχή σε ζητήματα ασφάλειας, δεδομένης της ευαισθησίας των δεδομένων και των πιθανών κινδύνων παραβίασης. Η εργασία καταλήγει στο συμπέρασμα ότι η επιλογή μεταξύ παραδοσιακής και νεφοϋπολογιστικής αποθήκης δεδομένων εξαρτάται από τις ανάγκες

του εκάστοτε οργανισμού, το κόστος υποδομής και τη στρατηγική διαχείρισης των δεδομένων.

Τέλος, η μελέτη προτείνει μελλοντικές κατευθύνσεις για την περαιτέρω έρευνα, όπως η ενσωμάτωση τεχνητής νοημοσύνης στις αποθήκες δεδομένων, η ανάπτυξη βελτιωμένων μεθόδων ασφαλείας και η διερεύνηση νέων τεχνολογιών που θα επιτρέψουν την πιο αποτελεσματική διαχείριση δεδομένων σε περιβάλλοντα νέφους.

### **Λέξεις – Κλειδιά**

Βάσεις Δεδομένων, Αποθήκες Δεδομένων, Υπολογιστικό Νέφος, Νεφοϋπολογιστικές Αποθήκες Δεδομένων

# Cloud data warehouse design, development and management.

George T. Triftis

## Abstract

This paper examines the significance and technological advancements in data warehouses, both in traditional environments and cloud computing. It explores the historical development of data storage and management, presenting the fundamental principles of data warehouses and their comparison with data lakes. At the same time, cloud computing is analyzed, including different deployment and service models, as well as the challenges and security concerns associated with it.

Initially, the study focuses on the presentation of data warehouses, analyzing their architecture, core functions, and data processing and analysis techniques. The advantages and challenges organizations face when adopting data warehouses are also examined. Subsequently, the transition of data warehouses to cloud computing is analyzed, offering scalability, flexibility, and increased efficiency. The paper compares traditional and cloud-based data warehouses, evaluating their performance, security, and maintenance costs.

The analysis also includes a practical implementation and evaluation of both a traditional data warehouse and a cloud-based data warehouse using the Apache Druid platform. The design steps, implementation process, and performance assessment of both approaches are presented. The study compares the two implementations based on data processing speed, ease of management, storage requirements, and overall efficiency.

One of the key findings of the study is that cloud-based data warehouses provide greater flexibility and scalability compared to traditional solutions. However, transitioning to the cloud requires careful consideration of security issues, given the sensitivity of data and the potential risks of breaches. The study concludes that the choice between a traditional and a cloud-based data warehouse depends on the specific needs of an organization, infrastructure costs, and data management strategies.

Finally, the study suggests future directions for further research, such as integrating artificial intelligence into data warehouses, developing enhanced security methods, and exploring new technologies that will enable more efficient data management in cloud environments.

### **Keywords**

DataBase, Data Warehouse, Cloud Computing, Cloud-Based Data Warehouse



## Περιεχόμενα

Περίληψη.....	v
Abstract .....	vii
Περιεχόμενα .....	ix
Κατάλογος Εικόνων / Σχημάτων .....	xi
Συντομογραφίες & Ακρωνύμια.....	xiii
1. Εισαγωγή.....	1
1.1 Σκοπός της εργασίας .....	1
1.2 Στόχοι.....	2
1.3 Μεθοδολογία.....	5
1.4 Διάρθρωση της εργασίας.....	5
2. Υπολογιστικό Νέφος (Cloud Computing) .....	9
2.1 Ορισμός.....	9
2.2 Ιστορική αναδρομή .....	10
2.3 Μοντέλα Ανάπτυξης Υπολογιστικού Νέφους .....	12
2.3.1 Δημόσιο Νέφος (Public Cloud).....	13
2.3.2 Ιδιωτικό Νέφος (Private Cloud).....	13
2.3.3 Κοινοτικό Νέφος (Community Cloud) .....	13
2.3.4 Υβριδικό Νέφος (Hybrid Cloud): Συνδυασμός δημόσιου και ιδιωτικού νέφους .....	14
2.4 Μοντέλα υπηρεσιών του Υπολογιστικού Νέφους .....	14
2.4.1 Υποδομή ως Υπηρεσία (IaaS).....	14
2.4.2 Πλατφόρμα ως Υπηρεσία (PaaS).....	15
2.4.3 Λογισμικό ως Υπηρεσία (SaaS).....	15
2.5 Οφέλη Υπολογιστικού Νέφους.....	16
2.6 Προκλήσεις .....	17
2.7 Ζητήματα ασφάλειας.....	19
2.7.1 Εμπιστευτικότητα.....	20
2.7.2 Ακεραιότητα.....	23
2.7.3 Διαθεσιμότητα.....	25
2.7.4 Λύσεις ασφάλειας .....	27
2.8 Ρύθμιση συμμόρφωσης ασφάλειας στο Cloud .....	30
2.9 Σύγχρονες τάσεις στο Cloud Computing .....	32
3. Παραδοσιακές Αποθήκες Δεδομένων .....	41
3.1 Ορισμός και χαρακτηριστικά των παραδοσιακών αποθηκών δεδομένων .....	41
3.2 Διαφορά μεταξύ Αποθηκών Δεδομένων και Λιμνών Δεδομένων .....	43
3.3 Αρχιτεκτονική Αποθήκης Δεδομένων .....	44
3.3.1 Βάση δεδομένων αποθήκης δεδομένων .....	48
3.3.2 Εργαλεία Εξαγωγής, Μετασχηματισμού και Φόρτωσης (ETL) .....	49
3.3.3 Εργαλεία Ερωτημάτων.....	49
3.4 Σχεδιασμός Αποθήκης Δεδομένων για Επιχειρησιακές Ανάγκες.....	50

3.5 Δημοφιλή Εργαλεία και Υπηρεσίες Αποθήκευσης Δεδομένων .....	53
3.6 Προκλήσεις Υλοποίησης Αποθήκης Δεδομένων.....	55
3.7 Αποθήκες Δεδομένων: Ευκαιρίες και Μελλοντικές Κατευθύνσεις.....	56
3.7.1 Νέες Μέθοδοι Αναλύσεων Δεδομένων.....	57
3.7.2 Διαδίκτυο των Πραγμάτων (IoT) .....	57
3.7.3 Μελλοντικές Δυνατότητες Αποθηκών Δεδομένων.....	58
4. Αποθήκες Δεδομένων στο υπολογιστικό νέφος.....	59
4.1 Ορισμός και σημασία της αποθήκης δεδομένων στο υπολογιστικό νέφος.....	59
4.2 Βασικά χαρακτηριστικά της αποθήκης δεδομένων στο υπολογιστικό νέφος.....	59
4.3 Σύγκριση παραδοσιακής και νεφοϋπολογιστικής αποθήκης δεδομένων.....	61
4.4 Αρχιτεκτονική αποθήκης δεδομένων στο υπολογιστικό νέφος .....	63
4.5 Οφέλη της αποθήκης δεδομένων στο υπολογιστικό νέφος .....	65
4.6 Προκλήσεις της αποθήκης δεδομένων στο υπολογιστικό νέφος .....	68
4.7 Λύσεις αποθήκευσης δεδομένων στο υπολογιστικό νέφος για επιχειρήσεις.....	70
4.7.1 Microsoft Azure Synapse Analytics.....	70
4.7.2 Amazon Redshift.....	71
4.7.3 Snowflake.....	72
4.7.4 Google BigQuery .....	73
5. Σχεδιασμός και υλοποίηση παραδοσιακής αποθήκης δεδομένων .....	75
5.1 Αρχιτεκτονική αποθήκης δεδομένων .....	75
5.2 Δημιουργία αποθήκης δεδομένων.....	77
5.3 Ερωτήματα προς εκτέλεση για αξιολόγηση.....	80
5.4 Screenshots Λειτουργικότητας.....	83
6. Σχεδιασμός και υλοποίηση αποθήκης δεδομένων στο υπολογιστικό νέφος. ....	89
6.1 Εισαγωγή.....	89
6.2 Δημιουργία αποθήκης δεδομένων στο Apache Druid .....	90
6.3 Υλοποίηση ερωτημάτων προς εκτέλεση και αξιολόγηση.....	100
7. Σύγκριση υλοποιήσεων .....	109
8. Συμπεράσματα .....	113
Βιβλιογραφία - Πηγές .....	117

## Κατάλογος Εικόνων / Σχημάτων

Εικόνα 1: Αρχιτεκτονική υπολογιστικού νέφους .....	10
Εικόνα 2: Data Warehouse vs Data Lakes .....	44
Εικόνα 3: Αρχιτεκτονική Αποθήκης Δεδομένων .....	46
Εικόνα 4: On promises Data Warehouse vs Cloud Data Warehouse .....	63
Εικόνα 5: Σχεσιακό σχήμα .....	75
Εικόνα 6: Πίνακες που έχουν δημιουργηθεί - MySQL Server .....	83
Εικόνα 7: Περιεχόμενα πίνακα fact_checks - MySQL Server.....	83
Εικόνα 8: Αποτελέσματα 1ου ερωτήματος - MySQL Server .....	84
Εικόνα 9: Αποτελέσματα 2ου ερωτήματος - MySQL Server.....	84
Εικόνα 10: Αποτελέσματα 3ου ερωτήματος - MySQL Server .....	85
Εικόνα 11: Αποτελέσματα 4ου ερωτήματος - MySQL Server .....	85
Εικόνα 12: Αποτελέσματα 5ου ερωτήματος - MySQL Server .....	86
Εικόνα 13: Αποτελέσματα 6ου ερωτήματος - MySQL Server .....	86
Εικόνα 14: Αποτελέσματα 7ου ερωτήματος - MySQL Server .....	87
Εικόνα 15: Αποτελέσματα 8ου ερωτήματος - MySQL Server .....	87
Εικόνα 16: Αποτελέσματα 9ου ερωτήματος - MySQL Server .....	88
Εικόνα 17: Αποτελέσματα 10ου ερωτήματος - MySQL Server .....	88
Εικόνα 18: Περιβάλλον Apache Druid .....	89
Εικόνα 19: Δημιουργία αποθήκης Apache Druid - Βήμα 1.....	92
Εικόνα 20: Δημιουργία αποθήκης Apache Druid - Βήμα 2.....	92
Εικόνα 21: Δημιουργία αποθήκης Apache Druid - Βήμα 3.....	93
Εικόνα 22: Δημιουργία αποθήκης Apache Druid - Βήμα 4.....	94
Εικόνα 23: Δημιουργία αποθήκης Apache Druid - Βήμα 5.....	95
Εικόνα 24: Δημιουργία αποθήκης Apache Druid - Βήμα 6.....	95
Εικόνα 25: Δημιουργία αποθήκης Apache Druid - Βήμα 7.....	96
Εικόνα 26: Δημιουργία αποθήκης Apache Druid - Βήμα 8.....	97
Εικόνα 27: Δημιουργία αποθήκης Apache Druid - Βήμα 9.....	98
Εικόνα 28: Δημιουργία αποθήκης Apache Druid - Βήμα 10.....	99
Εικόνα 29: Δημιουργία αποθήκης Apache Druid - Βήμα 11.....	100
Εικόνα 30: Αποτελέσματα 1ου ερωτήματος – Apache Druid.....	104
Εικόνα 31: Αποτελέσματα 2ου ερωτήματος – Apache Druid.....	104
Εικόνα 32: Αποτελέσματα 3ου ερωτήματος – Apache Druid.....	105
Εικόνα 33: Αποτελέσματα 4ου ερωτήματος – Apache Druid.....	105
Εικόνα 34: Αποτελέσματα 5ου ερωτήματος – Apache Druid.....	106
Εικόνα 35: Αποτελέσματα 6ου ερωτήματος – Apache Druid.....	106
Εικόνα 36: Αποτελέσματα 7ου ερωτήματος – Apache Druid.....	107
Εικόνα 37: Αποτελέσματα 8ου ερωτήματος – Apache Druid.....	107
Εικόνα 38: Αποτελέσματα 9ου ερωτήματος – Apache Druid.....	108
Εικόνα 39: Αποτελέσματα 10ου ερωτήματος – Apache Druid.....	108

## **Κατάλογος Πινάκων**

Πίνακας 1: Συγκριτικός πίνακας χρόνων εκτέλεσης .....	109
---	-----

## Συντομογραφίες & Ακρωνύμια

ACID	Atomicity, Consistency, Isolation, Durability
API	Application Programming Interface
ARDS	Amazon Relational Database Service
AWS	Amazon Web Services
AWS S3	Amazon Simple Storage Service
BI	Business Intelligence
CCPA	California Consumer Privacy Act
CIA	Confidentiality, Integrity, Availability
CP-ABE	Cipher Policy - Attributed Based Encryption
CRM	Customer Relationship Management
CSP	Cloud Service Provider
CSV	Comma Separated Values
DDoS	Distributed Denial of Service
DIL	Data Ingestion Layer
DoS	Denial of Service
DS	Data Sources
DW	Data Warehouse
ECPA	Electronic Communications Privacy Act
EHR	Electronic Health Record
ELT	Extract, Load, Transform
ERP	Enterprise Resource Planning
ETL	Extract, Transform, Load
FCRA	Fair Credit Reporting Act
FPGA	Field Programmable Gate Array
GCP	Google Cloud Platform
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability And Accountability
HPC	High Performance Computing

IaaS	Infrastructure as a Service
IoT	Internet of Things
IP	Internet Protocol
IPSec	Internet Protocol Security
IT	Information Technology
JDBC	Java DataBase Connectivity
JVM	Java Virtual Machine
LDAP	Lightweight Directory Access Protocol
LDSS	Lightweight Secure Data Sharing Scheme
MDDBs	Multi Dimensional DataBase
MFA	Multi Function Authentication
ODBC	Open DataBase Connectivity
OLAP	On Line Analytical Processing
OLTP	On Line Transaction Processing Systems
OTP	One Time Password
PaaS	Platform as a Service
RDBMS	Relational Database Management Systems
SaaS	Software as a Service
SLA	Service Level Agreement
SOAP	Simple Object Access Protocol
SQL	Structured Query Language
SSL	Secure Sockets Layer
SSO	Single Sign-On
TLS	Transport Layer Security
VLAN	Virtual Local Area Network
VM	Virtual Machine
VMI	Virtual Machine Image
VMM	Virtual Machine Manager
WAN	Wide Area Network
WSDL	Web Services Description Language

## 1. Εισαγωγή

### 1.1 Σκοπός της εργασίας

Η παρούσα εργασία αποσκοπεί στη μελέτη και σύγκριση των παραδοσιακών αποθηκών δεδομένων με τις νεφοϋπολογιστικές (cloud-based) αποθήκες δεδομένων, προκειμένου να αναδειχθούν οι βασικές διαφορές, τα πλεονεκτήματα και οι προκλήσεις που αντιμετωπίζει κάθε προσέγγιση. Η αποθήκευση και διαχείριση δεδομένων αποτελούν κρίσιμες διαδικασίες για τις επιχειρήσεις, καθώς οι οργανισμοί βασίζονται σε αυτές για τη λήψη στρατηγικών αποφάσεων και τη βελτιστοποίηση των λειτουργιών τους.

Η παραδοσιακή προσέγγιση περιλαμβάνει φυσικούς διακομιστές και τοπικές υποδομές, που αν και προσφέρουν έλεγχο και ασφάλεια, συνοδεύονται από υψηλό κόστος συντήρησης και περιορισμένη ευελιξία. Από την άλλη, η αποθήκευση στο cloud παρέχει δυναμικότητα, επεκτασιμότητα και προσβασιμότητα από οπουδήποτε, καθιστώντας την μια ελκυστική επιλογή για τις σύγχρονες επιχειρήσεις.

Η εργασία εστιάζει στη σχεδίαση, ανάπτυξη και ανάλυση των δύο προσεγγίσεων, λαμβάνοντας υπόψη τις ανάγκες αποθήκευσης δεδομένων διαφορετικών κλάδων της βιομηχανίας. Θα εξεταστούν οι τεχνολογικές υποδομές που υποστηρίζουν κάθε μέθοδο, καθώς και οι παράγοντες που επηρεάζουν την απόδοσή τους, όπως η διαθεσιμότητα, η ασφάλεια, το κόστος και η ευκολία διαχείρισης.

Επιπλέον, θα διερευνηθεί ο ρόλος της νεφοϋπολογιστικής αποθήκευσης δεδομένων σε περιβάλλοντα μεγάλης κλίμακας, όπου η ανάγκη για ευέλικτες και οικονομικά αποδοτικές λύσεις γίνεται όλο και πιο έντονη. Η ανάλυση θα περιλαμβάνει και τη μελέτη των τεχνολογικών εξελίξεων που έχουν βελτιώσει την αποδοτικότητα και τη λειτουργικότητα των cloud-based αποθηκών δεδομένων, όπως η χρήση τεχνητής νοημοσύνης, αυτοματοποιημένων διαδικασιών και προηγμένων μεθόδων κρυπτογράφησης.

Μέσα από αυτή την εργασία, επιδιώκεται η εις βάθος κατανόηση των χαρακτηριστικών που καθιστούν τη νεφοϋπολογιστική αποθήκευση δεδομένων

ελκυστική, καθώς και η ανάλυση των δομικών διαφορών της σε σχέση με τις παραδοσιακές μεθόδους αποθήκευσης. Η εξέταση των πλεονεκτημάτων και των προκλήσεων που παρουσιάζουν οι δύο προσεγγίσεις θα συμβάλει στην αξιολόγηση της καταλληλότητάς τους για διαφορετικές επιχειρηματικές ανάγκες. Τέλος, η εργασία στοχεύει να αναδείξει τις τάσεις και τις προοπτικές του μέλλοντος στον τομέα της αποθήκευσης δεδομένων, παρέχοντας πολύτιμες πληροφορίες για τις επιχειρήσεις που επιδιώκουν να υιοθετήσουν πιο αποδοτικές και ευέλικτες λύσεις.

## 1.2 Στόχοι

Η παρούσα διπλωματική εργασία έχει ως κύριο σκοπό τη σε βάθος ανάλυση και σύγκριση των παραδοσιακών αποθηκών δεδομένων με τις νεφοϋπολογιστικές (cloud-based) αποθήκες δεδομένων. Για την επίτευξη αυτού του στόχου, η εργασία επικεντρώνεται σε τρεις βασικούς άξονες, οι οποίοι αναλύονται ως εξής:

### 1. Κατανόηση των βασικών εννοιών αποθήκευσης δεδομένων

Η πρώτη φάση της εργασίας εστιάζει στην ανάλυση και κατανόηση των θεμελιωδών εννοιών που αφορούν την αποθήκευση δεδομένων. Σε αυτό το πλαίσιο, θα εξεταστούν οι βασικοί τύποι αποθήκευσης, όπως:

- **Αποθήκες Δεδομένων (Data Warehouses):** Πρόκειται για παραδοσιακές λύσεις που οργανώνουν και επεξεργάζονται δεδομένα σε δομημένη μορφή, παρέχοντας προηγμένες δυνατότητες ανάλυσης και επιχειρησιακής νοημοσύνης BI (Business Intelligence).
- **Τεχνολογίες αποθήκευσης στο υπολογιστικό νέφος:** Η νεφοϋπολογιστική αποθήκευση δεδομένων χρησιμοποιεί πλατφόρμες cloud για την κλιμακούμενη διαχείριση δεδομένων, εξασφαλίζοντας υψηλή διαθεσιμότητα, δυναμική επεκτασιμότητα και μειωμένο κόστος συντήρησης.

Η ανάλυση αυτή είναι απαραίτητη για τη δημιουργία ενός θεωρητικού υπόβαθρου που θα επιτρέψει τη σε βάθος κατανόηση των πλεονεκτημάτων και των



περιορισμών κάθε προσέγγισης, καθώς και τη διερεύνηση των κατάλληλων σεναρίων χρήσης για κάθε μία από αυτές.

## **2. Υλοποίηση των δύο προσεγγίσεων**

Για την καλύτερη κατανόηση των διαφορών μεταξύ των παραδοσιακών και των cloud-based αποθηκών δεδομένων, η εργασία θα περιλαμβάνει την πρακτική υλοποίηση των δύο προσεγγίσεων.

Πιο συγκεκριμένα, θα πραγματοποιηθεί:

- **Δημιουργία ενός παραδείγματος παραδοσιακής αποθήκης δεδομένων:**  
Αυτή η αποθήκη δεδομένων θα βασίζεται σε κλασικές αρχιτεκτονικές βάσεων δεδομένων και θα χρησιμοποιεί τεχνολογίες όπως ο MySQL Server. Η υλοποίηση θα επικεντρωθεί στη διαχείριση μεγάλων όγκων δεδομένων, στη δημιουργία αποδοτικών ερωτημάτων και στην ανάλυση της απόδοσης σε πραγματικά σενάρια χρήσης.
- **Δημιουργία μιας cloud-based αποθήκης δεδομένων:** Η αποθήκη αυτή θα χρησιμοποιήσει τεχνολογίες όπως το Apache Druid, προκειμένου να αξιολογηθεί η επεκτασιμότητα, η απόδοση και η αποδοτικότητα της αποθήκευσης δεδομένων στο cloud. Ιδιαίτερη έμφαση θα δοθεί στη διαχείριση της υπολογιστικής ισχύος, στο κόστος λειτουργίας και στην ασφάλεια των δεδομένων.
- **Σύγκριση των δύο υλοποιήσεων:** Η αξιολόγηση θα περιλαμβάνει τη σύγκριση μεταξύ των δύο συστημάτων αποθήκευσης, βασισμένη σε διαφορετικά κριτήρια, όπως η ταχύτητα ερωτημάτων, η δυνατότητα επέκτασης, η διαθεσιμότητα και η απόδοση υπό διαφορετικά φορτία εργασίας.

## **3. Σύγκριση των προσεγγίσεων**

Το τρίτο και πιο σημαντικό στάδιο της εργασίας περιλαμβάνει τη σύγκριση των δύο μεθόδων αποθήκευσης δεδομένων, με στόχο την εξαγωγή συμπερασμάτων σχετικά με τις προκλήσεις και τα οφέλη που προσφέρει κάθε λύση. Η αξιολόγηση αυτή θα βασιστεί σε συγκεκριμένα κριτήρια, όπως:

- **Απόδοση:** Πόσο γρήγορα μπορεί κάθε προσέγγιση να εκτελέσει αναζητήσεις και να επεξεργαστεί δεδομένα; Ποιες είναι οι επιπτώσεις της μεγάλης κλίμακας δεδομένων στην απόδοση των δύο συστημάτων;
- **Επεκτασιμότητα:** Πώς κάθε προσέγγιση αντιμετωπίζει την αυξανόμενη ζήτηση για αποθήκευση δεδομένων; Μπορούν οι παραδοσιακές αποθήκες δεδομένων να προσαρμοστούν στην αυξανόμενη ανάγκη για περισσότερους πόρους, σε σύγκριση με τις cloud-based λύσεις;
- **Ευελιξία και κόστος:** Ενώ οι παραδοσιακές λύσεις απαιτούν σημαντική επένδυση σε υλικό και συντήρηση, οι cloud-based αποθήκες δεδομένων παρέχουν μοντέλα χρέωσης pay-as-you-go, μειώνοντας το αρχικό κόστος εγκατάστασης. Ποιες είναι οι μακροπρόθεσμες οικονομικές επιπτώσεις κάθε επιλογής;
- **Ασφάλεια και κανονιστική συμμόρφωση:** Ποιες είναι οι προκλήσεις που αντιμετωπίζει κάθε προσέγγιση όσον αφορά την προστασία δεδομένων, την ιδιωτικότητα και την συμμόρφωση με κανονισμούς όπως το GDPR (General Data Protection Regulation) καθώς και με το HIPAA (Health Insurance Portability and Accountability);

Μέσα από αυτή τη συγκριτική ανάλυση, η εργασία θα προσπαθήσει να δώσει σαφείς απαντήσεις σχετικά με την καταλληλότητα της κάθε προσέγγισης για διαφορετικά σενάρια επιχειρησιακής χρήσης. Επιπλέον, θα προσφέρει χρήσιμες πληροφορίες για εταιρείες και οργανισμούς που επιδιώκουν να μεταβούν από παραδοσιακές λύσεις αποθήκευσης σε σύγχρονες cloud-based αρχιτεκτονικές.

Η διπλωματική εργασία θα παρέχει μια ολοκληρωμένη εικόνα της εξέλιξης των τεχνολογιών αποθήκευσης δεδομένων, παρουσιάζοντας τις βασικές αρχές, τις πρακτικές εφαρμογές και τις επιπτώσεις των διαφορετικών προσεγγίσεων. Μέσα από τη θεωρητική και πρακτική ανάλυση, θα αναδειχθούν οι βέλτιστες πρακτικές για την επιλογή της κατάλληλης αποθηκευτικής λύσης, ανάλογα με τις ανάγκες και τις απαιτήσεις κάθε οργανισμού.

### **1.3 Μεθοδολογία**

Η μεθοδολογία που θα ακολουθηθεί για την επίτευξη των παραπάνω στόχων περιλαμβάνει τα εξής βήματα:

1. Επισκόπηση Βιβλιογραφίας: Μελέτη και ανάλυση της βιβλιογραφίας σχετικά με τις αποθήκες δεδομένων και τις cloud-based λύσεις αποθήκευσης, προκειμένου να κατανοηθούν οι βασικές έννοιες, τα μοντέλα ανάπτυξης, τα πλεονεκτήματα και οι προκλήσεις των δύο τύπων αποθήκευσης δεδομένων.
2. Σχεδίαση Παραδοσιακής Αποθήκης Δεδομένων: Δημιουργία ενός παραδείγματος παραδοσιακής αποθήκης δεδομένων με χρήση σχετικών εργαλείων και τεχνολογιών (MySQL Server).
3. Σχεδίαση και Δημιουργία Cloud-Based Αποθήκης Δεδομένων: Υλοποίηση cloud-based αποθήκης δεδομένων με τη χρήση πλατφορμών νεφοϋπολογιστικού τύπου, όπως το Apache Druid.
4. Σύγκριση και Αξιολόγηση: Στη συνέχεια, η απόδοση των δύο προσεγγίσεων θα συγκριθεί με βάση τα κριτήρια απόδοσης, κόστους, ευελιξίας, ασφάλειας και κλίμακας, για να κατανοηθούν τα οφέλη και οι περιορισμοί κάθε προσέγγισης.

Η συνολική μεθοδολογία επιδιώκει να συνδυάσει τη θεωρητική ανάλυση με την πρακτική υλοποίηση και αξιολόγηση, προσφέροντας μια ολοκληρωμένη εικόνα για τη σύγκριση των δύο μοντέλων αποθήκευσης δεδομένων.

### **1.4 Διάρθρωση της εργασίας**

Η παρούσα εργασία επικεντρώνεται στη μελέτη των αποθηκών δεδομένων και της εφαρμογής τους στο υπολογιστικό νέφος. Η δομή της εργασίας ακολουθεί μια μεθοδική προσέγγιση, ξεκινώντας από τις βασικές έννοιες του υπολογιστικού

νέφους και καταλήγοντας στην ανάλυση και τη σύγκριση παραδοσιακών και νεφοϋπολογιστικών αποθηκών δεδομένων. Η ανάλυση ακολουθεί μια σταδιακή πορεία, αρχικά εστιάζοντας στο θεωρητικό υπόβαθρο και στη συνέχεια εξετάζοντας πρακτικές υλοποιήσεις και σύγκριση τεχνολογιών. Η ερευνητική μεθοδολογία στηρίζεται τόσο στη βιβλιογραφική επισκόπηση όσο και στην πρακτική εφαρμογή συγκεκριμένων τεχνολογιών, με στόχο τη βαθύτερη κατανόηση του θέματος και τη διαμόρφωση ασφαλών συμπερασμάτων.

Το πρώτο κεφάλαιο της εργασίας αποτελεί την εισαγωγή και περιλαμβάνει την παρουσίαση του σκοπού και των στόχων της μελέτης, καθώς και την αναλυτική περιγραφή της μεθοδολογίας που ακολουθήθηκε. Στοχεύει στη δημιουργία μιας στέρεης θεωρητικής βάσης, η οποία θα επιτρέψει την κατανόηση των βασικών αρχών και της σημασίας των αποθηκών δεδομένων στο υπολογιστικό νέφος. Παράλληλα, γίνεται επισκόπηση των ερευνητικών ερωτημάτων που θα απαντηθούν και ορίζονται τα βασικά κριτήρια ανάλυσης και αξιολόγησης των τεχνολογιών.

Το δεύτερο κεφάλαιο ασχολείται με το υπολογιστικό νέφος, έναν από τους σημαντικότερους τομείς της σύγχρονης τεχνολογίας. Αρχικά, δίνεται ένας σαφής ορισμός του όρου και στη συνέχεια παρουσιάζεται η ιστορική του εξέλιξη. Εξετάζονται οι διάφορες προσεγγίσεις ανάπτυξης υπολογιστικού νέφους, όπως τα ιδιωτικά, δημόσια, υβριδικά και κοινοτικά νέφη, καθώς και οι τρεις κύριες κατηγορίες υπηρεσιών που προσφέρει το νέφος: IaaS (Infrastructure as a Service), PaaS (Platform as a Service) και SaaS (Software as a Service).

Ιδιαίτερη έμφαση δίνεται στα οφέλη του υπολογιστικού νέφους, όπως η ευελιξία, η αποδοτικότητα και η μείωση κόστους, ενώ ταυτόχρονα αναλύονται οι προκλήσεις και τα ζητήματα ασφάλειας που το συνοδεύουν. Εξετάζονται θέματα εμπιστευτικότητας, ακεραιότητας και διαθεσιμότητας, καθώς και οι στρατηγικές αντιμετώπισής τους. Τέλος, παρουσιάζονται οι σύγχρονες τάσεις στον τομέα του υπολογιστικού νέφους, που περιλαμβάνουν την ενσωμάτωση της τεχνητής νοημοσύνης, την υιοθέτηση τεχνολογιών blockchain και τη χρήση serverless computing.

Το τρίτο κεφάλαιο εστιάζει στις αποθήκες δεδομένων, προσφέροντας μια εις βάθος ανάλυση του αντικειμένου. Ξεκινά με τον ορισμό και την παρουσίαση των διαφορών μεταξύ αποθηκών δεδομένων και λιμνών δεδομένων, δύο εννοιών που συχνά συγχέονται αλλά έχουν διακριτούς ρόλους στη διαχείριση μεγάλων όγκων δεδομένων. Ακολουθεί η ανάλυση της αρχιτεκτονικής των αποθηκών δεδομένων, με έμφαση στις βασικές συνιστώσες, όπως η διαδικασία ETL (Extract, Transform, Load) και η διαχείριση των μεταδεδομένων. Επιπλέον, εξετάζονται δημοφιλή εργαλεία και υπηρεσίες αποθήκευσης δεδομένων, με αναφορές σε τεχνολογίες όπως το Amazon Redshift, το Google BigQuery και το Snowflake. Το κεφάλαιο ολοκληρώνεται με τη διερεύνηση των προκλήσεων υλοποίησης αποθηκών δεδομένων και τις μελλοντικές κατευθύνσεις της έρευνας στον τομέα.

Το τέταρτο κεφάλαιο επικεντρώνεται στις αποθήκες δεδομένων στο υπολογιστικό νέφος, αναλύοντας τη σημασία και τα χαρακτηριστικά τους. Παρουσιάζεται μια σύγκριση μεταξύ παραδοσιακών και νεφοϋπολογιστικών αποθηκών δεδομένων, δίνοντας έμφαση στα πλεονεκτήματα που προσφέρει το νέφος, όπως η επεκτασιμότητα και η υψηλή διαθεσιμότητα. Εξετάζεται η αρχιτεκτονική των αποθηκών δεδομένων στο νέφος και παρουσιάζονται οι κύριες προκλήσεις, όπως η ασφάλεια, η διαχείριση κόστους και η ανάγκη για απόδοση σε πραγματικό χρόνο. Επιπλέον, διερευνώνται λύσεις αποθήκευσης δεδομένων που χρησιμοποιούνται από επιχειρήσεις, με αναφορές σε πλατφόρμες cloud, όπως το Microsoft Azure Synapse Analytics και το Snowflake.

Στο πέμπτο κεφάλαιο περιγράφεται η διαδικασία σχεδιασμού και υλοποίησης μιας παραδοσιακής αποθήκης δεδομένων. Παρουσιάζεται η αρχιτεκτονική της βάσης δεδομένων, η δημιουργία των βασικών δομών αποθήκευσης, καθώς και η υλοποίηση ερωτημάτων για την αξιολόγηση της αποδοτικότητας του συστήματος. Η ενότητα αυτή συνοδεύεται από στιγμιότυπα οθόνης (screenshots) που αποτυπώνουν τη λειτουργικότητα της αποθήκης δεδομένων.

Το έκτο κεφάλαιο εστιάζει στη σχεδίαση και υλοποίηση μιας αποθήκης δεδομένων στο υπολογιστικό νέφος, χρησιμοποιώντας την πλατφόρμα Apache Druid. Αναλύεται η διαδικασία δημιουργίας της αποθήκης δεδομένων, καθώς και η

υλοποίηση συγκεκριμένων ερωτημάτων που επιτρέπουν την αξιολόγηση της απόδοσης και της ευχρηστίας της λύσης. Η σύγκριση με την παραδοσιακή αποθήκη δεδομένων δίνει πολύτιμες πληροφορίες για τα πλεονεκτήματα και τα μειονεκτήματα των δύο προσεγγίσεων.

Το έβδομο κεφάλαιο παρουσιάζει μια συγκριτική ανάλυση μεταξύ των δύο υλοποιήσεων, παραδοσιακής και νεφοϋπολογιστικής αποθήκης δεδομένων. Η σύγκριση βασίζεται σε κριτήρια όπως η ταχύτητα απόκρισης, η ευελιξία, η ασφάλεια και η συνολική απόδοση των συστημάτων. Μέσα από αυτή τη διαδικασία, εξάγονται κρίσιμα συμπεράσματα σχετικά με τη βέλτιστη επιλογή αποθήκευσης δεδομένων ανάλογα με τις ανάγκες κάθε επιχείρησης.

Το όγδοο και τελευταίο κεφάλαιο περιλαμβάνει τα συμπεράσματα της μελέτης, συνοψίζοντας τα κύρια ευρήματα της έρευνας και προτείνοντας πιθανές κατευθύνσεις για μελλοντική έρευνα στον τομέα των αποθηκών δεδομένων στο υπολογιστικό νέφος. Τέλος, η εργασία ολοκληρώνεται με τη βιβλιογραφία, η οποία περιλαμβάνει όλες τις πηγές που χρησιμοποιήθηκαν κατά τη διάρκεια της έρευνας.

## 2. Υπολογιστικό Νέφος (Cloud Computing)

### 2.1 Ορισμός

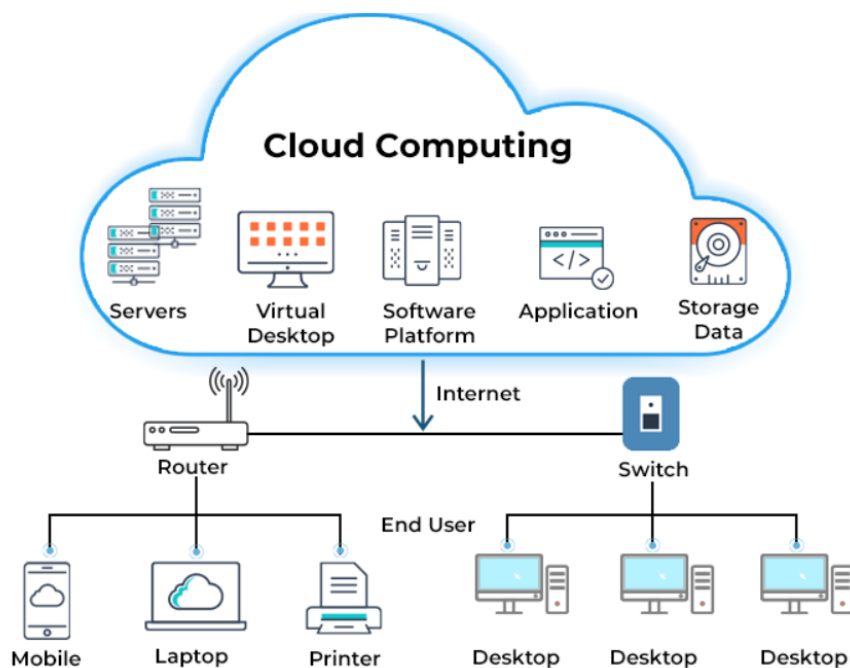
Η υπολογιστική νέφους [1] αναφέρεται στη χρήση φιλοξενούμενων υπηρεσιών, όπως αποθήκευση δεδομένων, διακομιστές, βάσεις δεδομένων, δικτύωση και λογισμικό, μέσω του διαδικτύου. Τα δεδομένα αποθηκεύονται σε φυσικούς διακομιστές που συντηρούνται από έναν πάροχο υπηρεσιών νέφους. Οι υπολογιστικοί πόροι, ιδιαίτερα η αποθήκευση δεδομένων και η υπολογιστική ισχύς, είναι διαθέσιμοι κατ' απαίτηση, χωρίς την άμεση διαχείριση από τον χρήστη.

Αντί να αποθηκεύει αρχεία σε μια συσκευή αποθήκευσης ή σκληρό δίσκο, ο χρήστης μπορεί να τα αποθηκεύει στο νέφος, διευκολύνοντας έτσι την πρόσβαση στα αρχεία από οπουδήποτε, αρκεί να υπάρχει πρόσβαση στο διαδίκτυο. Οι υπηρεσίες που φιλοξενούνται στο νέφος διακρίνονται ευρέως σε Υποδομή ως Υπηρεσία (IaaS), Πλατφόρμα ως Υπηρεσία (PaaS) και Λογισμικό ως Υπηρεσία (SaaS). Ανάλογα με το μοντέλο ανάπτυξης, το νέφος μπορεί επίσης να ταξινομηθεί ως δημόσιο, ιδιωτικό ή υβριδικό νέφος.

Η αρχιτεκτονική του νέφους χωρίζεται σε δύο επίπεδα: το front - end και το back-end:

- **Front – end επίπεδο:** Αυτό είναι το επίπεδο με το οποίο αλληλεπιδρούν οι χρήστες και επιτρέπει την πρόσβαση στα δεδομένα που είναι αποθηκευμένα στο νέφος μέσω λογισμικού υπολογιστικής νέφους.
- **Back – end επίπεδο:** Το επίπεδο αυτό αποτελείται από το λογισμικό και το υλικό, δηλαδή υπολογιστές, διακομιστές, κεντρικούς διακομιστές και βάσεις δεδομένων. Αποτελεί το κύριο συστατικό του νέφους και είναι υπεύθυνο για την ασφαλή αποθήκευση πληροφοριών. Για την ομαλή συνδεσιμότητα μεταξύ των συσκευών που συνδέονται μέσω του νέφους, οι κεντρικοί διακομιστές χρησιμοποιούν ένα λογισμικό που ονομάζεται

"middleware," το οποίο λειτουργεί ως γέφυρα μεταξύ των βάσεων δεδομένων και των εφαρμογών.



Εικόνα 1: Αρχιτεκτονική υπολογιστικού νέφους

[https://www.mdpi.com/BDCC/BDCC-06-00132/article\\_deploy/html/images/BDCC-06-00132-g002-550.jpg](https://www.mdpi.com/BDCC/BDCC-06-00132/article_deploy/html/images/BDCC-06-00132-g002-550.jpg)

## 2.2 Ιστορική αναδρομή

Το 1963, η DARPA (Defense Advanced Research Projects Agency) [3],[5] προσέφερε στο MIT χρηματοδότηση ύψους 2 εκατομμυρίων δολαρίων για το Project MAC. Η χρηματοδότηση απαιτούσε την ανάπτυξη τεχνολογίας που θα επέτρεπε τη χρήση ενός υπολογιστή από δύο ή περισσότερα άτομα ταυτόχρονα. Ένας από τους πρώιμους και μεγάλους υπολογιστές που χρησιμοποιούσε μαγνητική ταινία ως μνήμη αποτέλεσε τον πρόδρομο αυτού που σήμερα γνωρίζουμε ως υπολογιστική νέφος. Αυτή η τεχνολογία αποτέλεσε ένα πρωτόγονο νέφος με τη χρήση δύο ή τριών ατόμων ταυτόχρονα και περιγράφηκε ως «εικονικοποίηση», αν και η σημασία του όρου εξελίχθηκε αργότερα.



Το 1969, ο J.C.R. Licklider συνέβαλε στην ανάπτυξη του ARPANET, μιας πολύ πρώιμης εκδοχής του Διαδικτύου. Ο Licklider, ψυχολόγος και επιστήμονας υπολογιστών, προώθησε το όραμα του "Διαγαλαξιακού Δικτύου Υπολογιστών", όπου όλοι θα ήταν διασυνδεδεμένοι και θα είχαν πρόσβαση σε πληροφορίες από παντού, κάτι που αποτελεί βασική προϋπόθεση για την πρόσβαση στο νέφος.

Κατά τη δεκαετία του 1970, η σημασία της εικονικοποίησης άρχισε να αλλάζει, περιγράφοντας πλέον τη δημιουργία εικονικών μηχανών που λειτουργούν ως πλήρως αυτόνομοι υπολογιστές. Αυτή η ιδέα εξελίχθηκε με το Διαδίκτυο, καθώς εταιρείες άρχισαν να προσφέρουν εικονικά ιδιωτικά δίκτυα ως υπηρεσία. Στη δεκαετία του 1990, η χρήση εικονικών υπολογιστών έγινε δημοφιλής, οδηγώντας στην ανάπτυξη της σύγχρονης υποδομής της υπολογιστικής νέφους.

Στα αρχικά της στάδια, η υπολογιστική νέφος περιέγραφε τον «κενό χώρο» μεταξύ του τελικού χρήστη και του παρόχου. Το 1997, ο καθηγητής Ramnath Chellapa από το Πανεπιστήμιο Emory όρισε την υπολογιστική νέφος ως το νέο «υπολογιστικό παράδειγμα, όπου τα όρια της υπολογιστικής θα καθορίζονται από οικονομικούς παράγοντες παρά από τεχνικούς περιορισμούς».

Το 1999, η Salesforce χρησιμοποίησε την υπολογιστική νέφος για να προσφέρει λογισμικό μέσω Διαδικτύου στους τελικούς χρήστες, καθιερώνοντας έτσι την ιδέα της παροχής λογισμικού ως υπηρεσία.

Το 2002, η Amazon εισήγαγε τις διαδικτυακές της υπηρεσίες, προσφέροντας ένα μοντέλο υποδομής που επέτρεψε την αποδοτικότερη χρήση της υπολογιστικής της ικανότητας. Το 2006, η Amazon λάνσαρε τις Υπηρεσίες Ιστού της (AWS), που περιλάμβαναν την Elastic Compute Cloud (EC2), επιτρέποντας στους χρήστες να ενοικιάζουν εικονικούς υπολογιστές. Την ίδια χρονιά, η Google λάνσαρε τις υπηρεσίες Google Docs, αρχικά βασισμένες σε προϊόντα όπως τα Google Spreadsheets και Writely, επιτρέποντας στους χρήστες να αποθηκεύουν και να επεξεργάζονται έγγραφα διαδικτυακά.

Το 2007, η IBM και η Google συνεργάστηκαν με διάφορα πανεπιστήμια για την ανάπτυξη μιας υποδομής για ερευνητικά έργα που απαιτούσαν γρήγορους επεξεργαστές και τεράστια σύνολα δεδομένων. Την ίδια χρονιά, το Netflix

λάνσαρε την υπηρεσία συνεχούς ροής του μέσω νέφους, ενισχύοντας την πρακτική του "binge-watching." Το 2008, η NASA παρουσίασε το OpenNebula, το πρώτο λογισμικό ανοικτού κώδικα για την ανάπτυξη ιδιωτικών και υβριδικών νεφών. Το 2010, με την άνοδο των ανησυχιών για την ασφάλεια, εταιρείες όπως AWS, Microsoft και OpenStack ανέπτυξαν ιδιωτικά νέφη.

Το 2011, η έννοια του υβριδικού νέφους καθιερώθηκε, επιτρέποντας τη διαλειτουργικότητα μεταξύ ιδιωτικών και δημόσιων νεφών, ενώ το ίδιο έτος η Apple λάνσαρε το iCloud για προσωπική αποθήκευση.

Το 2012, η Oracle εισήγαγε την Oracle Cloud, παρέχοντας τις βασικές υπηρεσίες IaaS, PaaS και SaaS. Το 2013, η φιλοσοφία της χρήσης πολλαπλών νεφών (multi-cloud) έγινε δημοφιλής, επιτρέποντας στις επιχειρήσεις να επιλέγουν διαφορετικά νέφη για τις υπηρεσίες τους. Μέχρι το 2014, η υπολογιστική νέφους είχε καθιερώσει τα βασικά χαρακτηριστικά της, και η ασφάλεια είχε γίνει μία από τις κύριες ανησυχίες των πελατών.

Το 2016, η υπολογιστική νέφους μετασχηματίστηκε ώστε να επικεντρώνεται στις ανάγκες των προγραμματιστών εφαρμογών, καθώς οι πάροχοι νέφους ανέπτυξαν εργαλεία που υποστηρίζουν πλήρως τις ανάγκες των προγραμματιστών. Η υπολογιστική νέφους έχει πλέον ωριμάσει, προσφέροντας ολοκληρωμένες λύσεις τόσο για επιχειρήσεις όσο και για προσωπική χρήση, ενώ η ασφάλεια παραμένει προτεραιότητα, με υπηρεσίες που παρέχουν προστασία συγκρίσιμη με τα παραδοσιακά συστήματα IT.

## 2.3 Μοντέλα Ανάπτυξης Υπολογιστικού Νέφους

Τα μοντέλα ανάπτυξης [4] χαρακτηρίζουν το είδος της προσβασιμότητας στο περιβάλλον του νέφους, δηλαδή, πώς βρίσκεται το νέφος. Το νέφος χωρίζεται την προσβασιμότητα σε τέσσερις τύπους: Ιδιωτικό, Δημόσιο, Υβριδικό και Κοινοτικό.

### 2.3.1 Δημόσιο Νέφος (Public Cloud)

Αυτό το νέφος [19] είναι προσβάσιμο σε όλους τους εξωτερικούς χρήστες μέσω του διαδικτύου, οι οποίοι μπορούν να εγγραφούν και να χρησιμοποιήσουν τους πόρους του νέφους με μοντέλο πληρωμής ανά χρήση. Αυτό το νέφος δεν είναι τόσο ασφαλές όσο το ιδιωτικό νέφος, καθώς είναι ανοιχτό για όλους τους χρήστες του διαδικτύου. Είναι σχετικά λιγότερο προσαρμόσιμο σε σχέση με το ιδιωτικό νέφος. Η υποδομή του νέφους ανήκει και διαχειρίζεται από έναν μεγάλο Πάροχο Υπηρεσιών Νέφους CSP (Cloud Service Provider). Ο πάροχος νέφους είναι υπεύθυνος για τη δημιουργία και τη συνεχή υποστήριξη του δημόσιου νέφους και των πόρων του. Το δημόσιο νέφος ονομάζεται επίσης εξωτερικό νέφος, όπου οι πόροι παρέχονται δυναμικά σε αυτοεξυπηρέτηση μέσω του διαδικτύου.

### 2.3.2 Ιδιωτικό Νέφος (Private Cloud)

Αυτό το νέφος [6] δημιουργείται ειδικά για έναν οργανισμό μέσα στο δικό του κέντρο δεδομένων.

Οι οργανισμοί διαχειρίζονται όλους τους πόρους του νέφους, οι οποίοι ανήκουν σε αυτούς. Το ιδιωτικό νέφος προσφέρει μεγαλύτερη ασφάλεια σε σύγκριση με το δημόσιο ή το υβριδικό νέφος. Οι πόροι του ιδιωτικού νέφους δεν είναι τόσο οικονομικοί όσο εκείνοι του δημόσιου νέφους, αλλά προσφέρουν περισσότερη παραγωγικότητα από το δημόσιο νέφος.

Το νέφος αυτό διαχειρίζεται από τον οργανισμό και εξυπηρετεί αποκλειστικά αυτόν, μπορεί να βρίσκεται είτε εντός είτε εκτός της περιμέτρου του οργανισμού. Το ιδιωτικό νέφος ονομάζεται επίσης εσωτερικό ή εταιρικό νέφος, το οποίο παρέχει φιλοξενούμενες συσκευές για έναν περιορισμένο αριθμό χρηστών πίσω από ένα firewall.

### 2.3.3 Κοινοτικό Νέφος (Community Cloud)

Μερικοί οργανισμοί δημιουργούν και μοιράζονται μια κοινή υποδομή νέφους, μαζί με πολιτικές, απαιτήσεις, αξίες και ανησυχίες. Το κοινοτικό νέφος [11] δημιουργεί ένα επίπεδο οικονομικής κλίμακας και δημοκρατικής ισορροπίας. Η υποδομή του νέφους μπορεί να φιλοξενείται από έναν τρίτο πάροχο ή εντός ενός

από τους οργανισμούς της κοινότητας. Το νέφος διαχειρίζεται από πολλούς οργανισμούς και υποστηρίζει μια συγκεκριμένη κοινότητα με κοινά ενδιαφέροντα. Το κοινοτικό νέφος είναι πιο ασφαλές από το δημόσιο νέφος.

#### **2.3.4 Υβριδικό Νέφος (Hybrid Cloud): Συνδυασμός δημόσιου και ιδιωτικού νέφους**

Πρόκειται για έναν συνδυασμό δημόσιου, ιδιωτικού και κοινοτικού νέφους [6]. Ωστόσο, οι κρίσιμες λειτουργίες πραγματοποιούνται από το ιδιωτικό νέφος, ενώ οι μη κρίσιμες λειτουργίες από το δημόσιο νέφος. Το δημόσιο νέφος είναι πιο δαπανηρό από το ιδιωτικό νέφος, επομένως το υβριδικό νέφος μπορεί να προσφέρει εξοικονόμηση κόστους. Τα υβριδικά μοντέλα νέφους βασίζονται στην εσωτερική υποδομή, επομένως είναι σημαντικό να εξασφαλίζεται πλεονασμός στα κέντρα δεδομένων.

### **2.4 Μοντέλα υπηρεσιών του Υπολογιστικού Νέφους**

Ένα υπολογιστικό νέφος μπορεί να συνδεθεί με έναν πελάτη (χρήστη ή εφαρμογή) με διάφορους τρόπους μέσω δυνατοτήτων που ονομάζονται υπηρεσίες. Τα μοντέλα υπηρεσιών αποτελούν τα λειτουργικά μοντέλα στα οποία βασίζεται το Υπολογιστικό Νέφος.

Στο διαδίκτυο έχουν αναδειχθεί τρεις κύριοι τύποι υπηρεσιών ή μοντέλων υπηρεσιών [6]:

- Υποδομή ως Υπηρεσία (IaaS).
- Πλατφόρμα ως Υπηρεσία (PaaS).
- Λογισμικό ως Υπηρεσία (SaaS).

#### **2.4.1 Υποδομή ως Υπηρεσία (IaaS)**

Οι πάροχοι υπολογιστικού νέφους προσφέρουν φυσικούς και εικονικούς υπολογιστές, πρόσθετη χωρητικότητα, δικτυακές συσκευές κ.ά.. Οι εικονικές μηχανές ελέγχονται από υπερεπόπτες (hypervisors) που οργανώνονται σε ομάδες

και υποστηρίζονται από συστήματα υποστήριξης λειτουργίας. Οι πελάτες νέφους είναι υπεύθυνοι να εγκαταστήσουν λειτουργικά συστήματα στις εικονικές μηχανές, καθώς και το λογισμικό εφαρμογών τους. Το IaaS επιτρέπει στον πάροχο νέφους να προσφέρει την υποδομή στο διαδίκτυο με πρακτικό τρόπο. Οι πόροι IaaS, όπως αποθηκευτικός χώρος, εύρος ζώνης, υπηρεσίες παρακολούθησης, διευθύνσεις IP, firewalls και εικονικές μηχανές, παρέχονται σε ενοικίαση στον πελάτη, ο οποίος πληρώνει με βάση τον χρόνο χρήσης του πόρου.

#### **2.4.2 Πλατφόρμα ως Υπηρεσία (PaaS)**

Πρόκειται για την παροχή πλατφόρμας ανάπτυξης [31] και ανάπτυξης εφαρμογών μέσω διαδικτύου ως υπηρεσία προς προγραμματιστές, οι οποίοι μπορούν να χρησιμοποιήσουν την πλατφόρμα για να δημιουργούν, να αναπτύσσουν και να διαχειρίζονται εύκολα εφαρμογές SaaS. Προσφέρει επίσης εργαλεία βελτίωσης και ανάπτυξης που είναι απαραίτητα για τη δημιουργία εφαρμογών. Το κύριο χαρακτηριστικό του PaaS περιλαμβάνει ένα εργαλείο σημείου και κλικ που επιτρέπει και σε μη προγραμματιστές να δημιουργούν διαδικτυακές εφαρμογές. Ο χρήστης δεν χρειάζεται να αγοράσει ακριβούς διακομιστές, εξοπλισμό, ενέργεια και χώρο αποθήκευσης δεδομένων. Έτσι, είναι εύκολο να προσαρμόζεται προς τα πάνω ή προς τα κάτω με βάση τις ανάγκες πόρων της εφαρμογής.

#### **2.4.3 Λογισμικό ως Υπηρεσία (SaaS)**

Πρόκειται για την παροχή εφαρμογών (π.χ., ERP ή CRM) ως βοήθεια προς τους τελικούς χρήστες μέσω του διαδικτύου μέσω προγραμμάτων περιήγησης [31].

Οι χρήστες νέφους μπορούν να χρησιμοποιήσουν τις εφαρμογές που είναι ήδη εγκατεστημένες και σε λειτουργία στην υποδομή του νέφους. Με αυτόν τον τρόπο, δεν υπάρχει ανάγκη εγκατάστασης και λειτουργίας της εφαρμογής λογισμικού στους προσωπικούς τους υπολογιστές. Επιπλέον, η ανάγκη για συντήρηση και υποστήριξη του λογισμικού μειώνεται. Ορισμένες εφαρμογές SaaS δεν είναι πλήρως προσαρμόσιμες, όπως μια σουίτα Office. Παρ' όλα αυτά, το SaaS παρέχει Application Programming Interface (API), που επιτρέπει στους προγραμματιστές να δημιουργήσουν προσαρμοσμένες εφαρμογές.

## 2.5 Οφέλη Υπολογιστικού Νέφους

Η υπολογιστική νέφους έχει ορισμένα χαρακτηριστικά που ικανοποιούν τις απαιτήσεις των πελατών ή χρηστών και παρέχουν ποιοτικές υπηρεσίες [7], [8].

- **Υψηλή κλιμάκωση:** Παρέχει πόρους κατά απαίτηση σε μεγάλη κλίμακα, χωρίς να απαιτείται ανθρώπινη παρέμβαση από κάθε πάροχο υπηρεσιών.
- **Υψηλή διαθεσιμότητα και αξιοπιστία:** Η διαθεσιμότητα των διακομιστών είναι πιο αξιόπιστη και υψηλή, περιορίζοντας έτσι τις πιθανότητες αποτυχίας στην υποδομή.
- **Ευελιξία:** Διαμοιράζει τους πόρους μεταξύ των χρηστών και λειτουργεί πολύ γρήγορα.
- **Υψηλό επίπεδο διαμοίρασης:** Διάφοροι πελάτες και εφαρμογές λειτουργούν πιο αποτελεσματικά και με μικρότερο κόστος, μοιραζόμενοι βασική υποδομή μέσω της κατανεμημένης υπολογιστικής.
- **Συντήρηση:** Η συντήρηση των εφαρμογών υπολογιστικής νέφους είναι ευκολότερη, καθώς δεν χρειάζεται να εγκαθίστανται σε κάθε υπολογιστή και είναι προσβάσιμες από διάφορα μέρη, μειώνοντας έτσι το κόστος.
- **Χαμηλό κόστος:** Είναι οικονομικά αποδοτικό, καθώς η εταιρεία δεν χρειάζεται να εγκαταστήσει τη δική της υποδομή. Πληρώνει μόνο για τους πόρους που έχει καταναλώσει.
- **Υπηρεσίες με χρέωση ανά χρήση:** Παρέχονται APIs στους πελάτες για πρόσβαση στις υπηρεσίες του νέφους και πληρώνουν με βάση την κατανάλωση της υπηρεσίας.
- **Αυτόνομη παροχή κατά απαίτηση:** Η υπολογιστική νέφους επιτρέπει στους πελάτες να χρησιμοποιούν υπηρεσίες και πόρους κατά απαίτηση, χωρίς ανθρώπινη αλληλεπίδραση με τους παρόχους υπηρεσιών νέφους. Ο καθένας μπορεί να συνδεθεί σε έναν ιστότοπο και να τις χρησιμοποιήσει οποιαδήποτε στιγμή. Οι υπολογιστικοί πόροι περιλαμβάνουν εικονικές μηχανές, επεξεργαστική ισχύ, αποθηκευτικό χώρο κ.ά.

- **Ευρεία πρόσβαση δικτύου:** Πόροι όπως εικονικές μηχανές, αποθηκευτικός χώρος και επεξεργαστική ισχύς είναι προσβάσιμοι μέσω του διαδικτύου, χρησιμοποιώντας ετερογενείς συσκευές όπως κινητά τηλέφωνα, φορητούς υπολογιστές, υπολογιστές κ.λπ. Δεδομένου ότι η υπολογιστική νέφους βασίζεται στο διαδίκτυο, είναι προσβάσιμη ανά πάσα στιγμή και από οπουδήποτε.
- **Δεξαμενή πόρων:** Η υπολογιστική νέφους επιτρέπει σε πολλαπλούς χρήστες να μοιράζονται μια δεξαμενή πόρων. Κάποιος μπορεί να μοιραστεί μία φυσική βάση δεδομένων, εξοπλισμό και βασική υποδομή. Για παράδειγμα, ένας φυσικός διακομιστής μπορεί να φιλοξενήσει αρκετές εικονικές μηχανές που ανήκουν σε διαφορετικούς χρήστες.
- **Ταχεία ελαστικότητα:** Είναι πολύ εύκολο να κλιμακωθούν οι πόροι προς τα πάνω ή προς τα κάτω οποιαδήποτε στιγμή. Οι πόροι που χρησιμοποιούνται από τους πελάτες ή που τους έχουν αποδοθεί παρακολουθούνται αυτόματα.
- **Μετρημένη υπηρεσία:** Στην περίπτωση της μετρημένης υπηρεσίας, ο πάροχος νέφους ελέγχει και παρακολουθεί κάθε πτυχή της υπηρεσίας νέφους. Εξαρτάται από τον προγραμματισμό της χωρητικότητας, τη χρέωση των πόρων και τη βελτιστοποίηση.

## 2.6 Προκλήσεις

Παρά την ανάπτυξη και το δυναμικό των υπηρεσιών υπολογιστικού νέφους, οι οργανισμοί αντιμετωπίζουν ποικίλες προκλήσεις που πρέπει να αντιμετωπιστούν για την πλήρη αξιοποίηση των δυνατοτήτων του νέφους. Οι κύριες προκλήσεις είναι οι εξής [24], [26]:

**Ασφάλεια:** Η κύρια ανησυχία για την επένδυση στις υπηρεσίες νέφους είναι η ασφάλεια [89],[96], καθώς τα δεδομένα αποθηκεύονται και επεξεργάζονται από τρίτους.

Συχνά παρουσιάζονται θέματα όπως παραβιάσεις αυθεντικοποίησης και διαρροές δεδομένων, που κάνουν τους χρήστες πιο επιφυλακτικούς. Οι εταιρείες νέφους έχουν αρχίσει να ενισχύουν τις δυνατότητες ασφαλείας, και είναι σημαντικό να βεβαιωθείτε ότι ο πάροχος εφαρμόζει ασφαλείς διαδικασίες διαχείρισης ταυτότητας και ελέγχου πρόσβασης.

**Ασφάλεια κωδικών πρόσβασης:** Όσο περισσότεροι χρήστες χρησιμοποιούν έναν λογαριασμό νέφους, τόσο πιο ευάλωτος γίνεται. Για την ασφάλεια, πρέπει να χρησιμοποιούνται πολλαπλά επίπεδα αυθεντικοποίησης, να αλλάζονται τακτικά οι κωδικοί και να διασφαλίζεται η ασφαλής διαχείριση των διαπιστευτηρίων.

**Διαχείριση κόστους:** Αν και οι υπηρεσίες νέφους μειώνουν τα κόστη υλικού και λογισμικού, η προσαρμογή στις ανάγκες μιας επιχείρησης είναι δαπανηρή. Ένα ακόμη έξοδο είναι η μεταφορά δεδομένων σε δημόσιο νέφος, που μπορεί να είναι ιδιαίτερα ακριβή για μικρές επιχειρήσεις.

**Συνδεσιμότητα στο διαδίκτυο:** Η χρήση των υπηρεσιών νέφους απαιτεί γρήγορη σύνδεση στο διαδίκτυο. Οι επιχειρήσεις που αντιμετωπίζουν προβλήματα συνδεσιμότητας θα πρέπει να επενδύσουν σε αξιόπιστη σύνδεση για να αποφύγουν διακοπές λειτουργίας.

**Έλλειψη εξειδίκευσης:** Η διαχείριση των υπολογιστικών τεχνολογιών νέφους είναι περίπλοκη και απαιτεί εξειδικευμένο προσωπικό. Οι επιχειρήσεις πρέπει να εκπαιδεύουν το IT προσωπικό τους για να ανταποκριθούν στις απαιτήσεις της τεχνολογίας νέφους.

**Συμμόρφωση:** Η τήρηση των κανονισμών είναι σημαντική στην υπολογιστική νέφους. Οι επιχειρήσεις πρέπει να συμμορφώνονται με τους κανόνες των κρατικών φορέων, διασφαλίζοντας ότι η αποστολή και διαχείριση των δεδομένων γίνεται σύμφωνα με τα πρότυπα.

**Διαχείριση πόρων:** Είναι σημαντικό να υπάρχει σωστή διαχείριση των πόρων και της συντήρησης. Μια αφιερωμένη ομάδα θα πρέπει να διασφαλίζει ότι τα στοιχεία που χρησιμοποιούνται για τις υπηρεσίες νέφους αξιοποιούνται σύμφωνα με τις πολιτικές και τις διαδικασίες.



**Δημιουργία ιδιωτικού νέφους:** Η υλοποίηση ενός ιδιωτικού νέφους, όπου τα δεδομένα παραμένουν ασφαλή εντός της εταιρείας, απαιτεί από την ομάδα IT να δημιουργήσει και να συντηρήσει τα πάντα από την αρχή. Απαιτεί επίσης αυτοματοποίηση πολλών διαδικασιών για την ομαλή λειτουργία του νέφους.

**Απόδοση:** Όταν οι επιχειρηματικές εφαρμογές μεταφέρονται στο νέφος, η απόδοση εξαρτάται από τον πάροχο. Η επιλογή ενός κατάλληλου παρόχου υπηρεσιών νέφους, με σύγχρονες τεχνολογίες και λύσεις για πραγματικά προβλήματα, είναι κρίσιμη για την επίτευξη της μέγιστης απόδοσης.

**Διαλειτουργικότητα και φορητότητα:** Οι εφαρμογές πρέπει να μπορούν να μετακινούνται εύκολα μεταξύ παρόχων νέφους χωρίς περιορισμούς. Η μεταφορά δεδομένων από έναν πάροχο σε άλλον συχνά έχει περιορισμούς ευελιξίας και απαιτεί προσαρμογές για την ασφαλή δικτύωση και παρακολούθηση.

**Υψηλή διαθεσιμότητα και αξιοπιστία:** Η σταθερότητα και η διαθεσιμότητα είναι κρίσιμα στοιχεία. Εφόσον οι περισσότερες επιχειρήσεις βασίζονται σε τρίτους, η διαθεσιμότητα 24/7 είναι απαραίτητη για την αποφυγή διακοπών λειτουργίας και οικονομικών απωλειών. Είναι σημαντικό να παρακολουθούνται οι SLA (Service Level Agreement) η ανθεκτικότητα και η απόδοση.

**Πολυπλοκότητα του υβριδικού νέφους:** Το υβριδικό περιβάλλον νέφους συνδυάζει δημόσιες και ιδιωτικές εφαρμογές, δημιουργώντας συχνά πολύπλοκα οικοσυστήματα νέφους. Προκλήσεις όπως η κλιμάκωση, η ενσωμάτωση και η ανάκτηση από καταστροφή είναι ιδιαίτερα εμφανείς σε τέτοια πλαίσια.

## 2.7 Ζητήματα ασφάλειας

Η εμπιστοσύνη στον Πάροχο Υπηρεσιών Νέφους (CSP) και τις προσφερόμενες υπηρεσίες του αποτελεί μια από τις βασικές κινητήριες δυνάμεις πίσω από την απόφαση ενός χρήστη να μεταβεί σε ένα σύστημα νέφους ή να παραμείνει στο παραδοσιακό του σύστημα. Η εμπιστοσύνη βασίζεται στην αξιολόγηση του κατά πόσο ο πάροχος καλύπτει όλους τους κινδύνους, συμπεριλαμβανομένων των τομέων της ασφάλειας δεδομένων, της ασφάλειας των Εικονικών Μηχανών (VM),

καθώς και άλλων θεμάτων συμμόρφωσης και κανονιστικών απαιτήσεων. Οι τρεις παράγοντες που λαμβάνονται υπόψη για την αξιολόγηση της ασφάλειας του υπολογιστικού νέφους είναι η Εμπιστευτικότητα, η Ακεραιότητα και η Διαθεσιμότητα CIA (Confidentiality, Integrity, Availability). Δεδομένου ότι το CIA είναι ένα ευρέως χρησιμοποιούμενο πρότυπο για τον προσδιορισμό των ανησυχιών ασφαλείας ενός παραδοσιακού πληροφοριακού συστήματος, αυτή η ενότητα εστιάζει στη γενίκευση των απαιτήσεων ασφαλείας σε ένα υπάρχον σύστημα νέφους στο πλαίσιο αυτού του μοντέλου.

Παρουσιάζεται επίσης περαιτέρω υποκατηγοριοποίηση και ταξινόμηση των θεμάτων ασφαλείας για την ευκολότερη κατανόηση, χαρτογράφηση και αξιολόγηση των ειδικών επιθέσεων νέφους και των προτεινόμενων λύσεων στις επόμενες ενότητες [9], [25], [27], [28], [89], [96].

### 2.7.1 Εμπιστευτικότητα

Η εμπιστευτικότητα [10], [12], [13], [14], [15], [16], [17], [29] αναφέρεται στην προστασία ορισμένων πόρων μιας επιχείρησης από τη διαρροή σε μη εξουσιοδοτημένους χρήστες. Σε ένα σύστημα νέφους, τέτοιοι χρήστες μπορεί να είναι πελάτες που επιδιώκουν να αποκτήσουν μη εξουσιοδοτημένη πρόσβαση σε δεδομένα άλλου ατόμου που είναι αποθηκευμένα στον ίδιο πίνακα με τα δεδομένα του εισβολέα από τον CSP. Υπάρχει επίσης η πιθανότητα ο CSP να έχει ανέντιμα ή περίεργα μέλη που θα μπορούσαν να δουν ή ακόμα και να αλλοιώσουν τα προσωπικά και πολύτιμα δεδομένα του πελάτη. Πέρα από τα δεδομένα του πελάτη, το δίκτυο της Εικονικής Μηχανής και οι εικόνες της Εικονικής Μηχανής έχουν επίσης αναπόφευκτες απαιτήσεις εμπιστευτικότητας.

Σύμφωνα με τις διάφορες απαιτήσεις εμπιστευτικότητας του συστήματος νέφους, συζητούνται οι εξής κατηγορίες:

#### ***Εμπιστευτικότητα Δεδομένων***

Τα δεδομένα που βρίσκονται στο τέλος του CSP συχνά αποθηκεύονται και επεξεργάζονται σε απλό κείμενο. Έτσι, ο CSP (SaaS) είναι υπεύθυνος για τη

διατήρηση της εμπιστευτικότητας των δεδομένων των πελατών καθ' όλη τη διάρκεια ζωής τους. Ορισμένα ειδικά ζητήματα εμπιστευτικότητας δεδομένων περιλαμβάνουν:

- Πολλοί πάροχοι υπηρεσιών αποθήκευσης νέφους επιτρέπουν την κοινή πρόσβαση σε διαδικτυακούς φακέλους που αποθηκεύουν δεδομένα χρήστη, κάτι που μπορεί να οδηγήσει σε πιθανή απώλεια εμπιστευτικότητας. Ακόμα και όταν ένα αρχείο μοιράζεται σε μια ομάδα μέσω υπηρεσίας αποθήκευσης νέφους, ο ιδιοκτήτης πρέπει να λαμβάνει περιοδικές ενημερώσεις για οποιεσδήποτε αλλαγές στην ομάδα.
- Η πραγματική γεωγραφική τοποθεσία των δεδομένων ενός χρήστη είναι ένας άλλος παράγοντας που επηρεάζει την εμπιστευτικότητα των δεδομένων. Ο CSP μπορεί να μετακινεί τα δεδομένα από ένα κέντρο δεδομένων σε άλλο, το οποίο σε πολλές περιπτώσεις αλλάζει το σύνολο των νομικών κανονισμών που εφαρμόζονται (ειδικά αν τα δεδομένα περάσουν τα σύνορα μιας χώρας).
- Η ακατάλληλη ή ελλιπής διαγραφή δεδομένων από τον CSP μπορεί να είναι επικίνδυνη για τους καταναλωτές που ζήτησαν την αφαίρεση της υπηρεσίας ή για όσους η περίοδος συνδρομής έχει λήξει, καθώς τα υπόλοιπα των διαγραμμένων δεδομένων μπορεί να προκαλέσουν παραβίαση εμπιστευτικότητας.
- Ο CSP ενδέχεται να χρησιμοποιήσει τρίτους για υπηρεσίες δημιουργίας αντιγράφων ασφαλείας δεδομένων. Αυτοί οι μη αξιόπιστοι τρίτοι πάροχοι υπηρεσιών μπορεί να αξιοποιήσουν τα προσωπικά δεδομένα του πελάτη με άδικους τρόπους.
- Οι καταναλωτές νέφους συχνά ζητούν περισσότερα δεδομένα παρακολούθησης ή καταγραφής για τη δική τους ευκολία και ασφάλεια. Τα δεδομένα καταγραφής περιέχουν ευαίσθητα δεδομένα υποδομής του παρόχου υπηρεσιών, τα οποία δεν πρέπει να τίθενται σε κίνδυνο από το νέφος. Ως εκ τούτου, πρέπει να γίνονται αρκετές διαπραγματεύσεις μεταξύ του CSP και των χρηστών σχετικά με τα δεδομένα καταγραφής που

μπορούν να μοιραστούν χωρίς να τίθεται σε κίνδυνο η εμπιστευτικότητα του CSP.

- Οι πάροχοι υπηρεσιών νέφους που δεν επιτρέπουν στους ιδιοκτήτες δεδομένων να κρυπτογραφούν τα δεδομένα τους πριν τα αναπτύξουν στο νέφος, θέτουν σε σοβαρό κίνδυνο την εμπιστευτικότητα των δεδομένων των χρηστών. Ευαίσθητες πληροφορίες όπως ιατρικά ή κρατικά δεδομένα δεν πρέπει να αποθηκεύονται στο νέφος, αν δεν είναι διαθέσιμες επιλογές κρυπτογράφησης.

### ***Εμπιστευτικότητα στην Εικονικοποίηση***

Στην Υποδομή ως Υπηρεσία (IaaS), ο CSP φιλοξενεί εικονικές μηχανές όπου εκτελούνται οι εφαρμογές του χρήστη. Σε ένα σύστημα νέφους, όποιος έχει δικαιώματα πρόσβασης στο κεντρικό σύστημα μπορεί να διαβάσει ή να τροποποιήσει τις υπηρεσίες που βρίσκονται σε κάθε Εικονική Μηχανή (VM). Συνεπώς, οι χρήστες δεν μπορούν να προστατεύσουν από μόνοι τους την εμπιστευτικότητα των VMs.

- Ένας διαχειριστής συστήματος με πρόσβαση root θα μπορούσε να εισέλθει εξ αποστάσεως σε οποιαδήποτε μηχανή και να ανακατευθύνει την εικονική μηχανή σε άλλη, εκτός του περιβάλλοντος ασφαλείας της IaaS, προκαλώντας παραβίαση εμπιστευτικότητας.
- Η μετανάστευση εικονικών μηχανών, ειδικά σε πραγματικό χρόνο, αποτελεί ένα χρήσιμο χαρακτηριστικό των συστημάτων υπολογιστικού νέφους για την εξισορρόπηση φόρτου, την ελαστική κλιμάκωση, την ανοχή σε σφάλματα και τη συντήρηση υλικού. Ο CSP πρέπει να λάβει τα απαραίτητα μέτρα για τη διατήρηση της εμπιστευτικότητας των περιπτώσεων VMs και των μεταδεδομένων τους κατά τη μετανάστευση.
- Στο εικονικό περιβάλλον ενός συστήματος νέφους, πολλοί φόρτοι εργασίας μοιράζονται το ίδιο υλικό, κάτι που απαιτεί κατάλληλους κανόνες για τη διασφάλιση της απομόνωσης των πόρων και τη διατήρηση της ασφάλειας δεδομένων.

- Ο Διαχειριστής Εικονικών Μηχανών (VMM ή Hypervisor) είναι ένα λογισμικό χαμηλού επιπέδου που ελέγχει και παρακολουθεί τις εικονικές μηχανές. Εφόσον, όπως κάθε παραδοσιακό λογισμικό, μπορεί να παρουσιάσει αδυναμίες ασφάλειας, η διατήρηση του VMM όσο το δυνατόν πιο απλού μειώνει τους κινδύνους.
- Οι Εικόνες Εικονικών Μηχανών (VMI) μπορεί να περιέχουν κακόβουλο κώδικα, που αν κατέβει από νόμιμους χρήστες, μπορεί να επηρεάσει τα συστήματά τους, ή να αποθηκεύσουν πληροφορίες του προηγούμενου ιδιοκτήτη, που μπορούν να χρησιμοποιηθούν κακόβουλα από άλλους χρήστες.
- Τα Εικονικά Δίκτυα VLANs (Virtual Local Area Network) πρέπει επίσης να απομονώνονται για την αποφυγή μη εξουσιοδοτημένης ροής δεδομένων μέσω αυτών και οι πιθανότητες επιθέσεων sniffing ή spoofing να ελαχιστοποιούνται.

### 2.7.2 Ακεραιότητα

Η ακεραιότητα [27] αναφέρεται στην ιδιότητα ασφαλείας ενός πόρου που διασφαλίζει ότι δεν έχει τροποποιηθεί από μη εξουσιοδοτημένο τρίτο μέρος. Αυτό το χαρακτηριστικό εξασφαλίζει την ακρίβεια και ορθότητα ενός πόρου αναφορικά με τον ιδιοκτήτη του. Συνήθως, οι λειτουργίες προσθήκης, διαγραφής ή επεξεργασίας θεωρείται ότι επηρεάζουν την ακεραιότητα ενός πόρου.

Δεδομένου ότι οι υπηρεσίες νέφους προσπελούνται μέσω προγραμμάτων περιήγησης, όλες οι επιθέσεις που σχετίζονται με τον ιστό είναι εξαιρετικά συχνές σε ένα περιβάλλον νέφους και μπορούν να τροποποιήσουν τα περιεχόμενα των αρχείων χρηστών και των βάσεων δεδομένων, των μεταδεδομένων, των Εικονικών Μηχανών ακόμα και των αρχείων WSDL (Web Services Description Language).

Οι διάφορες απαιτήσεις ακεραιότητας για το σύστημα νέφους περιλαμβάνουν τις παρακάτω κατηγορίες:

### Ακεραιότητα Δεδομένων

Το σύστημα νέφους διαχειρίζεται μεγάλο όγκο δεδομένων με απαιτήσεις που φτάνουν σε επίπεδα Terabyte (TB) ή και Petabyte (PB). Έτσι, οι προκλήσεις ακεραιότητας δεδομένων που σχετίζονται με υπηρεσίες όπως Data as a Service, Software as a Service και Platform as a Service πρέπει να αντιμετωπίζονται προσεκτικά.

- **Απειλές ελλιπών δεδομένων:** Τα δεδομένα που αποθηκεύονται στον CSP ενέχουν κίνδυνο για την ακεραιότητά τους. Ο CSP θα μπορούσε να διαγράψει ορισμένα έγκυρα δεδομένα που σχετίζονται με πελάτες ή να στείλει ελλιπή σύνολα δεδομένων χωρίς ο πελάτης να το αντιληφθεί.
- **Επίθεση SQL Injection:** Είναι μια σημαντική διαδικτυακή επίθεση που θα μπορούσε να τροποποιήσει το περιεχόμενο των βάσεων δεδομένων πελατών εκμεταλλευόμενη τα κενά ασφαλείας των διακομιστών.
- **Επιθέσεις Cross-Scripting:** Οι χάκερ εισάγουν κακόβουλα scripts (JavaScript, VBScript, HTML) σε ευάλωτες δυναμικές ιστοσελίδες ώστε ο κακόβουλος κώδικας να εκτελείται στον περιηγητή του χρήστη, παραβιάζοντας την ακεραιότητα των δεδομένων.
- **Επίθεση Spoofing Μεταδεδομένων:** Τροποποιεί το περιεχόμενο των αρχείων περιγραφής WSDL για εκτέλεση μη εξουσιοδοτημένων ενεργειών.
- **Επίθεση Wrapping:** Αυτή η επίθεση επηρεάζει τις διαδικτυακές υπηρεσίες νέφους και γίνεται πιο πιθανή σε συστήματα νέφους. Η μετάφραση των μηνυμάτων SOAP (Simple Object Access Protocol) στην υπηρεσία TLS (Transport Layer Security) μπορεί να αντιγραφεί και να σταλεί στον διακομιστή ως αυθεντικός χρήστης, επιτρέποντας σε έναν εισβολέα να παρεμβαίνει στο νέφος και να εκτελεί κακόβουλο κώδικα.

### Ακεραιότητα Εικονικοποίησης

Η εικονικοποίηση εισάγει ορισμένα θέματα ασφαλείας που εκτείνονται πέρα από την εμπιστευτικότητα, καθώς πρέπει να ληφθεί μέριμνα και για την ακεραιότητα των Εικονικών Μηχανών (VM) και των εικόνων τους (VMIs).

- **Επιθέσεις εκ των έσω:** Οι διαχειριστές του CSP έχουν πλήρη πρόσβαση στις Εικονικές Μηχανές, κάτι που απαιτεί μέτρα προστασίας για την ακεραιότητα των VMs.
- **Δημιουργία κακόβουλης υπηρεσίας:** Ένας εισβολέας μπορεί να προσθέσει κακόβουλες VMs στο σύστημα νέφους, θέτοντας σε κίνδυνο την ακεραιότητα των υπηρεσιών και VMs.
- **Αναπαραγωγή Εικονικών Μηχανών:** Η αναπαραγωγή μιας VM μπορεί να προκαλέσει διαρροή δεδομένων, αν δεν αντιμετωπιστεί σωστά. Εφαρμογή πολιτικών για τον περιορισμό της αναπαραγωγής ευαίσθητων VMs είναι απαραίτητη.
- **Αναίρεση Εικονικών Μηχανών (VM Rollback):** Αυτή η διαδικασία επανενεργοποιεί τρωτά σημεία ασφαλείας ή απενεργοποιημένους λογαριασμούς. Για την προστασία από αυτό το πρόβλημα απαιτείται η συντήρηση στιγμιοτύπων των VMs.
- **Μετανάστευση VMs:** Η μεταφορά VMs από ένα κέντρο δεδομένων σε άλλο είναι χρήσιμη για αύξηση πόρων, αλλά απαιτείται σωστή εφαρμογή πολιτικών ασφαλείας για να διασφαλιστεί η ακεραιότητα των μεταφερόμενων VMs.

### 2.7.3 Διαθεσιμότητα

Η διαθεσιμότητα [28] αποτελεί έναν από τους πιο σημαντικούς τομείς ασφάλειας που πρέπει να διασφαλίζεται από τον Πάροχο Υπηρεσιών Νέφους (CSP). Οι επιχειρήσεις που χρησιμοποιούν υπηρεσίες νέφους για να εξυπηρετούν τους πελάτες τους πρέπει να εξασφαλίζουν τη συνεχή διαθεσιμότητα αυτών των υπηρεσιών, καθώς ακόμη και ο ελάχιστος χρόνος διακοπής μπορεί να οδηγήσει σε σημαντικές, ανεπανόρθωτες οικονομικές απώλειες. Μια τυπική Συμφωνία Επίπεδου Υπηρεσίας SLA (Service Level Agreement) δηλώνει την υποχρέωση του παρόχου για τη διαθεσιμότητα και την ανταπόκριση στη ζήτηση.

Για παράδειγμα, μπορεί να ορίζεται ότι οι πόροι θα είναι διαθέσιμοι στο 99,999% του χρόνου και ότι θα παρέχονται επιπλέον πόροι δυναμικά όταν η χρήση τους

υπερβαίνει το 80%. Θέματα που σχετίζονται με τη διαθεσιμότητα δεδομένων και Εικονικών Μηχανών αναλύονται παρακάτω:

### **Διαθεσιμότητα Δεδομένων/Υπηρεσιών**

Στην συγκεκριμένη υποκατηγορία αναφέρονται επίσης οι παρακάτω περιπτώσεις:

- Η επίθεση άρνησης υπηρεσίας DoS (Denial of Service) στο σύστημα νέφους αποτελεί μία από τις κύριες αιτίες μη διαθεσιμότητας δεδομένων ή υπηρεσιών. Ο επιτιθέμενος στέλνει μεγάλο αριθμό αιτημάτων σε μια συγκεκριμένη υπηρεσία, προκαλώντας υψηλή φόρτιση. Το σύστημα του νέφους παρέχει περισσότερη υπολογιστική ισχύ για να ανταποκριθεί στο επιπλέον φορτίο, γεγονός που αφενός αποτρέπει την πρόσβαση σε άλλους χρήστες και αφετέρου βοηθά τον επιτιθέμενο να εκμεταλλευτεί τους πόρους.
- Έμμεση επίθεση άρνησης υπηρεσίας είναι επίσης πιθανή σε ένα σύστημα νέφους, όπου άλλες υπηρεσίες που λειτουργούν στον ίδιο διακομιστή με μια υπηρεσία που δέχεται επίθεση ενδέχεται επίσης να καταστούν μη διαθέσιμες. Καθώς οι πόροι του διακομιστή εξαντλούνται, οι υπόλοιπες υπηρεσίες δεν μπορούν να ολοκληρώσουν τις λειτουργίες τους, και το πρόβλημα μπορεί να επιδεινωθεί καθώς το σύστημα προσπαθεί να μεταφέρει τις υπηρεσίες σε άλλους διακομιστές, επιβαρύνοντας τους υπόλοιπους πόρους του νέφους.
- Έλεγχοι διείσδυσης πελατών μπορούν επίσης να επηρεάσουν τη διαθεσιμότητα, καθώς μπορεί να σταματήσουν ορισμένες υπηρεσίες για συγκεκριμένο χρονικό διάστημα, προκαλώντας προσωρινή διακοπή.
- Αναστολή υπηρεσιών από τρίτους παρόχους WAN (Wide Area Network) μπορεί να οδηγήσει σε προσωρινές διακοπές υπηρεσιών, ενώ σφάλματα λογισμικού ενδέχεται να επηρεάσουν πολλαπλές εκδόσεις δεδομένων ταυτόχρονα, καθιστώντας τα μη διαθέσιμα στους ιδιοκτήτες τους.



- Φυσικές καταστροφές, όπως πυρκαγιά ή πλημμύρα σε ένα κέντρο δεδομένων, μπορεί να επηρεάσουν τόσο τα κύρια όσο και τα εφεδρικά αντίγραφα των δεδομένων, απειλώντας σοβαρά τη διαθεσιμότητα.

### **Διαθεσιμότητα Εικονικοποίησης**

Όπως έχουμε αναφέρει, η διασφάλιση υψηλής διαθεσιμότητας περιλαμβάνει διάφορους τομείς, όπως η τρωτότητα του δικτύου, η πολυτοπική εφεδρεία και η αποφυγή αστοχίας αποθήκευσης. Ωστόσο, δεδομένου ότι η εικονικοποίηση αποτελεί έναν από τους βασικότερους τομείς του συστήματος νέφους, η διαθεσιμότητα στο νέφος πρέπει να λαμβάνει υπόψη και την εικονικοποίηση.

Μία από τις κύριες προκλήσεις εδώ είναι η μεταφορά IP (IP failover). Η ανάγκη προστασίας μιας εφαρμογής από αστοχία οποιουδήποτε κόμβου δεν είναι νέα, ωστόσο πολλά από τα προϊόντα λογισμικού που αντιμετωπίζουν αυτό το πρόβλημα είναι εν μέρει ασύμβατα με τις παρεχόμενες υπηρεσίες νέφους, και οι περισσότεροι πάροχοι δημοσίου νέφους αδυνατούν να παρέχουν τη συγκεκριμένη λειτουργικότητα. Αυτό έχει ως αποτέλεσμα οι πελάτες να βασίζονται σε λύσεις υψηλής διαθεσιμότητας εκτός νέφους. Συνεπώς, οι εικονικές μηχανές πρέπει να προστατεύονται από αστοχίες, έτσι ώστε η αποτυχία μιας εικονικής μηχανής να μπορεί να αντισταθμιστεί άμεσα από άλλη μέσω αποδοτικών μεθόδων.

#### **2.7.4 Λύσεις ασφάλειας**

Η απώλεια δεδομένων επηρεάζει τις επιχειρήσεις και μειώνει την εμπιστοσύνη στους οργανισμούς. Ένα αποτελεσματικό σχέδιο ασφαλείας αποτελεί την καλύτερη προστασία απέναντι σε παραβιάσεις δεδομένων. Ο έλεγχος ταυτότητας πολλαπλών παραγόντων MFA (Multi Factor Authentication) καθώς και η κρυπτογράφηση είναι δύο σημαντικοί μηχανισμοί ασφαλείας που συμβάλλουν στην προστασία των επιχειρήσεων στο cloud. Ο MFA αποτελεί ένα μέτρο ενίσχυσης της ασφάλειας του δικτύου μέσω «απλών διαδικασιών IT», ενώ οι βέλτιστες πρακτικές περιλαμβάνουν τακτικές σαρώσεις ευπαθειών και γρήγορη ανίχνευση των απειλών στις συσκευές.

Για την προστασία των διαπιστευτηρίων και της διαχείρισης πρόσβασης, απαιτούνται μηχανισμοί ελέγχου ταυτότητας πολλαπλών παραγόντων, όπως έξυπνες κάρτες, κωδικοί OTP (One Time Password) και έλεγχος ταυτότητας μέσω κινητού, καθώς καθιστούν δυσκολότερη την πρόσβαση σε μη εξουσιοδοτημένους χρήστες που χρησιμοποιούν κλεμμένους κωδικούς πρόσβασης. Η χρήση κρυπτογράφησης συμβάλλει στην αποτροπή της κλοπής διαπιστευτηρίων, περιορίζοντας τον κίνδυνο ανεξέλεγκτης πρόσβασης σε πόρους.

Σε δίκτυα με ανεπαρκή προστασία, η κλοπή κωδικών πρόσβασης μπορεί να οδηγήσει σε επιθέσεις πλευρικής μετακίνησης.

Επιπλέον, οι μη ασφαλείς διεπαφές και API αποτελούν μία ακόμη σοβαρή απειλή, και για την αντιμετώπισή τους απαιτείται η επιθεώρηση του πηγαίου κώδικα μιας εφαρμογής, ώστε να διασφαλιστεί ότι περιλαμβάνει επαρκείς ελέγχους ασφαλείας και ότι αυτοί λειτουργούν σωστά. Οι δοκιμές διείσδυσης βοηθούν στον εντοπισμό των αδυναμιών ενός συστήματος, αξιολογώντας την επάρκεια των μέτρων ασφαλείας και ανιχνεύοντας πιθανές παρακάμψεις αυτών.

Η κατάληψη λογαριασμών και οι επιθέσεις άρνησης υπηρεσίας DoS αποτελούν επίσης σημαντικές απειλές για τις επιχειρήσεις, οι οποίες θα πρέπει να εφαρμόζουν τεχνικές ασφάλειας πολλαπλών επιπέδων για να περιορίσουν τις επιπτώσεις και να μειώσουν τον κίνδυνο νομικών συνεπειών. Για την αποφυγή τέτοιων επιθέσεων, συνιστάται η αποτροπή της χρήσης κοινών διαπιστευτηρίων μεταξύ χρηστών και υπηρεσιών, καθώς και η υιοθέτηση ισχυρών τεχνικών ελέγχου ταυτότητας δύο παραγόντων.

Η αγορά αυξημένου εύρους ζώνης μπορεί να περιορίσει την επιτυχία μιας DoS επίθεσης, καθώς ένας εισβολέας θα χρειαστεί μεγαλύτερο όγκο δεδομένων για να προκαλέσει συμφόρηση. Επιπλέον, τα σύγχρονα συστήματα πρόληψης εισβολών και οι νέες τεχνολογίες τείχους προστασίας περιλαμβάνουν λειτουργίες ανίχνευσης υπογραφών και επαλήθευσης συνδέσεων, οι οποίες μειώνουν την πιθανότητα επιτυχίας τέτοιων επιθέσεων. Οι προηγμένες κυβερνοεπιθέσεις στοχεύουν κυρίως επιχειρήσεις και κρατικούς οργανισμούς. Ένα από τα πιο αποτελεσματικά μέτρα αντιμετώπισης αυτών των απειλών είναι η συχνή ενημέρωση των μηχανισμών

ανίχνευσης, καθώς πολλές επιθέσεις εκμεταλλεύονται την αλληλεπίδραση των χρηστών. Η συνεχής επιτήρηση και η προληπτική αντιμετώπιση των πιθανών κινδύνων συμβάλλουν στην προστασία των δεδομένων και τη διασφάλιση της αξιοπιστίας των υπηρεσιών cloud.

Η αξιόπιστη υπολογιστική στο cloud εφαρμόζεται για τη διασφάλιση, την προστασία, την εμπιστευτικότητα και την ακεραιότητα των δεδομένων, με αποδοτικές υπολογιστικές διαδικασίες. Για τον σκοπό αυτό, χρησιμοποιείται μια επαναπρογραμματιζόμενη πύλη λογικής FPGA (Field Programmable Gate Array) για την ανάπτυξη ασφαλούς ταυτοποίησης των υπολογισμών, ενσωματωμένη στην αρχιτεκτονική λογικής.

Η FPGA αποθηκεύει συστηματική και συμμετρική κρυπτογράφηση στη μνήμη της και, αφού προετοιμαστεί, αναπτύσσεται στους cloud servers.

Τα κλειδιά και οι υπογραφές της FPGA χρησιμοποιούνται για την κρυπτογράφηση από αξιόπιστους διαχειριστές και αρχές, εξασφαλίζοντας έτσι την ασφαλή διεξαγωγή των εφαρμογών.

Για την περαιτέρω ενίσχυση της ασφάλειας του cloud, χρησιμοποιούνται καταναμεμημένοι πίνακες κατακερματισμού, οι οποίοι αποτρέπουν εισβολές και επιθέσεις καταναμεμμένης άρνησης υπηρεσίας DDoS (Distributed Denial of Service). Επιπλέον, η προστασία των δεδομένων πραγματοποιείται μέσω υδατογραφήματος βασισμένου στη χρωματική σήμανση. Η μέθοδος αυτή επιτρέπει την ένδειξη διαφορετικών επιπέδων ασφάλειας μέσω διακριτών χρωμάτων και ιδιοτήτων, τις οποίες γνωρίζουν μόνο οι ιδιοκτήτες των δεδομένων. Οι πάροχοι cloud δεν μπορούν να εντοπίσουν τους χρήστες, εκτός εάν έχουν γνώση αυτών των ιδιοτήτων.

Επιπρόσθετα, εφαρμόζεται ένα σύστημα διαχείρισης εμπιστοσύνης για την προστασία της επικοινωνίας μεταξύ των πελατών και των παρόχων υπηρεσιών cloud. Χρησιμοποιείται ένας "Trust Integrator" για τη διατήρηση της εμπιστοσύνης μεταξύ των παρόχων και των χρηστών, διαχειριζόμενος τις παραμέτρους διαπραγμάτευσης, την εύρεση παρόχων υπηρεσιών και τη δημιουργία ομάδων για την εκτέλεση των απαιτούμενων λειτουργιών. Μια άλλη

προσέγγιση για την ασφάλεια του περιβάλλοντος εκτέλεσης των cloud συστημάτων είναι η χρήση υδατογραφήματος, ιδιαίτερα σε εφαρμογές Java. Οι εφαρμογές Java προστατεύονται μέσω διαφόρων αλγορίθμων υδατογραφήματος, και εκτελούνται μόνο σε μια προσαρμοσμένη έκδοση της Εικονικής Μηχανής Java (JVM), η οποία επιτρέπει την αναγνώριση και την προστασία του κώδικα. Μόνο τα προγράμματα με έγκυρο υδατογράφημα εκτελούνται, ενώ όσα έχουν μη αναγνωρισμένο υδατογράφημα απορρίπτονται και δημιουργείται προειδοποιητικό μήνυμα για τον διαχειριστή, αποτρέποντας έτσι μη εξουσιοδοτημένη εκτέλεση.

Για την προστασία της εμπιστευτικότητας των δεδομένων, εφαρμόζεται συμμετρική και ασύμμετρη κρυπτογράφηση. Αυτός ο μηχανισμός χρησιμοποιείται από μια αξιόπιστη τρίτη οντότητα που λειτουργεί βάσει χαρακτηριστικών ή ιδιοτήτων. Η αξιόπιστη αυτή οντότητα διασφαλίζει την επικοινωνία και τη μεταφορά δεδομένων μεταξύ των διαφορετικών συστατικών του συστήματος.

Η προστασία και η ιδιωτικότητα των επικοινωνιών διασφαλίζονται μέσω των πρωτοκόλλων IPSec (Internet Protocol Security) και SSL (Secure Sockets Layer), τα οποία εφαρμόζονται στις επικοινωνίες μεταξύ των μηχανών. Επιπλέον, η μέθοδος SSO (Single Sign-On) και το πρωτόκολλο LDAP (Lightweight Directory Access Protocol), χρησιμοποιούνται μαζί με τις ψηφιακές υπογραφές για την εκτέλεση εξουσιοδοτήσεων.

## 2.8 Ρύθμιση συμμόρφωσης ασφάλειας στο Cloud

Η ασφάλεια και η ακεραιότητα των δεδομένων αποτελούν βασικές απαιτήσεις.

Για να διασφαλιστούν αυτές οι προϋποθέσεις, οι αρμόδιες αρχές και οι ρυθμιστικοί φορείς έχουν θεσπίσει κανονισμούς και κανόνες που εγγυώνται την προστασία των δεδομένων και τη διατήρηση της ιδιωτικότητας. Οι κανονισμοί αυτοί εφαρμόζονται ως εξής:

Η αξιολόγηση των προϊόντων ασφαλείας παγκοσμίως πραγματοποιείται βάσει των κοινών κριτηρίων. Σύμφωνα με αυτήν τη συμμόρφωση, πραγματοποιείται ανάλυση προστασίας, η οποία ορίζει με σαφήνεια τις απαιτήσεις ασφαλείας,

ανεξάρτητα από την ίδια την αξιολόγηση. Το επίπεδο διασφάλισης της αξιολόγησης κυμαίνεται από το 1 έως το 7, με την κατηγοριοποίηση να βασίζεται στα μέτρα που λαμβάνει η εκάστοτε εταιρεία. Επιπλέον, τα περιβάλλοντα cloud αξιολογούνται και πιστοποιούνται ως προς τη συμμόρφωσή τους με τις απαιτήσεις ασφαλείας. Παραδείγματα πιστοποιημένων λύσεων περιλαμβάνουν τον Citrix Xen Server, τον KVM Hypervisor και τον VMware ESXi.

Με την αυξανόμενη χρήση της τεχνολογίας cloud, δημόσια και ιδιωτικά δεδομένα αποθηκεύονται συχνά σε αυτά τα περιβάλλοντα. Συνεπώς, είναι απαραίτητο να ενισχυθεί η αξιοπιστία και η ασφάλεια του cloud, ώστε να διασφαλιστεί η προστασία των δεδομένων και να αποφευχθούν νομικά ή ηθικά ζητήματα. Μερικοί από τους βασικούς κανονισμούς που σχετίζονται με την ιδιωτικότητα περιλαμβάνουν:

Σύμφωνα με τον νόμο περί Φορητότητας και Λογοδοσίας της Ασφάλισης Υγείας HIPAA (Health Insurance Portability And Accountability), τα δεδομένα που σχετίζονται με την υγεία θεωρούνται προσωπικά και η ιδιωτικότητά τους πρέπει να διασφαλίζεται.

Κατά τις συναλλαγές και την επεξεργασία αυτών των δεδομένων, η προστασία τους από απειλές και ιούς είναι κρίσιμη. Δεδομένης της ευαισθησίας αυτών των πληροφοριών, η ασφαλής αποθήκευση, η δημιουργία αντιγράφων ασφαλείας και η ύπαρξη σχεδίου ανάκαμψης σε περίπτωση καταστροφής αποτελούν βασικές προϋποθέσεις. Επιπλέον, οι πληροφορίες υγείας ενός ασθενούς δεν μπορούν να κοινοποιηθούν χωρίς τη συγκατάθεσή του, εκτός εάν ζητηθεί από αρμόδια αρχή και η πρόσβαση είναι σύμφωνη με τους σχετικούς κανονισμούς.

Επιπλέον, σύμφωνα με τον Νόμο περί Ιδιωτικότητας Ηλεκτρονικών Επικοινωνιών ECPA (Electronic Communications Privacy Act), η ασφάλεια και η προστασία των ηλεκτρονικών δεδομένων είναι υποχρεωτικές. Ο νόμος ορίζει ότι οι μεταδόσεις δεδομένων πρέπει να παραμένουν ιδιωτικές και η μη εξουσιοδοτημένη πρόσβαση σε ηλεκτρονικά δεδομένα απαγορεύεται. Στο πλαίσιο του ECPA, η μη εξουσιοδοτημένη πρόσβαση σε δεδομένα που αποθηκεύονται μέσω τρίτων φορέων δεν επιτρέπεται, ενώ επιτρέπεται μόνο η εξουσιοδοτημένη πρόσβαση.

Μια ένωση, γνωστή ως **Digital 4th**, εργάζεται για τη βελτίωση της αξιοπιστίας του ECPA, προτείνοντας αλλαγές στον νόμο.

Επιπροσθέτως, ο Νόμος περί Δίκαιης Πιστωτικής Αναφοράς FCRA (Fair Credit Reporting Act) καθορίζει τους κανόνες και τους κανονισμούς που σχετίζονται με την εμπιστευτικότητα των δεδομένων πιστωτικών καρτών των χρηστών. Οι εταιρείες αναφοράς πιστωτικών στοιχείων αποθηκεύουν συνήθως τα δεδομένα τους στο cloud, γεγονός που καθιστά απαραίτητη την εφαρμογή μέτρων ασφαλείας για τη συμμόρφωση με τους κανονισμούς του FCRA.

Τέλος, ο Νόμος Gramm–Leach–Bliley (GLBA) απαιτεί από τα χρηματοπιστωτικά ιδρύματα να διατηρούν τα δεδομένα των πελατών εμπιστευτικά και ασφαλή. Τα ιδρύματα αυτά είναι υποχρεωμένα να επιλέγουν μόνο παρόχους υπηρεσιών που συμμορφώνονται με τους κανονισμούς ιδιωτικότητας και ακεραιότητας των δεδομένων.

## 2.9 Σύγχρονες τάσεις στο Cloud Computing

Η κοινοποίηση δεδομένων στο υπολογιστικό νέφος είναι ένα πολύ σημαντικό ζήτημα. Πολλές τεχνολογίες ασφαλούς κοινοποίησης δεδομένων έχουν εισαχθεί σε αυτόν τον τομέα και έχουν γίνει δημοφιλείς, όπως η κοινοποίηση δεδομένων με λεπτομερή έλεγχο πρόσβασης και η κρυπτογράφηση με βάση χαρακτηριστικά [97], [99].

Ωστόσο, οι περισσότερες από τις υπάρχουσες λύσεις αντιμετωπίζουν μειονεκτήματα, όπως το υψηλό υπολογιστικό κόστος και η ασθενής ασφάλεια δεδομένων, γεγονός που περιορίζει σημαντικά τις συσκευές με περιορισμένους πόρους να προσαρμόσουν την υπηρεσία στις ανάγκες τους. Το ζήτημα της ταυτόχρονης επίτευξης αποδοτικότητας από την πλευρά του κατόχου των δεδομένων και της τυπικής εμπιστευτικότητας της κοινοποίησης δεδομένων στο υπολογιστικό νέφος παραμένει ακόμη άλυτο. Με την αυξανόμενη δημοτικότητα του υπολογιστικού νέφους, η επίτευξη ασφαλούς και αποδοτικής κοινοποίησης δεδομένων σε περιβάλλοντα νέφους αποτελεί ένα επείγον πρόβλημα που πρέπει να

λυθεί. Επιπλέον, η επίτευξη τόσο της ανωνυμίας όσο και της δυνατότητας ιχνηλάτησης είναι επίσης μια πρόκληση στην κοινοποίηση δεδομένων στο νέφος.

Ένα νέο ερευνητικό άρθρο προτείνει μια νέα μέθοδο κοινοποίησης δεδομένων βασισμένη σε χαρακτηριστικά, η οποία είναι κατάλληλη για χρήστες κινητών συσκευών με περιορισμένους πόρους στο υπολογιστικό νέφος.

Η προτεινόμενη μέθοδος εξαλείφει το μεγαλύτερο μέρος του υπολογιστικού φόρτου με την προσθήκη δημόσιων παραμέτρων συστήματος, καθώς και με τη μεταφορά μέρους της κρυπτογραφικής επεξεργασίας σε εκτός σύνδεσης διαδικασίες. Επιπλέον, πραγματοποιείται ένας δημόσιος έλεγχος του κρυπτοκειμένου πριν από τη φάση αποκρυπτογράφησης, γεγονός που μειώνει σημαντικά το υπολογιστικό κόστος λόγω μη έγκυρων κρυπτοκειμένων. Για την ασφάλεια των δεδομένων, χρησιμοποιείται μια συνάρτηση κατακερματισμού τύπου Χαμαιλέον για τη δημιουργία ενός ενδιάμεσου κρυπτοκειμένου, το οποίο θα συγκαλύπτεται από τα κρυπτοκείμενα εκτός σύνδεσης για την απόκτηση του τελικού κρυπτοκειμένου. Η προτεινόμενη μέθοδος αποδεικνύεται ασφαλής έναντι προσαρμοσμένων επιθέσεων με επιλεγμένο κρυπτοκείμενο, οι οποίες θεωρούνται ως ένα ευρέως αποδεκτό πρότυπο ασφαλείας. Επίσης, παρουσιάζεται εκτενής ανάλυση απόδοσης, η οποία δείχνει ότι η προτεινόμενη μέθοδος είναι ασφαλής και αποδοτική.

Η κοινοποίηση δεδομένων σε ομάδες σε περιβάλλοντα νέφους αποτελεί επίσης μια πρόκληση και έχει γίνει ένα σημαντικό ζήτημα τα τελευταία χρόνια. Μια άλλη ερευνητική εργασία επικεντρώνεται στην ασφαλή και αποδοτική κοινοποίηση και αποθήκευση δεδομένων για μια ομάδα στο υπολογιστικό νέφος με ανώνυμο τρόπο. Μια νέα μέθοδος ιχνηλάσιμης κοινοποίησης δεδομένων ομάδας προτείνεται, αξιοποιώντας τη συμφωνία κλειδιών και την ομαδική υπογραφή για την υποστήριξη ανώνυμων πολλαπλών χρηστών σε δημόσια νέφη.

Τα μέλη της ομάδας μπορούν να επικοινωνούν ανώνυμα μέσω της ομαδικής υπογραφής, αλλά οι πραγματικές τους ταυτότητες μπορούν να εντοπιστούν αν κριθεί απαραίτητο. Μέσω της συμφωνίας κλειδιών, δημιουργείται ένα κοινό κλειδί συνεδρίου, επιτρέποντας στα μέλη να κοινοποιούν και να αποθηκεύουν τα



δεδομένα τους με ασφάλεια. Ένα σημαντικό χαρακτηριστικό αυτής της μεθόδου είναι η χρήση ενός συμμετρικού σχεδιασμού ατελούς μπλοκ για τη δημιουργία του κλειδιού, μειώνοντας σημαντικά το φορτίο των μελών στην εξαγωγή ενός κοινού κλειδιού συνεδρίου. Θεωρητικές και πειραματικές αναλύσεις δείχνουν ότι η προτεινόμενη μέθοδος είναι ασφαλής και αποδοτική για την κοινοποίηση δεδομένων ομάδας στο υπολογιστικό νέφος.

Οι κινητές συσκευές μπορούν να αποθηκεύουν και να ανακτούν προσωπικά δεδομένα από οπουδήποτε και οποτεδήποτε. Με την αυξανόμενη δημοτικότητα του υπολογιστικού νέφους, το πρόβλημα της ασφάλειας δεδομένων στο κινητό νέφος γίνεται ολοένα και πιο σοβαρό, αποτρέποντας περαιτέρω την ανάπτυξη του κινητού νέφους. Ως εκ τούτου, έχουν διεξαχθεί πολλές ερευνητικές μελέτες για τη βελτίωση της ασφάλειας του νέφους. Ωστόσο, οι περισσότερες από τις νέες προσεγγίσεις δεν είναι εφαρμόσιμες στο κινητό νέφος, καθώς οι κινητές συσκευές διαθέτουν περιορισμένους υπολογιστικούς πόρους και ισχύ. Λύσεις για αυτό το πρόβλημα είναι εξαιρετικά απαραίτητες για τις εφαρμογές του κινητού νέφους.

Σε μια νέα ερευνητική εργασία, οι συγγραφείς προτείνουν ένα ελαφρύ σχήμα κοινοποίησης δεδομένων LDSS (Lightweight Secure Data Sharing Scheme) για το κινητό υπολογιστικό νέφος. Το LDSS υιοθετεί την κρυπτογράφηση με βάση χαρακτηριστικά πολιτικής CP-ABE (Cipher Policy - Attributed Based Encryption), μια τεχνολογία ελέγχου πρόσβασης που χρησιμοποιείται σε τυπικά περιβάλλοντα νέφους, αλλά τροποποιεί τη δομή του δέντρου ελέγχου πρόσβασης ώστε να είναι κατάλληλη για περιβάλλοντα κινητού νέφους. Το LDSS μεταφέρει ένα μεγάλο μέρος των υπολογιστικά απαιτητικών διεργασιών μετασχηματισμού του δέντρου ελέγχου πρόσβασης από τις κινητές συσκευές σε εξωτερικούς διαμεσολαβητικούς διακομιστές. Επιπλέον, για τη μείωση του κόστους ανάκλησης χρηστών, το LDSS εισάγει πεδία περιγραφής χαρακτηριστικών για την υλοποίηση μιας διαδικασίας καθυστερημένης ανάκλησης, η οποία αποτελεί ένα ιδιαίτερα δύσκολο ζήτημα στα συστήματα CP-ABE που βασίζονται σε προγράμματα.



Τα πειραματικά αποτελέσματα δείχνουν ότι το LDSS μπορεί να μειώσει αποτελεσματικά το φορτίο στις κινητές συσκευές όταν οι χρήστες κοινοποιούν δεδομένα σε περιβάλλοντα κινητού νέφους.

Στο περιβάλλον του υπολογιστικού νέφους, η κοινοποίηση δεδομένων με πολλαπλούς χρήστες από διαφορετικούς τομείς έχει ωφεληθεί σημαντικά. Για να διασφαλιστεί ότι τα κοινοποιημένα αρχεία δεν θα εκτεθούν σε μη εξουσιοδοτημένους χρήστες ή παρόχους νέφους, είναι εξαιρετικά επιθυμητό να πραγματοποιείται η κοινοποίηση με έναν εξαιρετικά ασφαλή τρόπο.

Δυστυχώς, εξακολουθούν να υπάρχουν προκλήσεις που καθιστούν το ζήτημα δύσκολο να επιλυθεί. Για την αντιμετώπιση αυτών των προκλήσεων, προτείνεται ένα νέο ανώνυμο σχήμα κρυπτογράφησης μετάδοσης βασισμένο σε χαρακτηριστικά ( $A^2B^2E$ ), το οποίο διαθέτει την ιδιότητα της κρυφής πολιτικής πρόσβασης και επιτρέπει στον κάτοχο των δεδομένων να τα κοινοποιεί σε πολλαπλούς συμμετέχοντες που ανήκουν σε ένα προκαθορισμένο σύνολο παραληπτών και πληρούν την πολιτική πρόσβασης. Έχει προταθεί πρώτα ένα συγκεκριμένο σχήμα  $A^2B^2E$ , συνοδευόμενο από αυστηρή και επίσημη απόδειξη ασφάλειας χωρίς τη χρήση του μοντέλου τυχαίου χρησμού.

Στη συνέχεια, σχεδιάζουν ένα αποδοτικό και ασφαλές σύστημα κοινοποίησης δεδομένων, ενσωματώνοντας το σχήμα  $A^2B^2E$ , μια τεχνική επαληθεύσιμης αποκρυπτογράφησης μέσω αντιπροσώπευσης για κρυπτογράφηση βασισμένη σε χαρακτηριστικά, καθώς και την ιδέα της διαδικτυακής/εκτός σύνδεσης κρυπτογράφησης βασισμένης σε χαρακτηριστικά. Εκτεταμένες αναλύσεις ασφάλειας και αξιολογήσεις απόδοσης καταδεικνύουν ότι το προτεινόμενο σύστημα κοινοποίησης δεδομένων είναι ασφαλές και πρακτικό.

Είναι βολικό να κοινοποιούνται δεδομένα μεγάλης κλίμακας μεταξύ διαφόρων χρηστών στο υπολογιστικό νέφος. Ως μια μορφή κρυπτογράφησης βασισμένης σε χαρακτηριστικά, η κρυπτογράφηση με βάση πολιτική κρυπτοκειμένου CP-ABE αποτελεί μια αποδοτική τεχνική για την υλοποίηση λεπτομερούς ελέγχου πρόσβασης σε κοινοποιημένα δεδομένα. Ωστόσο, η παραδοσιακή CP-ABE δεν είναι κατάλληλη για το κινητό υπολογιστικό νέφος, όπου οι κινητοί χρήστες έχουν

περιορισμένους πόρους και η ιδιωτικότητα είναι ευάλωτη. Σε μια νέα ερευνητική εργασία, οι συγγραφείς προτείνουν ένα αποδοτικό και φιλικό προς την ιδιωτικότητα σύστημα κοινοποίησης δεδομένων βασισμένο σε χαρακτηριστικά, το οποίο υποστηρίζει τη δημιουργία κλειδιών και την κρυπτογράφηση σε λειτουργία εκτός σύνδεσης. Στο προτεινόμενο σύστημα, οι ευαίσθητες τιμές χαρακτηριστικών που καθορίζονται σε μια δομή πρόσβασης δεν αποστέλλονται ρητά μαζί με το κρυπτοκείμενο.

Ο μηχανισμός διαδικτυακής/εκτός σύνδεσης κρυπτογράφησης μειώνει το υπολογιστικό βάρος των κινητών χρηστών, εκτελώντας το μεγαλύτερο μέρος των εργασιών κρυπτογράφησης χωρίς να καταναλώνεται η μπαταρία της συσκευής. Επιπλέον, ο μηχανισμός διαδικτυακής/εκτός σύνδεσης δημιουργίας κλειδιών επιτρέπει στην αρχή χαρακτηριστικών να ολοκληρώνει το μεγαλύτερο μέρος των διαδικασιών δημιουργίας κλειδιών εκ των προτέρων, διευκολύνοντας την αποδοτική εγγραφή των κινητών χρηστών. Το προτεινόμενο σύστημα αποδεικνύεται πλήρως ασφαλές και αποδοτικό στο τυπικό μοντέλο, ενώ η ανάλυση απόδοσης καταδεικνύει την αποτελεσματικότητά του στο κινητό υπολογιστικό νέφος.

Στην πραγματικότητα, το πρόβλημα της ασφάλειας αποτελεί ένα από τα μεγαλύτερα εμπόδια στην υιοθέτηση του υπολογιστικού νέφους. Έχουν προταθεί πολλές αποδοτικές τεχνολογίες για την προστασία των δεδομένων στο νέφος. Ωστόσο, η προστασία των δεδομένων εξακολουθεί να παρουσιάζει σημαντικές προκλήσεις στην αποθήκευση και την κοινοποίηση δεδομένων στο νέφος.

Παρακάτω παρουσιάζονται ορισμένες νέες προτάσεις από διάφορους ερευνητές σε αυτόν τον τομέα.

Οι κρυπτογραφικές τεχνικές εφαρμόζονται συνήθως για την προστασία του απορρήτου των κοινοποιούμενων ευαίσθητων δεδομένων στο υπολογιστικό νέφος. Μεταξύ αυτών, η προστασία και η ανάκληση του κρυπτογραφικού κλειδιού αποτελεί θεμελιώδη πρόκληση. Για την αντιμετώπιση αυτού του ζητήματος, προτείνεται ένας νέος μηχανισμός προστασίας δεδομένων για την αποθήκευση στο νέφος.

Οι ερευνητές πάνω στο συγκεκριμένο ζήτημα αναφέρουν τις ακόλουθες ιδιότητες :

1. Το κρυπτογραφικό κλειδί προστατεύεται μέσω δύο παραγόντων. Μόνο αν λειτουργήσει ένας από τους δύο παράγοντες, διατηρείται το απόρρητο του κλειδιού.
2. Το κρυπτογραφικό κλειδί μπορεί να ανακληθεί αποδοτικά, μέσω της ενσωμάτωσης της τεχνικής επανακρυπτογράφησης μέσω αντιπροσώπου και της τεχνικής διαχωρισμού κλειδιών.
3. Τα δεδομένα προστατεύονται με λεπτομερή τρόπο μέσω της τεχνικής κρυπτογράφησης βασισμένης σε χαρακτηριστικά.

Έχουν διεξαχθεί αναλύσεις ασφάλειας και αξιολογήσεις απόδοσης, οι οποίες δείχνουν ότι η προτεινόμενη λύση είναι ασφαλής και αποδοτική. Μια εργασία προτείνει έναν λεπτομερή μηχανισμό προστασίας δεδομένων δύο παραγόντων για την αποθήκευση στο νέφος. Το μυστικό κλειδί διαχωρίζεται σε δύο μέρη (τους δύο παράγοντες), εκ των οποίων το ένα μπορεί να αποθηκευτεί σε έναν ενδεχομένως μη ασφαλή χώρο, ενώ το άλλο αποθηκεύεται σε μια συσκευή ανθεκτική σε παραβιάσεις.

Μια ενδιαφέρουσα πρόταση παρουσιάζεται από τους Langmead, Ben, και Abhinav Nellore. Στην εργασία τους, εξηγούν πώς το υπολογιστικό νέφος χρησιμοποιείται για μεγάλες συνεργασίες και έρευνες στον τομέα της γονιδιωματικής. Επιπλέον, αναλύουν πώς το υπολογιστικό νέφος πιθανότατα θα αποτελέσει βασικό θεμέλιο για μελλοντικές μεγάλες συνεργασίες στη γονιδιωματική, καθώς και προσπάθειες επανεπεξεργασίας αρχειοθετημένων δεδομένων συμπεριλαμβανομένων δεδομένων που προστατεύονται από πολιτικές ιδιωτικότητας.

Μια άλλη προσέγγιση προτείνεται μέσω της υλοποίησης ενός βιομετρικού συστήματος. Η προτεινόμενη μέθοδος παρουσιάζει μια νέα πλατφόρμα νέφους, σχεδιασμένη για την υποστήριξη βασικών διαδικτυακών εφαρμογών που κοινοποιούνται από μικρές και μεσαίες επιχειρήσεις. Η πλατφόρμα εγγυάται ασφαλή πρόσβαση πολλαπλών χρηστών και πλήρη λογικό διαχωρισμό των

υπολογιστικών και δεδομένων πόρων που σχετίζονται με διαφορετικές επιχειρήσεις.

Έχουν διασφαλίσει υψηλό επίπεδο προστασίας των δεδομένων που αποθηκεύονται στο νέφος, χρησιμοποιώντας μια ιδιαίτερη προσέγγιση κατακερματισμού δεδομένων. Η αρχιτεκτονική OpenStack χρησιμοποιείται για την κατασκευή της πλατφόρμας. Η ταυτοποίηση των χρηστών βασίζεται σε μια πρωτότυπη βιομετρική προσέγγιση, η οποία ενσωματώνει εύκολα τη χρήση δακτυλικών αποτυπωμάτων και αναγνώρισης προσώπου.

Δύο ακόμη ερευνητικές εργασίες παρουσιάζουν νέες προσεγγίσεις που συνδέουν το υπολογιστικό νέφος με τον τομέα της υγειονομικής περίθαλψης. Στην πρώτη εργασία, οι συγγραφείς προτείνουν ένα νέο σχήμα που παρέχει έναν συνδυασμένο μηχανισμό λεπτομερούς ελέγχου πρόσβασης σε δεδομένα πολλαπλών διακομιστών στο νέφος, μαζί με έναν αποδεδειγμένα ασφαλή μηχανισμό ταυτοποίησης κινητών χρηστών για τη Βιομηχανία Υγείας 4.0.

Το προτεινόμενο σχήμα είναι το πρώτο που επιδιώκει λεπτομερή έλεγχο πρόσβασης σε δεδομένα αποθηκευμένα σε πολλαπλούς διακομιστές νέφους σε περιβάλλον κινητού υπολογιστικού νέφους.

Στη δεύτερη εργασία, οι συγγραφείς ασχολούνται με την ασφάλεια και την αποδοτικότητα των συστημάτων ηλεκτρονικών ιατρικών αρχείων EHR (Electronic Health Record), τα οποία επιτρέπουν τη χρήση διαφόρων συσκευών για την παρακολούθηση της υγείας των ασθενών και την αποστολή των δεδομένων αυτών σε διακομιστές νέφους για κοινοποίηση. Παρόλα αυτά, τα συστήματα EHR αντιμετωπίζουν σημαντικές προκλήσεις όσον αφορά την ιδιωτικότητα και την αποδοτικότητα, με ένα από τα πιο κρίσιμα ζητήματα για τους ασθενείς να είναι η εμπιστευτικότητα των δεδομένων των ιατρικών τους αρχείων.

Για την αντιμετώπιση αυτών των προκλήσεων, προτείνεται ένα νέο σχήμα λεπτομερούς ελέγχου πρόσβασης στα EHR, το οποίο αποδεικνύεται ασφαλές στο τυπικό μοντέλο. Στο προτεινόμενο σχήμα, ένας κάτοχος EHR μπορεί να δημιουργήσει εκτός σύνδεσης κρυπτοκείμενα πριν καν καθοριστούν τα ιατρικά δεδομένα και οι πολιτικές πρόσβασης, μειώνοντας έτσι το υπολογιστικό κόστος.

Επιπλέον, η διαδικτυακή φάση επιτρέπει τη γρήγορη συναρμολόγηση των τελικών κρυπτοκειμένων μόλις τα δεδομένα των EHR και οι πολιτικές πρόσβασης γίνουν διαθέσιμες.

Ο Ning, Zhaolong, και οι συνεργάτες του παρουσίασαν το «Πράσινο και Βιώσιμο Νέφος των Πραγμάτων». Το άρθρο αρχικά αναλύει διάφορες ερευνητικές κατευθύνσεις, προσφέροντας μια ολοκληρωμένη κατανόηση της υπολογιστικής στην άκρη του δικτύου, υποστηριζόμενη από την ενσωμάτωση του Διαδικτύου των Πραγμάτων IoT (Internet of Things) και του υπολογιστικού νέφους. Οι συγγραφείς προτείνουν ένα πράσινο και βιώσιμο πλαίσιο εικονικής ενσωμάτωσης δικτύου για συνεργατική υπολογιστική στην άκρη του δικτύου σε ασύρματα-οπτικά ευρυζωνικά δίκτυα πρόσβασης.

Ένα ανοιχτού κώδικα λογισμικό πλαίσιο για το υπολογιστικό νέφος παρουσιάζεται ως EUCALYPTUS. Το EUCALYPTUS υλοποιείται ως σύστημα Υποδομής ως Υπηρεσία (IaaS), παρέχοντας στους χρήστες τη δυνατότητα να εκτελούν και να διαχειρίζονται ολόκληρες εικονικές μηχανές. Η εργασία περιγράφει τις βασικές αρχές σχεδιασμού του EUCALYPTUS, αναλύει σημαντικές επιχειρησιακές πτυχές του συστήματος και συζητά συμβιβασμούς στην αρχιτεκτονική του.

Τέλος, οι συγγραφείς παρέχουν στοιχεία που δείχνουν ότι το EUCALYPTUS επιτρέπει στους χρήστες που είναι εξοικειωμένοι με υπάρχοντα συστήματα Grid και Υπολογιστικής Υψηλών Επιδόσεων HPC (High Performance Computing) να εξερευνήσουν νέες λειτουργίες του υπολογιστικού νέφους, διατηρώντας παράλληλα πρόσβαση σε υπάρχοντα εργαλεία ανάπτυξης εφαρμογών και middleware Grid.

Έχουν δημοσιευτεί πολλές ερευνητικές εργασίες βασισμένες στο υπολογιστικό νέφος. Καθημερινά, οι ερευνητές εισάγουν νέες τεχνολογίες για τη βελτίωση της ασφάλειας, της εμπιστοσύνης και της ταχύτητας κοινοποίησης δεδομένων στο υπολογιστικό νέφος. Ορισμένοι συγγραφείς έχουν συνεισφέρει με άρθρα επισκόπησης σχετικά με το υπολογιστικό νέφος.

Ένα άρθρο επισκόπησης από τη Sakshi Kathuria υποστηρίζει ότι η ασφάλεια είναι υψηλότερη σε περιβάλλον πολλαπλού νέφους σε σύγκριση με ένα ενιαίο νέφος. Σε

ένα άλλο άρθρο επισκόπησης, οι συγγραφείς ανέλυσαν συνοπτικά τους παράγοντες και τις λειτουργικές πτυχές των μοντέλων εμπιστοσύνης στο υπολογιστικό νέφος. Επιπλέον, συνέκριναν διαφορετικούς τύπους μοντέλων εμπιστοσύνης, υπολογίζοντας τις τιμές εμπιστοσύνης με βάση διάφορες παραμέτρους, όπως η ακεραιότητα δεδομένων, η διαθεσιμότητα υπηρεσιών και η αποδοτικότητα επεξεργασίας δεδομένων.

Συνοπτικά, στις μέρες μας, σχεδόν κάθε τομέας χρησιμοποιεί τεχνολογίες νέφους. Ακόμη και οι άνθρωποι που δεν έχουν επίγνωση των τεχνολογιών νέφους τις χρησιμοποιούν καθημερινά. Το υπολογιστικό νέφος έχει γίνει αναπόσπαστο μέρος της ζωής μας, διευκολύνοντάς μας και αποτρέποντας τη σπατάλη πολύτιμου χρόνου.

### 3. Παραδοσιακές Αποθήκες Δεδομένων

#### 3.1 Ορισμός και χαρακτηριστικά των παραδοσιακών αποθηκών δεδομένων

Η έννοια των αποθηκών δεδομένων (Data Warehouses - DWs) [32] εισήχθη στα τέλη της δεκαετίας του 1980 από τους ερευνητές της IBM Barry Devlin και Paul Murphy, με στόχο την παροχή ενός αρχιτεκτονικού μοντέλου για την επίλυση της ροής δεδομένων προς περιβάλλοντα υποστήριξης αποφάσεων.

Σύμφωνα με τον ορισμό του Inmon, «μια αποθήκη δεδομένων είναι μια συλλογή δεδομένων που είναι προσανατολισμένη σε θέματα, αμετάβλητη, ενσωματωμένη και μεταβλητή στο χρόνο, υποστηρίζοντας τη λήψη διοικητικών αποφάσεων». Τυπικά, μια αποθήκη δεδομένων DW είναι ένα μεγάλο αποθετήριο δεδομένων, όπου τα δεδομένα μπορούν να αποθηκευτούν και να ενσωματωθούν από διάφορες πηγές με καλά δομημένο τρόπο, υποβοηθώντας τη διαδικασία λήψης αποφάσεων μέσω κατάλληλων αναλύσεων δεδομένων. Η διαδικασία συλλογής πληροφοριών σε μια αποθήκη δεδομένων ονομάζεται αποθήκευση δεδομένων (data warehousing).

Στη διαχείριση επιχειρησιακών δεδομένων, η αποθήκευση δεδομένων αναφέρεται σε ένα σύνολο συστημάτων υποστήριξης αποφάσεων, που στοχεύουν στην ενδυνάμωση του ειδικού της πληροφορίας (ηγέτη, διαχειριστή ή αναλυτή) για τη βελτίωση της λήψης αποφάσεων και τη γρηγορότερη λήψη αποφάσεων. Επομένως, τα συστήματα DW λειτουργούν ως σημαντικά εργαλεία επιχειρησιακής ευφυΐας, χρησιμοποιούμενα στη διαχείριση επιχειρησιακών δεδομένων από τις περισσότερες μεσαίες και μεγάλες οργανώσεις. Η τελευταία δεκαετία έχει γνωρίσει πρωτοφανή ανάπτυξη τόσο στον αριθμό των προϊόντων και υπηρεσιών που προσφέρονται, όσο και στην ευρεία κλίμακα υιοθέτηση αυτών των εξελίξεων από τη βιομηχανία. Σύμφωνα με λεπτομερή ερευνητική αναφορά της Market Research Future (MRFR), με τίτλο «Η αγορά υπηρεσιών Data Warehouse ως υπηρεσία ανά χρήση, ανά υλοποίηση, ανά εφαρμογή και μέγεθος οργανισμού -

πρόβλεψη έως το 2028», το μέγεθος της αγοράς θα φτάσει τα 7,69 δισεκατομμύρια δολάρια, με ετήσιο ρυθμό ανάπτυξης 24,5%, έως το 2028.

Στο πλαίσιο της αποθήκης δεδομένων, τα δεδομένα εξάγονται περιοδικά από προγράμματα που υποστηρίζουν επιχειρησιακές λειτουργίες και αντιγράφονται σε εξειδικευμένες μονάδες επεξεργασίας.

Στη συνέχεια, τα δεδομένα μπορεί να εγκριθούν, να μετατραπούν, να αναδομηθούν και να εμπλουτιστούν με εισροές από διάφορες επιλογές. Η ανεπτυγμένη αποθήκη δεδομένων γίνεται τότε η κύρια πηγή δεδομένων για την παραγωγή, ανάλυση και παρουσίαση αναφορών μέσω άμεσων αναφορών, ηλεκτρονικών πυλών και ψηφιακών απεικονίσεων. Χρησιμοποιεί την τεχνολογία OLAP (Online Analytical Processing), της οποίας η χρησιμότητα και οι απαιτήσεις εκτέλεσης διαφέρουν από τις υλοποιήσεις OLTP (Online Transaction Processing), που συνήθως υποστηρίζονται από λειτουργικές βάσεις δεδομένων. Τα προγράμματα OLTP συχνά αυτοματοποιούν τη διαχείριση διοικητικών διαδικασιών δεδομένων, όπως η καταχώρηση παραγγελιών και οι τραπεζικές συναλλαγές, που είναι απαραίτητες δραστηριότητες ενός οργανισμού. Αντίθετα, οι αποθήκες δεδομένων επικεντρώνονται κυρίως στην υποστήριξη αποφάσεων.

Οι εξελίξεις στις αποθήκες δεδομένων έχουν ωφελήσει διάφορους τομείς, όπως την παραγωγή (για αποστολή εφοδίων και εξυπηρέτηση πελατών), την επιχειρηματικότητα (για προφίλ πελατών και διαχείριση αποθεμάτων), τις χρηματοοικονομικές υπηρεσίες (για ανάλυση απαιτήσεων, αξιολόγηση κινδύνων, εξέταση τιμολόγησης και εντοπισμό απάτης), τα logistics (για διαχείριση οχημάτων), τις τηλεπικοινωνίες (για ανάλυση κλήσεων), τις εταιρείες κοινής ωφέλειας (για ανάλυση χρήσης ενέργειας) και τις ιατρικές υπηρεσίες. Ο τομέας της αποθήκευσης δεδομένων έχει γνωρίσει τεράστια έρευνα και εξελίξεις τις τελευταίες δύο δεκαετίες σε διάφορες κατηγορίες, όπως η αρχιτεκτονική των αποθηκών δεδομένων, ο σχεδιασμός αποθηκών δεδομένων και η εξέλιξη αποθηκών δεδομένων.



### 3.2 Διαφορά μεταξύ Αποθηκών Δεδομένων και Λιμνών Δεδομένων

Αν και οι αποθήκες δεδομένων (Data Warehouses) και οι λίμνες δεδομένων (Data Lakes) χρησιμοποιούνται συχνά ως εναλλάξιμοι όροι, δεν είναι το ίδιο. Μία από τις κύριες διαφορές μεταξύ τους είναι η διαφορετική δομή των δεδομένων (δηλαδή, επεξεργασμένα έναντι ακατέργαστων δεδομένων) [33], [34], [35], [36].

Μια αποθήκη δεδομένων αποθηκεύει δεδομένα σε επεξεργασμένη και φιλτραρισμένη μορφή, ενώ οι λίμνες δεδομένων αποθηκεύουν ακατέργαστα ή μη επεξεργασμένα δεδομένα. Συγκεκριμένα, τα δεδομένα επεξεργάζονται και οργανώνονται σε ένα ενιαίο σχήμα πριν εισαχθούν στην αποθήκη, ενώ τα ακατέργαστα και μη δομημένα δεδομένα διοχετεύονται σε μια λίμνη δεδομένων. Η ανάλυση πραγματοποιείται στα καθαρισμένα δεδομένα της αποθήκης. Αντιθέτως, στη λίμνη δεδομένων, τα δεδομένα επιλέγονται και οργανώνονται όταν και όπως χρειάζεται.

Όσον αφορά την αποθήκευση επεξεργασμένων δεδομένων, μια αποθήκη δεδομένων είναι οικονομική. Αντίθετα, οι λίμνες δεδομένων έχουν συγκριτικά μεγαλύτερη χωρητικότητα από τις αποθήκες δεδομένων και είναι ιδανικές για ανάλυση ακατέργαστων και μη επεξεργασμένων δεδομένων, καθώς και για την εφαρμογή μηχανικής μάθησης. Μια άλλη βασική διαφορά είναι ο σκοπός ή η χρήση. Συνήθως, τα επεξεργασμένα δεδομένα που εισέρχονται στις αποθήκες δεδομένων χρησιμοποιούνται για συγκεκριμένους σκοπούς, με αποτέλεσμα να μην σπαταλάτε ο χώρος αποθήκευσης. Αντίθετα, ο σκοπός χρήσης για τις λίμνες δεδομένων δεν είναι καθορισμένος και μπορούν ιδανικά να χρησιμοποιηθούν για οποιονδήποτε σκοπό. Για τη χρήση επεξεργασμένων ή φιλτραρισμένων δεδομένων, δεν απαιτείται εξειδικευμένη τεχνογνωσία, καθώς η εξοικείωση με την παρουσίαση δεδομένων (π.χ. διαγράμματα, φύλλα, πίνακες και παρουσιάσεις) αρκεί. Επομένως, οι αποθήκες δεδομένων μπορούν να χρησιμοποιηθούν από οποιαδήποτε επιχείρηση ή άτομο. Αντίθετα, είναι συγκριτικά δύσκολο να αναλυθούν οι λίμνες δεδομένων χωρίς εξοικείωση με τα ακατέργαστα δεδομένα,

απαιτώντας έτσι επιστήμονες δεδομένων με τις κατάλληλες δεξιότητες ή εργαλεία για την κατανόησή τους για συγκεκριμένη επιχειρησιακή χρήση.

Η προσβασιμότητα ή η ευκολία χρήσης των αποθετηρίων δεδομένων αποτελεί επίσης μια πτυχή που διαφοροποιεί τις αποθήκες δεδομένων από τις λίμνες δεδομένων. Δεδομένου ότι η αρχιτεκτονική μιας λίμνης δεδομένων δεν έχει κατάλληλη δομή, διαθέτει ευελιξία στη χρήση. Αντίθετα, η δομή μιας αποθήκης δεδομένων διασφαλίζει ότι δεν εισβάλλουν ξένα στοιχεία, ενώ η τροποποίησή της είναι πολύ δαπανηρή.

Difference between Data Warehouse and Data Lake		
Basis of Comparison	Data Lake	Vs Data Warehouse
Data Structure	Raw	Structured
Purpose of Data	Still to be determined	Currently in use
Users	Data Scientists	Business Professionals
Accessibility	Highly accessible and updated quickly	Changes are more difficult and expensive to implement

Εικόνα 2: Data Warehouse vs Data Lakes

<https://www.beyondkey.com/blog/wp-content/uploads/2022/08/Difference-between-Data-Warehouse-and-Data-Lake.png>

### 3.3 Αρχιτεκτονική Αποθήκης Δεδομένων

Η αρχιτεκτονική μιας αποθήκης δεδομένων περιλαμβάνει ιστορικά και συνδυαστικά δεδομένα από πολλαπλές πηγές. Βασικά, υπάρχουν τρεις τύποι αρχιτεκτονικών [37]:

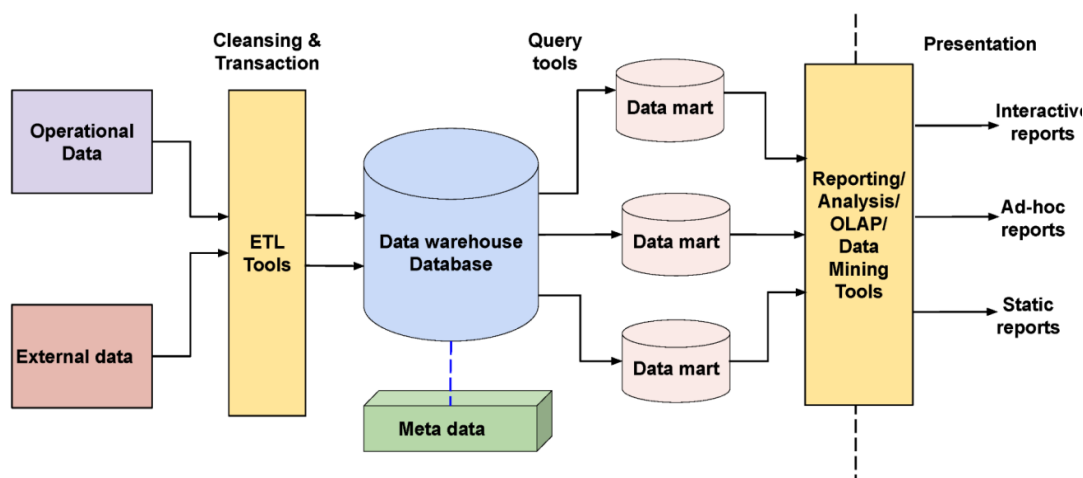
1. Αρχιτεκτονική μίας βαθμίδας: Αυτού του είδους το μονοεπίπεδο μοντέλο ελαχιστοποιεί την ποσότητα των αποθηκευμένων δεδομένων και βοηθά στην εξάλειψη της πλεοναστικής αποθήκευσης δεδομένων. Ωστόσο, το

μειονέκτημά του είναι η έλλειψη ενός στοιχείου που διαχωρίζει την αναλυτική από τη συναλλακτική επεξεργασία. Αυτή η αρχιτεκτονική δεν χρησιμοποιείται συχνά στην πράξη.

2. Αρχιτεκτονική δύο βαθμίδων: Αυτό το μοντέλο διαχωρίζει φυσικά τις διαθέσιμες πηγές δεδομένων από την αποθήκη μέσω μιας ενδιάμεσης περιοχής επεξεργασίας. Αυτή η αρχιτεκτονική διασφαλίζει ότι όλα τα δεδομένα που φορτώνονται στην αποθήκη βρίσκονται σε κατάλληλη καθαρισμένη μορφή. Ωστόσο, δεν επεκτείνεται εύκολα, ούτε μπορεί να υποστηρίξει μεγάλο αριθμό τελικών χρηστών. Επίσης, παρουσιάζει προβλήματα συνδεσιμότητας λόγω περιορισμών του δικτύου.
3. Αρχιτεκτονική τριών βαθμίδων: Αυτή είναι η πιο ευρέως χρησιμοποιούμενη αρχιτεκτονική για αποθήκες δεδομένων. Αποτελείται από τρία επίπεδα:
  - Το κάτω επίπεδο: Τα δεδομένα καθαρίζονται, μετασχηματίζονται και φορτώνονται μέσω εργαλείων back-end. Αυτό το επίπεδο λειτουργεί ως η βάση δεδομένων της αποθήκης δεδομένων.
  - Το μεσαίο επίπεδο: Είναι ένας διακομιστής OLAP που παρέχει μια αφηρημένη προβολή της βάσης δεδομένων, λειτουργώντας ως μεσολαβητής μεταξύ του τελικού χρήστη και της βάσης δεδομένων.
  - Το επάνω επίπεδο: Το επίπεδο πελάτη περιλαμβάνει εργαλεία και APIs που χρησιμοποιούνται για τη σύνδεση και την εξαγωγή δεδομένων από την αποθήκη (π.χ. εργαλεία ερωτημάτων, αναφορών, διαχείρισης ερωτημάτων, ανάλυσης και εξόρυξης δεδομένων).

Η αρχιτεκτονική μιας αποθήκης δεδομένων φαίνεται στο επόμενο σχήμα.

Περιλαμβάνει ένα κεντρικό αποθετήριο πληροφοριών, το οποίο περιβάλλεται από βασικά στοιχεία DW που καθιστούν το περιβάλλον λειτουργικό, διαχειρίσιμο και προσβάσιμο.



Εικόνα 3: Αρχιτεκτονική Αποθήκης Δεδομένων

[https://www.mdpi.com/BDCC/BDCC-06-00132/article\\_deploy/html/images/BDCC-06-00132-g002.png](https://www.mdpi.com/BDCC/BDCC-06-00132/article_deploy/html/images/BDCC-06-00132-g002.png)

Η διαδικασία ξεκινά με τη συλλογή δεδομένων από δύο κύριες πηγές: τα λειτουργικά δεδομένα (Operational Data) και τα εξωτερικά δεδομένα (External Data). Τα λειτουργικά δεδομένα περιλαμβάνουν πληροφορίες που προέρχονται από τα εσωτερικά συστήματα μιας επιχείρησης, όπως βάσεις δεδομένων συναλλαγών, εφαρμογές ERP, CRM και άλλες επιχειρησιακές εφαρμογές. Τα εξωτερικά δεδομένα αφορούν δεδομένα από εξωτερικές πηγές, όπως δημόσιες βάσεις δεδομένων, δεδομένα κοινωνικών δικτύων, πληροφορίες αγοράς ή ακόμα και δεδομένα που συλλέγονται από IoT συσκευές. Η συλλογή δεδομένων από διαφορετικές πηγές επιτρέπει την ολοκληρωμένη ανάλυση της πληροφορίας και τη βελτιστοποίηση της λήψης αποφάσεων.

Τα δεδομένα που συλλέγονται δεν είναι έτοιμα προς ανάλυση και γι' αυτόν τον λόγο υποβάλλονται σε διαδικασίες καθαρισμού και μετασχηματισμού μέσω των ETL (Extract, Transform, Load) εργαλείων. Τα εργαλεία ETL εκτελούν τρεις βασικές λειτουργίες: την εξαγωγή των δεδομένων από τις πηγές, τον μετασχηματισμό τους για να διασφαλιστεί η ομοιομορφία, η συνοχή και η ποιότητά τους, και τη φόρτωσή τους στην αποθήκη δεδομένων. Κατά τη διαδικασία μετασχηματισμού, ενδέχεται να πραγματοποιούνται λειτουργίες όπως η κανονικοποίηση δεδομένων, η αφαίρεση διπλότυπων, η συμπλήρωση ελλειπόντων τιμών και η τυποποίηση μορφών. Η χρήση ETL διαδικασιών διασφαλίζει ότι τα

δεδομένα που εισάγονται στην αποθήκη δεδομένων είναι έγκυρα, χρήσιμα και έτοιμα προς αξιοποίηση.

Ο πυρήνας της αρχιτεκτονικής αποτελείται από την βάση δεδομένων της αποθήκης δεδομένων (Data Warehouse Database), η οποία λειτουργεί ως το κεντρικό σημείο αποθήκευσης όλων των καθαρισμένων και δομημένων δεδομένων.

Η βάση δεδομένων της αποθήκης δεδομένων είναι βελτιστοποιημένη για λειτουργίες ανάγνωσης και ανάλυσης, σε αντίθεση με τις βάσεις δεδομένων συναλλαγών που είναι σχεδιασμένες για γρήγορες εγγραφές και ενημερώσεις. Για την αποτελεσματική αποθήκευση και ανάλυση μεγάλων δεδομένων, συχνά χρησιμοποιούνται τεχνολογίες όπως αποθήκες δεδομένων στήλης (columnar databases) ή υβριδικές αρχιτεκτονικές αποθήκευσης. Ένα σημαντικό τμήμα της αποθήκης δεδομένων είναι τα μεταδεδομένα (Metadata), τα οποία περιγράφουν τη δομή των δεδομένων, τις σχέσεις τους και τις πολιτικές πρόσβασης. Τα μεταδεδομένα επιτρέπουν την αποδοτική διαχείριση των δεδομένων και τη βελτιστοποίηση των ερωτημάτων που εκτελούνται στην αποθήκη δεδομένων.

Για τη βελτίωση της απόδοσης και της προσβασιμότητας των δεδομένων, η αποθήκη δεδομένων συχνά υποδιαιρείται σε Data Marts, τα οποία είναι εξειδικευμένα υποσύνολα δεδομένων που έχουν σχεδιαστεί για συγκεκριμένες επιχειρησιακές ανάγκες ή τμήματα μιας εταιρείας. Για παράδειγμα, ένα Data Mart μπορεί να επικεντρώνεται αποκλειστικά στις πωλήσεις, στα οικονομικά ή στη διαχείριση ανθρωπίνων πόρων. Η δημιουργία Data Marts βελτιώνει την απόδοση των ερωτημάτων, μειώνει τον χρόνο πρόσβασης στα δεδομένα και επιτρέπει στους αναλυτές να επικεντρωθούν σε συγκεκριμένες περιοχές ενδιαφέροντος χωρίς να επεξεργάζονται ολόκληρη την αποθήκη δεδομένων.

Η τελική φάση της διαδικασίας αφορά την ανάλυση, την εξόρυξη γνώσης και την παρουσίαση των δεδομένων μέσω των εργαλείων ανάλυσης, OLAP (Online Analytical Processing) και εξόρυξης δεδομένων (Data Mining). Τα εργαλεία αυτά επιτρέπουν τη διερεύνηση των δεδομένων, την εξαγωγή πολύτιμων πληροφοριών και τη δημιουργία προβλέψεων για μελλοντικές τάσεις. Τα OLAP εργαλεία επιτρέπουν τη γρήγορη εκτέλεση πολυδιάστατων αναλύσεων, δίνοντας τη

δυνατότητα στους χρήστες να εξετάζουν τα δεδομένα από διαφορετικές οπτικές γωνίες, όπως ανά γεωγραφική περιοχή, ανά χρονική περίοδο ή ανά κατηγορία προϊόντων.

Παράλληλα, τα εργαλεία εξόρυξης δεδομένων εφαρμόζουν αλγορίθμους μηχανικής μάθησης και τεχνικές ανάλυσης προτύπων για την ανακάλυψη χρήσιμων γνώσεων που δεν είναι άμεσα εμφανείς.

Τέλος, η παρουσίαση των δεδομένων γίνεται μέσω διαφορετικών τύπων αναφορών, ανάλογα με τις ανάγκες του οργανισμού. Οι διαδραστικές αναφορές (Interactive Reports) επιτρέπουν στους χρήστες να εξερευνούν τα δεδομένα δυναμικά, φιλτράροντας και εστιάζοντας σε συγκεκριμένες πληροφορίες.

Οι ad-hoc αναφορές επιτρέπουν στους χρήστες να δημιουργούν προσαρμοσμένες αναφορές βάσει συγκεκριμένων ερωτημάτων που προκύπτουν από την επιχειρηματική ανάλυση. Οι στατικές αναφορές παρέχουν συνοπτικά δεδομένα που παρουσιάζονται σε σταθερή μορφή και χρησιμοποιούνται κυρίως για επίσημες παρουσιάσεις και διοικητικές αποφάσεις.

### 3.3.1 Βάση δεδομένων αποθήκης δεδομένων

Η βάση δεδομένων αποτελεί το θεμέλιο της αρχιτεκτονικής μιας αποθήκης δεδομένων. Υλοποιείται με τη χρήση τεχνολογίας Σχεσιακών Συστημάτων Διαχείρισης Βάσεων Δεδομένων RDBMS (Relational DataBase Management Systems). Ωστόσο τα παραδοσιακά συστήματα RDBMS είναι βελτιστοποιημένα για συναλλακτική επεξεργασία και όχι για αποθήκες δεδομένων. Εναλλακτικές λύσεις περιλαμβάνουν:

- Χρήση παράλληλων σχεσιακών βάσεων δεδομένων που επιτρέπουν κοινή χρήση μνήμης σε διαμορφώσεις πολυεπεξεργαστών.
- Νέες δομές δεικτών για βελτίωση της ταχύτητας και αποφυγή σάρωσης πινάκων.
- Πολυδιάστατες βάσεις δεδομένων MDDBs (Multi Dimensional DataBase) για υπέρβαση των περιορισμών των σχεσιακών μοντέλων δεδομένων.

### 3.3.2 Εργαλεία Εξαγωγής, Μετασχηματισμού και Φόρτωσης (ETL)

Όλες οι μετατροπές, οι συνοψίσεις και οι αλλαγές που απαιτούνται για τη μετατροπή των δεδομένων σε ενιαία μορφή στην αποθήκη δεδομένων πραγματοποιούνται μέσω εργαλείων ETL. Η διαδικασία ETL περιλαμβάνει [39], [40], [41]:

- **Εξαγωγή:** Σύνδεση με συστήματα και συλλογή των δεδομένων για αναλυτική επεξεργασία.
- **Μετασχηματισμός:** Μετατροπή των εξαγόμενων δεδομένων σε τυποποιημένη μορφή.
- **Φόρτωση:** Εισαγωγή των μετασχηματισμένων δεδομένων στην αποθήκη δεδομένων.

Τα εργαλεία ETL εξασφαλίζουν συμμόρφωση με κανονιστικές απαιτήσεις και ανωνυμοποιούν ευαίσθητες πληροφορίες πριν τη φόρτωση. Επιπλέον, εξαλείφουν ανεπιθύμητα δεδομένα από λειτουργικές βάσεις δεδομένων και διασφαλίζουν τη διατήρηση των μεταδεδομένων.

Τα μεταδεδομένα είναι τα δεδομένα που περιγράφουν την αποθήκη δεδομένων. Παίζουν σημαντικό ρόλο στη μετατροπή των δεδομένων σε γνώση, αφού καθορίζουν την πηγή, τη χρήση, τις τιμές και τα χαρακτηριστικά των δεδομένων της αποθήκης.

Υπάρχουν δύο κατηγορίες:

- **Τεχνικά μεταδεδομένα:** Πληροφορίες που χρησιμοποιούνται από σχεδιαστές και διαχειριστές.
- **Επιχειρησιακά μεταδεδομένα:** Πληροφορίες που βοηθούν τους τελικούς χρήστες να κατανοήσουν τα δεδομένα.

### 3.3.3 Εργαλεία Ερωτημάτων

Τα εργαλεία ερωτημάτων επιτρέπουν στους χρήστες να αλληλεπιδρούν με την αποθήκη δεδομένων για τη συλλογή πληροφοριών. Υπάρχουν διάφοροι τύποι εργαλείων [42], [43], [44]:

- Εργαλεία αναφορών και ερωτημάτων: Χρησιμοποιούνται για τη δημιουργία λειτουργικών αναφορών και υποστηρίζουν μεγάλα φορτία εργασίας, όπως εκτυπώσεις και υπολογισμούς.
- Εργαλεία ανάπτυξης εφαρμογών: Επιπλέον των ενσωματωμένων εργαλείων, ικανοποιούν αναλυτικές ανάγκες οργανισμών.
- Εργαλεία εξόρυξης δεδομένων: Αυτοματοποιούν τη διαδικασία ανακάλυψης συσχετίσεων σε μεγάλες ποσότητες δεδομένων.
- Εργαλεία OLAP: Εκμεταλλεύονται πολυδιάστατες βάσεις δεδομένων για σύνθετες αναλύσεις δεδομένων.

Τύποι OLAP Εργαλείων:

1. MOLAP: Δημιουργεί έναν κύβο δεδομένων από τη σχεσιακή πηγή δεδομένων, προσφέροντας γρήγορη απόκριση σε αναφορές χρηστών.
2. ROLAP: Λειτουργεί ως έξυπνος SQL generator, όπου οι διαχειριστές ορίζουν συνδέσεις μεταξύ σχεσιακών πινάκων, χαρακτηριστικών και ιεραρχιών.

### **3.4 Σχεδιασμός Αποθήκης Δεδομένων για Επιχειρησιακές Ανάγκες**

Για να σχεδιαστεί μια επιτυχημένη αποθήκη δεδομένων, είναι απαραίτητο να κατανοηθούν οι απαιτήσεις της επιχείρησης και να αναπτυχθεί ένα κατάλληλο πλαίσιο που να τις εξυπηρετεί. Μερικά από τα βασικά κριτήρια που πρέπει να ληφθούν υπόψη κατά την επιλογή και τον σχεδιασμό μιας αποθήκης δεδομένων είναι τα εξής [45], [46]:

1. Ανάγκες χρηστών και κατάλληλο μοντέλο δεδομένων: Ο πρώτος παράγοντας σχεδιασμού είναι οι επιχειρησιακές και χρήσιμες ανάγκες. Στη φάση του σχεδιασμού, είναι απαραίτητη η ενσωμάτωση της αποθήκης δεδομένων με τις υπάρχουσες επιχειρησιακές διαδικασίες και η εξασφάλιση συμβατότητας με μακροπρόθεσμες στρατηγικές. Οι επιχειρήσεις πρέπει να κατανοήσουν ξεκάθαρα τον σκοπό της αποθήκης



δεδομένων, τις τεχνικές απαιτήσεις, τα οφέλη για τους τελικούς χρήστες, τα βελτιωμένα μέσα αναφοράς για την επιχειρηματική ευφυΐα (BI) και τις αναλύσεις. Στο πλαίσιο αυτό, η αναγνώριση του ποια πληροφορία είναι σημαντική για την επιχείρηση είναι καθοριστικής σημασίας για την επιτυχία της αποθήκης δεδομένων. Η δημιουργία ενός κατάλληλου μοντέλου δεδομένων είναι βασική πτυχή του σχεδιασμού, όπως και η χρήση διαγραμμάτων ροής δεδομένων για την απεικόνιση της ροής των δεδομένων εντός της επιχείρησης.

2. Υιοθέτηση πρότυπης αρχιτεκτονικής και μεθοδολογίας αποθήκης δεδομένων: Κατά τον σχεδιασμό μιας αποθήκης δεδομένων, είναι σημαντικό να επιλέγεται ένα αναγνωρισμένο πρότυπο μοντελοποίησης (π.χ. 3NF, αστέρι (star schema), Data Vault). Η διατήρηση μιας συγκεκριμένης αρχιτεκτονικής ενισχύει την αποτελεσματικότητα της ανάπτυξης της αποθήκης δεδομένων. Παράλληλα, μια ευέλικτη μεθοδολογία αποθήκης δεδομένων είναι ζωτικής σημασίας.

Με σωστό προγραμματισμό, τα έργα αποθήκης δεδομένων μπορούν να χωριστούν σε μικρότερα τμήματα που παραδίδονται γρηγορότερα, προσαρμοσμένα στις μεταβαλλόμενες ανάγκες της επιχείρησης.

3. Αποθήκευση στο cloud έναντι τοπικής αποθήκευσης: Οι επιχειρήσεις μπορούν να επιλέξουν είτε τοπική αρχιτεκτονική είτε αποθήκη δεδομένων στο cloud. Η τοπική αρχιτεκτονική απαιτεί εγκατάσταση φυσικού εξοπλισμού, συμπεριλαμβανομένων διακομιστών για τις διαδικασίες ETL, την αποθήκευση και τις αναλυτικές λειτουργίες.

Αντίθετα, η αποθήκευση στο cloud αποφεύγει αυτή τη διαδικασία. Παρόλα αυτά, υπάρχουν περιπτώσεις όπου η τοπική αποθήκευση έχει νόημα, όπως όταν οι περισσότερες κρίσιμες βάσεις δεδομένων είναι παλιές και δεν συνεργάζονται καλά με λύσεις cloud ή όταν ισχύουν αυστηρές κανονιστικές απαιτήσεις που απαγορεύουν την αποθήκευση δεδομένων εκτός χώρας.

4. Οικοσύστημα εργαλείων δεδομένων και μοντελοποίηση δεδομένων: Το οικοσύστημα της επιχείρησης παίζει καθοριστικό ρόλο. Η υιοθέτηση εργαλείων αυτοματοποίησης αποθήκης δεδομένων διασφαλίζει την αποδοτική χρήση των πόρων πληροφορικής, την ταχύτερη υλοποίηση έργων και την υποστήριξη μέσω τυποποιημένων κωδικών. Η φάση σχεδιασμού της μοντελοποίησης δεδομένων παρέχει λεπτομερή και επαναχρησιμοποιήσιμη τεκμηρίωση για την υλοποίηση μιας αποθήκης δεδομένων. Αυτή περιλαμβάνει την αξιολόγηση των δομών δεδομένων, την ανάλυση των απαιτήσεων OLAP και άλλα.
5. Σχεδιασμός ETL ή ELT: Η επιλογή της κατάλληλης λύσης ETL ή ELT είναι άλλος ένας κρίσιμος παράγοντας σχεδιασμού. Στις επιχειρήσεις που χρησιμοποιούν ακριβά εσωτερικά αναλυτικά συστήματα, η προετοιμασία των δεδομένων, συμπεριλαμβανομένων των μετασχηματισμών, μπορεί να γίνει κατά την επεξεργασία ETL. Ωστόσο, το ELT είναι καλύτερη προσέγγιση για αποθήκες δεδομένων στο cloud, όπου οι ενσωματώσεις και οι μετασχηματισμοί πραγματοποιούνται μέσω της ισχύος ενός κεντρικού μηχανισμού στο cloud.
6. Στρώσεις σημασιολογίας και αναφορών: Με βάση τα τεκμηριωμένα μοντέλα δεδομένων, ο διακομιστής OLAP εφαρμόζεται για να διευκολύνει τις αναλυτικές ερωτήσεις των χρηστών και να ενισχύσει τα συστήματα BI. Οι μηχανικοί δεδομένων πρέπει να λάβουν υπόψη τους τον χρόνο για ανάλυση και τις απαιτήσεις καθυστέρησης για την αξιολόγηση των δυνατοτήτων αναλυτικής επεξεργασίας της αποθήκης δεδομένων. Κατά τον σχεδιασμό της στρώσης αναφορών, είναι απαραίτητο να καθοριστούν οι μέθοδοι αναφοράς και τα επίπεδα πρόσβασης.
7. Ευκολία επεκτασιμότητας: Η κατανόηση των τρεχουσών αναγκών της επιχείρησης είναι κρίσιμη για την επιχειρηματική ευφυΐα και τη λήψη αποφάσεων. Αυτό περιλαμβάνει την ποσότητα των δεδομένων που διαθέτει η επιχείρηση και το πόσο γρήγορα οι ανάγκες της μπορεί να αυξηθούν. Οι

απαιτήσεις σε προσωπικό και τα κόστη προμηθευτών πρέπει επίσης να ληφθούν υπόψη κατά την απόφαση για την κλίμακα ανάπτυξης.

### 3.5 Δημοφιλή Εργαλεία και Υπηρεσίες Αποθήκευσης Δεδομένων

Μια επιχειρησιακή αποθήκη δεδομένων αποτελεί ένα από τα βασικά στοιχεία της επιχειρηματικής ευφυΐας. Αποθηκεύει δεδομένα από μία ή περισσότερες ετερογενείς πηγές, τα αναλύει και εξάγει χρήσιμες πληροφορίες για τη λήψη αποφάσεων. Ορισμένα από τα πιο δημοφιλή εργαλεία αποθήκευσης δεδομένων περιγράφονται παρακάτω [47], [48], [49]:

1. Εργαλεία αποθήκευσης δεδομένων της Amazon Web Services (AWS): Η AWS είναι ένας από τους κορυφαίους παρόχους λύσεων αποθήκευσης δεδομένων. Παρέχει πολλές υπηρεσίες, όπως το AWS Redshift, το AWS S3 και το Amazon RDS, που την καθιστούν μια οικονομικά αποδοτική και εξαιρετικά επεκτάσιμη πλατφόρμα. Το AWS Redshift είναι κατάλληλο για επιχειρήσεις που χρειάζονται προηγμένες δυνατότητες, χρησιμοποιώντας εξειδικευμένα εργαλεία.

Το Amazon Simple Storage Service (AWS S3) προσφέρει χαμηλού κόστους λύση αποθήκευσης με κορυφαία χαρακτηριστικά επεκτασιμότητας, απόδοσης και ασφάλειας. Το Amazon Relational Database Service (Amazon RDS) διευκολύνει την κλιμάκωση σχεσιακών βάσεων δεδομένων, παρέχοντας οικονομική και εύελικτη τεχνολογία διαχείρισης.

2. Εργαλεία αποθήκευσης δεδομένων της Google: Η Google φημίζεται για τις ικανότητές της στη διαχείριση δεδομένων και τις λύσεις αποθήκευσης cloud. Το Google BigQuery είναι μια πλατφόρμα επιχειρησιακής αποθήκευσης δεδομένων που επιτρέπει την αποθήκευση και αναζήτηση μεγάλων συνόλων δεδομένων με ταχύτητα SQL queries, προσφέροντας πραγματική γνώση δεδομένων σε πραγματικό χρόνο. Το Google Cloud Data Fusion είναι μια διαχειριζόμενη λύση ETL που διευκολύνει την ενοποίηση δεδομένων μέσω μιας οπτικής διεπαφής point-and-click. Το Google Data Studio επιτρέπει τη μετατροπή δεδομένων σε εξατομικευμένες αναφορές και dashboards.

3. Εργαλεία αποθήκευσης δεδομένων της Microsoft Azure: Η Microsoft Azure προσφέρει λύσεις υποδομής (IaaS), πλατφόρμας (PaaS) και λογισμικού ως υπηρεσία (SaaS). Η βάση δεδομένων Azure SQL Database υποστηρίζει εφαρμογές αποθήκευσης δεδομένων με μεγάλο όγκο δεδομένων και πολλούς ενεργούς χρήστες. Το Azure Synapse Analytics συνδυάζει δυνατότητες ενσωμάτωσης δεδομένων, ανάλυσης μεγάλων δεδομένων και επιχειρησιακής αποθήκευσης, ενσωματώνοντας τεχνολογίες μηχανικής μάθησης.

4. Oracle Autonomous Data Warehouse: Το Oracle Autonomous Data Warehouse είναι μια υπηρεσία αποθήκευσης δεδομένων στο cloud που αυτοματοποιεί τη διαχείριση της αποθήκης, την προστασία των δεδομένων και την ανάπτυξη εφαρμογών. Είναι εύκολο στη χρήση, ασφαλές, γρήγορο και επεκτάσιμο.

5. Snowflake: Το Snowflake είναι ένα εργαλείο αποθήκευσης δεδομένων στο cloud που προσφέρει γρήγορη, εύχρηστη και ευέλικτη πλατφόρμα αποθήκευσης. Χρησιμοποιεί μια ολοκληρωμένη αρχιτεκτονική SaaS, διευκολύνοντας την επεξεργασία δεδομένων με χρήση μιας ενιαίας γλώσσας, της SQL, για συνδυασμό, ανάλυση και μετασχηματισμούς δεδομένων. Η πολλαπλών εννοιών αρχιτεκτονική του Snowflake επιτρέπει την ανταλλαγή δεδομένων σε πραγματικό χρόνο χωρίς την ανάγκη μετακίνησης δεδομένων.

6. Εργαλεία αποθήκευσης δεδομένων της IBM: Η IBM είναι ιδιαίτερα δημοφιλής στις μεγάλες επιχειρήσεις λόγω των κάθετων μοντέλων δεδομένων, των λύσεων διαχείρισης και των δυνατοτήτων ανάλυσης σε πραγματικό χρόνο. Το IBM DB2 Warehouse είναι μια λύση cloud που επιτρέπει αυτοματοποιημένη κλιμάκωση και ευελιξία ανάπτυξης.

Το IBM Datastage επιτρέπει τη λήψη δεδομένων από διάφορα συστήματα, τη μετατροπή τους και τη μεταφορά τους σε στοχευμένα συστήματα, διευκολύνοντας την ενοποίηση δεδομένων μέσω αρχιτεκτονικής που μπορεί να είναι τοπική ή βασισμένη στο cloud.

### 3.6 Προκλήσεις Υλοποίησης Αποθήκης Δεδομένων

Η υλοποίηση μιας αποθήκης δεδομένων απαιτεί σωστό σχεδιασμό και εκτέλεση με βάση κατάλληλες μεθόδους. Ορισμένες από τις σημαντικότερες προκλήσεις που προκύπτουν κατά την υλοποίηση περιλαμβάνουν το σχεδιασμό, την κατασκευή και την εφαρμογή [50], [51], [52], [53], [54], [55], [56].

1. **Ποιότητα και συνέπεια δεδομένων:** Η αποδοτικότητα και η λειτουργικότητα μιας αποθήκης δεδομένων εξαρτώνται από την ποιότητα των δεδομένων που υποστηρίζουν τις λειτουργίες της. Λανθασμένα ή πλεοναστικά δεδομένα καθιστούν δύσκολη τη μέτρηση του πραγματικού κόστους από τους διαχειριστές της αποθήκης. Μια βασική λύση είναι η αυτοματοποίηση του συστήματος για τη βελτίωση της ποιότητας των δεδομένων και η διασφάλιση ότι η ομάδα πωλήσεων λαμβάνει πλήρεις, σωστές και συνεπείς πληροφορίες. Ο έλεγχος ποιότητας των δεδομένων (δηλαδή η ποιότητα και η συνέπεια των δεδομένων) είναι μια από τις κύριες ανησυχίες που πρέπει να αντιμετωπιστούν. Η διαδικασία επιχειρηματικής ευφυΐας μπορεί να βελτιωθεί με την ενσωμάτωση ευελιξίας για την αποδοχή αναλύσεων, καθώς και την αναβάθμιση του σχήματος της αποθήκης για την αντιμετώπιση μελλοντικών αλλαγών.
2. **Ετερογενείς πηγές δεδομένων:** Ένα άλλο βασικό πρόβλημα είναι οι διαφορές στις ονοματολογίες, τους ορισμούς τομέων και τους αριθμούς ταυτοποίησης από ετερογενείς πηγές. Η αποθήκη δεδομένων πρέπει να σχεδιαστεί ώστε να μπορεί να προσαρμόζεται στην προσθήκη ή αφαίρεση πηγών δεδομένων και να διαχειρίζεται την εξέλιξη των δεδομένων, αποφεύγοντας μεγάλες ανασχεδιάσεις. Επιπλέον, η προσαρμογή των δεδομένων των διαθέσιμων πηγών στο μοντέλο δεδομένων της αποθήκης αποτελεί σημαντική πρόκληση, καθώς οι δυνατότητες μιας αποθήκης δεδομένων ενδέχεται να αλλάξουν με την πάροδο του χρόνου, λόγω των τεχνολογικών εξελίξεων.

3. **Διαχείριση και δεξιότητες:** Η διαχείριση μιας αποθήκης δεδομένων σε μεγάλες οργανώσεις απαιτεί ευρύτερες δεξιότητες από αυτές που απαιτούνται στη διαχείριση παραδοσιακών βάσεων δεδομένων. Ο σχεδιασμός της λειτουργίας διαχείρισης, η επιλογή της ομάδας διαχείρισης και η γενικότερη διαχείριση της αποθήκης δεδομένων είναι σημαντικές πτυχές που πρέπει να ληφθούν υπόψη.
4. **Ασφάλεια δεδομένων:** Η ασφάλεια δεδομένων αποτελεί κρίσιμη προϋπόθεση, καθώς τα επιχειρησιακά δεδομένα είναι εξαιρετικά ευαίσθητα και μπορεί να διαρρεύσουν εύκολα. Το τυπικό μοντέλο ασφάλειας που βασίζεται σε πίνακες, γραμμές και χαρακτηριστικά είναι ασύμβατο με τις αποθήκες δεδομένων. Το μοντέλο πρέπει να προσαρμοστεί ώστε να ενσωματωθεί πλήρως με το σχετικό πολυδιάστατο μοντέλο που επικεντρώνεται σε έννοιες όπως τα γεγονότα, οι διαστάσεις και τα μέτρα. Η ασφάλεια πληροφοριών πρέπει να λαμβάνεται υπόψη σε όλα τα στάδια της διαδικασίας ανάπτυξης, από την ανάλυση των απαιτήσεων έως την υλοποίηση και τη συντήρηση.
5. **Διακυβέρνηση αποθήκης δεδομένων:** Ένας άλλος σημαντικός παράγοντας είναι η διακυβέρνηση της αποθήκης δεδομένων, που περιλαμβάνει την έγκριση προτύπων μοντελοποίησης και μεταδεδομένων, το σχεδιασμό πολιτικής πρόσβασης στα δεδομένα και τη στρατηγική δημιουργίας αντιγράφων ασφαλείας. Αυτές οι διαδικασίες συμβάλλουν στη συνολική οργάνωση και προστασία της αποθήκης δεδομένων.

### 3.7 Αποθήκες Δεδομένων: Ευκαιρίες και Μελλοντικές Κατευθύνσεις

Το τοπίο της επιχειρηματικής διαχείρισης έχει αλλάξει ριζικά με την εμφάνιση των αποθηκών δεδομένων. Οι εξελίξεις στην τεχνολογία cloud, το Διαδίκτυο των Πραγμάτων (IoT) και τις αναλύσεις μεγάλων δεδομένων έχουν φέρει

αποτελεσματικές λύσεις δεδομένων στις σύγχρονες αποθήκες δεδομένων [57], [58], [59], [60], [61].

Με την ταχεία εξέλιξη της τεχνολογίας, πολλές επιχειρήσεις έχουν μεταφέρει τα δεδομένα τους στο cloud για να επεκτείνουν τα δίκτυα και τις αγορές τους. Οι αποθήκες δεδομένων στο cloud συμβάλλουν στην αντιμετώπιση του υψηλού κόστους αγοράς, υποδομής και εγκατάστασης. Τα επόμενα χρόνια προβλέπεται η ανάπτυξη πιο προηγμένων τεχνολογιών στις αποθήκες δεδομένων cloud, με στόχο τη δημιουργία ισχυρών, εύχρηστων και οικονομικών υπηρεσιών δεδομένων στο cloud.

Τα μακροπρόθεσμα οφέλη από την υιοθέτηση αποθηκών δεδομένων στο cloud περιλαμβάνουν κυρίως τη διαθεσιμότητα και την επεκτασιμότητα των δεδομένων. Η ευελιξία αποθήκευσης ποικίλων μορφών δεδομένων, σε συνδυασμό με την εγγενή ευελιξία των υπηρεσιών cloud, επιτρέπει μια ευρεία κατανομή υπηρεσιών cloud.

### 3.7.1 Νέες Μέθοδοι Αναλύσεων Δεδομένων

Μια άλλη μεγάλη αλλαγή είναι η πρόοδος στον τρόπο ανάλυσης δεδομένων. Σε αντίθεση με το παρελθόν, όπου οι αναλύσεις δεδομένων και η επιχειρηματική ευφυΐα ανήκαν σε ξεχωριστές διακριτές ενότητες, προκαλώντας καθυστερήσεις στην αποδοτικότητα, οι σύγχρονες αποθήκες δεδομένων προσφέρουν προηγμένη δομή αποθήκευσης και ταχύτερη ροή δεδομένων, καθιστώντας τα εύκολα προσβάσιμα στους επιχειρηματικούς χρήστες [62], [63].

Ένα τέτοιο μοντέλο ευελιξίας επιτρέπει την ανάλυση και την πρόσβαση σε δεδομένα σε πραγματικό χρόνο (Real Time) σε ολόκληρη την επιχείρηση, βελτιώνοντας σημαντικά την αποτελεσματικότητα.

### 3.7.2 Διαδίκτυο των Πραγμάτων (IoT)

Μια άλλη σημαντική πρόοδος είναι οι πλατφόρμες του Διαδικτύου των Πραγμάτων (IoT) για την κοινή χρήση και αποθήκευση δεδομένων. Αυτό έχει αλλάξει δραματικά τη διαχείριση δεδομένων ροής, επιτρέποντας στους χρήστες να αποθηκεύουν και να έχουν πρόσβαση σε δεδομένα από πολλαπλές συσκευές. Η

έννοια του IoT είναι ιδιαίτερα σχετική με τον πραγματικό κόσμο, λόγω της αυξανόμενης δημοτικότητας των κινητών συσκευών, των ενσωματωμένων τεχνολογιών επικοινωνίας, της υπολογιστικής στο cloud και των αναλύσεων δεδομένων. Όπως και με το Διαδίκτυο, το IoT επιτρέπει τη συνύπαρξη συσκευών σε πολλαπλά περιβάλλοντα και διευκολύνει εφαρμογές από τις πιο απλές μέχρι τις πιο κρίσιμες. Τεχνολογίες όπως η υπολογιστική ευφυΐα και τα μεγάλα δεδομένα μπορούν να συνδυαστούν με το IoT για τη βελτίωση της διαχείρισης δεδομένων και της ανακάλυψης γνώσης σε μεγάλη κλίμακα [64].

### **3.7.3 Μελλοντικές Δυνατότητες Αποθηκών Δεδομένων**

Συνοψίζοντας, το μέλλον των αποθηκών δεδομένων περιλαμβάνει χαρακτηριστικά που επιτρέπουν τα εξής [65], [66], [67], [68], [69], [70], [71], [72], [73], [74], [75]:

- Πρόσβαση σε όλα τα δεδομένα από μία κεντρική τοποθεσία.
- Δυνατότητα ανάθεσης της διατήρησης υψηλής διαθεσιμότητας της υπηρεσίας σε όλους τους πελάτες.
- Διακυβέρνηση βασισμένη σε πολιτικές.
- Πλατφόρμες με υψηλή εμπειρία χρήστη (UX) και εύκολη ανακάλυψη.
- Πλατφόρμες που καλύπτουν τις ανάγκες όλων των πελατών.



## **4. Αποθήκες Δεδομένων στο υπολογιστικό νέφος**

### **4.1 Ορισμός και σημασία της αποθήκης δεδομένων στο υπολογιστικό νέφος**

Μία αποθήκη δεδομένων στο υπολογιστικό νέφος (Cloud Data Warehouse) [1], [6], αποτελεί μια σύγχρονη προσέγγιση στη διαχείριση και ανάλυση δεδομένων, η οποία βασίζεται σε διαδικτυακή υποδομή. Πρόκειται για μια κεντρική βάση δεδομένων που φιλοξενείται σε δημόσιο υπολογιστικό νέφος, επιτρέποντας τη συλλογή, αποθήκευση, επεξεργασία, ενοποίηση και διαχείριση μεγάλων όγκων δεδομένων, τόσο δομημένων όσο και ημιδομημένων. Το κύριο χαρακτηριστικό που διακρίνει τις αποθήκες δεδομένων στο υπολογιστικό νέφος από τις παραδοσιακές λύσεις είναι η αποδέσμευση από τη φυσική διαχείριση εξυπηρετητών και υποδομών. Όλες οι λειτουργίες πραγματοποιούνται σε απομακρυσμένους διακομιστές μέσω διαδικτύου, επιτρέποντας στις επιχειρήσεις να έχουν πρόσβαση στα δεδομένα και στα εργαλεία ανάλυσης χωρίς την ανάγκη εγκατάστασης και συντήρησης εξειδικευμένου λογισμικού ή εξοπλισμού.

Η χρήση μιας αποθήκης δεδομένων στο υπολογιστικό νέφος είναι κρίσιμη για τη λήψη ταχύτατων, τεκμηριωμένων αποφάσεων που βασίζονται σε δεδομένα. Μέσω της αυξημένης υπολογιστικής ισχύος και της απλοποιημένης διαχείρισης δεδομένων, οι επιχειρήσεις μπορούν να εξάγουν πολύτιμες πληροφορίες από δεδομένα που είναι επικαιροποιημένα, ακριβή και εμπλουτισμένα.

### **4.2 Βασικά χαρακτηριστικά της αποθήκης δεδομένων στο υπολογιστικό νέφος**

Οι αποθήκες δεδομένων στο υπολογιστικό νέφος διαθέτουν μια σειρά από χαρακτηριστικά που τις καθιστούν εξαιρετικά χρήσιμες για επιχειρήσεις που επιδιώκουν να αξιοποιήσουν τις δυνατότητες του νέφους [6], [76]. Η ισορροπία

μεταξύ ασφάλειας, κλιμάκωσης και προσβασιμότητας είναι καθοριστική για την επιτυχία αυτής της τεχνολογίας.

Η απόδοση αποτελεί ένα από τα σημαντικότερα πλεονεκτήματα της αποθήκης δεδομένων στο νέφος, καθώς έχει σχεδιαστεί για να επιτρέπει την εκτέλεση πολύπλοκων ερωτημάτων σε μεγάλα σύνολα δεδομένων με υψηλή ταχύτητα και αποτελεσματικότητα.

Οι επιχειρήσεις μπορούν να επεξεργάζονται μεγάλους όγκους πληροφοριών σχεδόν σε πραγματικό χρόνο, διευκολύνοντας τη λήψη στρατηγικών αποφάσεων και την εξαγωγή κρίσιμων επιχειρηματικών συμπερασμάτων.

Η ενσωμάτωση με διάφορα εργαλεία επιχειρηματικής ευφυΐας BI και αναλυτικής δεδομένων είναι ένα άλλο βασικό χαρακτηριστικό, καθώς επιτρέπει τη συγκέντρωση και ανάλυση δεδομένων από πολλαπλές πηγές. Αυτή η δυνατότητα διασφαλίζει ότι οι επιχειρήσεις μπορούν να δημιουργούν ολοκληρωμένες και λεπτομερείς αναφορές, αποκτώντας μια σαφή εικόνα της απόδοσής τους και των πιθανών ευκαιριών βελτίωσης.

Η ασφάλεια είναι πρωταρχικής σημασίας σε μια αποθήκη δεδομένων στο νέφος, με προηγμένα μέτρα προστασίας, όπως κρυπτογράφηση δεδομένων, αυστηρός έλεγχος πρόσβασης και συμμόρφωση με διεθνή πρότυπα προστασίας δεδομένων. Αυτές οι πρακτικές συμβάλλουν στην αποτροπή μη εξουσιοδοτημένης πρόσβασης και στη διασφάλιση της εμπιστευτικότητας και ακεραιότητας των δεδομένων.

Η διαχείριση κόστους (Cost Management) είναι ένα άλλο πλεονέκτημα, καθώς το μοντέλο πληρωμής ανάλογα με τη χρήση (pay-as-you-go) επιτρέπει στις επιχειρήσεις να προσαρμόζουν τις δαπάνες τους σύμφωνα με τις ανάγκες τους, εξαλείφοντας τις περιττές επενδύσεις σε υλικό εξοπλισμό και υποδομές. Αυτό καθιστά τις αποθήκες δεδομένων στο νέφος ιδιαίτερα ελκυστικές για εταιρείες κάθε μεγέθους, επιτρέποντας τους να επεκτείνουν ή να μειώσουν τις λειτουργίες τους ανάλογα με τις επιχειρηματικές απαιτήσεις.

Η κλιμακωσιμότητα επιτρέπει στις αποθήκες δεδομένων στο νέφος να προσαρμόζονται εύκολα στις αυξανόμενες απαιτήσεις σε όγκο δεδομένων και

υπολογιστική ισχύ. Οι επιχειρήσεις μπορούν να διαχειρίζονται αυξημένα δεδομένα χωρίς να ανησυχούν για ζητήματα απόδοσης ή περιορισμούς στην αποθηκευτική ικανότητα, εξασφαλίζοντας έτσι την ομαλή λειτουργία των πληροφοριακών τους συστημάτων.

Η προσβασιμότητα αποτελεί έναν ακόμα κρίσιμο παράγοντα, επιτρέποντας στους χρήστες να έχουν πρόσβαση στα δεδομένα τους από οποιοδήποτε σημείο με σύνδεση στο διαδίκτυο.

Αυτή η ευελιξία διευκολύνει τη συνεργασία μεταξύ ομάδων και την ανταλλαγή πληροφοριών σε πραγματικό χρόνο, ενισχύοντας την παραγωγικότητα και την αποδοτικότητα των επιχειρηματικών διαδικασιών.

Οι αυτόματες ενημερώσεις που παρέχονται από τους παρόχους αποθηκών δεδομένων στο υπολογιστικό νέφος διασφαλίζουν ότι οι χρήστες διαθέτουν πάντα τις πιο πρόσφατες λειτουργίες και τα νεότερα μέτρα ασφαλείας, εξαλείφοντας την ανάγκη για χειροκίνητες αναβαθμίσεις και συντήρηση.

### **4.3 Σύγκριση παραδοσιακής και νεφοϋπολογιστικής αποθήκης δεδομένων.**

Η επιλογή μεταξύ μιας παραδοσιακής αποθήκης δεδομένων (On-Premises Data Warehouse) και μιας αποθήκης δεδομένων στο υπολογιστικό νέφος (Cloud Data Warehouse) αποτελεί κρίσιμη απόφαση για τις επιχειρήσεις. Η σύγκριση των δύο λύσεων αποκαλύπτει ουσιώδεις διαφορές που επηρεάζουν την απόδοση, το κόστος και την ευελιξία της επιχείρησης [77], [78].

Η ανάπτυξη μιας παραδοσιακής αποθήκης δεδομένων απαιτεί την εγκατάσταση και διαχείριση φυσικών διακομιστών, γεγονός που αυξάνει την πολυπλοκότητα και το κόστος. Αντίθετα, οι αποθήκες δεδομένων στο υπολογιστικό νέφος βασίζονται σε εικονικούς διακομιστές μέσω διαδικτύου, μειώνοντας την ανάγκη για φυσική υποδομή και επιτρέποντας την απομακρυσμένη διαχείριση των δεδομένων.

Η επεκτασιμότητα αποτελεί ένα άλλο κρίσιμο σημείο διαφοροποίησης. Οι παραδοσιακές λύσεις έχουν περιορισμένη δυνατότητα επέκτασης, απαιτώντας πρόσθετες επενδύσεις σε εξοπλισμό. Από την άλλη, οι αποθήκες δεδομένων στο νέφος είναι ιδιαίτερα ευέλικτες, επιτρέποντας την προσαρμογή των πόρων βάσει των εκάστοτε αναγκών.

Όσον αφορά τη συντήρηση, οι on-premises λύσεις απαιτούν εξειδικευμένη διαχείριση από εσωτερικές ομάδες πληροφορικής, γεγονός που συνεπάγεται υψηλότερο λειτουργικό κόστος. Οι αποθήκες δεδομένων στο υπολογιστικό νέφος, αντίθετα, προσφέρουν διαχειριζόμενες υπηρεσίες, μειώνοντας σημαντικά το βάρος της συντήρησης.

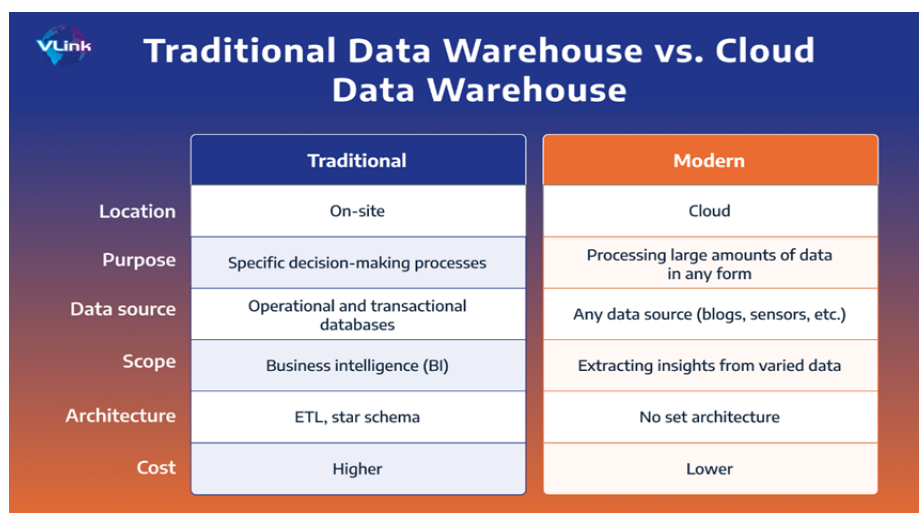
Η δομή κόστους διαφέρει επίσης σημαντικά. Οι παραδοσιακές αποθήκες δεδομένων συνεπάγονται αρχική δαπάνη κεφαλαίου για αγορά εξοπλισμού και υποδομών, ενώ οι λύσεις νέφους ακολουθούν μοντέλο λειτουργικών εξόδων με πληρωμή ανάλογα με τη χρήση, παρέχοντας μεγαλύτερη οικονομική ευελιξία.

Στον τομέα της ενσωμάτωσης και της προσβασιμότητας, οι on-premises αποθήκες δεδομένων έχουν περιορισμένη διασύνδεση με υπηρεσίες νέφους και είναι συνδεδεμένες με φυσικές τοποθεσίες, ενώ οι cloud αποθήκες δεδομένων επιτρέπουν την πρόσβαση από οπουδήποτε υπάρχει σύνδεση στο διαδίκτυο και υποστηρίζουν απρόσκοπτη ενσωμάτωση με σύγχρονες πλατφόρμες ανάλυσης δεδομένων.

Η ταχύτητα ανάπτυξης αποτελεί ακόμη ένα κρίσιμο πλεονέκτημα των λύσεων νέφους, καθώς επιτρέπουν την άμεση κλιμάκωση και την ταχύτερη υλοποίηση σε σχέση με τις παραδοσιακές εγκαταστάσεις που απαιτούν μακροχρόνιες διαδικασίες προμηθειών και ρυθμίσεων. Παράλληλα, οι αποθήκες δεδομένων στο υπολογιστικό νέφος επωφελούνται από αυτόματες ενημερώσεις και βελτιώσεις ασφάλειας, μειώνοντας τον κίνδυνο παρωχημένων συστημάτων και απρογραμμάτιστων διακοπών λειτουργίας.

Τέλος, στον τομέα της αποκατάστασης καταστροφών, οι παραδοσιακές αποθήκες δεδομένων βασίζονται σε εσωτερικές λύσεις δημιουργίας αντιγράφων ασφαλείας,

ενώ οι cloud λύσεις προσφέρουν ενσωματωμένες επιλογές αποκατάστασης, αυξάνοντας την ασφάλεια και τη διαθεσιμότητα των δεδομένων.



	Traditional	Modern
Location	On-site	Cloud
Purpose	Specific decision-making processes	Processing large amounts of data in any form
Data source	Operational and transactional databases	Any data source (blogs, sensors, etc.)
Scope	Business intelligence (BI)	Extracting insights from varied data
Architecture	ETL, star schema	No set architecture
Cost	Higher	Lower

Εικόνα 4: On promises Data Warehouse vs Cloud Data Warehouse

[https://backend.vlinkinfo.com/uploads/warehouse\\_solutions\\_img\\_5403ec9a73.png](https://backend.vlinkinfo.com/uploads/warehouse_solutions_img_5403ec9a73.png)

#### 4.4 Αρχιτεκτονική αποθήκης δεδομένων στο υπολογιστικό νέφος

Η αρχιτεκτονική μιας αποθήκης δεδομένων στο υπολογιστικό νέφος [76], [79], [80], αναφέρεται στον σχεδιασμό και την οργάνωση των στοιχείων που τη συνθέτουν, διασφαλίζοντας την αποδοτική επεξεργασία, αποθήκευση, ενσωμάτωση και ανάκτηση των δεδομένων. Οι αποθήκες δεδομένων στο νέφος βασίζονται σε ένα δομημένο σύστημα διαχείρισης, το οποίο περιλαμβάνει διαφορετικά επίπεδα επεξεργασίας και διαχείρισης δεδομένων, που συνεργάζονται αρμονικά για να παρέχουν υψηλή απόδοση και αξιοπιστία.

Οι πηγές δεδομένων (Data Sources) αποτελούν το πρώτο και βασικό στοιχείο της αρχιτεκτονικής, περιλαμβάνοντας διάφορες πηγές δεδομένων, όπως συναλλακτικές βάσεις δεδομένων, δεδομένα ροής, αρχεία καταγραφής, αισθητήρες IoT και εξωτερικά APIs.

Η μεγάλη δύναμη των αποθηκών δεδομένων στο υπολογιστικό νέφος έγκειται στην ικανότητά τους να χειρίζονται ποικίλες μορφές δεδομένων, είτε αυτά είναι δομημένα, ημιδομημένα ή αδόμητα, επιτρέποντας στις επιχειρήσεις να συλλέγουν και να ενοποιούν δεδομένα από διαφορετικές πηγές.

Το στρώμα εισαγωγής δεδομένων (Data Ingestion Layer) είναι υπεύθυνο για τη συλλογή, μετατροπή και φόρτωση δεδομένων στην αποθήκη. Αυτή η διαδικασία επιτυγχάνεται μέσω ETL ή ELT μεθόδων, που διασφαλίζουν την κατάλληλη προετοιμασία των δεδομένων για περαιτέρω ανάλυση. Ορισμένες αποθήκες δεδομένων στο νέφος υποστηρίζουν την εισαγωγή δεδομένων σε πραγματικό χρόνο, επιτρέποντας την ανάλυση των δεδομένων μόλις γίνουν διαθέσιμα. Αυτό είναι ιδιαίτερα χρήσιμο σε περιβάλλοντα όπου η γρήγορη λήψη αποφάσεων είναι κρίσιμη, όπως οι χρηματοοικονομικές υπηρεσίες και η διαχείριση αποθεμάτων.

Το στρώμα αποθήκευσης (Storage Layer) είναι ένα από τα πιο κρίσιμα στοιχεία της αρχιτεκτονικής, καθώς είναι υπεύθυνο για την οργάνωση και διατήρηση των δεδομένων σε αποδοτική μορφή για αναλυτική επεξεργασία. Συνήθως χρησιμοποιεί στηλοειδή αποθήκευση (columnar storage), η οποία διευκολύνει τη συμπίεση και επιτρέπει την ταχύτερη εκτέλεση ερωτημάτων, καθώς τα σχετικά δεδομένα αποθηκεύονται μαζί. Επιπλέον, πολλές αποθήκες δεδομένων στο νέφος αξιοποιούν καταμεμημένα συστήματα αρχείων, επιτρέποντας την αποθήκευση και διαχείριση δεδομένων σε πολλούς κόμβους, γεγονός που προσφέρει υψηλή επεκτασιμότητα και δυνατότητες παράλληλης επεξεργασίας.

Το στρώμα υπολογισμού (Compute Layer) διαδραματίζει κεντρικό ρόλο στην επεξεργασία των δεδομένων και την εκτέλεση αναλυτικών ερωτημάτων. Η διαχείριση των υπολογιστικών πόρων, όπως η CPU και η μνήμη, είναι δυναμική και προσαρμόζεται ανάλογα με τις απαιτήσεις των ερωτημάτων και των φορτίων εργασίας. Αυτό σημαίνει ότι οι αποθήκες δεδομένων στο υπολογιστικό νέφος μπορούν να αυξάνουν ή να μειώνουν τους υπολογιστικούς πόρους σε πραγματικό χρόνο, εξασφαλίζοντας τη βέλτιστη απόδοση και αποφεύγοντας περιττές δαπάνες.

Η βελτιστοποίηση και εκτέλεση ερωτημάτων (Query Optimization and Execution) είναι ένα βασικό χαρακτηριστικό που διαφοροποιεί τις αποθήκες δεδομένων στο υπολογιστικό νέφος από τις παραδοσιακές λύσεις. Οι σύγχρονοι μηχανισμοί βελτιστοποίησης SQL εφαρμόζουν στρατηγικές όπως ευρετηρίαση, προχωρημένες τεχνικές caching και cost-based optimization, επιτρέποντας τη βελτιστοποίηση των ερωτημάτων με τρόπο που μειώνει τους χρόνους απόκρισης. Το σύστημα ανάλυσης των ερωτημάτων εξετάζει διαφορετικά σενάρια εκτέλεσης και επιλέγει το πιο αποδοτικό, λαμβάνοντας υπόψη την πολυπλοκότητα των δεδομένων και τους διαθέσιμους υπολογιστικούς πόρους.

Τέλος, η ενσωμάτωση με εργαλεία επιχειρηματικής ανάλυσης (BI Tools Integration) αποτελεί βασικό παράγοντα για τη χρήση των δεδομένων στις επιχειρηματικές αποφάσεις. Οι αποθήκες δεδομένων στο υπολογιστικό νέφος παρέχουν υποστήριξη για JDBC, ODBC και RESTful APIs, καθιστώντας δυνατή την απρόσκοπτη διασύνδεση με εργαλεία ανάλυσης και οπτικοποίησης δεδομένων. Επιπλέον, υποστηρίζουν τεχνολογίες OLAP, που επιτρέπουν την πολυδιάστατη ανάλυση δεδομένων και τη δημιουργία δεδομένων κύβων (data cubes), διευκολύνοντας τη λήψη σύνθετων επιχειρηματικών αποφάσεων.

#### 4.5 Οφέλη της αποθήκης δεδομένων στο υπολογιστικό νέφος

Οι αποθήκες δεδομένων στο υπολογιστικό νέφος παρέχουν σημαντικά πλεονεκτήματα σε σύγκριση με τις παραδοσιακές λύσεις αποθήκευσης δεδομένων. Ο απλούστερος τρόπος εγκατάστασης και διαχείρισης, η επεκτασιμότητα, η βελτιωμένη προσβασιμότητα και η αυξημένη απόδοση αποτελούν ορισμένα από τα βασικά πλεονεκτήματα που τις καθιστούν ελκυστικές για επιχειρήσεις που επιθυμούν να αξιοποιήσουν στο έπακρο τα δεδομένα τους. Μια σύγχρονη αποθήκη δεδομένων στο υπολογιστικό νέφος μπορεί να αποθηκεύσει, να ενσωματώσει και να επεξεργαστεί μεγάλους όγκους δεδομένων από πολλαπλές πηγές, είτε αυτές προέρχονται από φυσικές εγκαταστάσεις είτε από το διαδίκτυο [78], [81].



Ένα από τα σημαντικότερα πλεονεκτήματα είναι η **ενισχυμένη προσβασιμότητα** που προσφέρει. Οι αποθήκες δεδομένων που φιλοξενούνται στο υπολογιστικό νέφος επιτρέπουν την πρόσβαση στα σχετικά δεδομένα από οπουδήποτε στον κόσμο, καθιστώντας τη λήψη αποφάσεων πιο ευέλικτη και αποδοτική. Παράλληλα, οι λύσεις αυτές διαθέτουν ισχυρούς μηχανισμούς ελέγχου πρόσβασης, διασφαλίζοντας ότι τα δεδομένα είναι διαθέσιμα μόνο στο εξουσιοδοτημένο προσωπικό.

Παρά το γεγονός ότι πολλοί εργαζόμενοι μπορούν να έχουν ταυτόχρονη πρόσβαση στην αποθήκη δεδομένων, η ακεραιότητα των δεδομένων παραμένει άθικτη, ενώ οι ενισχυμένοι μηχανισμοί διακυβέρνησης συμβάλλουν στη βελτίωση της συνολικής διαχείρισης ποιότητας των δεδομένων ενός οργανισμού.

Η **απεριόριστη επεκτασιμότητα** είναι ένα άλλο ουσιώδες πλεονέκτημα των αποθηκών δεδομένων στο νέφος. Η εικονική αρχιτεκτονική επιτρέπει στις επιχειρήσεις να προσαρμόζουν τους διαθέσιμους πόρους τους ανάλογα με τις εκάστοτε ανάγκες, χωρίς την ανάγκη υπερβολικών επενδύσεων σε υλικό εξοπλισμό. Αυτό είναι ιδιαίτερα χρήσιμο για εταιρείες με μεταβαλλόμενες απαιτήσεις, όπως αυτές του τουριστικού κλάδου, όπου κατά τις περιόδους υψηλής ζήτησης απαιτείται μεγαλύτερη υπολογιστική ισχύς για προηγμένες αναλύσεις, ενώ κατά τις περιόδους χαμηλής ζήτησης η ανάγκη αυτή μειώνεται αισθητά.

Ένα επιπλέον καθοριστικό πλεονέκτημα είναι η **υψηλή απόδοση** των αποθηκών δεδομένων στο υπολογιστικό νέφος. Όλα τα τμήματα μιας επιχείρησης μπορούν να έχουν ταυτόχρονη πρόσβαση στα απαραίτητα δεδομένα, χωρίς να επηρεάζεται η απόδοση του συστήματος. Αυτό επιτυγχάνεται μέσω της κατανομής των εργασιών σε πολλαπλούς διακομιστές, οι οποίοι εξασφαλίζουν ότι ακόμη και μεγάλοι όγκοι δεδομένων μπορούν να υποβληθούν σε επεξεργασία ταυτόχρονα, χωρίς καθυστερήσεις.

Επιπλέον, οι αποθήκες δεδομένων στο υπολογιστικό νέφος προσφέρουν **άφθονη χωρητικότητα αποθήκευσης**, επιτρέποντας στις επιχειρήσεις να διατηρούν τεράστιους όγκους δεδομένων με ελάχιστο κόστος. Σε αντίθεση με τις



παραδοσιακές λύσεις, οι οποίες απαιτούν ακριβό εξοπλισμό και πρόσθετους χώρους αποθήκευσης, οι πάροχοι αποθηκών δεδομένων στο υπολογιστικό νέφος ακολουθούν ένα μοντέλο πληρωμής ανάλογα με τη χρήση. Αυτό επιτρέπει στις εταιρείες να προσαρμόζουν τη χωρητικότητα της αποθήκης τους ανάλογα με τις πραγματικές τους ανάγκες, αποφεύγοντας την άσκοπη κατανάλωση πόρων.

Ένα ακόμη κρίσιμο χαρακτηριστικό είναι η **απρόσκοπτη ενσωμάτωση** με άλλες πλατφόρμες και εργαλεία ανάλυσης δεδομένων. Με δεδομένο ότι οι σύγχρονες επιχειρήσεις χρησιμοποιούν δεδομένα από εκατοντάδες διαφορετικές πηγές, η ενοποίηση αυτών των δεδομένων μπορεί να αποτελέσει πρόκληση. Οι αποθήκες δεδομένων στο νέφος έχουν σχεδιαστεί ώστε να επιτρέπουν τη συγκέντρωση δεδομένων από διάφορες πηγές, όπως εφαρμογές νέφους, βάσεις δεδομένων και διαφορετικές μορφές αρχείων.

Επιπλέον, προσφέρουν τη δυνατότητα ενοποίησης ημιδομημένων και αδόμητων δεδομένων, διευκολύνοντας την επιχειρηματική ανάλυση και την εξαγωγή χρήσιμων συμπερασμάτων.

Τέλος, η **αποκατάσταση καταστροφών** αποτελεί έναν από τους τομείς όπου οι αποθήκες δεδομένων στο νέφος υπερέχουν σημαντικά έναντι των παραδοσιακών λύσεων. Οι επιχειρήσεις που χρησιμοποιούν παλαιότερα εργαλεία συχνά πρέπει να δαπανούν μεγάλα χρηματικά ποσά για την απόκτηση επιπλέον υλικού για δημιουργία αντιγράφων ασφαλείας.

Οι αποθήκες δεδομένων στο υπολογιστικό νέφος επιλύουν αυτό το ζήτημα μέσω αυτοματοποιημένων μηχανισμών δημιουργίας αντιγράφων ασφαλείας, οι οποίοι διασφαλίζουν την προστασία των κρίσιμων δεδομένων ακόμη και σε περιπτώσεις καταστροφών ή αποτυχιών συστήματος. Επιπλέον, οι οργανισμοί που υιοθετούν λύσεις νέφους αποφεύγουν το υψηλό κόστος αγοράς και συντήρησης εξοπλισμού για τη διατήρηση των αντιγράφων ασφαλείας.

## 4.6 Προκλήσεις της αποθήκης δεδομένων στο υπολογιστικό νέφος

Παρόλο που οι αποθήκες δεδομένων στο υπολογιστικό νέφος προσφέρουν σημαντικά οφέλη, ιδιαίτερα όσον αφορά την επεκτασιμότητα και την ευελιξία, αντιμετωπίζουν και συγκεκριμένες προκλήσεις και πολυπλοκότητες που οι επιχειρήσεις πρέπει να λάβουν υπόψη [82], [83], [84].

Μία από τις πιο συχνές προκλήσεις είναι η **ενσωμάτωση δεδομένων (Data Integration)**. Οι δυσκολίες σε αυτόν τον τομέα προκύπτουν λόγω της ποικιλομορφίας των πηγών δεδομένων, της δυναμικής φύσης της υποδομής του νέφους και της ανάγκης αποτελεσματικής διαχείρισης και διακυβέρνησης των δεδομένων. Πολλές επιχειρήσεις λειτουργούν με έναν συνδυασμό τοπικών και νεφοϋπολογιστικών συστημάτων, γεγονός που καθιστά την ενοποίηση των δεδομένων πιο περίπλοκη. Η εξασφάλιση της συνοχής μεταξύ αυτών των περιβαλλόντων απαιτεί επιπλέον προσπάθειες, καθώς προκύπτουν ζητήματα όπως η ασφάλεια, η καθυστέρηση δεδομένων και η συνδεσιμότητα.

Η **ασφάλεια (Security)** αποτελεί μία από τις μεγαλύτερες προκλήσεις για τις αποθήκες δεδομένων στο νέφος. Ο συγχρονισμός των πρακτικών κρυπτογράφησης με τις ιδιαίτερες απαιτήσεις ενός οργανισμού μπορεί να είναι πολύπλοκος, καθώς οι επιχειρήσεις συχνά λειτουργούν σε μικτά περιβάλλοντα, συνδυάζοντας τοπικά και νεφοϋπολογιστικά συστήματα. Η ενοποίηση των πρακτικών κρυπτογράφησης και η δημιουργία ενός ενιαίου πλαισίου ασφαλείας απαιτεί προσεκτικό σχεδιασμό. Επιπλέον, η λειτουργία σε πολλαπλά περιβάλλοντα νέφους επιβάλλει την υιοθέτηση προτύπων ελέγχου πρόσβασης που είναι συμβατά μεταξύ διαφορετικών παρόχων. Η διατήρηση συνεπών πολιτικών πρόσβασης όταν τα δεδομένα είναι διανεμημένα σε πολλές πλατφόρμες μπορεί να είναι ιδιαίτερα απαιτητική, καθιστώντας την τυποποίηση απαραίτητη για την προστασία των δεδομένων.

Μία ακόμα κρίσιμη πρόκληση είναι η **συμμόρφωση με κανονισμούς**. Οι πάροχοι υπηρεσιών νέφους ακολουθούν ένα μοντέλο κατανομής ευθυνών, στο οποίο διαχειρίζονται ορισμένες πτυχές της ασφάλειας, ενώ οι πελάτες είναι υπεύθυνοι για άλλες. Η κατανόηση και η υλοποίηση αυτής της κοινής ευθύνης μπορεί να

είναι περίπλοκη. Ταυτόχρονα, η συνεχώς μεταβαλλόμενη και διαφοροποιημένη φύση των κανονιστικών πλαισίων, που επηρεάζουν διάφορους κλάδους και γεωγραφικές περιοχές, μπορεί να αποτελέσει σημαντικό εμπόδιο στη συμμόρφωση με τους κανονισμούς των ρυθμιστικών αρχών.

Η **διαχείριση κόστους (Cost Management)** είναι μία από τις βασικές ανησυχίες για τις επιχειρήσεις που χρησιμοποιούν αποθήκες δεδομένων στο νέφος. Αν και το μοντέλο πληρωμής ανά χρήση παρέχει ευελιξία και προσαρμοστικότητα, μπορεί να οδηγήσει σε απροσδόκητες δαπάνες εάν δεν υπάρχει κατάλληλη παρακολούθηση. Η πρόκληση έγκειται στην αποτελεσματική διαχείριση των πόρων, ώστε να ανταποκρίνονται στις μεταβαλλόμενες απαιτήσεις εργασιών επεξεργασίας δεδομένων. Η ακριβής πρόβλεψη του κόστους μπορεί να είναι δύσκολη, ειδικά όταν υπάρχουν διακυμάνσεις στον όγκο των δεδομένων ή όταν εκτελούνται πολύπλοκα αναλυτικά ερωτήματα. Επιπλέον, η πληθώρα υπηρεσιών και λειτουργιών που παρέχονται από τις αποθήκες δεδομένων στο νέφος μπορεί να οδηγήσει είτε σε υπερεκτίμηση είτε σε υποχρησιμοποίηση των πόρων, επηρεάζοντας αρνητικά την αποδοτικότητα του κόστους.

Ένα ακόμη σημαντικό ζήτημα είναι η **εξάρτηση από συγκεκριμένο πάροχο (Vendor Lock-In)**. Οι επιχειρήσεις που βασίζονται στις δυνατότητες και τις υπηρεσίες ενός συγκεκριμένου παρόχου αποθήκευσης δεδομένων στο νέφος διατρέχουν τον κίνδυνο να εξαρτηθούν υπερβολικά από τις ιδιόκτητες τεχνολογίες και τα APIs του. Παρόλο που αυτές οι τεχνολογίες αυξάνουν την αποδοτικότητα και τη λειτουργικότητα, δημιουργούν επίσης εξαρτήσεις που μπορεί να είναι δύσκολο να ξεπεραστούν.

Η μετάβαση σε άλλον πάροχο ή η υιοθέτηση μιας στρατηγικής πολλαπλών παρόχων μπορεί να αποδειχθεί σύνθετη, καθώς μπορεί να απαιτήσει την επανεγγραφή ερωτημάτων, την προσαρμογή μοντέλων δεδομένων και την αντιμετώπιση ζητημάτων συμβατότητας.

## 4.7 Λύσεις αποθήκευσης δεδομένων στο υπολογιστικό νέφος για επιχειρήσεις

Οι περισσότερες λύσεις αποθήκευσης δεδομένων στο υπολογιστικό νέφος [85], [88] λειτουργούν με το μοντέλο πληρωμής ανάλογα με τη χρήση, το οποίο προτιμάται από τις επιχειρήσεις, ιδίως τις νεοφυείς εταιρείες που εισέρχονται στον κόσμο της αποθήκευσης δεδομένων. Αυτή η επιλογή τιμολόγησης είναι ιδιαίτερα χρήσιμη για οργανισμούς που αναμένουν να προσθέσουν νέες πηγές και πλατφόρμες στην αρχιτεκτονική των δεδομένων τους, καθώς οι αποθήκες δεδομένων στο νέφος μπορούν να προσαρμόζονται δυναμικά για να καλύψουν τις ανάγκες αυτές. Επιπλέον, οι πιο δημοφιλείς λύσεις αποθήκευσης δεδομένων στο υπολογιστικό νέφος προσφέρουν παρόμοια αξία όσον αφορά την υψηλή απόδοση, την επεκτασιμότητα, την ευελιξία, την ευκολία χρήσης και την τιμολόγηση. Ωστόσο, οι διαφορές τους έγκεινται στον τρόπο με τον οποίο υλοποιούνται αυτές οι δυνατότητες. Οι επιχειρήσεις θα πρέπει να αξιολογούν προσεκτικά τα μοναδικά χαρακτηριστικά και τα πλεονεκτήματα κάθε λύσης αποθήκευσης δεδομένων στο νέφος, λαμβάνοντας υπόψη τις ειδικές τους απαιτήσεις και προτιμήσεις.

### 4.7.1 Microsoft Azure Synapse Analytics

Συνδυάζει την ανάλυση μεγάλων δεδομένων με την επιχειρησιακή αποθήκευση δεδομένων, επιταχύνοντας τον χρόνο εξαγωγής πληροφοριών. Συγκεκριμένα, χρησιμοποιεί SQL για αποθήκευση δεδομένων, τεχνολογίες Spark για τη διαχείριση μεγάλων δεδομένων και Pipelines για ενσωμάτωση δεδομένων μέσω ETL και ELT. Επίσης, ενσωματώνεται απρόσκοπτα με εργαλεία BI, όπως το Power BI. Αποτελεί μια αξιόλογη λύση αποθήκευσης δεδομένων για οργανισμούς που ασχολούνται με διαφορετικές πτυχές της διαχείρισης δεδομένων. Επιπλέον, αν η επιχείρησή σας χρησιμοποιεί ήδη πολλές άλλες υπηρεσίες της Microsoft, η ενσωμάτωση του Azure Synapse Analytics στην υπάρχουσα αρχιτεκτονική δεδομένων μπορεί να είναι μια ιδιαίτερα αποδοτική επιλογή, δεδομένης της άριστης διαλειτουργικότητας των υπηρεσιών της Microsoft [85].

Ένα από τα μεγαλύτερα πλεονεκτήματα του Azure Synapse Analytics είναι η απρόσκοπτη ενσωμάτωσή του με άλλες υπηρεσίες του Azure καθώς και με προηγμένες πλατφόρμες BI, αναλυτικής και μηχανικής μάθησης. Υποστηρίζει ποικίλους τύπους δεδομένων, συμπεριλαμβανομένων των μη δομημένων δεδομένων, ενώ η δυναμική διαχείριση των πόρων επιτρέπει την προσαρμογή στις ανάγκες επεξεργασίας δεδομένων. Παρέχει οικονομικά αποδοτική δυναμική ερώτηση μέσω serverless υποδομής και επιτρέπει την εύκολη κλιμάκωση για τη διαχείριση μεγάλων συνόλων δεδομένων. Παράλληλα, διαθέτει ισχυρά χαρακτηριστικά ασφαλείας για την προστασία των δεδομένων και επιτρέπει την αξιοποίηση προηγμένων τεχνικών ανάλυσης και πρόβλεψης, καθιστώντας το ιδανική λύση για οργανισμούς που ασχολούνται με ανάλυση δεδομένων σε πραγματικό χρόνο και προγνωστική ανάλυση.

Ωστόσο, υπάρχουν και ορισμένες προκλήσεις που πρέπει να ληφθούν υπόψη. Η υψηλή εξάρτηση από το οικοσύστημα του Azure μπορεί να αποτελέσει περιορισμό για επιχειρήσεις που χρησιμοποιούν διαφορετικές cloud υποδομές. Η βελτιστοποίηση των παραμέτρων για τη μέγιστη απόδοση μπορεί να είναι σύνθετη διαδικασία, καθώς απαιτεί συνεχή προσαρμογή και παρακολούθηση. Επιπλέον, η συχνότητα των ενημερώσεων σημαίνει ότι οι χρήστες πρέπει να είναι διαρκώς ενήμεροι για τις αλλαγές και να προσαρμόζονται ανάλογα. Το κόστος μπορεί να αυξηθεί σημαντικά με την εντατική χρήση, ενώ η μαθησιακή καμπύλη για τις ομάδες που δεν είναι εξοικειωμένες με την πλατφόρμα μπορεί να αποτελέσει μια επιπλέον πρόκληση, απαιτώντας εκπαίδευση των χρηστών.

#### **4.7.2 Amazon Redshift**

Είναι μια πλήρως διαχειριζόμενη υπηρεσία αποθήκευσης δεδομένων σε petabyte κλίμακα, που παρέχεται από το Amazon Web Services (AWS). Σχεδιάστηκε για τη διαχείριση μεγάλων συνόλων δεδομένων και την εκτέλεση αναλυτικών φορτίων εργασίας με υψηλή απόδοση. Προσφέρει μια κλιμακούμενη και οικονομικά αποδοτική λύση, κατάλληλη για επιχειρήσεις που χρειάζονται ισχυρή ανάλυση δεδομένων και εφαρμογές επιχειρηματικής ευφυΐας [86].

Ένα από τα κύρια πλεονεκτήματα του Amazon Redshift είναι η ικανότητά του να κλιμακώνεται εύκολα, υποστηρίζοντας τόσο μικρά όσο και μεγάλα σύνολα δεδομένων. Παρέχει γρήγορη εκτέλεση ερωτημάτων, ειδικά για αναλυτικά φορτία εργασίας, διασφαλίζοντας υψηλή απόδοση για απαιτητικές εφαρμογές. Επιπλέον, η ενσωμάτωσή του με άλλες υπηρεσίες AWS επιτρέπει τη δημιουργία ολοκληρωμένων λύσεων διαχείρισης δεδομένων. Η αυτοματοποιημένη δημιουργία αντιγράφων ασφαλείας και η διαχείριση συντήρησης μειώνουν το λειτουργικό βάρος των επιχειρήσεων, ενώ τα ισχυρά χαρακτηριστικά ασφαλείας συμβάλλουν στην προστασία ευαίσθητων δεδομένων.

Ωστόσο, το Amazon Redshift είναι βελτιστοποιημένο κυρίως για αναλυτικά ερωτήματα και λιγότερο κατάλληλο για συναλλακτικές εργασίες. Η διαθεσιμότητα των χαρακτηριστικών μπορεί να ποικίλει ανάλογα με την περιοχή, ενώ οι χρήστες που δεν είναι εξοικειωμένοι με την πλατφόρμα AWS ενδέχεται να χρειαστούν επιπλέον χρόνο εκμάθησης. Αν και είναι μια οικονομικά αποδοτική λύση, η εκτεταμένη χρήση σε μεγάλα δεδομένα μπορεί να οδηγήσει σε σημαντικές δαπάνες.

#### 4.7.3 Snowflake

Είναι μια πλατφόρμα αποθήκευσης δεδομένων στο υπολογιστικό νέφος που παρέχει μια πλήρως διαχειριζόμενη και κλιμακούμενη λύση για την αποθήκευση και ανάλυση δεδομένων. Λειτουργεί ως πλατφόρμα SaaS και έχει σχεδιαστεί για να προσφέρει απλότητα, ευελιξία και αποδοτικότητα στις επιχειρήσεις που αναζητούν μια σύγχρονη αποθήκη δεδομένων στο νέφος [87].

Ένα από τα βασικά πλεονεκτήματα του Snowflake είναι η δυνατότητα ανάπτυξης σε πολλαπλές cloud πλατφόρμες, παρέχοντας ευελιξία και αποτρέποντας τον εγκλωβισμό σε έναν συγκεκριμένο πάροχο. Η αυτόματη κλιμάκωση διασφαλίζει τη βέλτιστη απόδοση για μεταβαλλόμενα φορτία εργασίας, επιτρέποντας στις επιχειρήσεις να προσαρμόζουν τους πόρους τους με βάση τις ανάγκες τους. Το Snowflake διευκολύνει επίσης την ασφαλή και εύκολη κοινή χρήση δεδομένων

μεταξύ οργανισμών ή τμημάτων, ενώ η δυνατότητα κλωνοποίησης βάσεων δεδομένων ή πινάκων χωρίς επιπλέον κατανάλωση αποθηκευτικού χώρου (zero - copy cloning) συμβάλλει στην αποδοτική διαχείριση των δεδομένων.

Επιπλέον, οι χρήστες έχουν πρόσβαση σε ιστορικά δεδομένα και μπορούν να ανακτούν προηγούμενες εκδόσεις σε περίπτωση αλλαγών, ενώ η δυνατότητα ανεξάρτητης κλιμάκωσης της αποθήκευσης και των υπολογιστικών πόρων προσφέρει αυξημένη προσαρμοστικότητα.

Ωστόσο, το Snowflake συνοδεύεται και από ορισμένες προκλήσεις. Η μεταφορά δεδομένων μεταξύ διαφορετικών παρόχων cloud μπορεί να επιφέρει επιπλέον κόστη, καθιστώντας απαραίτητο τον προσεκτικό προγραμματισμό των μεταφορών δεδομένων. Οι νέοι χρήστες μπορεί να χρειαστούν χρόνο για να εξοικειωθούν με την πλατφόρμα και τη λειτουργικότητά της, ενώ ορισμένες σύνθετες εργασίες απαιτούν λεπτομερή ρύθμιση για τη βέλτιστη απόδοση. Παρόλο που το Snowflake προσφέρει οικονομικά αποδοτικές λύσεις, η εκτεταμένη χρήση σε μεγάλο όγκο δεδομένων μπορεί να οδηγήσει σε σημαντικές δαπάνες.

#### 4.7.4 Google BigQuery

Είναι μια πλήρως διαχειριζόμενη, serverless λύση αποθήκευσης δεδομένων στο υπολογιστικό νέφος, που παρέχεται από το Google Cloud Platform (GCP). Έχει σχεδιαστεί για να διαχειρίζεται μεγάλες αναλυτικές εργασίες και επιτρέπει την ανάλυση και επεξεργασία μεγάλων συνόλων δεδομένων σε πραγματικό χρόνο. Η βαθιά ενσωμάτωσή του με άλλες υπηρεσίες της Google καθιστά το BigQuery μια ολοκληρωμένη πλατφόρμα για διάφορες ανάγκες ανάλυσης δεδομένων [88].

Ένα από τα κύρια πλεονεκτήματα του Google BigQuery είναι η serverless λειτουργία του, που σημαίνει ότι η πλατφόρμα κλιμακώνεται αυτόματα χωρίς την ανάγκη διαχείρισης υποδομής. Είναι βελτιστοποιημένο για γρήγορη εκτέλεση ερωτημάτων, καθιστώντας το ιδανικό για αναλύσεις σε πραγματικό χρόνο και επιτρέπει την επεξεργασία μεγάλων συνόλων δεδομένων με αυτόματη κλιμάκωση ανάλογα με το φόρτο εργασίας. Η απρόσκοπτη ενσωμάτωσή του με άλλες

υπηρεσίες του Google Cloud διευκολύνει την ενσωμάτωσή του σε υπάρχοντα οικοσυστήματα δεδομένων, ενώ η χρήση γνώριμης SQL σύνταξης επιτρέπει την εύκολη προσαρμογή από αναλυτές δεδομένων και προγραμματιστές. Επιπλέον, υποστηρίζει τη ροή δεδομένων σε πραγματικό χρόνο, επιτρέποντας στις επιχειρήσεις να αναλύουν δεδομένα καθώς αυτά δημιουργούνται.

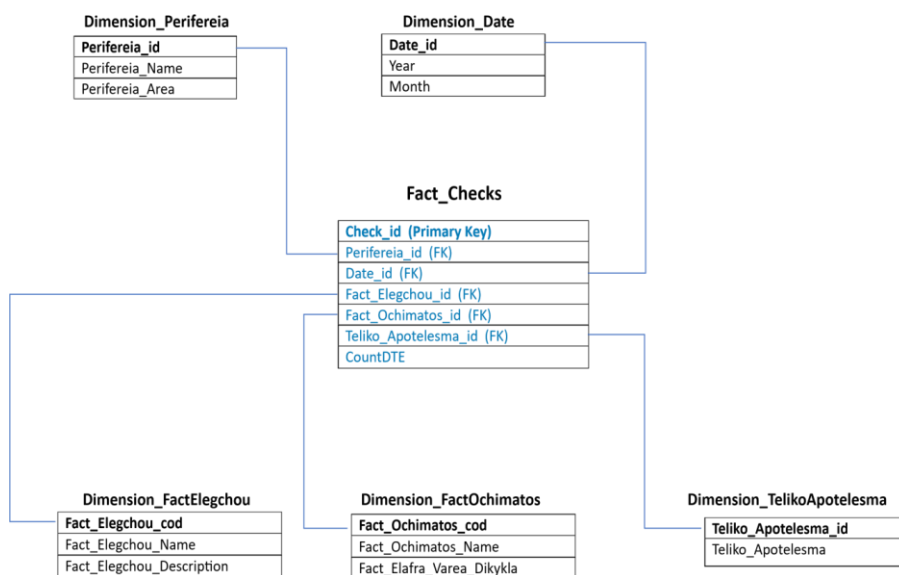
Ωστόσο, το Google BigQuery δεν είναι σχεδιασμένο για συναλλακτικές εργασίες, καθώς είναι βελτιστοποιημένο για αναλυτικά φορτία εργασίας. Παρόλο που η ενσωμάτωσή του με το GCP παρέχει σημαντικά οφέλη, μπορεί επίσης να δημιουργήσει έναν βαθμό εξάρτησης από τον πάροχο. Η λύση είναι ιδιαίτερα αποδοτική για μικρούς και μεσαίους φόρτους εργασίας, ωστόσο, το κόστος μπορεί να αυξηθεί σημαντικά σε περιπτώσεις μεγάλης κλίμακας χρήσης. Επιπλέον, οι χρήστες μπορεί να χρειαστούν χρόνο για να εξοικειωθούν με το οικοσύστημα του Google Cloud και να αξιοποιήσουν πλήρως τις δυνατότητές του.



## 5. Σχεδιασμός και υλοποίηση παραδοσιακής αποθήκης δεδομένων

### 5.1 Αρχιτεκτονική αποθήκης δεδομένων

Το παρακάτω σχεσιακό σχήμα έχει σχεδιαστεί για να υποστηρίξει τη διαχείριση και την ανάλυση δεδομένων που αφορούν ελέγχους οχημάτων. Ακολουθεί μια αρχιτεκτονική τύπου **αστέρα (Star Schema)**, η οποία περιλαμβάνει έναν κεντρικό πίνακα γεγονότων (**Fact Table**) και πολλαπλούς πίνακες διαστάσεων (**Dimension Tables**).



Εικόνα 5: Σχεσιακό σχήμα

Ο κεντρικός πίνακας **Fact\_Checks** περιέχει δεδομένα για κάθε έλεγχο που πραγματοποιείται. Περιλαμβάνει τις ακόλουθες πληροφορίες: τον μοναδικό αναγνωριστικό αριθμό του ελέγχου (**Check\_id**), την περιφέρεια όπου διεξήχθη ο έλεγχος (**Perifereia\_id**), την ημερομηνία πραγματοποίησης (**Date\_id**), τον τύπο του ελέγχου (**Fact\_Elegchou\_id**), τον τύπο του οχήματος που εξετάστηκε

(**Fact\_Ochimatots\_id**), το τελικό αποτέλεσμα του ελέγχου (**Teliko\_Apotelesma\_id**) και τον αριθμό των ελέγχων (**CountDTE**).

Οι εξωτερικές κλειδίες που περιέχονται στον πίνακα συνδέονται με τους πίνακες διαστάσεων, προσφέροντας δυνατότητες πολύπλοκης ανάλυσης.

Οι πίνακες διαστάσεων περιλαμβάνουν:

- **Dimension\_Perifereia:** Αποθηκεύει πληροφορίες για τις γεωγραφικές περιοχές, με πεδία το μοναδικό αναγνωριστικό της περιφέρειας (**Perifereia\_id**), το όνομα της περιφέρειας (**Perifereia\_Name**) και μια επιπλέον περιγραφή της περιοχής (**Perifereia\_Area**). Ο πίνακας αυτός επιτρέπει γεωγραφικές αναλύσεις και συγκρίσεις μεταξύ διαφορετικών περιοχών.
- **Dimension\_Date:** Επιτρέπει την οργάνωση των ελέγχων βάσει χρονικής περιόδου. Περιλαμβάνει το μοναδικό αναγνωριστικό της ημερομηνίας (**Date\_id**), το έτος (**Year**) και τον μήνα (**Month**) του ελέγχου. Αυτός ο πίνακας χρησιμοποιείται για την ανάλυση της εξέλιξης των ελέγχων σε διαφορετικές χρονικές περιόδους.
- **Dimension\_FactElegchou:** Καταγράφει πληροφορίες σχετικά με τους τύπους των ελέγχων που πραγματοποιούνται, περιλαμβάνοντας το μοναδικό αναγνωριστικό (**Fact\_Elegchou\_cod**), την ονομασία του τύπου ελέγχου (**Fact\_Elegchou\_Name**) και μια σύντομη περιγραφή (**Fact\_Elegchou\_Description**). Ο πίνακας αυτός συμβάλλει στην ανάλυση της συχνότητας και της φύσης των ελέγχων.
- **Dimension\_FactOchimatots:** Αποθηκεύει πληροφορίες για τα οχήματα που εξετάζονται κατά τη διάρκεια των ελέγχων, όπως το μοναδικό αναγνωριστικό του τύπου οχήματος (**Fact\_Ochimatots\_cod**), το όνομα του τύπου οχήματος (**Fact\_Ochimatots\_Name**) και την κατηγοριοποίησή του ως ελαφρύ, βαρύ ή δίκυκλο (**Fact\_Elafra\_Varea\_Dikykla**). Αυτή η διάσταση επιτρέπει την ανάλυση των ελέγχων βάσει τύπου οχήματος.

- **Dimension\_TelikoApotelesma:** Περιλαμβάνει δεδομένα σχετικά με την έκβαση του ελέγχου. Αποτελείται από το μοναδικό αναγνωριστικό του αποτελέσματος (**Teliko\_Apotelesma\_id**) και μια περιγραφή του τελικού αποτελέσματος (**Teliko\_Apotelesma**). Η ύπαρξη αυτού του πίνακα επιτρέπει τη διερεύνηση των επιτυχών και αποτυχημένων ελέγχων.

## 5.2 Δημιουργία αποθήκης δεδομένων

Για την υλοποίηση του σχεσιακού σχήματος επιλέχθηκε ο MySQL Server [100], [104] ως το σύστημα διαχείρισης της αποθήκης δεδομένων. Η επιλογή αυτή βασίστηκε σε διάφορους λόγους:

1. **Υψηλή Απόδοση:** Ο MySQL Server είναι βελτιστοποιημένος για ταχύτητα και αποδοτικότητα, επιτρέποντας γρήγορη επεξεργασία μεγάλου όγκου δεδομένων.
2. **Υποστήριξη ACID:** Διαθέτει υποστήριξη για ακεραιότητα συναλλαγών (Atomicity, Consistency, Isolation, Durability), διασφαλίζοντας την αξιοπιστία των δεδομένων.
3. **Συμβατότητα με το Star Schema:** Ο MySQL Server υποστηρίζει σχέσεις με ξένα κλειδιά (Foreign Keys), επιτρέποντας τη σωστή λειτουργία του σχεσιακού σχήματος.
4. **Επεκτασιμότητα:** Δίνει τη δυνατότητα κλιμάκωσης και υποστήριξης πολύπλοκων αναλύσεων, διευκολύνοντας τη διαχείριση των δεδομένων.
5. **Ευκολία στη Χρήση:** Διαθέτει εργαλεία όπως το MySQL Workbench, που επιτρέπουν τον εύκολο σχεδιασμό και διαχείριση της βάσης.

Η υλοποίηση του συγκεκριμένου σχεσιακού σχήματος σε MySQL απαιτεί τη δημιουργία μιας βάσης δεδομένων και των αντίστοιχων πινάκων με τη χρήση SQL. Παρακάτω παρουσιάζεται ο κώδικας για τη δημιουργία της βάσης και των σχετικών πινάκων.

```
CREATE DATABASE IF NOT EXISTS VehicleChecksDB;
USE VehicleChecksDB;

-- Δημιουργία των Πινάκων Διαστάσεων
CREATE TABLE Dimension_Perifereia (
    Perifereia_id INT PRIMARY KEY,
    Perifereia_Name VARCHAR(255),
    Perifereia_Area VARCHAR(255)
);

CREATE TABLE Dimension_Date (
    Date_id INT PRIMARY KEY,
    Year INT,
    Month INT
);

CREATE TABLE Dimension_FactElegchou (
    Fact_Elegchou_cod INT PRIMARY KEY,
    Fact_Elegchou_Name VARCHAR(255),
    Fact_Elegchou_Description TEXT
);

CREATE TABLE Dimension_FactOchimatatos (
    Fact_Ochimatatos_cod INT PRIMARY KEY,
    Fact_Ochimatatos_Name VARCHAR(255),
    Fact_Elafra_Varea_Dikykla VARCHAR(255)
);

CREATE TABLE Dimension_TelikoApotelesma (
    Teliko_Apotelesma_id INT PRIMARY KEY,
    Teliko_Apotelesma VARCHAR(255)
);

-- Δημιουργία του Πίνακα Γεγονότων
CREATE TABLE Fact_Checks (
    Check_id INT PRIMARY KEY,
    Perifereia_id INT,
    Date_id INT,
    Fact_Elegchou_id INT,
    Fact_Ochimatatos_id INT,
    Teliko_Apotelesma_id INT,
    CountDTE INT,
    FOREIGN KEY (Perifereia_id) REFERENCES
Dimension_Perifereia(Perifereia_id),
    FOREIGN KEY (Date_id) REFERENCES Dimension_Date(Date_id),
    FOREIGN KEY (Fact_Elegchou_id) REFERENCES
Dimension_FactElegchou(Fact_Elegchou_cod),
    FOREIGN KEY (Fact_Ochimatatos_id) REFERENCES
Dimension_FactOchimatatos(Fact_Ochimatatos_cod),
    FOREIGN KEY (Teliko_Apotelesma_id) REFERENCES
Dimension_TelikoApotelesma(Teliko_Apotelesma_id)
);
```

Ο αντίστοιχος κώδικας για την εισαγωγή εγγραφών στους πίνακες:

```
-- Εισαγωγή δεδομένων στον πίνακα Dimension_Perifereia
INSERT INTO Dimension_Perifereia (Perifereia_id, Perifereia_Name,
Perifereia_Area) VALUES
(1, 'Π.Ε ΘΕΣΣΑΛΟΝΙΚΗΣ', 3681),
(2, 'ΑΝΑΤΟΛΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ & ΘΡΑΚΗΣ', 14157),
(3, 'ΑΤΤΙΚΗΣ', 3808),
(4, 'ΒΟΡΕΙΟΥ ΑΙΓΑΙΟΥ', 3836),
(5, 'ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ', 11336),
(6, 'ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ', 9451),
(7, 'ΗΠΕΙΡΟΥ', 9203),
(8, 'ΘΕΣΣΑΛΙΑΣ', 14037),
(9, 'ΙΟΝΙΩΝ ΝΗΣΩΝ', 2307),
(10, 'ΚΕΝΤΡΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ', 30319),
(11, 'ΚΡΗΤΗΣ', 8303),
(12, 'ΝΟΤΙΟΥ ΑΙΓΑΙΟΥ', 5286),
(13, 'ΠΕΛΟΠΟΝΝΗΣΟΥ', 15490),
(14, 'ΣΤΕΡΕΑΣ ΕΛΛΑΔΟΣ', 15549);

-- Εισαγωγή δεδομένων στον πίνακα Dimension_Date
INSERT INTO Dimension_Date (Date_id, Year, Month) VALUES
(1, 2013, 1),
(2, 2014, 2),
(3, 2015, 3),
(4, 2016, 4),
(5, 2017, 5),
(6, 2018, 6),
(7, 2019, 7),
(8, 2020, 8),
(9, 2021, 9),
(10, 2022, 10),
(11, 2023, 11),
(12, 2024, 12);

-- Εισαγωγή δεδομένων στον πίνακα Dimension_FactElegchou
INSERT INTO Dimension_FactElegchou (Fact_Elegchou_cod,
Fact_Elegchou_Name) VALUES
(0, 'ΠΕΡΙΟΔΙΚΟΣ'),
(1, 'ΑΡΧΙΚΟΣ'),
(2, 'ΕΠΑΝΕΛΕΓΧΟΣ'),
(3, 'ΕΙΔΙΚΟΣ & ΠΕΡΙΟΔΙΚΟΣ ΕΠΑΝΕΛΕΓΧΟΣ ΕΚΠΑΙΔΕΥΤΙΚΟΥ ΟΧΗΜΑΤΟΣ');

-- Εισαγωγή δεδομένων στον πίνακα Dimension_FactOchimatatos
INSERT INTO Dimension_FactOchimatatos (Fact_Ochimatatos_cod,
Fact_Ochimatatos_Name, Fact_Elafra_Varea_Dikykla) VALUES
(3, 'ΑΣΘΕΝΟΦΟΡΟ-Τ<3.5-ΔΧ', 'ΕΛΑΦΡΑ'),
(4, 'ΑΣΘΕΝΟΦΟΡΟ-Τ<3.5-ΙΧ', 'ΕΛΑΦΡΑ'),
(5, 'ΕΠΙΒΑΤΙΚΟ ΕΩΣ 8 ΘΕΣΕΙΣ-ΙΧ', 'ΕΛΑΦΡΑ'),
(18, 'ΛΕΩΦΟΡΕΙΟ ΑΠΟ ΘΕΣΕΙΣ 8 ΚΑΙ ΑΝΩ-Τ<=10-ΔΧ', 'ΒΑΡΕΑ'),
(21, 'ΛΕΩΦΟΡΕΙΟ ΑΠΟ ΘΕΣΕΙΣ 8 ΚΑΙ ΑΝΩ-Τ>10-ΔΧ', 'ΒΑΡΕΑ'),
(219, 'ΕΚΠΑΙΔΕΥΤΙΚΟ ΟΧΗΜΑ ΚΑΤΗΓΟΡΙΑΣ Α1 (ΔΙΤΡΟΧΕΣ ΜΟΤΟΣΥΚΛΕΤΕΣ)',
'ΔΙΚΥΚΛΑ');
```

```
(220, 'ΕΚΠΑΙΔΕΥΤΙΚΟ ΟΧΗΜΑ ΚΑΤΗΓΟΡΙΑΣ Α2 (ΔΙΤΡΟΧΕΣ ΜΟΤΟΣΥΚΛΕΤΕΣ)',  
'ΔΙΚΥΚΛΑ');  
  
-- Εισαγωγή δεδομένων στον πίνακα Dimension_TelikoApotelesma  
INSERT INTO Dimension_TelikoApotelesma (Teliko_Apotelesma_id,  
Teliko_Apotelesma) VALUES  
(1, 'ΕΠΙΤΥΧΗΣ'),  
(2, 'ΔΕΥΤΕΡΕΥΟΥΣΕΣ ΕΛΛΕΙΨΕΙΣ'),  
(3, 'ΣΟΒΑΡΕΣ ΕΛΛΕΙΨΕΙΣ'),  
(4, 'ΕΠΙΚΥΝΔΥΝΕΣ ΕΛΛΕΙΨΕΙΣ');  
  
-- Εισαγωγή δεδομένων στον πίνακα Fact_Checks  
INSERT INTO Fact_Checks (Check_id, Perifereia_id, Date_id,  
Fact_Elegchou_id, Fact_Ochimatou_id, Teliko_Apotelesma_id,  
CountDTE) VALUES  
(1, 1, 1, 1, 3, 1, 10),  
(2, 2, 2, 2, 4, 2, 15),  
(3, 3, 3, 3, 5, 3, 8),  
(4, 4, 4, 0, 18, 4, 20);
```

### 5.3 Ερωτήματα προς εκτέλεση για αξιολόγηση

Παρακάτω, παρατίθενται 10 ερωτήματα προς την αποθήκη δεδομένων η οποία δημιουργήθηκε με σκοπό την μετέπειτα αξιολόγησή της:

```
-- 1. Προβολή όλων των περιφερειών  
SELECT Perifereia_Name FROM Dimension_Perifereia;  
  
-- 2. Εμφάνιση όλων των ετών που καταγράφονται στη βάση  
SELECT DISTINCT Year FROM Dimension_Date;  
  
-- 3. Προβολή όλων των τύπων ελέγχων  
SELECT * FROM Dimension_FactElegchou;  
  
-- 4. Σύνολο ελέγχων που έχουν πραγματοποιηθεί  
SELECT SUM(CountDTE) FROM Fact_Checks;  
  
-- 5. Πλήθος ελέγχων ανά τύπο οχήματος  
SELECT Fact_Ochimatou_Name, COUNT(*) FROM Fact_Checks JOIN  
Dimension_FactOchimatou ON Fact_Checks.Fact_Ochimatou_id =  
Dimension_FactOchimatou.Fact_Ochimatou_cod GROUP BY  
Fact_Ochimatou_Name;  
  
-- 6. Περιφέρεια με τους περισσότερους ελέγχους  
SELECT Perifereia_Name FROM Fact_Checks JOIN Dimension_Perifereia  
ON Fact_Checks.Perifereia_id = Dimension_Perifereia.Perifereia_id  
GROUP BY Perifereia_Name ORDER BY SUM(CountDTE) DESC LIMIT 1;
```

```
-- 7. Ποσοστό επιτυχίας ελέγχων
SELECT (SUM(CASE WHEN Teliko_Apotelesma_id = 1 THEN CountDTE ELSE
0 END) * 100.0 / SUM(CountDTE)) AS SuccessRate FROM Fact_Checks;

-- 8. Κατανομή τύπων ελέγχου ανά έτος
SELECT Year, Fact_Elegchou_Name, COUNT(*) FROM Fact_Checks JOIN
Dimension_Date ON Fact_Checks.Date_id = Dimension_Date.Date_id
JOIN Dimension_FactElegchou ON Fact_Checks.Fact_Elegchou_id =
Dimension_FactElegchou.Fact_Elegchou_cod GROUP BY Year,
Fact_Elegchou_Name;

-- 9. Τα οχήματα με τη μεγαλύτερη αποτυχία στους ελέγχους
SELECT Fact_Ochimatots_Name, COUNT(*) FROM Fact_Checks JOIN
Dimension_FactOchimatots ON Fact_Checks.Fact_Ochimatots_id =
Dimension_FactOchimatots.Fact_Ochimatots_cod WHERE
Teliko_Apotelesma_id IN (3,4) GROUP BY Fact_Ochimatots_Name ORDER
BY COUNT(*) DESC;

-- 10. Σύγκριση αριθμού ελέγχων σε διαφορετικές περιφέρειες μεταξύ
δύο ετών
SELECT p.Perifereia_Name, d.Year, SUM(fc.CountDTE) FROM
Fact_Checks fc JOIN Dimension_Perifereia p ON fc.Perifereia_id =
p.Perifereia_id JOIN Dimension_Date d ON fc.Date_id = d.Date_id
WHERE d.Year IN (2013, 2016) GROUP BY p.Perifereia_Name, d.Year
ORDER BY p.Perifereia_Name, d.Year;
```

Παρακάτω, παρατίθεται η επεξήγηση της λειτουργίας των ερωτημάτων:

- Το 1<sup>ο</sup> ερώτημα επιστρέφει όλα τα ονόματα των περιφερειών που υπάρχουν στον πίνακα Dimension\_Perifereia. Αυτή η διάσταση πιθανότατα αποθηκεύει πληροφορίες σχετικά με τις γεωγραφικές περιοχές που συμμετέχουν στους ελέγχους.
- Χρησιμοποιώντας DISTINCT, το 2<sup>ο</sup> ερώτημα εμφανίζει όλα τα διαφορετικά έτη που υπάρχουν στον πίνακα Dimension\_Date. Με αυτόν τον τρόπο, μπορούμε να δούμε για ποιες χρονιές υπάρχουν καταγεγραμμένα δεδομένα.
- Το 3<sup>ο</sup> ερώτημα επιστρέφει όλα τα δεδομένα από τον πίνακα Dimension\_FactElegchou, που πιθανώς περιέχει διαφορετικές κατηγορίες ελέγχων, όπως τεχνικοί έλεγχοι, περιβαλλοντικοί έλεγχοι κ.λπ.
- Στο 4<sup>ο</sup> ερώτημα, Το SUM (CountDTE) υπολογίζει το συνολικό αριθμό των ελέγχων που έχουν πραγματοποιηθεί, όπως αποθηκεύονται στον πίνακα Fact\_Checks. Αυτό μας δίνει μια γενική εικόνα του όγκου των ελέγχων.

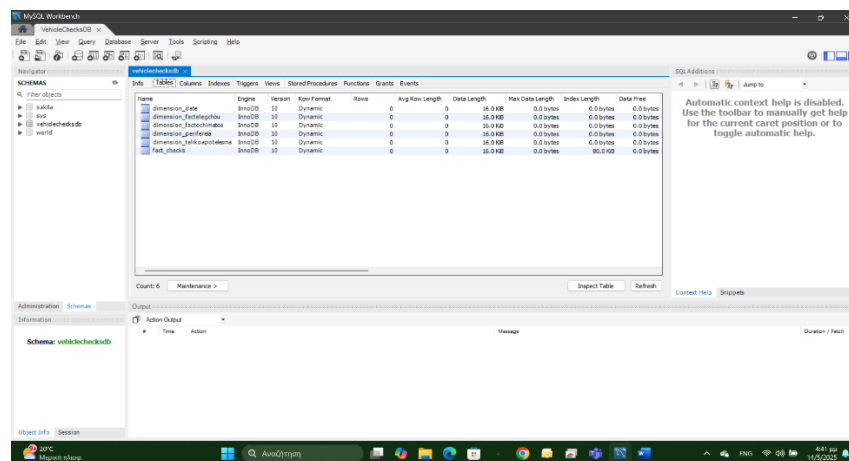
- Χρησιμοποιώντας JOIN στα Fact\_Checks και Dimension\_FactOchimatou, το 5<sup>ο</sup> ερώτημα υπολογίζει το πλήθος των ελέγχων για κάθε τύπο οχήματος (π.χ. αυτοκίνητα, μοτοσυκλέτες, φορτηγά). Η χρήση GROUP BY διασφαλίζει ότι τα αποτελέσματα ομαδοποιούνται κατά τύπο οχήματος.
- Το έκτο ερώτημα επιστρέφει την περιφέρεια με τον μεγαλύτερο αριθμό ελέγχων. Χρησιμοποιώντας SUM (CountDTE), υπολογίζει το σύνολο των ελέγχων ανά περιφέρεια, ταξινομεί τα αποτελέσματα φθίνουσα (ORDER BY DESC) και επιλέγει την πρώτη εγγραφή (LIMIT 1).
- Στο 7<sup>ο</sup> ερώτημα, το ποσοστό επιτυχίας υπολογίζεται ως ο λόγος των επιτυχημένων ελέγχων (Teliko\_Apotelesma\_id = 1) προς το συνολικό αριθμό ελέγχων, εκφρασμένο σε ποσοστό. Αυτό δείχνει την αναλογία των οχημάτων που πέρασαν επιτυχώς τους ελέγχους.
- Το 8<sup>ο</sup> ερώτημα εμφανίζει την κατανομή των διαφορετικών τύπων ελέγχων ανά έτος. Κάνοντας JOIN με τους πίνακες Dimension\_Date και Dimension\_FactElegchou, εξασφαλίζεται ότι για κάθε έτος εμφανίζεται το πλήθος των ελέγχων που ανήκουν σε κάθε κατηγορία.
- Το 9<sup>ο</sup> ερώτημα επιστρέφει τους τύπους οχημάτων που απέτυχαν περισσότερο στους ελέγχους (Teliko\_Apotelesma\_id IN (3,4), που μάλλον αντιπροσωπεύουν αποτυχίες). Τα αποτελέσματα ταξινομούνται φθίνουσα, ώστε τα οχήματα με τις περισσότερες αποτυχίες να εμφανίζονται πρώτα.
- Το 10<sup>ο</sup> ερώτημα συγκρίνει τον αριθμό ελέγχων που πραγματοποιήθηκαν σε διαφορετικές περιφέρειες μεταξύ των ετών 2013 και 2016. Ομαδοποιεί τα δεδομένα με βάση την περιφέρεια και το έτος (GROUP BY p.Perifereia\_Name, d.Year), δίνοντας τη δυνατότητα να δούμε πώς άλλαξε η συχνότητα των ελέγχων στις περιφέρειες μεταξύ των δύο ετών.



## 5.4 Screenshots Λειτουργικότητας

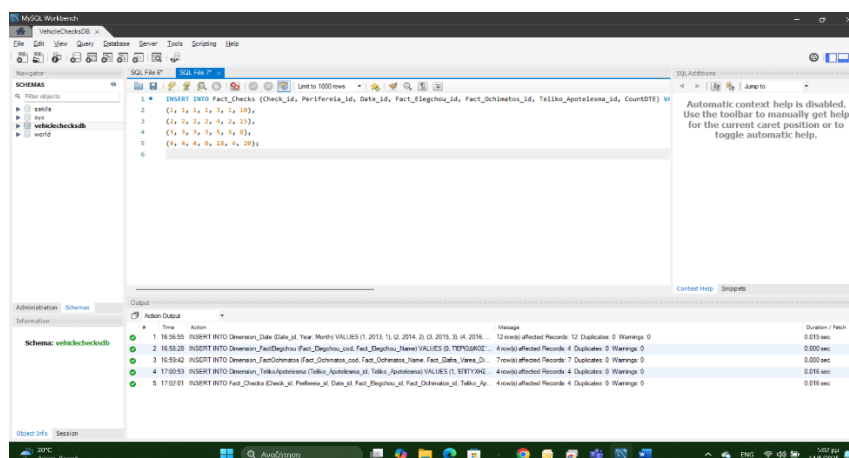
Στα πλαίσια της συγκεκριμένης ενότητας παρουσιάζονται ενδεικτικά screenshots από την βάση που έχει δημιουργηθεί, καθώς και από την εκτέλεση των ερωτημάτων SELECT (η υλοποίηση πραγματοποιήθηκε στον MySQL Server του εργαλείου XAMPP).

Αρχικά, παρατίθεται στιγμιότυπο με τους πίνακες που έχουν δημιουργηθεί:



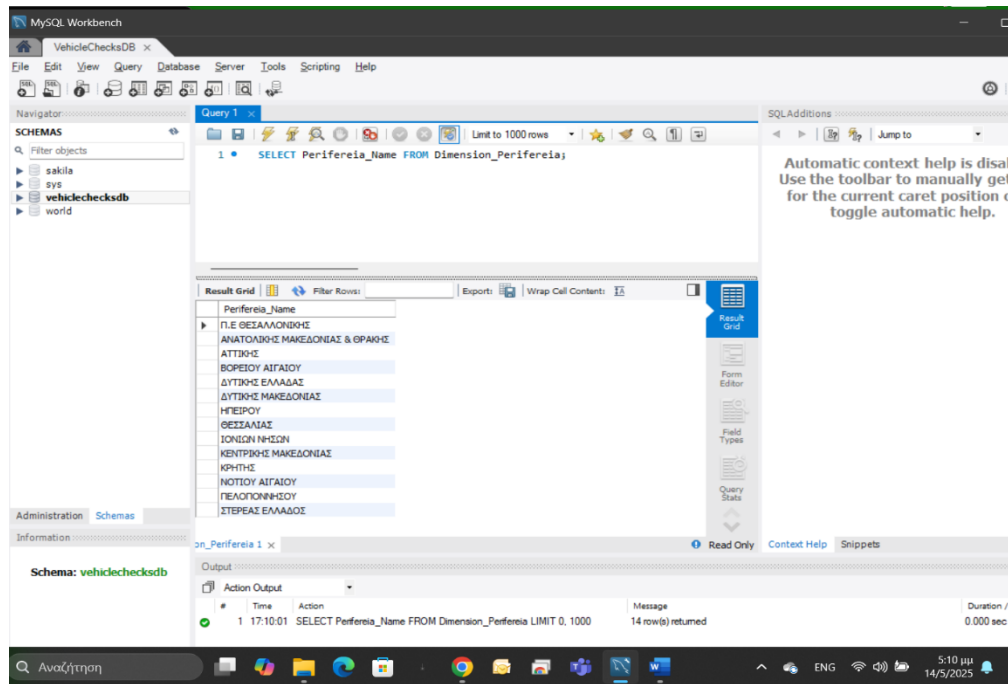
Εικόνα 6: Πίνακες που έχουν δημιουργηθεί - MySQL Server

Ακολουθεί ενδεικτικό στιγμιότυπο με τα περιεχόμενα του πίνακα fact\_checks:

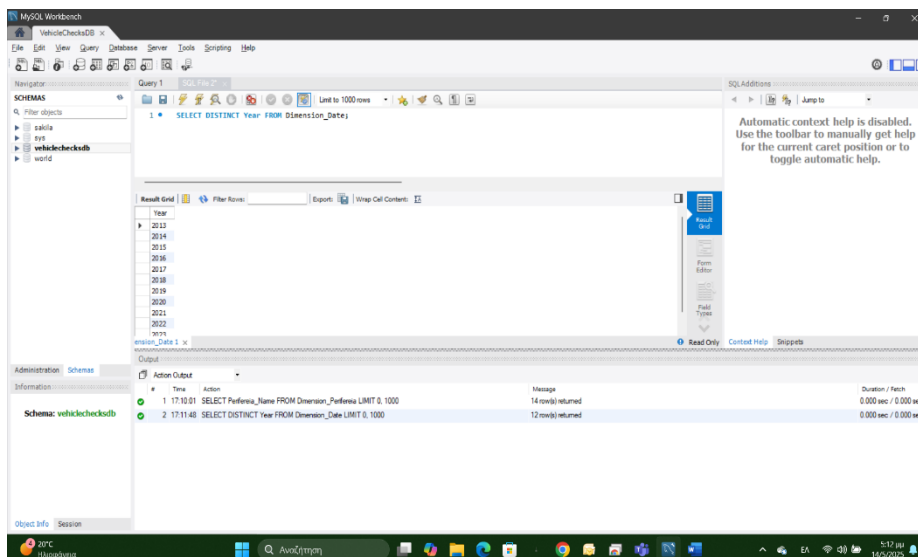


Εικόνα 7: Περιεχόμενα πίνακα fact\_checks - MySQL Server

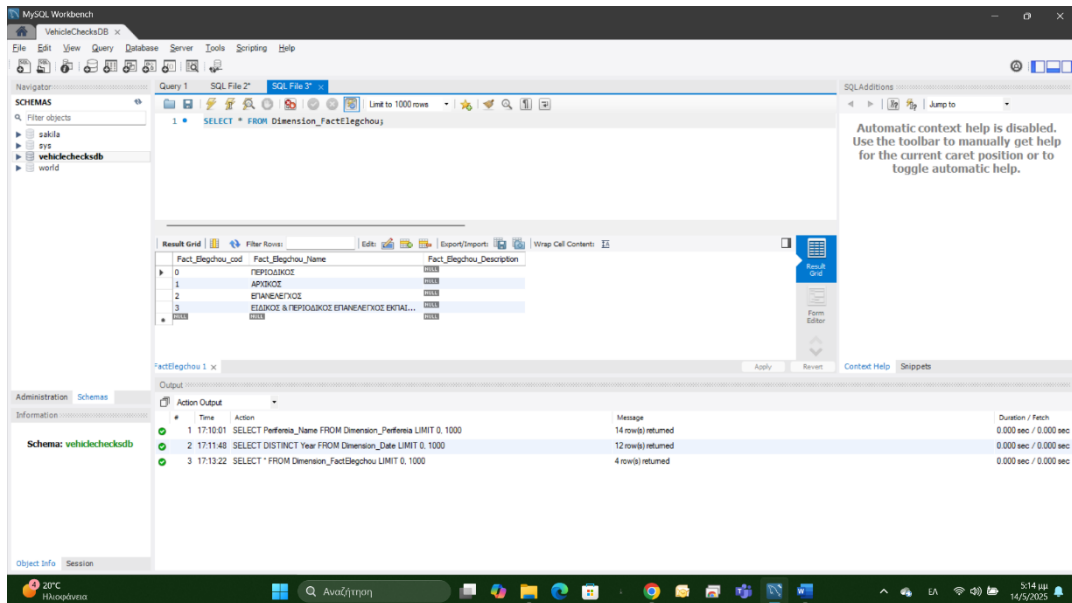
Ακολουθούν στιγμιότυπα από τα αποτελέσματα εκτέλεσης των 10 ερωτημάτων:



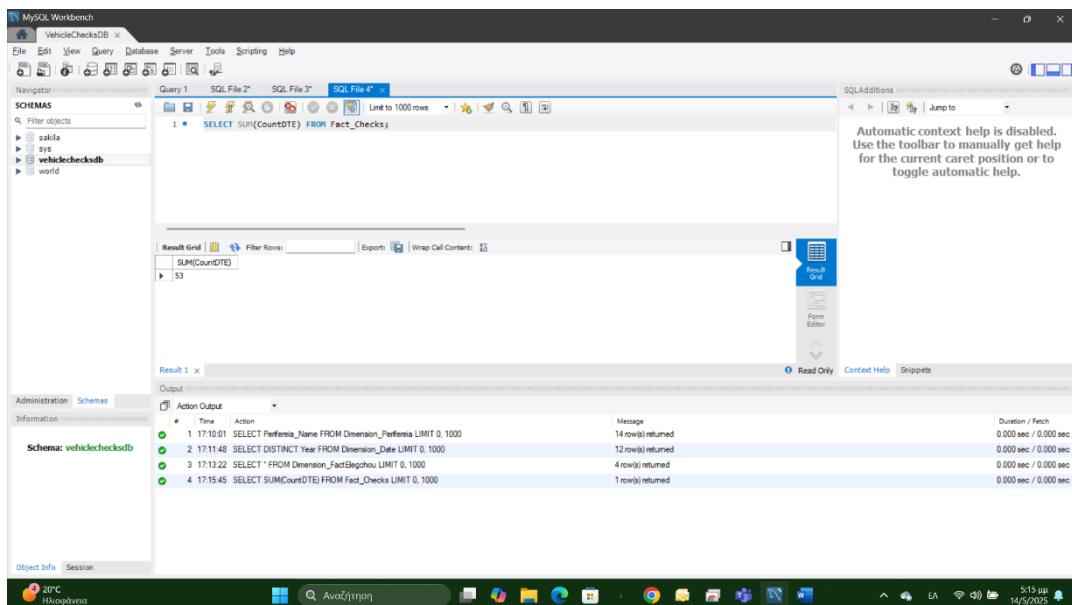
Εικόνα 8: Αποτελέσματα 1ου ερωτήματος - MySQL Server



Εικόνα 9: Αποτελέσματα 2ου ερωτήματος - MySQL Server



Εικόνα 10: Αποτελέσματα 3ου ερωτήματος - MySQL Server



Εικόνα 11: Αποτελέσματα 4ου ερωτήματος - MySQL Server

MySQL Workbench

VehicleChecksDB

Query 1

SQL File 2" SQL File 3" SQL File 4" SQL File 5" SQL File 6"

1. JOIN Fact\_Ochmatos ON Fact\_Checks.Fact\_Ochmatos\_Id = Dimension\_FactOchmatos.Fact\_Ochmatos\_cod GROUP BY Fact\_Ochmatos\_Name;

Result Grid

Fact_Ochmatos_Name	COUNT(*)
ΑΔΕΝΔΟΚΟΡΟ-Τ <3.5-ΔΧ	1
ΑΔΕΝΔΟΚΟΡΟ-Τ <3.5-ΔΧ	1
ΕΓΓΙΣΤΙΚΟ ΕΩΣ 8 ΒΕΛΤΗΣΗ	1
ΝΕΟΚΟΡΕΟ ΑΥΤΟ ΒΕΛΤΗΣΗ 8 ΚΑΙ ΑΝΩ-Τ <=10-ΔΧ	1

Result 1 x

Action Output

#	Time	Action	Message	Duration / Fetch
1	17:10:01	SELECT DISTINCT Year FROM Dimension_Periferea LIMIT 0, 1000	14 row(s) returned	0.000 sec / 0.000 sec
2	17:11:48	SELECT DISTINCT Year FROM Dimension_Date LIMIT 0, 1000	12 row(s) returned	0.000 sec / 0.000 sec
3	17:13:22	SELECT * FROM Dimension_FactBleghou LIMIT 0, 1000	4 row(s) returned	0.000 sec / 0.000 sec
4	17:15:45	SELECT SUM(COUNT(*) FROM Fact_Checks LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
5	17:16:40	SELECT Fact_Ochmatos_Name, COUNT(*) FROM Fact_Checks JOIN Dimension_FactOchmatos ON Fact_Ch...	4 row(s) returned	0.000 sec / 0.000 sec

Εικόνα 12: Αποτελέσματα 5ου ερωτήματος - MySQL Server

MySQL Workbench

VehicleChecksDB

Query 1

SQL File 2" SQL File 3" SQL File 4" SQL File 5" SQL File 6"

1. ON Fact\_Checks.Periferea\_Id = Dimension\_Periferea.Periferea\_Id GROUP BY Periferea\_Name ORDER BY SUM(COUNT(\*) DESC LIMIT 1;

Result Grid

Periferea_Name
ΡΟΠΕΡΟΥ ΑΥΤΑΙΟΥ

Result 1 x

Action Output

#	Time	Action	Message	Duration / Fetch
1	17:10:01	SELECT Periferea_Name FROM Dimension_Periferea LIMIT 0, 1000	14 row(s) returned	0.000 sec / 0.000 sec
2	17:11:48	SELECT DISTINCT Year FROM Dimension_Date LIMIT 0, 1000	12 row(s) returned	0.000 sec / 0.000 sec
3	17:13:22	SELECT * FROM Dimension_FactBleghou LIMIT 0, 1000	4 row(s) returned	0.000 sec / 0.000 sec
4	17:15:45	SELECT SUM(COUNT(*) FROM Fact_Checks LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
5	17:16:40	SELECT Fact_Ochmatos_Name, COUNT(*) FROM Fact_Checks JOIN Dimension_FactOchmatos ON Fact_Ch...	4 row(s) returned	0.000 sec / 0.000 sec
6	17:18:00	SELECT Periferea_Name FROM Fact_Checks JOIN Dimension_Periferea ON Fact_Checks.Periferea_Id = Dim...	1 row(s) returned	0.000 sec / 0.000 sec

Εικόνα 13: Αποτελέσματα 6ου ερωτήματος - MySQL Server

Query 7: `(SUM(CASE WHEN Teliko_Apotelesma_Id = 1 THEN CountDTE ELSE 0 END) * 100.0 / SUM(CountDTE)) AS SuccessRate FROM Fact_Checks;`

Result Grid:

SuccessRate
18.86792

Action Output:

#	Time	Action	Message	Duration / Fetch
1	17:10:01	SELECT Prefencia_Name FROM Dimension_Prefencia LIMIT 0, 1000	14 row(s) returned	0.000 sec / 0.000 sec
2	17:11:48	SELECT DISTINCT Year FROM Dimension_Date LIMIT 0, 1000	12 row(s) returned	0.000 sec / 0.000 sec
3	17:13:22	SELECT * FROM Dimension_FactBleghou LIMIT 0, 1000	4 row(s) returned	0.000 sec / 0.000 sec
4	17:15:45	SELECT SUM(CountDTE) FROM Fact_Checks LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
5	17:16:40	SELECT Fact_Ochmatos_Name, COUNT(*) FROM Fact_Checks JOIN Dimension_FactOchmatos ON Fact_Ch...	4 row(s) returned	0.000 sec / 0.000 sec
6	17:18:00	SELECT Prefencia_Name FROM Fact_Checks JOIN Dimension_Prefencia ON Fact_Checks Prefencia_id = Dim...	1 row(s) returned	0.000 sec / 0.000 sec
7	17:19:07	SELECT (SUM(CASE WHEN Teliko_Apotelesma_Id = 1 THEN CountDTE ELSE 0 END) * 100.0 / SUM(CountD...	1 row(s) returned	0.000 sec / 0.000 sec

Εικόνα 14: Αποτελέσματα 7ου ερωτήματος - MySQL Server

Query 8: `Dimension_FactBleghou ON Fact_Checks.Fact_Elegchou_id = Dimension_FactBleghou.Fact_Elegchou_cod GROUP BY Year, Fact_Elegchou_Name;`

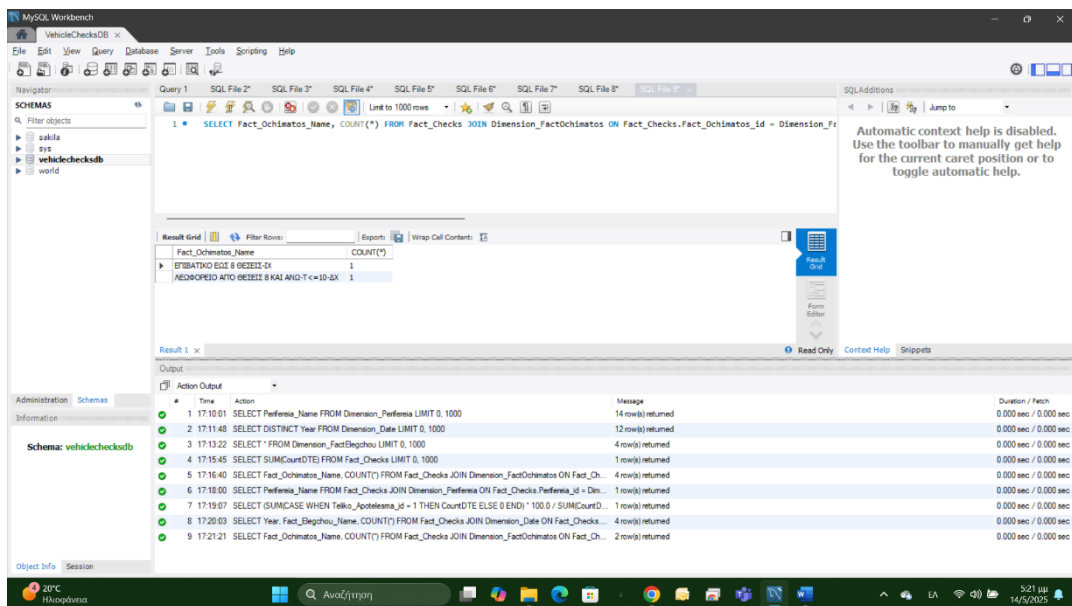
Result Grid:

Year	Fact_Elegchou_Name	COUNT(*)
2013	ΑΠΟΡΙΣΕ	1
2014	ΕΤΑΝΕΛΕΥΤΙΣ	1
2015	ΕΙΔΙΚΟΣ & ΠΕΡΙΟΔΙΚΟΣ ΕΤΑΝΕΛΕΥΤΙΣ ΕΠΙΛ...	1
2016	ΠΕΡΙΟΔΙΚΟΣ	1

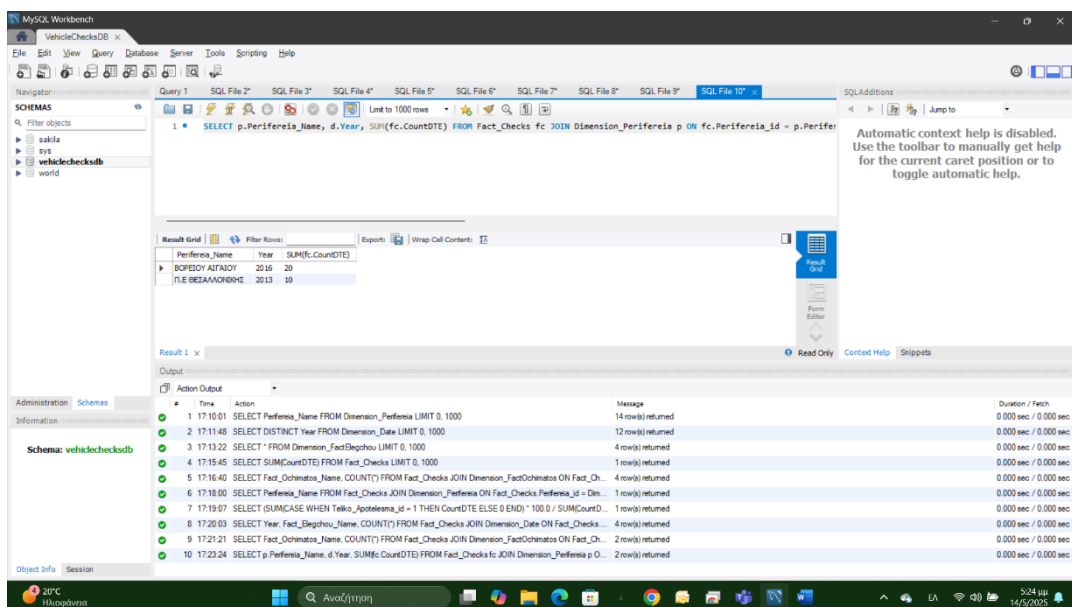
Action Output:

#	Time	Action	Message	Duration / Fetch
1	17:10:01	SELECT Prefencia_Name FROM Dimension_Prefencia LIMIT 0, 1000	14 row(s) returned	0.000 sec / 0.000 sec
2	17:11:48	SELECT DISTINCT Year FROM Dimension_Date LIMIT 0, 1000	12 row(s) returned	0.000 sec / 0.000 sec
3	17:13:22	SELECT * FROM Dimension_FactBleghou LIMIT 0, 1000	4 row(s) returned	0.000 sec / 0.000 sec
4	17:15:45	SELECT SUM(CountDTE) FROM Fact_Checks LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
5	17:16:40	SELECT Fact_Ochmatos_Name, COUNT(*) FROM Fact_Checks JOIN Dimension_FactOchmatos ON Fact_Ch...	4 row(s) returned	0.000 sec / 0.000 sec
6	17:18:00	SELECT Prefencia_Name FROM Fact_Checks JOIN Dimension_Prefencia ON Fact_Checks Prefencia_id = Dim...	1 row(s) returned	0.000 sec / 0.000 sec
7	17:19:07	SELECT (SUM(CASE WHEN Teliko_Apotelesma_Id = 1 THEN CountDTE ELSE 0 END) * 100.0 / SUM(CountD...	1 row(s) returned	0.000 sec / 0.000 sec
8	17:20:03	SELECT Year, Fact_Bleghou_Name, COUNT(*) FROM Fact_Checks JOIN Dimension_Date ON Fact_Checks...	4 row(s) returned	0.000 sec / 0.000 sec

Εικόνα 15: Αποτελέσματα 8ου ερωτήματος - MySQL Server



Εικόνα 16: Αποτελέσματα 9ου ερωτήματος - MySQL Server

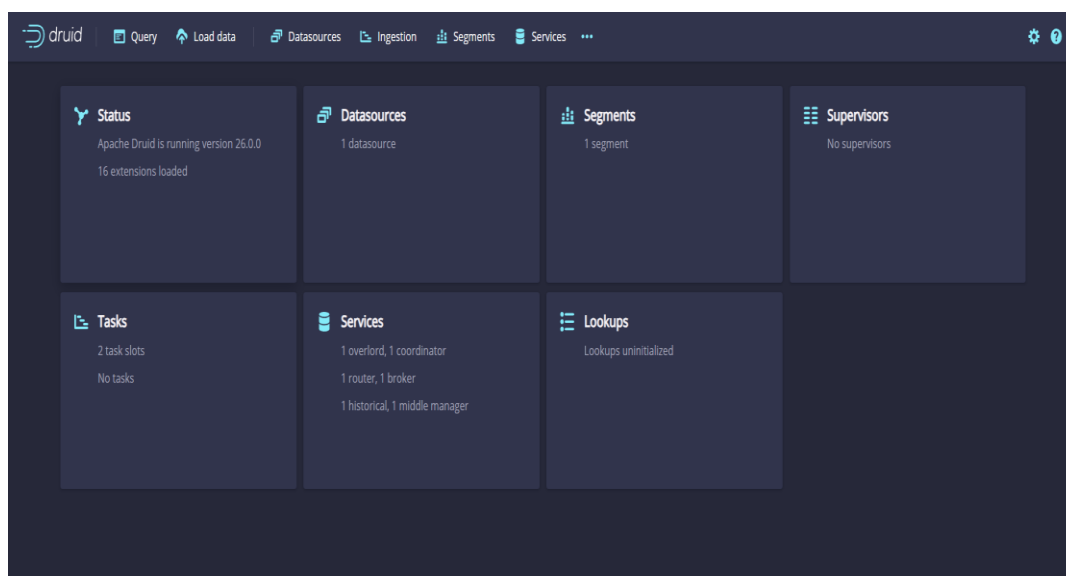


Εικόνα 17: Αποτελέσματα 10ου ερωτήματος - MySQL Server

## 6. Σχεδιασμός και υλοποίηση αποθήκης δεδομένων στο υπολογιστικό νέφος.

### 6.1 Εισαγωγή

Για την υλοποίηση της προηγούμενης αποθήκης δεδομένων με βάση τη λογική του υπολογιστικού νέφους χρησιμοποιήθηκε το λογισμικό Apache Druid [105], για την εγκατάσταση του οποίου ήταν απαραίτητη η χρήση του λειτουργικού συστήματος Linux και πιο συγκεκριμένα της έκδοσης Ubuntu 22.04.



Εικόνα 18: Περιβάλλον Apache Druid

Το Apache Druid είναι ένα ανοιχτού κώδικα, κατακευματισμένο σύστημα διαχείρισης βάσεων δεδομένων που έχει σχεδιαστεί για αναλυτικά ερωτήματα σε πραγματικό χρόνο και γρήγορη επεξεργασία μεγάλων όγκων δεδομένων. Χρησιμοποιείται κυρίως για OLAP εφαρμογές, καθώς υποστηρίζει αποδοτική αποθήκευση και ανάκτηση δεδομένων μέσω στήλης, καθώς και προ-υπολογισμούς για τη βελτίωση των επιδόσεων. Χρησιμοποιεί μια υβριδική αρχιτεκτονική που συνδυάζει batch και

streaming επεξεργασία, επιτρέποντας την άμεση ανάλυση δεδομένων που προέρχονται από πηγές όπως Apache Kafka, Amazon Kinesis και άλλα.

Ένα από τα κύρια πλεονεκτήματα του Apache Druid είναι η ικανότητά του να εκτελεί χαμηλής καθυστέρησης ερωτήματα ακόμα και σε τεράστια σύνολα δεδομένων, γεγονός που το καθιστά ιδανικό για εφαρμογές BI, παρακολούθηση συστημάτων, ανάλυση συμβάντων και διαφημιστική τεχνολογία. Επιπλέον, υποστηρίζει αυτόματη κλιμάκωση και κατανομή φορτίου, γεγονός που το καθιστά κατάλληλο για cloud-based περιβάλλοντα. Ο συνδυασμός της υψηλής απόδοσης, της διαθεσιμότητας και της ικανότητας διαχείρισης δεδομένων σε πραγματικό χρόνο καθιστά το Apache Druid μια δημοφιλή επιλογή για εταιρείες που επιθυμούν να αποκτήσουν γρήγορη και αξιόπιστη επιχειρηματική πληροφόρηση.

## 6.2 Δημιουργία αποθήκης δεδομένων στο Apache Druid

Για την δημιουργία της αποθήκης δεδομένων, ακολουθήθηκε διαφορετική διαδικασία σε σχέση με αυτή της παραδοσιακής αποθήκης δεδομένων. Για τον λόγο αυτό δημιουργήθηκε το παρακάτω .csv αρχείο:

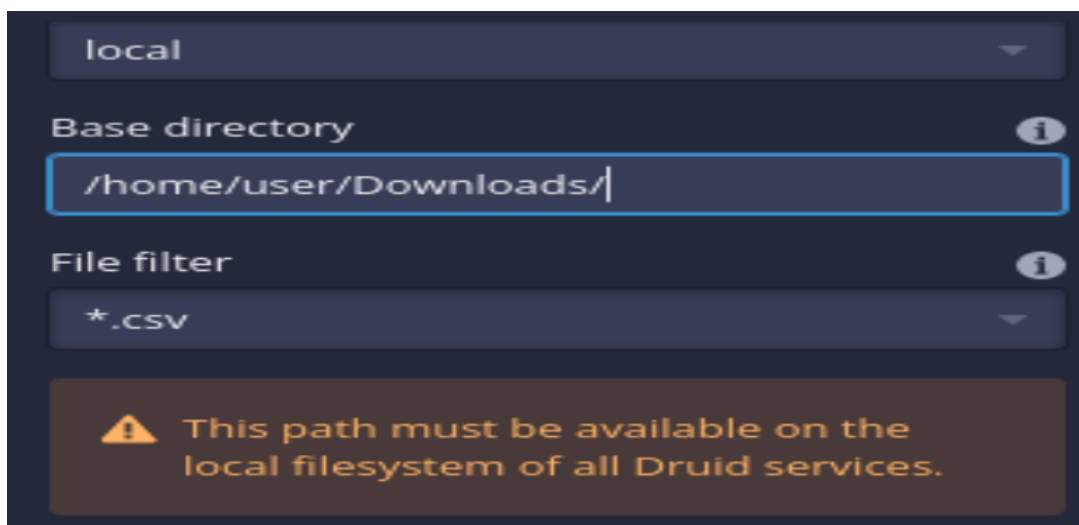
```
"Check_id","Perifereia_id","Perifereia_Name","Perifereia_Area","Date_id","Year","Month","Fact_Elegchou_id","Fact_Elegchou_Name","Fact_Elegchou_Description","Fact_Ochimatots_id","Fact_Ochimatots_Name","Fact_Elafra_Varea_Dikykla","Teliko_Apotelesma_id","Teliko_Apotelesma","CountDTE"
"1","1","Π.Ε  
ΘΕΣΣΑΛΛΟΝΙΚΗΣ","3681","1","2013","1","1","ΑΡΧΙΚΟΣ",NULL,"3","ΑΣΘΕΝΟ  
ΦΟΡΟ-Τ<3.5-ΔΧ","ΕΛΑΦΡΑ","1","ΕΠΙΤΥΧΗΣ","10"
"2","2","ΑΝΑΤΟΛΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ &  
ΘΡΑΚΗΣ","14157","2","2014","2","2","ΕΠΑΝΕΛΕΓΧΟΣ",NULL,"4","ΑΣΘΕΝΟΦ  
ΟΡΟ-Τ<3.5-ΙΧ","ΕΛΑΦΡΑ","2","ΔΕΥΤΕΡΕΥΟΥΣΕΣ ΕΛΛΕΙΨΕΙΣ","15"
"3","3","ΑΤΤΙΚΗΣ","3808","3","2015","3","3","ΕΙΔΙΚΟΣ & ΠΕΡΙΟΔΙΚΟΣ  
ΕΠΑΝΕΛΕΓΧΟΣ ΕΚΠΑΙΔΕΥΤΙΚΟΥ ΟΧΗΜΑΤΟΣ",NULL,"5","ΕΠΙΒΑΤΙΚΟ ΕΩΣ 8  
ΘΕΣΕΙΣ-ΙΧ","ΕΛΑΦΡΑ","3","ΣΟΒΑΡΕΣ ΕΛΛΕΙΨΕΙΣ","8"
"4","4","ΒΟΡΕΙΟΥ  
ΑΙΓΑΙΟΥ","3836","4","2016","4","0","ΠΕΡΙΟΔΙΚΟΣ",NULL,"18","ΛΕΩΦΟΡΕ  
ΙΟ ΑΠΟ ΘΕΣΕΙΣ 8 ΚΑΙ ΑΝΩ-Τ<=10-ΔΧ","ΒΑΡΕΑ","4","ΕΠΙΚΥΝΔΥΝΕΣ  
ΕΛΛΕΙΨΕΙΣ","20"
```



Οι στήλες του .csv αρχείου είναι:

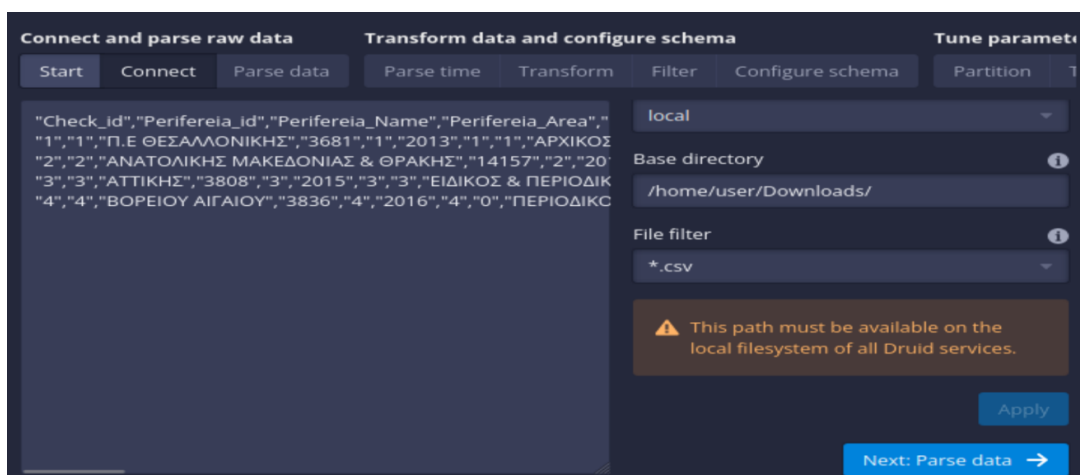
- Check\_id – Μοναδικός αναγνωριστικός αριθμός ελέγχου.
- Perifereia\_id – Αναγνωριστικός αριθμός περιφέρειας.
- Perifereia\_Name – Όνομα περιφέρειας.
- Perifereia\_Area – Έκταση περιφέρειας.
- Date\_id – Αναγνωριστικός αριθμός ημερομηνίας.
- Year – Έτος.
- Month – Μήνας.
- Fact\_Elegchou\_id – Αναγνωριστικός αριθμός κατηγορίας ελέγχου.
- Fact\_Elegchou\_Name – Όνομα κατηγορίας ελέγχου.
- Fact\_Elegchou\_Description – Περιγραφή κατηγορίας ελέγχου (έχει τιμές NULL).
- Fact\_Ochimatos\_id – Αναγνωριστικός αριθμός κατηγορίας οχήματος.
- Fact\_Ochimatos\_Name – Όνομα κατηγορίας οχήματος.
- Fact\_Elafra\_Varea\_Dikykla – Κατηγορία οχήματος (Ελαφρά, Βαρέα, Δίκυκλα).
- Teliko\_Apotelesma\_id – Αναγνωριστικός αριθμός τελικού αποτελέσματος ελέγχου.
- Teliko\_Apotelesma – Τελικό αποτέλεσμα ελέγχου.

Παρακάτω παρατίθενται τα βήματα της διαδικασίας που ακολουθήθηκαν για την εισαγωγή του .csv αρχείου με σκοπό την δημιουργία της cloud-based αποθήκης δεδομένων, παραθέτοντας τα αντίστοιχα στιγμιότυπα οθόνης. Η παρακάτω εικόνα δείχνει τη διεπαφή ρύθμισης για την εισαγωγή δεδομένων στο **Apache Druid**, όπου έχει οριστεί η διαδρομή "/home/user/Downloads/" ως τοπικός φάκελος αποθήκευσης αρχείων CSV. Ωστόσο, εμφανίζεται ένα μήνυμα προειδοποίησης που αναφέρει ότι η συγκεκριμένη διαδρομή πρέπει να είναι διαθέσιμη στο τοπικό σύστημα αρχείων όλων των υπηρεσιών του Druid:



Εικόνα 19: Δημιουργία αποθήκης Apache Druid - Βήμα 1

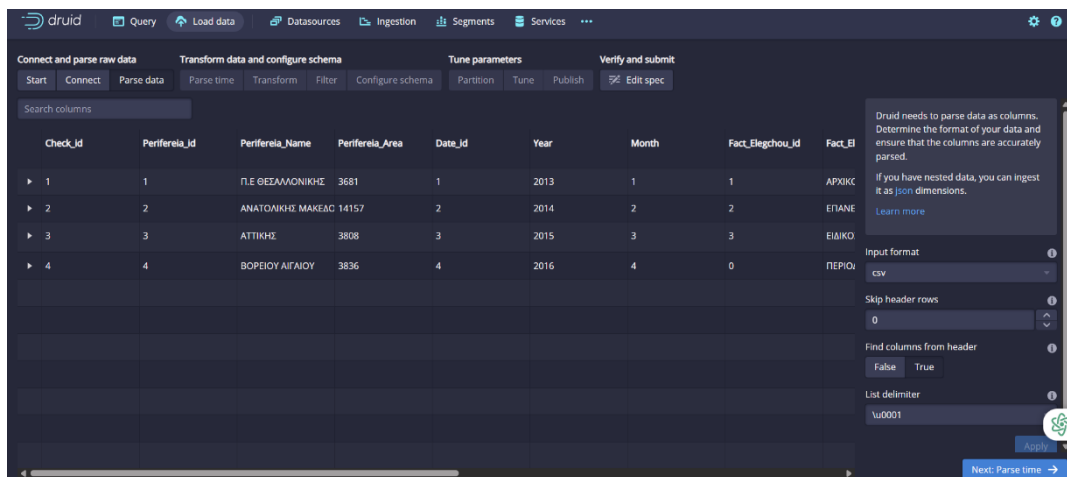
Η επόμενη εικόνα δείχνει τη διεπαφή χρήστη του Apache Druid για τη φόρτωση και επεξεργασία δεδομένων από ένα αρχείο CSV. Η πρώτη στήλη εμφανίζει τα ακατέργαστα δεδομένα, τα οποία περιέχουν πληροφορίες όπως "Check\_id", "Perifereia\_id", "Perifereia\_Name". Η δεξιά πλευρά της εικόνας περιλαμβάνει επιλογές για τον ορισμό της διαδρομής αρχείων, όπου φαίνεται ότι το σύστημα αναζητά αρχεία .csv στον φάκελο "/home/user/Downloads/":



Εικόνα 20: Δημιουργία αποθήκης Apache Druid - Βήμα 2

Η επόμενη εικόνα δείχνει ένα βήμα της διαδικασίας φόρτωσης δεδομένων στο Apache Druid, όπου τα δεδομένα από το αρχείο CSV έχουν αναλυθεί και εμφανίζονται σε στήλες. Στη δεξιά πλευρά εμφανίζεται ένα ενημερωτικό μήνυμα που εξηγεί ότι το Druid πρέπει να αναλύσει τα δεδομένα σε μορφή στηλών και ότι οι χρήστες μπορούν να διαμορφώσουν τον τρόπο με τον οποίο γίνεται η ανάλυση. Παρέχεται επίσης η επιλογή εισαγωγής nested data σε μορφή JSON, αν τα δεδομένα περιέχουν πιο πολύπλοκες δομές. Ορίστηκε σωστά το CSV ως μορφή εισαγωγής, αλλά υπάρχει η επιλογή να παραλειφθούν γραμμές κεφαλίδας (Skip header rows), αν το αρχείο περιέχει επικεφαλίδες που δεν χρειάζεται να αναλυθούν ως δεδομένα.

Από αυτή τη φάση, το επόμενο βήμα είναι η διαμόρφωση του χρόνου ανάλυσης δεδομένων, που θα επιτρέψει στο Druid να διαχειριστεί σωστά χρονικές σφραγίδες ή άλλα χρονικά δεδομένα. Εάν δεν υπάρχει κατάλληλη μορφοποίηση στηλών, μπορεί να χρειαστούν μετασχηματισμοί ή ρυθμίσεις στο σχήμα δεδομένων.

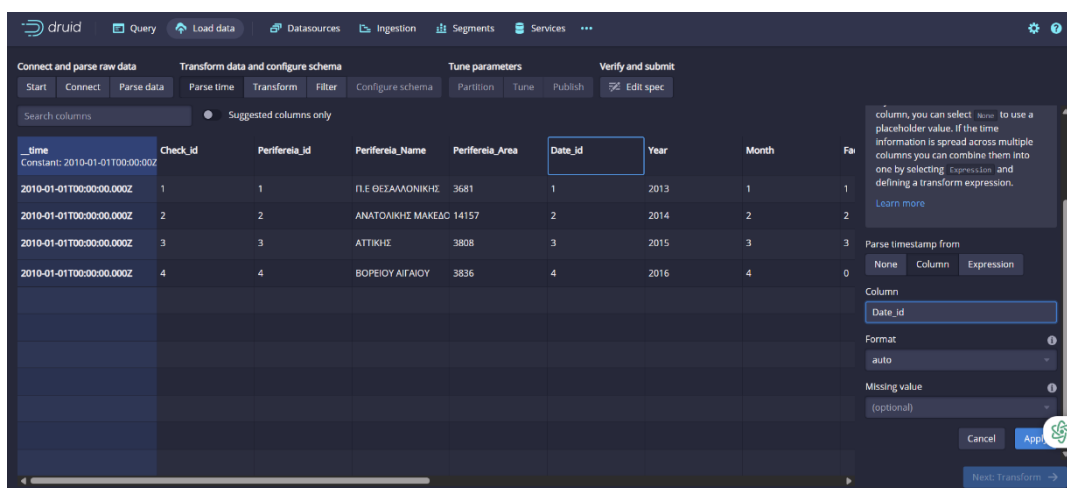


Εικόνα 21: Δημιουργία αποθήκης Apache Druid - Βήμα 3

Η επόμενη εικόνα δείχνει τη διαδικασία ρύθμισης της χρονοσήμανσης στο Apache Druid, κατά τη μετατροπή και διαμόρφωση του σχήματος δεδομένων. Στη συγκεκριμένη περίπτωση, έχει επιλεγεί η στήλη "Date\_id" ως πιθανή χρονική

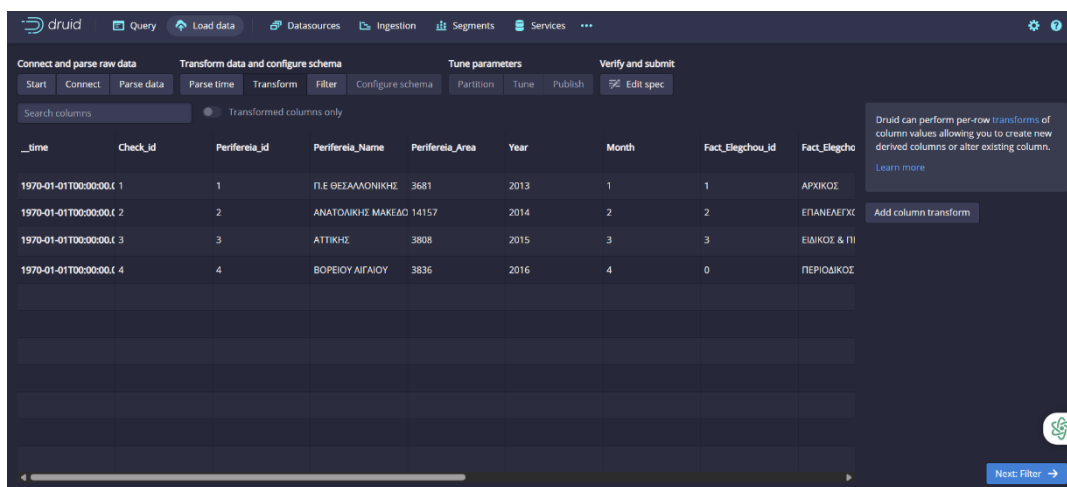
αναφορά, ωστόσο, η μορφή είναι ορισμένη ως "auto", που σημαίνει ότι το Druid θα προσπαθήσει να εντοπίσει αυτόματα τη μορφή της ημερομηνίας.

Εάν η στήλη Date\_id περιέχει ημερομηνίες σε μη αναγνωρίσιμη μορφή (π.χ. αριθμητική αναπαράσταση χωρίς καθορισμένη μορφή ημερομηνίας), μπορεί να χρειαστεί να οριστεί χειροκίνητα μια συγκεκριμένη μορφή, όπως "yyyy-MM-dd" ή "timestamp", ανάλογα με τη δομή των δεδομένων. Επιπλέον, η επιλογή "Missing value" παρέχει τη δυνατότητα καθορισμού προεπιλεγμένης τιμής, εάν κάποια εγγραφή δεν έχει έγκυρη τιμή για τη στήλη ημερομηνίας.



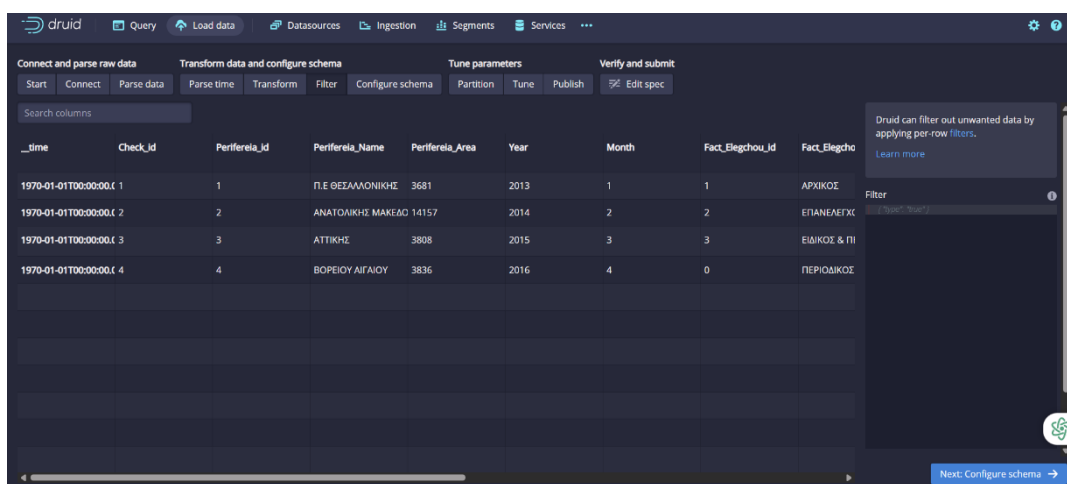
Εικόνα 22: Δημιουργία αποθήκης Apache Druid - Βήμα 4

Στην επόμενη εικόνα, ο χρήστης μπορεί να εφαρμόσει μετασχηματισμούς (transforms) σε στήλες δεδομένων, επιτρέποντας τη δημιουργία νέων ή την τροποποίηση υπαρχουσών στηλών. Στην κορυφή υπάρχουν επιλογές όπως Transform, Filter, Configure schema και Partition, που σχετίζονται με τη διαχείριση δεδομένων. Υπάρχει επίσης ένα κουμπί "Add column transform", το οποίο επιτρέπει την προσθήκη νέων μετασχηματισμών, και ένα κουμπί "Next: Filter", που υποδηλώνει τη μετάβαση στο επόμενο βήμα της διαδικασίας.



Εικόνα 23: Δημιουργία αποθήκης Apache Druid - Βήμα 5

Η επόμενη εικόνα δείχνει την καρτέλα Filter, όπου ο χρήστης μπορεί να εφαρμόσει φίλτρα για να αποκλείσει ανεπιθύμητα δεδομένα σε επίπεδο γραμμής. Το μήνυμα στην κορυφή εξηγεί αυτή τη λειτουργία και παρέχει έναν σύνδεσμο "Learn more" για περισσότερες πληροφορίες. Στο κάτω μέρος, υπάρχει ένα κουμπί "Next: Configure schema", που υποδηλώνει το επόμενο βήμα της διαδικασίας διαμόρφωσης του σχήματος των δεδομένων.



Εικόνα 24: Δημιουργία αποθήκης Apache Druid - Βήμα 6

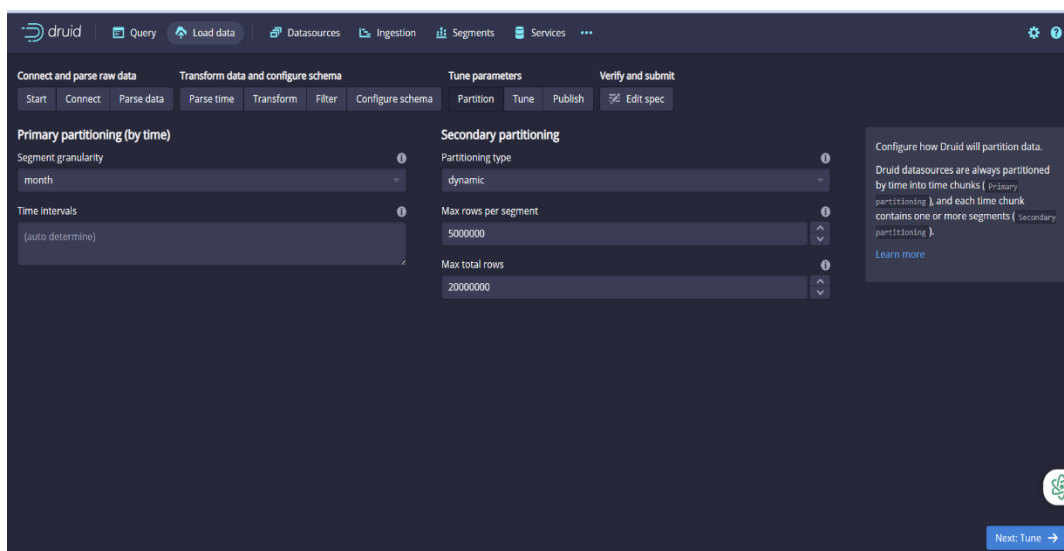
Η επόμενη εικόνα δείχνει την καρτέλα Configure schema, όπου ο χρήστης μπορεί να διαμορφώσει το σχήμα των δεδομένων. Η επιλογή "Explicitly specify schema" είναι ενεργοποιημένη, επιτρέποντας στον χρήστη να ορίσει με ακρίβεια τη δομή των δεδομένων. Το "Rollup" είναι απενεργοποιημένο, που σημαίνει ότι δεν εφαρμόζεται συγκέντρωση δεδομένων κατά την εισαγωγή. Η "Query granularity" έχει οριστεί σε "day", υποδηλώνοντας ότι τα δεδομένα θα ερωτώνται σε ημερήσια ανάλυση. Υπάρχουν επιλογές για προσθήκη διαστάσεων ("Add dimension") και μετρικών ("Add metric"), καθώς και ένα κουμπί εργαλείων ("Tools"). Στο κάτω μέρος, το κουμπί "Next: Partition" οδηγεί στο επόμενο βήμα, που αφορά τη διαμόρφωση της διαμέρισης των δεδομένων.



Εικόνα 25: Δημιουργία αποθήκης Apache Druid - Βήμα 7

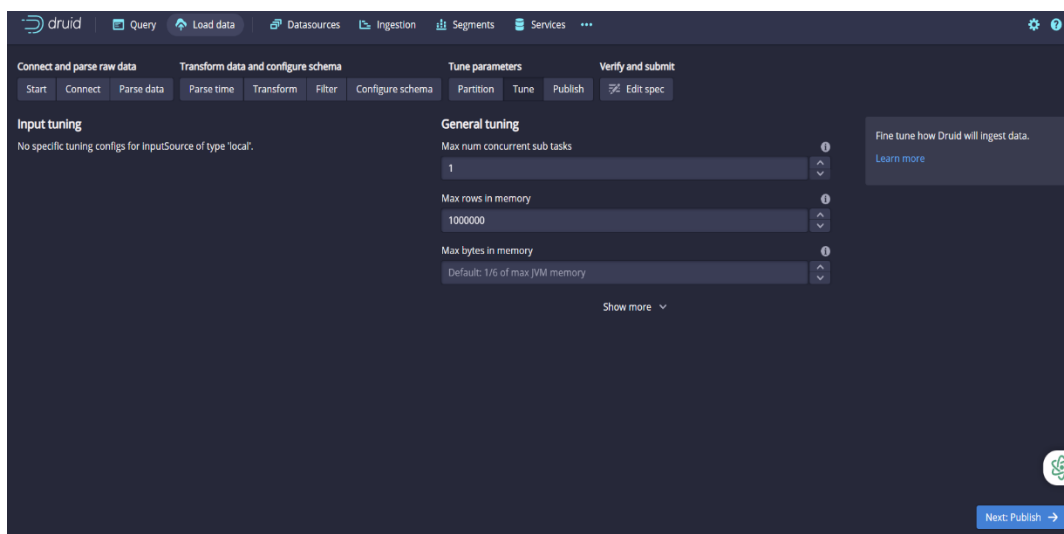
Η επόμενη εικόνα δείχνει Partition, όπου ο χρήστης μπορεί να διαμορφώσει τη διαμέριση των δεδομένων. Η πρωτεύουσα διαμέριση (Primary partitioning) γίνεται βάσει χρόνου, με τη "Segment granularity" ρυθμισμένη σε "month", ενώ το πεδίο των χρονικών διαστημάτων είναι στη ρύθμιση "auto determine". Η δευτερεύουσα διαμέριση (Secondary partitioning) είναι ρυθμισμένη σε "dynamic", με μέγιστο αριθμό γραμμών ανά τμήμα 5.000.000 και συνολικό μέγιστο αριθμό γραμμών 20.000.000. Δεξιά, υπάρχει μια περιγραφή για το πώς γίνεται η διαμέριση των

δεδομένων στο Druid, εξηγώντας ότι η πρωτεύουσα διαμέριση βασίζεται στο χρόνο, ενώ κάθε χρονικό τμήμα περιέχει ένα ή περισσότερα τμήματα μέσω δευτερεύουσας διαμέρισης. Το κουμπί "Next: Tune" οδηγεί στο επόμενο βήμα της διαδικασίας.



Εικόνα 26: Δημιουργία αποθήκης Apache Druid - Βήμα 8

Η επόμενη εικόνα δείχνει την καρτέλα Tune, όπου ο χρήστης μπορεί να προσαρμόσει τις ρυθμίσεις εισαγωγής δεδομένων. Στην ενότητα "Input tuning", αναφέρεται ότι δεν υπάρχουν συγκεκριμένες ρυθμίσεις συντονισμού για πηγές εισόδου τύπου 'local'. Στην ενότητα "General tuning", υπάρχουν διάφορες παράμετροι για τον έλεγχο της μνήμης και της ταχύτητας εισαγωγής δεδομένων. Ο μέγιστος αριθμός ταυτόχρονων υποεργασιών έχει οριστεί σε 1, οι μέγιστες γραμμές στη μνήμη είναι 100.000, ενώ το μέγιστο μέγεθος σε bytes στη μνήμη είναι στην προεπιλεγμένη τιμή (1/6 της συνολικής διαθέσιμης μνήμης). Δεξιά, υπάρχει μια περιγραφή που εξηγεί ότι αυτές οι ρυθμίσεις επιτρέπουν τη λεπτομερή διαμόρφωση της διαδικασίας εισαγωγής δεδομένων στο Druid. Στο κάτω μέρος, υπάρχει το κουμπί "Next: Publish", που οδηγεί στο επόμενο βήμα της διαδικασίας.

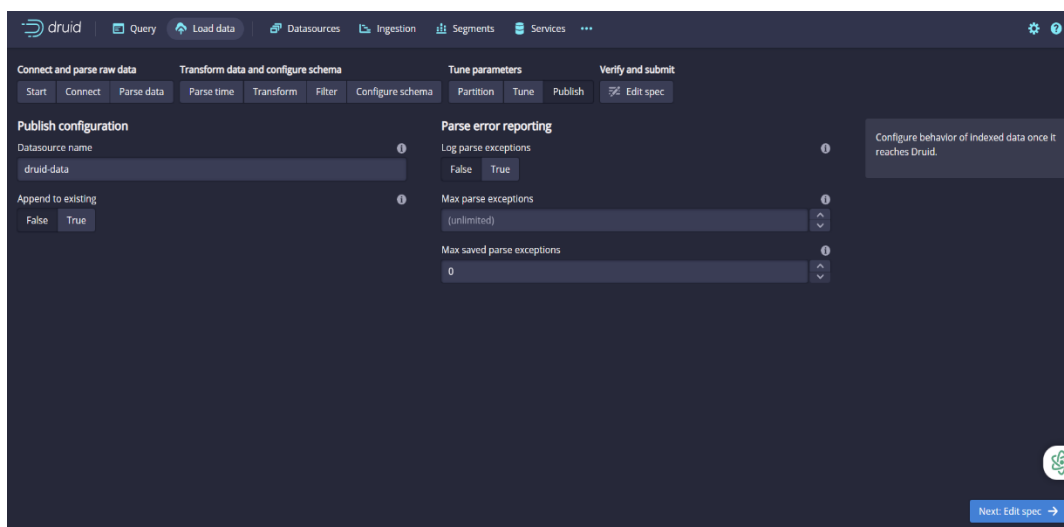


Εικόνα 27: Δημιουργία αποθήκης Apache Druid - Βήμα 9

Η επόμενη εικόνα δείχνει την καρτέλα Publish, όπου ο χρήστης μπορεί να διαμορφώσει τη συμπεριφορά των δεδομένων μετά την εισαγωγή τους στο σύστημα. Στην ενότητα "Publish configuration", το όνομα της πηγής δεδομένων έχει οριστεί ως "Downloads". Η επιλογή "Append to existing" είναι ρυθμισμένη σε "False", που σημαίνει ότι τα νέα δεδομένα δεν θα προστεθούν σε υπάρχοντα σύνολα. Υπάρχει επίσης η επιλογή "Use concurrent locks", η οποία είναι απενεργοποιημένη. Επιπλέον, στην ενότητα "Parse error reporting", υπάρχουν ρυθμίσεις για τη διαχείριση σφαλμάτων ανάλυσης δεδομένων. Η καταγραφή εξαιρέσεων είναι απενεργοποιημένη ("False"), ο μέγιστος αριθμός εξαιρέσεων είναι απεριόριστος, και ο μέγιστος αριθμός αποθηκευμένων εξαιρέσεων έχει οριστεί σε 0.

Στα δεξιά, υπάρχει μια περιγραφή που εξηγεί ότι αυτή η ενότητα επιτρέπει τη διαμόρφωση της συμπεριφοράς των δεδομένων μετά την ευρετηρίασή τους στο Druid. Στο κάτω μέρος, υπάρχει το κουμπί "Next: Edit spec", που οδηγεί στο επόμενο βήμα της διαδικασίας.





Εικόνα 28: Δημιουργία αποθήκης Apache Druid - Βήμα 10

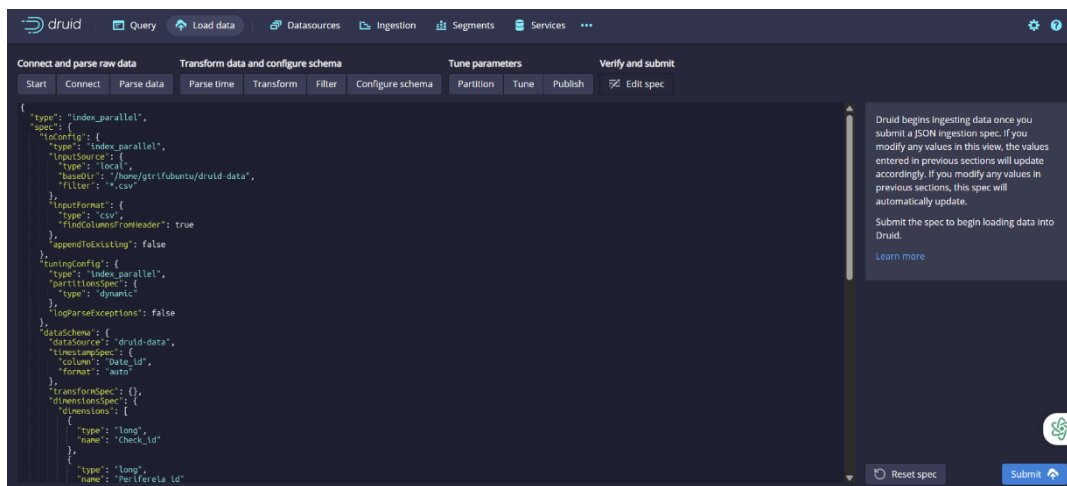
Η επόμενη εικόνα δείχνει το περιβάλλον χρήστη του **Apache Druid**, συγκεκριμένα την καρτέλα **Edit spec**, όπου ο χρήστης μπορεί να επεξεργαστεί το JSON Ingestion Spec πριν από την εισαγωγή των δεδομένων στο σύστημα.

Στην αριστερή πλευρά, εμφανίζεται το JSON configuration, το οποίο περιλαμβάνει τις εξής βασικές ρυθμίσεις :

- Το dataSource είναι "Downloads".
- Στην ενότητα timestampSpec, η στήλη "Date\_id" έχει οριστεί ως η χρονική σήμανση, με τη μορφή "auto".
- Το dimensionsSpec περιέχει διάφορες διαστάσεις όπως "Check\_id", "Periferia\_id", "Periferia\_Name", "Periferia\_Area" και "Year", όλες δηλωμένες ως τύπου "long".
- Η ενότητα transformSpec είναι κενή, που σημαίνει ότι δεν εφαρμόζονται μετασχηματισμοί στα δεδομένα κατά την εισαγωγή.

Στη δεξιά πλευρά, υπάρχει μια περιγραφή που εξηγεί ότι το Druid ξεκινά την εισαγωγή δεδομένων μόλις υποβληθεί αυτό το spec. Αν γίνουν αλλαγές σε προηγούμενα βήματα, αυτές θα ενημερωθούν αυτόματα εδώ. Στο κάτω μέρος,

υπάρχει το κουμπί "Submit task", το οποίο επιτρέπει την έναρξη της διαδικασίας φόρτωσης των δεδομένων στο Druid.



Εικόνα 29: Δημιουργία αποθήκης Apache Druid - Βήμα 11

## 6.3 Υλοποίηση ερωτημάτων προς εκτέλεση και αξιολόγηση

Ακολουθεί ο κώδικας των ερωτημάτων, όπως αυτά τροποποιήθηκαν ώστε να είναι συμβατά με την υλοποίηση στο Apache Druid:

```
-- 1. Προβολή όλων των περιφερειών
SELECT DISTINCT Perifereia_Name
FROM druid-data
ORDER BY Perifereia_Name;

-- 2. Εμφάνιση όλων των ετών που καταγράφονται στη βάση
SELECT DISTINCT Year
FROM druid-data
ORDER BY Year;

-- 3. Προβολή όλων των τύπων ελέγχων
SELECT DISTINCT Fact_Elegchou_Name
FROM druid-data
ORDER BY Fact_Elegchou_Name;

-- 4. Σύνολο ελέγχων που έχουν πραγματοποιηθεί
SELECT SUM(CountDTE) AS Total_Checks
FROM druid-data;
```

```
-- 5. Πλήθος ελέγχων ανά τύπο οχήματος
SELECT Fact_Ochimatosis_Name, SUM(*) AS Record_Count
FROM druid-data
GROUP BY Fact_Ochimatosis_Name
ORDER BY Record_Count DESC;

-- 6. Περιφέρεια με τους περισσότερους ελέγχους
SELECT Periferia_Name, SUM(CountDTE) AS Total_Checks
FROM druid-data
GROUP BY Periferia_Name
ORDER BY Total_Checks DESC
LIMIT 1;

-- 7. Ποσοστό επιτυχίας ελέγχων
SELECT
    (SUM(CASE WHEN Teliko_Apotelesma = 'ΕΠΙΤΥΧΗΣ' THEN CAST(CountDTE
AS DOUBLE) ELSE 0 END) * 100.0)
    / SUM(CAST(CountDTE AS DOUBLE)) AS Success_Rate
FROM druid-data;

-- 8. Κατανομή τύπων ελέγχου ανά έτος
SELECT Year, Fact_Elegchou_Name, COUNT(*) AS Total_Records
FROM druid-data
GROUP BY Year, Fact_Elegchou_Name
ORDER BY Year, Total_Records DESC;

-- 9. Τα οχήματα με τη μεγαλύτερη αποτυχία στους ελέγχους
SELECT Fact_Ochimatosis_Name, COUNT(*) AS Failed_Checks
FROM druid-data
WHERE Teliko_Apotelesma IN ('ΣΟΒΑΡΕΣ ΕΛΛΕΙΨΕΙΣ', 'ΕΠΙΚΥΝΔΥΝΕΣ
ΕΛΛΕΙΨΕΙΣ')
GROUP BY Fact_Ochimatosis_Name
ORDER BY Failed_Checks DESC;

--10. Σύγκριση αριθμού ελέγχων σε διαφορετικές περιφέρειες μεταξύ
2013 και 2016
SELECT Periferia_Name,
    SUM(CASE WHEN Year = 2013 THEN CAST(CountDTE AS DOUBLE)
ELSE 0 END) AS Checks_2013,
    SUM(CASE WHEN Year = 2016 THEN CAST(CountDTE AS DOUBLE)
ELSE 0 END) AS Checks_2016
FROM druid-data
WHERE Year IN (2013, 2016)
GROUP BY Periferia_Name
ORDER BY Periferia_Name;
```

Το πρώτο ερώτημα επιστρέφει όλες τις μοναδικές περιφέρειες που υπάρχουν στην αποθήκη δεδομένων. Αυτό γίνεται με τη χρήση της λειτουργίας DISTINCT, η οποία εξασφαλίζει ότι δεν υπάρχουν διπλότυπες εγγραφές. Η ταξινόμηση των αποτελεσμάτων γίνεται αλφαβητικά μέσω ORDER BY, ώστε η λίστα των περιφερειών να εμφανίζεται με οργανωμένο τρόπο.

Το δεύτερο ερώτημα εμφανίζει όλα τα διαφορετικά έτη που καταγράφονται στην αποθήκη δεδομένων. Χρησιμοποιεί επίσης το DISTINCT για την αποφυγή διπλότυπων τιμών και ταξινομεί τα έτη σε αύξουσα σειρά με τη χρήση ORDER BY. Αυτό το ερώτημα βοηθά στην κατανόηση του χρονικού εύρους των δεδομένων και στη διερεύνηση των διαθέσιμων περιόδων για περαιτέρω ανάλυση. Το τρίτο ερώτημα επιστρέφει όλους τους μοναδικούς τύπους ελέγχων που υπάρχουν στην αποθήκη δεδομένων. Η χρήση του DISTINCT διασφαλίζει ότι κάθε τύπος εμφανίζεται μόνο μία φορά, ενώ η ταξινόμηση με ORDER BY επιτρέπει την εύκολη αναζήτηση και κατανόηση των διαφόρων τύπων ελέγχου που πραγματοποιούνται στα οχήματα.

Το τέταρτο ερώτημα υπολογίζει το συνολικό αριθμό ελέγχων που έχουν πραγματοποιηθεί. Αυτό επιτυγχάνεται με τη χρήση της συνάρτησης SUM ( ), η οποία αθροίζει όλες τις τιμές της στήλης CountDTE. Το αποτέλεσμα παρέχει μια συνολική εικόνα της δραστηριότητας των ελέγχων, δίνοντας μια γενική εκτίμηση του συνολικού αριθμού τους στην αποθήκη δεδομένων.

Το πέμπτο ερώτημα ομαδοποιεί τους ελέγχους με βάση τον τύπο οχήματος και υπολογίζει το σύνολο των ελέγχων για κάθε τύπο. Αυτό γίνεται μέσω GROUP BY, που διαχωρίζει τα δεδομένα ανά τύπο οχήματος, και SUM ( ), που αθροίζει τον αριθμό των ελέγχων. Επιπλέον, τα αποτελέσματα ταξινομούνται σε φθίνουσα σειρά με ORDER BY, ώστε να εμφανίζονται πρώτα τα οχήματα με τους περισσότερους ελέγχους.

Το έκτο ερώτημα εντοπίζει την περιφέρεια με τους περισσότερους ελέγχους. Χρησιμοποιεί GROUP BY για την ομαδοποίηση των δεδομένων ανά περιφέρεια, SUM ( ) για τον υπολογισμό του συνόλου των ελέγχων και ORDER BY DESC για την κατάταξη από την υψηλότερη στη χαμηλότερη τιμή. Με το LIMIT 1, επιστρέφεται μόνο η περιφέρεια με τον μεγαλύτερο αριθμό ελέγχων.

Το έβδομο ερώτημα υπολογίζει το ποσοστό επιτυχίας των ελέγχων. Χρησιμοποιεί τη συνάρτηση SUM ( ) μαζί με τη συνθήκη CASE WHEN, ώστε να υπολογιστεί το πλήθος των επιτυχημένων ελέγχων. Το ποσοστό προκύπτει διαιρώντας το

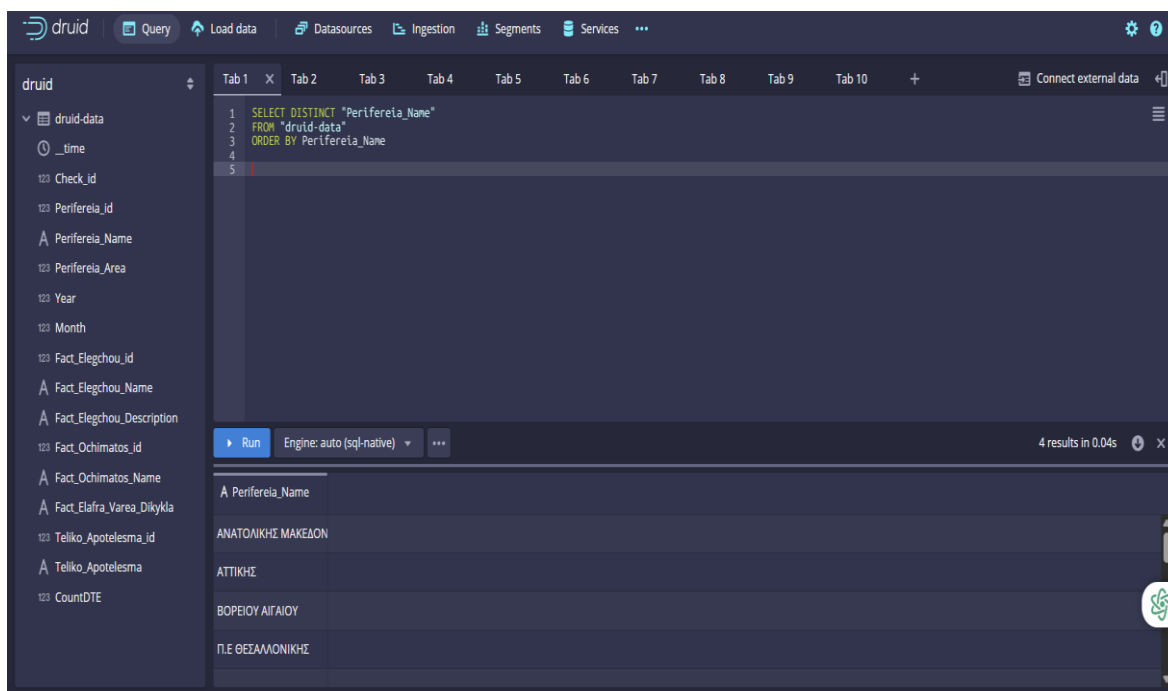
σύνολο των επιτυχημένων ελέγχων με το σύνολο όλων των ελέγχων και πολλαπλασιάζοντας με 100 για να μετατραπεί σε ποσοστό.

Το όγδοο ερώτημα εμφανίζει την κατανομή των τύπων ελέγχου ανά έτος. Αυτό επιτυγχάνεται μέσω GROUP BY, που ομαδοποιεί τα δεδομένα με βάση το έτος και τον τύπο ελέγχου, και SUM ( ), που αθροίζει τον αριθμό των ελέγχων για κάθε κατηγορία. Η ταξινόμηση των αποτελεσμάτων γίνεται με ORDER BY, ώστε να παρουσιάζονται πρώτα τα έτη και στη συνέχεια οι πιο συχνοί τύποι ελέγχου.

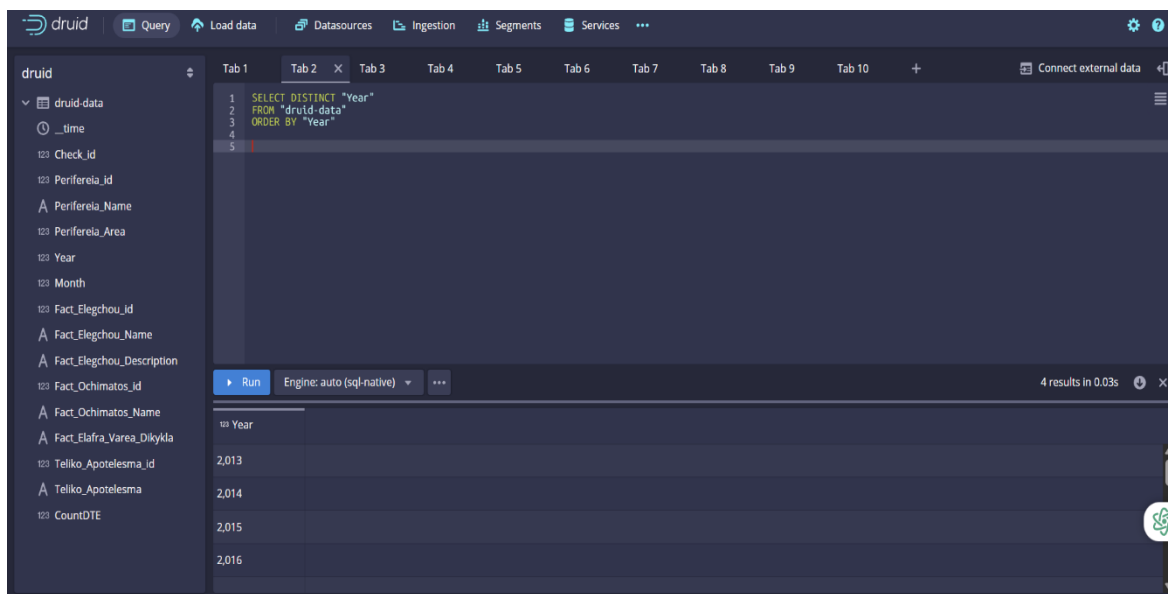
Το ένατο ερώτημα βρίσκει τα οχήματα με τις περισσότερες αποτυχίες στους ελέγχους. Φιλτράρει τα δεδομένα με WHERE Teliko\_Apotelesma = 'ΣΟΒΑΡΕΣ ΕΛΛΕΙΨΕΙΣ', 'ΣΟΒΑΡΕΣ ΕΛΛΕΙΨΕΙΣ' ώστε να συμπεριλάβει μόνο τις αποτυχημένες επιθεωρήσεις. Στη συνέχεια, με GROUP BY και SUM ( ), υπολογίζει το πλήθος των αποτυχιών για κάθε τύπο οχήματος, ενώ το ORDER BY DESC ταξινομεί τα αποτελέσματα από τις περισσότερες προς τις λιγότερες αποτυχίες.

Το δέκατο ερώτημα συγκρίνει τον αριθμό των ελέγχων που πραγματοποιήθηκαν σε διάφορες περιφέρειες μεταξύ των ετών 2013 και 2016. Χρησιμοποιεί SUM ( ), με συνθήκες CASE WHEN για να υπολογίσει ξεχωριστά τον αριθμό ελέγχων για κάθε έτος. Το WHERE Year IN (2013, 2016), διασφαλίζει ότι λαμβάνονται υπόψη μόνο τα δύο επιλεγμένα έτη, ενώ το GROUP BY επιτρέπει τη σύγκριση ανά περιφέρεια. Τα αποτελέσματα ταξινομούνται αλφαβητικά με ORDER BY, ώστε η σύγκριση να είναι πιο ευανάγνωστη.

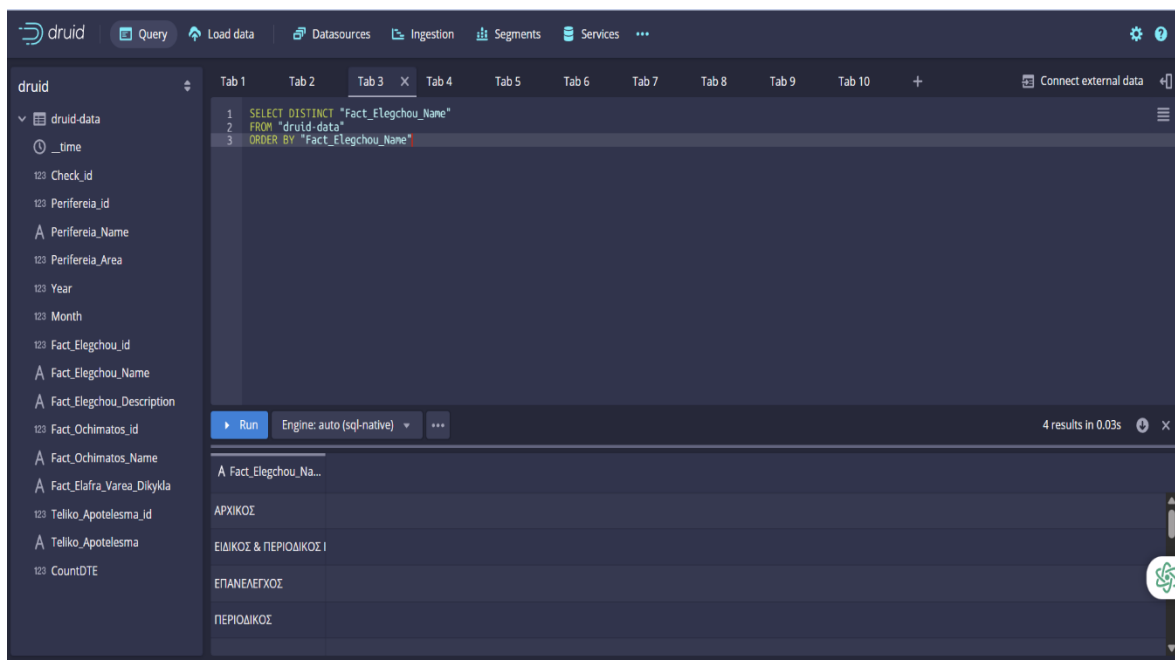
Η εκτέλεση των ερωτημάτων επέφερε τα ίδια ακριβώς αποτελέσματα με την παραδοσιακή αποθήκη δεδομένων. Ακολουθούν τα αποτελέσματα των ερωτημάτων στον Apache Druid.



Εικόνα 30: Αποτελέσματα 1ου ερωτήματος – Apache Druid



Εικόνα 31: Αποτελέσματα 2ου ερωτήματος – Apache Druid



druid

Query Load data Datasources Ingestion Segments Services

druid-data

\_time

Check\_id

Perifereia\_id

Perifereia\_Name

Perifereia\_Area

Year

Month

Fact\_Elegchou\_id

Fact\_Elegchou\_Name

Fact\_Elegchou\_Description

Fact\_Ochimatos\_id

Fact\_Ochimatos\_Name

Fact\_Elafra\_Varea\_Dikykla

Teliko\_Apotelesma\_id

Teliko\_Apotelesma

CountDTE

1 SELECT DISTINCT 'Fact\_Elegchou\_Name'

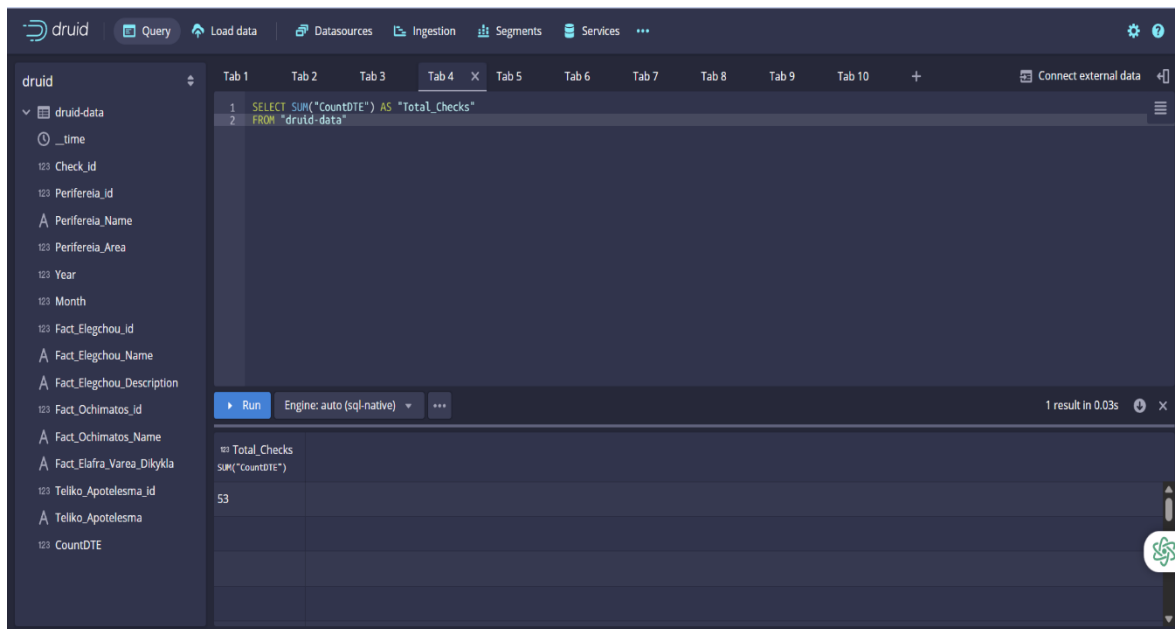
2 FROM 'druid-data'

3 ORDER BY 'Fact\_Elegchou\_Name'

Run Engine: auto (sql-native) 4 results in 0.03s

Fact_Elegchou_Name
ΑΡΧΙΚΟΣ
ΕΙΔΙΚΟΣ & ΠΕΡΙΟΔΙΚΟΣ Ι
ΕΠΑΝΕΛΕΓΧΟΣ
ΠΕΡΙΟΔΙΚΟΣ

Εικόνα 32: Αποτελέσματα 3ου ερωτήματος – Apache Druid



druid

Query Load data Datasources Ingestion Segments Services

druid-data

\_time

Check\_id

Perifereia\_id

Perifereia\_Name

Perifereia\_Area

Year

Month

Fact\_Elegchou\_id

Fact\_Elegchou\_Name

Fact\_Elegchou\_Description

Fact\_Ochimatos\_id

Fact\_Ochimatos\_Name

Fact\_Elafra\_Varea\_Dikykla

Teliko\_Apotelesma\_id

Teliko\_Apotelesma

CountDTE

1 SELECT SUM('CountDTE') AS 'Total\_Checks'

2 FROM 'druid-data'

Run Engine: auto (sql-native) 1 result in 0.03s

Total_Checks	SUM('CountDTE')
53	

Εικόνα 33: Αποτελέσματα 4ου ερωτήματος – Apache Druid

The screenshot shows the Apache Druid interface with a query executed on the 'druid-data' source. The query is: `SELECT Fact_Ochmatos_Name, SUM(CountDTE) AS Total_Checks FROM "druid-data" GROUP BY Fact_Ochmatos_Name ORDER BY Total_Checks DESC`. The results are displayed in a table with 4 rows.

Fact_Ochmatos_Name	Total_Checks
ΛΕΩΦΟΡΕΙΟ ΑΙΤΟ ΘΕΣΣΕΙ	20
ΑΣΘΕΝΟΦΟΡΟ-Τ<3.5-ΙΧ	15
ΑΣΘΕΝΟΦΟΡΟ-Τ<3.5-Δ	10
ΕΠΙΒΑΤΙΚΟ ΕΩΣ 8 ΘΕΣΣΕΙ	8

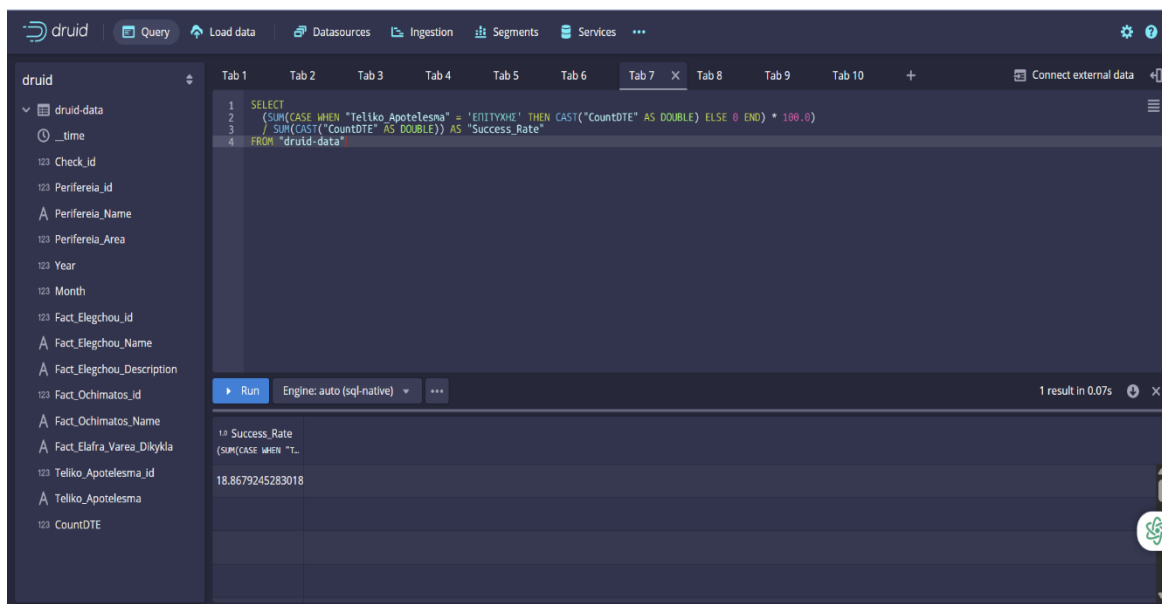
Εικόνα 34: Αποτελέσματα 5ου ερωτήματος – Apache Druid

The screenshot shows the Apache Druid interface with a query executed on the 'druid-data' source. The query is: `SELECT Perifereia_Name, SUM(CountDTE) AS Total_Checks FROM "druid-data" GROUP BY Perifereia_Name ORDER BY Total_Checks DESC LIMIT 1`. The results are displayed in a table with 1 row.

Perifereia_Name	Total_Checks
ΒΟΡΕΙΟΥ ΑΙΓΑΙΟΥ	20

Εικόνα 35: Αποτελέσματα 6ου ερωτήματος – Apache Druid

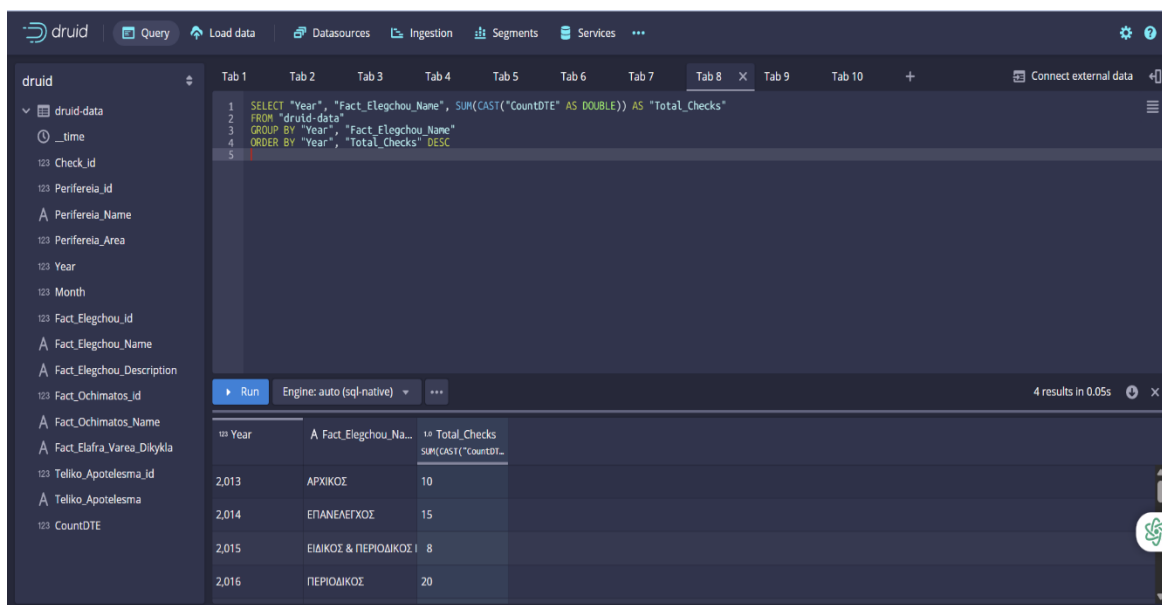




The screenshot shows the Apache Druid web console. On the left, a sidebar lists various data sources and dimensions. The main panel displays a SQL query in a text editor. Below the editor, a 'Run' button is visible. The results section shows a single row of data.

Success_Rate
18.8679245283018

Εικόνα 36: Αποτελέσματα 7ου ερωτήματος – Apache Druid



The screenshot shows the Apache Druid web console with a different SQL query. The results section displays a table with four columns: Year, Fact\_Elegchou\_Name, and Total\_Checks. The data is grouped by Year and ordered by Total\_Checks in descending order.

Year	Fact_Elegchou_Name	Total_Checks
2,013	ΑΡΧΙΚΟΣ	10
2,014	ΕΠΑΝΕΛΕΓΧΟΣ	15
2,015	ΕΙΔΙΚΟΣ & ΠΕΡΙΟΔΙΚΟΣ	8
2,016	ΠΕΡΙΟΔΙΚΟΣ	20

Εικόνα 37: Αποτελέσματα 8ου ερωτήματος – Apache Druid

druid

Query

Load data

Datasources

Ingestion

Segments

Services

Tab 1 Tab 2 Tab 3 Tab 4 Tab 5 Tab 6 Tab 7 Tab 8 Tab 9 X Tab 10 + Connect external data

```

1 SELECT "Fact_Ochmatos_Name", SUM(CAST("CountDTE" AS DOUBLE)) AS "Failed_Checks"
2 FROM "druid-data"
3 WHERE "Teliko_Apotelesma" IN ('ΙΟΒΑΡΕΕ ΕΛΛΕΙΨΕΙΣ', 'ΕΠΙΚΥΝΔΥΝΕΣ ΕΛΛΕΙΨΕΙΣ')
4 GROUP BY "Fact_Ochmatos_Name"
5 ORDER BY "Failed_Checks" DESC
6 LIMIT 5

```

Run Engine: auto (sql-native) 2 results in 0.04s

A Fact_Ochmatos_...	1.0 Failed_Checks
ΛΕΩΦΟΡΕΙΟ ΑΠΟ ΘΕΣΣΕΙ	20
ΕΠΙΒΑΤΙΚΟ ΕΩΣ 8 ΘΕΣΣΕΙ	8

Εικόνα 38: Αποτελέσματα 9ου ερωτήματος – Apache Druid

druid

Query

Load data

Datasources

Ingestion

Segments

Services

Tab 1 Tab 2 Tab 3 Tab 4 Tab 5 Tab 6 Tab 7 Tab 8 Tab 9 Tab 10 X + Connect external data

```

1 SELECT "Periferia_Name",
2 SUM(CASE WHEN "Year" = 2013 THEN CAST("CountDTE" AS DOUBLE) ELSE 0 END) AS "Checks_2013",
3 SUM(CASE WHEN "Year" = 2016 THEN CAST("CountDTE" AS DOUBLE) ELSE 0 END) AS "Checks_2016",
4 FROM "druid-data"
5 WHERE "Year" IN (2013, 2016)
6 GROUP BY "Periferia_Name"
7 ORDER BY "Periferia_Name"
8

```

Run Engine: auto (sql-native) 2 results in 0.07s

A Periferia_Name	1.0 Checks_2013	1.0 Checks_2016
ΙΟΒΑΡΕΕ ΕΛΛΕΙΨΕΙΣ	0	20
ΕΠΙΚΥΝΔΥΝΕΣ ΕΛΛΕΙΨΕΙΣ	10	0

Εικόνα 39: Αποτελέσματα 10ου ερωτήματος – Apache Druid

## 7. Σύγκριση υλοποιήσεων

Στην πειραματική αξιολόγηση που πραγματοποιήθηκε, δεν παρατηρήθηκε σημαντική διαφορά ταχύτητας μεταξύ του MySQL Server και του Apache Druid. Τα ίδια SQL ερωτήματα εκτελέστηκαν με παρόμοιους χρόνους απόκρισης και στα δύο συστήματα, πράγμα που σημαίνει ότι και οι δύο αποθήκες δεδομένων ανταποκρίθηκαν εξίσου αποδοτικά για το συγκεκριμένο workload. Και τα δύο συστήματα επέστρεφαν αποτελέσματα γρήγορα, υποδεικνύοντας συγκρίσιμη απόδοση υπό τις δοκιμασμένες συνθήκες.

Ερώτημα	Χρόνος Εκτέλεσης σε MySQL Server (secs)	Χρόνος Εκτέλεσης σε Apache Druid (secs)
1	0,0002	0,00015
2	0,0003	0,00032
3	0,0002	0,00018
4	0,0002	0,00017
5	0,0009	0,00084
6	0,0003	0,00025
7	0,0003	0,00031
8	0,0003	0,00026
9	0,0009	0,00087
10	0,0003	0,00033

Πίνακας 1: Συγκριτικός πίνακας χρόνων εκτέλεσης

Λαμβάνοντας υπόψιν των Μ.Ο απόκρισης των υλοποιήσεων στα SQL ερωτήματα που τέθηκαν, παρατηρούμε ότι η προσέγγιση στο νέφος παρουσιάζει μικρή βελτίωση στο χρόνο εκτέλεσης έναντι της συμβατικής. Συγκεκριμένα το 0,000368 (Μ.Ο. Apache Druid) είναι περίπου 6% καλύτερο (χαμηλότερο) από το 0,00039 (Μ.Ο. MySQL Server). Η βελτίωση του χρόνου μπορεί να είναι μικρή στο συγκεκριμένο workload, αλλά σε μεγαλύτερου όγκου δεδομένα αναμένεται πιο σημαντική.

Όσον αφορά το κόστος υλοποίησης και συντήρησης, ο MySQL Server διαπιστώθηκε ότι είναι απλούστερος στην εγκατάσταση και συντήρηση. Απαιτεί μόνο έναν διακομιστή με ελάχιστες εξαρτήσεις, γεγονός που καθιστά το αρχικό στήσιμο και τη διαχείρισή του πιο εύκολη. Αντίθετα, το Apache Druid χρειάστηκε περίπλοκη υποδομή για να λειτουργήσει. Στην πράξη απαιτεί ένα cluster με πολλαπλές υπηρεσίες και εξωτερικές εξαρτήσεις (όπως μια ξεχωριστή βάση δεδομένων για μεταδεδομένα και υπηρεσία ZooKeeper για συντονισμό). Αυτό μεταφράζεται σε υψηλότερο κόστος υλοποίησης και συντήρησης, η ρύθμιση του Druid ήταν πιο χρονοβόρα και απαιτεί περισσότερη διαχειριστική προσπάθεια σε σύγκριση με τον απλό server του MySQL.

Σχετικά με την ευελιξία και επεκτασιμότητα, και τα δύο συστήματα εκτέλεσαν επιτυχώς τα απαιτούμενα SQL queries, όμως παρουσιάζουν διαφορές όταν μεγαλώνει η κλίμακα των δεδομένων. Ο MySQL Server είναι ευέλικτος για γενική χρήση και υποστηρίζει πλήρες SQL, όμως η επεκτασιμότητά του βασίζεται κυρίως σε κάθετη αναβάθμιση του ίδιου server (προσθήκη πόρων) ή σε περίπλοκες λύσεις οριζόντιας κλιμάκωσης (replication, sharding). Στην πειραματική διαδικασία, ο MySQL Server χειρίστηκε τον όγκο δεδομένων που του ανατέθηκε χωρίς προβλήματα, αλλά γίνεται σαφές ότι αν τα δεδομένα αυξηθούν δραματικά, θα απαιτηθεί πρόσθετη προσπάθεια (π.χ. ρύθμιση πολλαπλών replicated instances). Αντίθετα, το Apache Druid σχεδιάστηκε εξ αρχής για οριζόντια κλιμάκωση και διαμοιρασμό φόρτου. Στις δοκιμές, το Druid φάνηκε ικανό να διατηρήσει σταθερό χρόνο απόκρισης, ακόμα και καθώς αυξανόταν ο όγκος των δεδομένων, αξιοποιώντας το καταναμημένο του αρχιτεκτονικό μοντέλο. Με απλή προσθήκη κόμβων, το Druid μπορεί να κλιμακωθεί γραμμικά για μεγαλύτερα σύνολα δεδομένων και περισσότερους ταυτόχρονους χρήστες, ενώ ο MySQL Server θα απαιτούσε πιο σύνθετη αρχιτεκτονική για αντίστοιχη κλιμάκωση.

Όσον αφορά τη διαθεσιμότητα και αξιοπιστία, το Apache Druid επέδειξε υψηλή ανθεκτικότητα και σταθερότητα κατά τη διάρκεια των δοκιμών με μεγάλα workloads. Η αρχιτεκτονική του είναι εγγενώς fault-tolerant έχει σχεδιαστεί έτσι

ώστε ένα cluster Druid να συνεχίζει να λειτουργεί ακόμα κι αν χαθεί ένας κόμβος, χωρίς διακοπή της υπηρεσίας.

Αυτό σημαίνει ότι σε σενάρια παραγωγής, το Druid μπορεί να προσφέρει υψηλή διαθεσιμότητα με κατάλληλη διαμόρφωση (π.χ. πολλαπλούς κόμβους ιστορικών δεδομένων και replicas για ανοχή σε σφάλματα). Παράλληλα, το σύστημα διαχείρισης δεδομένων του Druid εξασφαλίζει ότι δεν χάνονται δεδομένα, ακόμη και σε σοβαρές αστοχίες. Από την άλλη, το MySQL Server, στη βασική του μορφή (μονός server), αποτελεί σημείο αποτυχίας αν πέσει ο server, η υπηρεσία διακόπτεται. Στην πράξη, για να προσεγγίσει την αξιοπιστία του Druid, ο MySQL Server μπορεί να ρυθμιστεί με αναπαραγωγή (replication) ή clustering. Πράγματι, υποστηρίζει master-slave ή master-master replication για βελτίωση της ανθεκτικότητας και διαθεσιμότητας, όμως αυτό απαιτεί επιπλέον σχεδίαση και κόστη. Στην πειραματική σύγκριση, και οι δύο αποθήκες ήταν σταθερές υπό φόρτο, όμως το Druid απέδειξε ότι μπορεί να χειριστεί μεγάλους φόρτους με συνεχόμενη διαθεσιμότητα λόγω του κατανεμημένου σχεδιασμού του, ενώ ο MySQL Server θα χρειαζόταν μηχανισμούς υψηλής διαθεσιμότητας για να ανταπεξέλθει ισοδύναμα.

Αν και στην πειραματική αξιολόγηση ο χρόνος απόκρισης για τα βασικά ερωτήματα ήταν παρόμοιος, η απόδοση των δύο συστημάτων μπορεί να διαφέρει όταν εκτελούνται πιο σύνθετα αναλυτικά ερωτήματα. Ο MySQL Server λειτουργεί ως μια γενικού σκοπού σχεσιακή βάση δεδομένων, προσφέροντας πλήρη υποστήριξη SQL, αλλά δεν είναι βελτιστοποιημένος για αναλυτικά ερωτήματα σε μεγάλους όγκους δεδομένων. Από την άλλη, το Apache Druid έχει σχεδιαστεί για OLAP σενάρια και προσφέρει βελτιστοποιημένες αποδόσεις σε πολύπλοκες αναλύσεις με μεγάλο όγκο δεδομένων. Χρησιμοποιώντας columnar storage και indexing τεχνικές, το Druid μπορεί να προσφέρει καλύτερη απόδοση σε queries που απαιτούν ομαδοποίηση και αθροιστικές συναρτήσεις σε μεγάλα datasets, ενώ η MySQL ενδέχεται να επιβαρύνεται από το row-based storage της.

Γενικότερα μιλώντας και πέρα από τα πλαίσια της παρούσας εργασίας, ένα σημαντικό πλεονέκτημα του Apache Druid σε σχέση με τον MySQL Server είναι η εγγενής υποστήριξη real-time ingestion και ανάλυσης δεδομένων.

Το Druid επιτρέπει τη συνεχή ροή δεδομένων από διάφορες πηγές, καθιστώντας το ιδανικό για περιβάλλοντα όπου απαιτείται ανάλυση σε πραγματικό χρόνο. Αντίθετα, ο MySQL Server βασίζεται σε παραδοσιακές τεχνικές batch εισαγωγής δεδομένων και δεν είναι εξίσου αποδοτικός για workloads που απαιτούν real-time analytics. Αυτό σημαίνει ότι αν η εφαρμογή απαιτεί άμεση ανάλυση δεδομένων που ρέουν συνεχώς, το Druid αποτελεί καλύτερη επιλογή σε σχέση με τον MySQL Server.

Επιπλέον, η επιλογή μιας αποθήκης δεδομένων δεν εξαρτάται μόνο από την απόδοση αλλά και από το οικοσύστημα εργαλείων και την υποστήριξη της κοινότητας. Ο MySQL Server είναι ένα από τα πιο δημοφιλή συστήματα διαχείρισης βάσεων δεδομένων, με εκτενή υποστήριξη από την κοινότητα και πληθώρα εργαλείων για διαχείριση, backup, και monitoring. Υπάρχει μεγάλη τεκμηρίωση, ενώ πολλές εφαρμογές υποστηρίζουν τη MySQL ως προεπιλεγμένη επιλογή. Το Apache Druid, αν και αναπτύσσεται ραγδαία και υποστηρίζεται από μια ενεργή κοινότητα, δεν διαθέτει ακόμα το ίδιο επίπεδο ωριμότητας και ευκολίας στην ενσωμάτωση με υπάρχοντα συστήματα. Επομένως, η επιλογή μεταξύ των δύο μπορεί να επηρεαστεί και από το διαθέσιμο οικοσύστημα εργαλείων και την ανάγκη για εξειδικευμένη τεχνογνωσία.

Συνοψίζοντας, η πειραματική διαδικασία έδειξε ότι, παρόλο που ο MySQL Server είναι πιο εύκολος στην εγκατάσταση και διαχείριση, το Apache Druid προσφέρει σημαντικά πλεονεκτήματα σε επίπεδο επεκτασιμότητας και διαθεσιμότητας, ιδιαίτερα όταν τα δεδομένα μεγαλώνουν και οι ανάγκες της ανάλυσης είναι πιο απαιτητικές. Ως εκ τούτου, η επιλογή μεταξύ των δύο συστημάτων εξαρτάται από το συγκεκριμένο use case και τις απαιτήσεις σε επεκτασιμότητα και ανθεκτικότητα.

## 8. Συμπεράσματα

Η παρούσα εργασία ανέλυσε τις βασικές έννοιες της αποθήκευσης δεδομένων και προχώρησε στην υλοποίηση και σύγκριση δύο διαφορετικών προσεγγίσεων για τη διαχείριση των δεδομένων: της παραδοσιακής αποθήκης δεδομένων και της cloud-based (νεφούπολογιστικής) αποθήκης δεδομένων. Μέσα από τη μελέτη και την υλοποίηση, καταδείχθηκε ότι οι παραδοσιακές αποθήκες δεδομένων, αν και προσφέρουν αξιόπιστες λύσεις με αυστηρούς κανόνες ασφαλείας και ελέγχου, αντιμετωπίζουν περιορισμούς όσον αφορά την ευελιξία και την κλίμακα.

Η δημιουργία, η συντήρηση και η αναβάθμιση μιας παραδοσιακής αποθήκης δεδομένων απαιτούν σημαντικούς πόρους και ένα αυστηρό σύστημα υποδομών, που ενδέχεται να καταστήσει την επένδυση ακριβή και δύσκολα κλιμακούμενη, ειδικά σε μεγάλες ή δυναμικές επιχειρησιακές ανάγκες. Από την άλλη, οι νεφούπολογιστικές αποθήκες δεδομένων αποδείχθηκαν ότι παρέχουν μεγαλύτερη ευελιξία και δυνατότητα κλιμάκωσης, επιτρέποντας στις επιχειρήσεις να επωφεληθούν από τις δυνατότητες του υπολογιστικού νέφους, όπως είναι η αποθήκευση δεδομένων με μικρότερο κόστος και η ταχύτερη πρόσβαση. Παρά τα προφανή πλεονεκτήματα, ωστόσο, ενδέχεται να προκύψουν προκλήσεις γύρω από ζητήματα ασφαλείας και απόδοσης, ιδίως όταν οι όγκοι δεδομένων γίνονται ιδιαίτερα μεγάλοι ή όταν οι απαιτήσεις για απόδοση είναι υψηλές.

Η εργασία ανέδειξε, επίσης, ότι η επιλογή της καταλληλότερης λύσης εξαρτάται σε μεγάλο βαθμό από τις συγκεκριμένες ανάγκες και στρατηγικές της εκάστοτε επιχείρησης, όπως ο προϋπολογισμός, η ανάγκη για ευελιξία, και οι απαιτήσεις ασφαλείας. Συνεπώς, η σύγκριση των δύο προσεγγίσεων ανέδειξε τη σημασία του να εξετάσει κάθε επιχείρηση τις ιδιαιτερότητες της, προκειμένου να επιλέξει την πιο κατάλληλη στρατηγική αποθήκευσης δεδομένων.

Η έρευνα αυτή ανοίγει το δρόμο για περαιτέρω μελέτες και εξελίξεις στον τομέα της αποθήκευσης δεδομένων και του υπολογιστικού νέφους. Ένας από τους βασικούς τομείς που απαιτεί περαιτέρω διερεύνηση είναι η συνεχής ανάπτυξη των νέων τεχνολογιών και εργαλείων που σχετίζονται με την αποθήκευση δεδομένων.

Οι σύγχρονες τεχνολογίες big data και τα εργαλεία όπως το Apache Kafka, το Apache Spark και άλλες πλατφόρμες που επιτρέπουν την ανάλυση δεδομένων σε πραγματικό χρόνο, προσφέρουν νέες δυνατότητες και καινοτομίες που μπορούν να βελτιώσουν τις δυνατότητες κλιμάκωσης και απόδοσης των νεφοϋπολογιστικών αποθηκών δεδομένων. Επιπλέον, η συνεχής πρόοδος στον τομέα των υποδομών του υπολογιστικού νέφους ενδέχεται να φέρει νέες λύσεις σε προβλήματα που αφορούν την ασφάλεια και τη διαχείριση των δεδομένων. Καθώς οι τεχνολογίες αυτές εξελίσσονται, είναι απαραίτητο να γίνονται έρευνες που να εξετάζουν την αποτελεσματικότητά τους σε διάφορες βιομηχανίες και επιχειρησιακά περιβάλλοντα, προκειμένου να αξιοποιηθούν στο μέγιστο οι δυνατότητες τους.

Αναφορικά με τη διαχείριση της ασφάλειας των δεδομένων, ο τομέας αυτός απαιτεί συνεχιζόμενη προσοχή, καθώς η προστασία των δεδομένων στον ψηφιακό κόσμο γίνεται ολοένα και πιο περίπλοκη. Παρά την ανάπτυξη λύσεων ασφαλείας, όπως η κρυπτογράφηση δεδομένων και οι προηγμένες μέθοδοι ελέγχου ταυτότητας, οι νέες απειλές στον κυβερνοχώρο συνεχώς εξελίσσονται. Στο μέλλον, οι ερευνητές και οι επαγγελματίες του χώρου πρέπει να επικεντρωθούν στη βελτίωση των πολιτικών και των τεχνικών ασφαλείας για να διασφαλίσουν την εμπιστευτικότητα, την ακεραιότητα και τη διαθεσιμότητα των δεδομένων σε κάθε στάδιο της αποθήκευσης και της διαχείρισής τους. Επίσης, η ενσωμάτωση πιο αποτελεσματικών λύσεων διαχείρισης και ανάλυσης των δεδομένων με παράλληλη διασφάλιση της ασφάλειας θα αποτελέσει σημαντική πρόκληση και πεδίο έρευνας στο μέλλον.

Οι μελλοντικές έρευνες θα μπορούσαν να επικεντρωθούν στην ανάπτυξη εξατομικευμένων λύσεων αποθήκευσης δεδομένων που να καλύπτουν τις ανάγκες συγκεκριμένων τομέων, όπως ο χρηματοοικονομικός τομέας, η υγειονομική περίθαλψη ή η εκπαίδευση. Καθώς οι ανάγκες και οι προτεραιότητες διαφέρουν σημαντικά από κλάδο σε κλάδο, η ανάπτυξη εξειδικευμένων λύσεων για κάθε τομέα θα μπορούσε να βελτιώσει σημαντικά την αποτελεσματικότητα της αποθήκευσης και ανάλυσης των δεδομένων, ενώ παράλληλα να μειώσει το κόστος και να ενισχύσει την ασφάλεια.



Συνολικά, οι νέες τεχνολογίες αποθήκευσης δεδομένων και οι λύσεις υπολογιστικού νέφους συνεχίζουν να εξελίσσονται και να προσφέρουν αμέτρητες δυνατότητες και προοπτικές για τις επιχειρήσεις, αλλά και για τον τρόπο που προσεγγίζουμε την αποθήκευση και ανάλυση δεδομένων σε παγκόσμιο επίπεδο.

Ένα ακόμα σημαντικό στοιχείο που αξίζει να διερευνηθεί είναι η βιωσιμότητα και η περιβαλλοντική επίδραση των διαφόρων προσεγγίσεων στην αποθήκευση δεδομένων. Η αυξανόμενη ζήτηση για αποθηκευτικούς πόρους οδηγεί σε μεγαλύτερη κατανάλωση ενέργειας, γεγονός που καθιστά αναγκαία την ανάπτυξη πιο αποδοτικών και φιλικών προς το περιβάλλον τεχνολογιών.

Οι πάροχοι υπηρεσιών cloud έχουν αρχίσει να επενδύουν σε πράσινες τεχνολογίες, όπως ενεργειακά αποδοτικά κέντρα δεδομένων και χρήση ανανεώσιμων πηγών ενέργειας, προκειμένου να μειώσουν το οικολογικό τους αποτύπωμα. Από την άλλη, οι παραδοσιακές αποθήκες δεδομένων, αν και προσφέρουν πλήρη έλεγχο της υποδομής, μπορεί να απαιτούν υψηλή κατανάλωση πόρων, γεγονός που αυξάνει το συνολικό κόστος λειτουργίας και τη δυσκολία προσαρμογής σε περιβαλλοντικές προκλήσεις.

Επιπλέον, ένας τομέας που δεν έχει μελετηθεί εκτενώς είναι ο αντίκτυπος της αποθήκευσης δεδομένων στις κανονιστικές ρυθμίσεις και στη συμμόρφωση με νομικά πλαίσια. Οι επιχειρήσεις που αποθηκεύουν και διαχειρίζονται δεδομένα, ειδικά σε διεθνές επίπεδο, καλούνται να τηρούν διαφορετικούς κανονισμούς, όπως το GDPR στην Ευρώπη ή τον CCPA στην Καλιφόρνια.

Η πολυπλοκότητα των κανονισμών μπορεί να επηρεάσει την επιλογή της κατάλληλης τεχνολογίας αποθήκευσης, αφού η διατήρηση της συμμόρφωσης απαιτεί συγκεκριμένα μέτρα προστασίας, ελέγχου πρόσβασης και διαχείρισης των δεδομένων. Η εναρμόνιση μεταξύ τεχνολογικής και νομικής προσαρμογής αποτελεί ένα συνεχές ζήτημα που απαιτεί βαθύτερη έρευνα και καινοτομία.

Παράλληλα, η χρήση τεχνητής νοημοσύνης και μηχανικής μάθησης για τη βελτιστοποίηση των συστημάτων αποθήκευσης δεδομένων αποτελεί ένα πεδίο με ιδιαίτερο ερευνητικό ενδιαφέρον. Η αυτοματοποιημένη διαχείριση των δεδομένων,

μέσω έξυπνων αλγορίθμων, μπορεί να προσφέρει βελτιωμένη απόδοση, πρόβλεψη αναγκών αποθήκευσης και δυναμική κατανομή πόρων.

Η αξιοποίηση της τεχνητής νοημοσύνης μπορεί επίσης να συμβάλει στην έγκαιρη ανίχνευση απειλών ασφαλείας, ενισχύοντας την προστασία των δεδομένων και μειώνοντας τον κίνδυνο επιθέσεων. Σε αυτό το πλαίσιο, η ενσωμάτωση έξυπνων συστημάτων διαχείρισης αποθήκευσης δεδομένων θα μπορούσε να αποτελέσει μια καινοτόμο προσέγγιση που βελτιώνει τόσο την αποδοτικότητα όσο και την ασφάλεια των πληροφοριών.

Ένα άλλο ζήτημα που αξίζει να εξεταστεί είναι η σημασία της αποκεντρωμένης αποθήκευσης δεδομένων, η οποία κερδίζει έδαφος ως εναλλακτική προσέγγιση στην παραδοσιακή και cloud-based αποθήκευση.

Με τη χρήση τεχνολογιών blockchain και peer-to-peer συστημάτων αποθήκευσης, οι επιχειρήσεις μπορούν να διαμοιράζουν τα δεδομένα τους σε πολλαπλούς κόμβους, μειώνοντας την εξάρτηση από κεντρικούς διακομιστές και ενισχύοντας την ανθεκτικότητα των δεδομένων έναντι κυβερνοεπιθέσεων. Η αποκεντρωμένη αποθήκευση προσφέρει επίσης μεγαλύτερη διαφάνεια και έλεγχο στα δεδομένα, γεγονός που μπορεί να αποτελέσει πλεονέκτημα σε περιβάλλοντα που απαιτούν αυξημένη ασφάλεια και ακεραιότητα των πληροφοριών.

Τέλος, ένα θέμα που θα μπορούσε να διερευνηθεί περαιτέρω είναι η εμπειρία του τελικού χρήστη σε σχέση με τις διαφορετικές προσεγγίσεις αποθήκευσης δεδομένων. Οι απαιτήσεις για ταχύτητα, ευχρηστία και αξιοπιστία των συστημάτων επηρεάζουν την παραγωγικότητα και την αποτελεσματικότητα των επιχειρησιακών διαδικασιών. Η μελέτη της αλληλεπίδρασης των χρηστών με τις διαφορετικές τεχνολογίες αποθήκευσης μπορεί να οδηγήσει στην ανάπτυξη βελτιωμένων εργαλείων και πλατφορμών που ενισχύουν την απόδοση και μειώνουν την πολυπλοκότητα της διαχείρισης των δεδομένων.

## Βιβλιογραφία - Πηγές

- [1]. Yang, L., Sun, X., Yang, Y., & Zhang, W. (2018). A remotely keyed file encryption scheme under mobile cloud computing. *Journal of Network and Computer Applications*, 106, 90-99.
- [2]. Surbiryala, J., & Rong, C. (2019, August). Cloud computing: History and overview. In *2019 IEEE Cloud Summit* (pp. 1-7). IEEE.
- [3]. Bairagi, S. I., & Bang, A. O. (2015, March). Cloud computing: History, architecture, security issues. In *National Conference "CONVERGENCE* (Vol. 2015, p. 28).
- [4]. Arutyunov, V. V. (2012). Cloud computing: Its history of development, modern state, and future considerations. *Scientific and Technical Information Processing*, 39, 173 - 178.
- [5]. Malik, M. I., Wani, S. H., & Rashid, A. (2018). CLOUD COMPUTING-TECHNOLOGIES. *International Journal of Advanced Research in Computer Science*, 9 (2).
- [6]. Islam, R., Patamsetti, V., Gadhi, A., Gondu, R. M., Bandaru, C. M., Kesani, S. C., & Abiona, O. (2023). The future of cloud computing: benefits and challenges. *International Journal of Communications, Network and System Sciences*, 16 (4), 53 - 65.
- [7]. Langmead, B., & Nellore, A. (2018). Cloud computing for genomic data analysis and collaboration. *Nature Reviews Genetics*, 19 (4), 208.
- [8]. Basu, S., Bardhan, A., Gupta, K., Saha, P., Pal, M., Bose, M., & Sarkar, P. (2018, January). Cloud computing security challenges & solutions - A survey. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 347 - 356). IEEE.
- [9]. Nambiar, A., & Mundra, D. (2022). An overview of data warehouse and data lake in modern enterprise data management. *Big Data and Cognitive Computing*, 6 (4), 132.
- [10]. Chandra, P., Gupta, M. K. Comprehensive survey on data warehousing research. *Int. J. Inf. Technol.* 2018, 10, 217–224.
- [11]. Simões, D. M. Enterprise Data Warehouses: A conceptual framework for a successful implementation. In *Proceedings of the Canadian Council for Small*

- Business & Entrepreneurship Annual Conference, Calgary, AL, Canada, 28–30 October 2010.
- [12]. Al - Debei, M. M. Data Warehouse as a Backbone for Business Intelligence: Issues and Challenges. Eur. J. Econ. Financ. Adm. Sci. 2011, 33, 153–166.
- [13]. Report by Market Research Future (MRFR). Available online: <https://finance.yahoo.com/news/data - warehouse - dwaas - market - predicted - 153000649. html> (accessed on 27 October 2022).
- [14]. Chaudhuri, S., Dayal, U. An overview of data warehousing and OLAP technology. ACM Sigmod Rec. 1997, 26, 65–74. [Green Version].
- [15]. Codd, E. F., Codd, S. B., Salley, C. T. Providing OLAP to User - Analysts: An IT Mandate; Codd & Associates: Ladera Ranch, CA, USA, 1993, pp. 1-26.
- [16]. The Best Applications of Data Warehousing. 2020. Available online: <https://datachannel.co/blogs/best - applications - of - data - warehousing/> (accessed on 27 October 2022).
- [17]. Hai, R., Quix, C., Jarke, M. Data lake concept and systems: A survey. arXiv 2021, arXiv: 2106. 09592.
- [18]. Zagan, E., Danubianu, M. Data Lake Approaches: A Survey. In Proceedings of the 2020 International Conference on Development and Application Systems (DAS), Suceava, Romania, 21–23 May 2020; pp.189–193.
- [19]. Cherradi, M., El Haddadi, A. Data Lakes: A Survey Paper. In Innovations in Smart Cities Applications; Ben Ahmed, M., Boudhir, A. A., Karaş, R., Jain, V., Mellouli, S., Eds., Lecture Notes in Networks and Systems; Springer International Publishing: Cham, Switzerland, 2022; Volume 5, pp. 823–835.
- [20]. Dixon, J. Pentaho, Hadoop, and Data Lakes. 2010. Available online: <https://jamesdixon.wordpress.com/2010/10/14/pentaho - hadoop - and - data - lakes/> (accessed on 27 October 2022).
- [21]. King, T. The Emergence of Data Lake: Pros and Cons. 2016. Available online: <https://solutionsreview.com/data - integration/the - emergence - of - data - lake - pros - and - cons/> (accessed on 27 October 2022).
- [22]. Zuo, C., Ren, Y., & Zhang, M. (2018). Fine-grained two-factor protection mechanism for data sharing in cloud storage. IEEE Transactions on Information Forensics and Security, 13 (1), 186-196.

- [23]. Jouini, M., & Ben Arfa Rabai, L. (2019). A security framework for secure cloud computing environments. In *Cloud Security: Concepts, Methodologies, Tools, and Applications* (pp. 249-263). IGI Global.
- [24]. Alrehamy, H., Walker, C. Personal Data Lake with Data Gravity Pull. In *Proceedings of the IEEE Fifth International Conference on Big Data and Cloud Computing 2015, Beijing, China, 26–28 August 2015*.
- [25]. Yang, Q., Ge, M., Helfert, M. Analysis of Data Warehouse Architectures: Modeling and Classification. In *Proceedings of the 21st International Conference on Enterprise Information Systems, Heraklion, Greece, 3–5 May 2019*; pp. 604–611.
- [26]. Yessad, L., Labiod, A. Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault. In *Proceedings of the 2016 International Conference on System Reliability and Science (ICSRS), Paris, France, 15–18 November 2016*, pp. 95–99.
- [27]. Oueslati, W., Akaichi, J., A Survey on Data Warehouse Evolution. *Int. J. Database Manag. Syst.* 2010, 2, 11–24.
- [28]. Ali, F. S. E. A Survey of Real - Time Data Warehouse and ETL. *Int. J. Sci. Eng. Res.* 2014, 5, 3–9.
- [29]. Roy, S., De, D., Buyya, R., & Ghosh, S. (2019). Provably secure fine-grained data access control over multiple cloud servers in mobile cloud computing-based healthcare applications. *IEEE Transactions on Industrial Informatics*, 15 (1), 457-468.
- [30]. Liu, Y., Zhang, K., Zhang, X., Zhang, J., & Xie, S. (2018). Secure and fine-grained access control on e-healthcare records in mobile cloud computing. *Future Generation Computer Systems*, 78, 1020-1026.
- [31]. Aftab, U., Siddiqui, G. F. Big Data Augmentation with Data Warehouse: A Survey. In *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018*, pp. 2785–2794.
- [32]. Alsqour, M., Matouk, K., Owoc, M. A survey of data warehouse architectures—Preliminary results. In *Proceedings of the Federated Conference on Computer Science and Information Systems, Wroclaw, Poland, 9–12 September 2012*, pp. 1121–1126.
- [33]. Rizzi, S., Abelló, A., Lechtenböcker, J., Trujillo, J. Research in data warehouse modeling and design: Dead or alive? In *Proceedings of the 9th ACM international*

- workshop on Data warehousing and OLAP, DOLAP '06, Arlington, VA, USA, 10 November 2006, Association for Computing Machinery: New York, NY, USA, 2006, pp. 3–10.
- [34]. Maccioni, A., Torlone, R. KAYAK: A Framework for Just - in - Time Data Preparation in a Data Lake. In Advanced Information Systems Engineering, Krogstie, J., Reijers, H. A., Eds., Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2018, pp. 474–489.
- [35]. Gao, Y., Huang, S., Parameswaran, A. Navigating the Data Lake with DATAMARAN: Automatically Extracting Structure from Log Datasets. In Proceedings of the 2018 International Conference on Management of Data, Houston, TX, USA, 10–15 June 2018, ACM: Houston, TX, USA, 2018, pp. 943–958.
- [36]. Astriani, W., Trisminingsih, R., Extraction, Transformation, and Loading (ETL) Module for Hotspot Spatial Data Warehouse Using Geokettle. *Procedia Environ. Sci.* 2016, 33, 626–634. [Green Version]
- [37]. Halevy, A. V., Korn, F., Noy, N. F., Olston, C., Polyzotis, N., Roy, S., Whang, S. E. Managing Google's data lake: An overview of the Goods system. *IEEE Data Eng. Bull.* 2016, 39, 5–14.
- [38]. Dehne, F., Robillard, D., Rau - Chaplin, A., Burke, N. VOLAP: A Scalable Distributed System for Real - Time OLAP with High Velocity Data. In Proceedings of the 2016 IEEE International Conference on Cluster Computing (CLUSTER), Taipei, Taiwan, 13–15 September 2016, pp. 354–363.
- [39]. Hurtado, C. A., Gutierrez, C., Mendelzon, A. O. Capturing summarizability with integrity constraints in OLAP. *ACM Trans. Database Syst.* 2005, 30, 854–886. [Green Version]
- [40]. Farid, M., Roatis, A., Ilyas, I. F., Hoffmann, H. F., Chu, X. CLAMS: Bringing Quality to Data Lakes. In Proceedings of the 2016 International Conference on Management of Data, SIGMOD '16, San Francisco, CA, USA, 26 June–1 July 2016, Association for Computing Machinery: New York, NY, USA, 2016, pp. 2089–2092.
- [41]. Zhang, Y., Ives, Z. G. Juneau: Data lake management for Jupyter. *Proc. VLDB Endow.* 2019, 12, 1902–1905.
- [42]. Zhu, E., Deng, D., Nargesian, F., Miller, R. J. JOSIE: Overlap Set Similarity Search for Finding Joinable Tables in Data Lakes. In Proceedings of the 2019 International Conference on Management of Data, SIGMOD '19, Amsterdam, The

- Netherlands, 30 June–5 July 2019, Association for Computing Machinery: New York, NY, USA, 2019, pp. 847–864. [Green Version]
- [43]. Beheshti, A., Benatallah, B., Nouri, R., Chhieng, V. M., Xiong, H., Zhao, X., CoreDB: A Data Lake Service. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM '17, Singapore, 6–10 November 2017, Association for Computing Machinery: New York, NY, USA, 2017, pp. 2451–2454.
- [44]. Hai, R., Geisler, S., Quix, C., Constance: An Intelligent Data Lake System. In Proceedings of the 2016 International Conference on Management of Data, SIGMOD '16, San Francisco, CA, USA, 26 June–1 July 2016, Association for Computing Machinery: New York, NY, USA, 2016, pp. 2097–2100.
- [45]. Ahmed, A. S., Salem, A. M., Alhabibi, Y. A., Combining the Data Warehouse and Operational Data Store. In Proceedings of the Eighth International Conference on Enterprise Information Systems, Paphos, Cyprus, 23–27 May 2006, pp. 282–288.
- [46]. Software Architecture: N Tier, 3 Tier, 1 Tier, 2 Tier Architecture. Available online: <https://www.appsierra.com/blog/url> (accessed on 27 October 2022).
- [47]. Han, S. W., Three - Tier Architecture for Sentinel Applications and Tools: Separating Presentation from Functionality. Ph. D., Thesis, University of Florida, Gainesville, FL, USA, 1997.
- [48]. What Is Three - Tier Architecture. Available online: <https://www.ibm.com/in-en/cloud/learn/three-tier-architecture> (accessed on 27 October 2022).
- [49]. Phaneendra, S. V., Reddy, E. M., Big Data-Solutions for RDBMS Problems-A Survey. Int. J. Adv. Res., Comput. Commun. Eng. 2013, 2, 3686–3691.
- [50]. Simitsis, A., Vassiliadis, P., Sellis, T. Optimizing ETL processes in data warehouses. In Proceedings of the 21st International Conference on Data Engineering (ICDE'05), Tokyo, Japan, 5–8 April 2005, pp. 564–575. [Green Version]
- [51]. Prasser, F., Spengler, H., Bild, R., Eicher, J., Kuhn, K. A., Privacy - enhancing ETL - processes for biomedical data. In J. Med. Inform. 2019, 126, 72–81.
- [52]. Rousidis, D., Garoufallou, E., Balatsoukas, P., Sicilia, M. A., Metadata for Big Data: A preliminary investigation of metadata quality issues in research data repositories. Inf. Serv. Use 2014, 34, 279–286. [Green Version]
- [53]. Mailvaganam, H., Introduction to OLAP-Slice, Dice and Drill! 2007. Data Warehousing Review. Retrieved on 18 March 2008.



- [54]. Masala, G. L., Ruiu., P., & Grosso, E. (2018). Biometric authentication and data security in cloud computing. In Computer and Network Security Essentials (pp. 337-353). Springer, Cham.
- [55]. Pendse, N., What is OLAP? Available online: <https://dssresources.Com / papers / features / pendse04072002. htm> (accessed on 27 October 2022).
- [56]. Xu, J., Luo, Y. Q., Zhou, X. X. Solution for Data Growth Problem of MOLAP. Appl. Mech. Mater. 2013, 321–324, 2551–2556.
- [57]. Dehne, F., Eavis, T., Rau - Chaplin, A. Parallel multi - dimensional ROLAP indexing. In Proceedings of the CCGrid 2003. 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid, Tokyo, Japan, 12–15 May 2003, pp. 86–93.
- [58]. Shvachko, K., Kuang, H., Radia, S., Chansler, R. The Hadoop Distributed File System. In Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), Incline Village, NV, USA, 3–7 May 2010, pp. 1-10.
- [59]. Luo, Z., Niu, L., Korukanti, V., Sun, Y., Basmanova, M., He, Y., Wang, B., Agrawal, D., Luo, H., Tang, C., et al. From Batch Processing to Real Time Analytics: Running Presto® at Scale. In Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE), Kuala Lumpur, Malaysia, 9–12 May 2022, pp. 1598–1609.
- [60]. Sethi, R., Traverso, M., Sundstrom, D., Phillips, D., Xie, W., Sun, Y., Yegitbasi, N., Jin, H., Hwang, E., Shingte, N., et al. Presto: SQL on Everything. In Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE), Macao, China, 8–1 April 2019, pp. 1802–1813.
- [61]. Overview of Amazon Web Services: AWS Whitepaper. 2022. Available online: <https://d1.awsstatic.com/whitepapers/aws - overview.pdf>
- [62]. Pandis, I., The evolution of Amazon redshift. Proc. VLDB Endow. 2021, 14, 3162–3174.
- [63]. Microsoft Azure Documentation. Available online: <http://azure.microsoft.com/en-us/documentation>.
- [64]. Automate Your Data Warehouse. Available online: <https://www.oracle.com/autonomous - database/autonomous - data - warehouse/>
- [65]. Dageville, B., Cruanes, T., Zukowski, M., Antonov, V., Avanes, A., Bock, J., Claybaugh, J., Engovatov, D., Hentschel, M., Huang, J., et al., The Snowflake Elastic



- Data Warehouse. In Proceedings of the 2016 International Conference on Management of Data, San Francisco, CA, USA, 26 June–1 July 2016, ACM: San Francisco, CA, USA, 2016, pp. 215–226.
- [66]. Zagan, E., Danubianu, M., Cloud DATA LAKE: The new trend of data storage. In Proceedings of the 2021 3rd International Congress on Human - Computer Interaction, Optimization and Robotic Applications (HORA), Online, 11–13 June 2021, IEEE: Ankara, Turkey, 2021, pp. 1–4.
- [67]. Ramakrishnan, R., Sridharan, B., Douceur, J. R., Kasturi, P., Krishnamachari - Sampath, B., Krishnamoorthy, K., Li, P., Manu, M., Michaylov, S., Ramos, R., et al. Azure Data Lake Store: A Hyperscale Distributed File Service for Big Data Analytics. In Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD '17, Chicago, IL, USA, 14–19 May 2017, Association for Computing Machinery: New York, NY, USA, 2017, pp. 51–63.
- [68]. Agrawal, R., Nyamful, C. Challenges of big data storage and management. Glob. J. Inf. Technol. Emerg. Technol. 2016, 6, 1-10.
- [69]. Kadadi, A., Agrawal, R., Nyamful, C., Atiq, R. Challenges of data integration and interoperability in big data. In Proceedings of the 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 27–30 October 2014, IEEE: Washington, DC, USA, 2014, pp. 38-40.
- [70]. Chen, E. T. Implementation issues of enterprise data warehousing and business intelligence in the healthcare industry. Commun. IIMA 2012, 12, 3.
- [71]. Cuzzocrea, A., Bellatreche, L., Song, I. Y. Data warehousing and OLAP over big data: Current challenges and future research directions. In Proceedings of the Sixteenth International Workshop on Data Warehousing and OLAP, DOLAP'13, San Francisco, CA, USA, 28 October 2013, Association for Computing Machinery: New York, NY, USA, 2013, pp. 67–70.
- [72]. Singh, R., Singh, K., A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing. Int. J. Comput. Sci. Issues 2010, 7, 41.
- [73]. Longbottom, C., Bamforth, R., Optimising the Data Warehouse. 2013. Available online: [https://www.it-daily.net/downloads/WP\\_Optimising - the - data - warehouse. Pdf](https://www.it-daily.net/downloads/WP_Optimising_the_data_warehouse.Pdf).
- [74]. Santos, R. J., Bernardino, J., Vieira, M. A survey on data security in data warehousing: Issues, challenges and opportunities. In Proceedings of the 2011 IEEE

- EUROCON-International Conference on Computer as a Tool, Lisbon, Portugal, 27–29 April 2011, pp. 1-4.
- [75]. Responsibilities of a Data Warehouse Governance Committee. Available online: [https://docs.oracle.com/cd/E29633\\_01/CDMOG/GUID-7E43F311-4510-4F1E-A17E-693F94BD0EC7.htm](https://docs.oracle.com/cd/E29633_01/CDMOG/GUID-7E43F311-4510-4F1E-A17E-693F94BD0EC7.htm)
- [76]. Kumar, N. Cloud Data Warehouse Is the Future of Data Storage. 2020. Available online: <https://www.sigmoid.com/blogs/cloud-data-warehouse-is-the-future-of-data-storage/>
- [77]. Li, R., Chen, K., Chen, L., Zhang, Z., & Wei, X. (2018). A lightweight secure data sharing scheme for mobile cloud computing. *IEEE Transactions on Cloud Computing*, 6 (2), 344-357.
- [78]. Kahn, M. G., Mui, J. Y., Ames, M. J., Yamsani, A. K., Pozdeyev, N., Rafaels, N., Brooks, I. M. Migrating a research data warehouse to a public cloud: Challenges and opportunities. *J. Am. Med. Inform. Assoc.* 2022, 29, 592–600.
- [79]. Armbrust, M., Ghodsi, A., Xin, R., Zaharia, M. Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. In *Proceedings of the Conference on Innovative Data Systems Research*, Virtual Event, 11–15 January 2021.
- [80]. Rehman, K. U. U., Ahmad, U., & Mahmood, S. (2018). A Comparative Analysis of Traditional and Cloud Data Warehouse. *VAWKUM Trans. Comput. Sci*, 6, 34 - 40.
- [81]. Shakhovska, N., Boyko, N., & Pukach, P. (2018, September). The information model of cloud data warehouses. In *Conference on Computer Science and Information Technologies* (pp. 182 - 191). Cham: Springer International Publishing.
- [82]. Golec, D., Strugar, I., & Belak, D. (2021). The benefits of enterprise data warehouse implementation in cloud vs. on-premises. *ENTRENOVA - ENTERPRISE RESEARCH INNOVATION*, 7 (1), 66 - 74.
- [83]. Nookala, G. (2022). Emerging Techniques in Cloud - Based Data Warehousing: Benefits and Limitations. *Journal of Computing and Information Technology*, 2 (1).
- [84]. Shiyal, B. (2021). Azure Synapse Analytics Use Cases and Reference Architecture. In *Beginning Azure Synapse Analytics: Transition from Data Warehouse to Data Lakehouse* (pp. 225 - 241). Berkeley, CA: Apress.

- [85]. Gupta, A., Agarwal, D., Tan, D., Kulesza, J., Pathak, R., Stefani, S., & Srinivasan, V. (2015, May). Amazon redshift and the case for simpler data warehouses. In Proceedings of the 2015 ACM SIGMOD international conference on management of data (pp. 1917 - 1923).
- [86]. Kashyap, R. (2023). Data Sharing, Disaster Management, and Security Capabilities of Snowflake a Cloud Datawarehouse. International Journal of Computer Trends and Technology, 71 (2), 78 - 86.
- [87]. Lakshmanan, V., & Tigani, J. (2019). Google Bigquery: the definitive guide: data warehousing, analytics, and machine learning at scale. O'Reilly Media.
- [88]. Butt, U. A., Amin, R., Mehmood, M., Aldabbas, H., Alharbi, M. T., & Albaqami, N. (2023). Cloud security threats and solutions: A survey. Wireless Personal Communications, 128 (1), 387-413.
- [89]. Xiong, H., Zhang, H., & Sun, J. (2018). Attribute-based privacy-preserving data sharing for dynamic groups in cloud computing. IEEE Systems Journal, 99, 1-22.
- [90]. Zhang, Y., Wu, A., & Zheng, D. (2018). Efficient and privacy-aware attribute-based data sharing in mobile cloud computing. Journal of Ambient Intelligence and Humanized Computing, 9 (4), 1039-1048.
- [91]. Ali, M., Khan, S. U., & Vasilakos, A. V. (2015). Security in cloud computing: Opportunities and challenges. Information sciences, 305, 357-383.
- [92]. Fernandes, D. A., Soares, L. F., Gomes, J. V., Freire, M. M., & Inácio, P. R. (2014). Security issues in cloud environments: a survey. International journal of information security, 13, 113-170.
- [93]. Coppolino, L., D'Antonio, S., Mazzeo, G., & Romano, L. (2017). Cloud security: Emerging threats and current solutions. Computers & Electrical Engineering, 59, 126-140.
- [94]. Winkler, V. J. (2011). Securing the Cloud: Cloud computer Security techniques and tactics. Elsevier.
- [95]. Hashizume, K., Rosado, D. G., Fernández-Medina, E., & Fernandez, E. B. (2013). An analysis of security issues for cloud computing. Journal of internet services and applications, 4, 1-13.

- [96]. Mollah, M. B., Islam, K. R., & Islam, S. S. (2012, April). Next generation of computing through cloud computing technology. In 2012 25th IEEE Canadian conference on electrical and computer engineering (CCECE) (pp. 1-6). IEEE.
- [97]. Varghese, B., & Buyya, R. (2018). Next generation cloud computing: New trends and research directions. *Future Generation Computer Systems*, 79, 849-861.
- [98]. Durao, F., Carvalho, J. F. S., Fonseca, A., & Garcia, V. C. (2014). A systematic review on cloud computing. *The Journal of Supercomputing*, 68, 1321-1346.
- [99]. Widenius, M., & Axmark, D. (2002). *MySQL reference manual: documentation from the source*. " O'Reilly Media, Inc."
- [100]. Dvorski, D. D. (2007). *Installing, configuring, and developing with Xampp*. Skills Canada, 492.
- [101]. Ahmed, M., Uddin, M. M., Azad, M. S., & Haseeb, S. (2010, April). MySQL performance analysis on a limited resource server: Fedora vs. Ubuntu Linux. In *Proceedings of the 2010 Spring Simulation Multiconference* (pp. 1-7).
- [102]. Schwartz, B., Zaitsev, P., & Tkachenko, V. (2012). *High performance MySQL: optimization, backups, and replication*. " O'Reilly Media, Inc."
- [103]. Yang, F., Tschetter, E., Léauté, X., Ray, N., Merlino, G., & Ganguli, D. (2014, June). *Druid: A real-time analytical data store*. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data* (pp. 157-168).

**Υπεύθυνη Δήλωση Συγγραφέα:**

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν.1599/1986, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης.