Boosting Research for a Smart and Carbon Neutral Built Environment with Digital Twins –
SmartWins

CERTH
CENTRE FOR
RESEARCH & TECHNOLOGY
HELLAS

Data Storage Types

(Hierarchical, network, relational, graph, object-oriented databases and their representation: SQL, NoSQL, RDF, LPG, etc.)

Paraskevas Koukaras, CERTH

CERTH SmartWins Training Sessions: Day2
23 April 2024
Thessaloniki

Funded by
the European Union

This project has received funding from the European Union's Horizon research and innovation programme under grant agreement No 101078997

SmartWins

Introduction to Data Storage

- Evolution of data storage from physical files to complex digital systems
- Importance of understanding different data models for efficient data management and retrieval
- Types of data storage models to be discussed:
  - Hierarchical
  - Network
  - Relational
  - Graph
  - Object-Oriented

From physical filing systems to modern digital databases

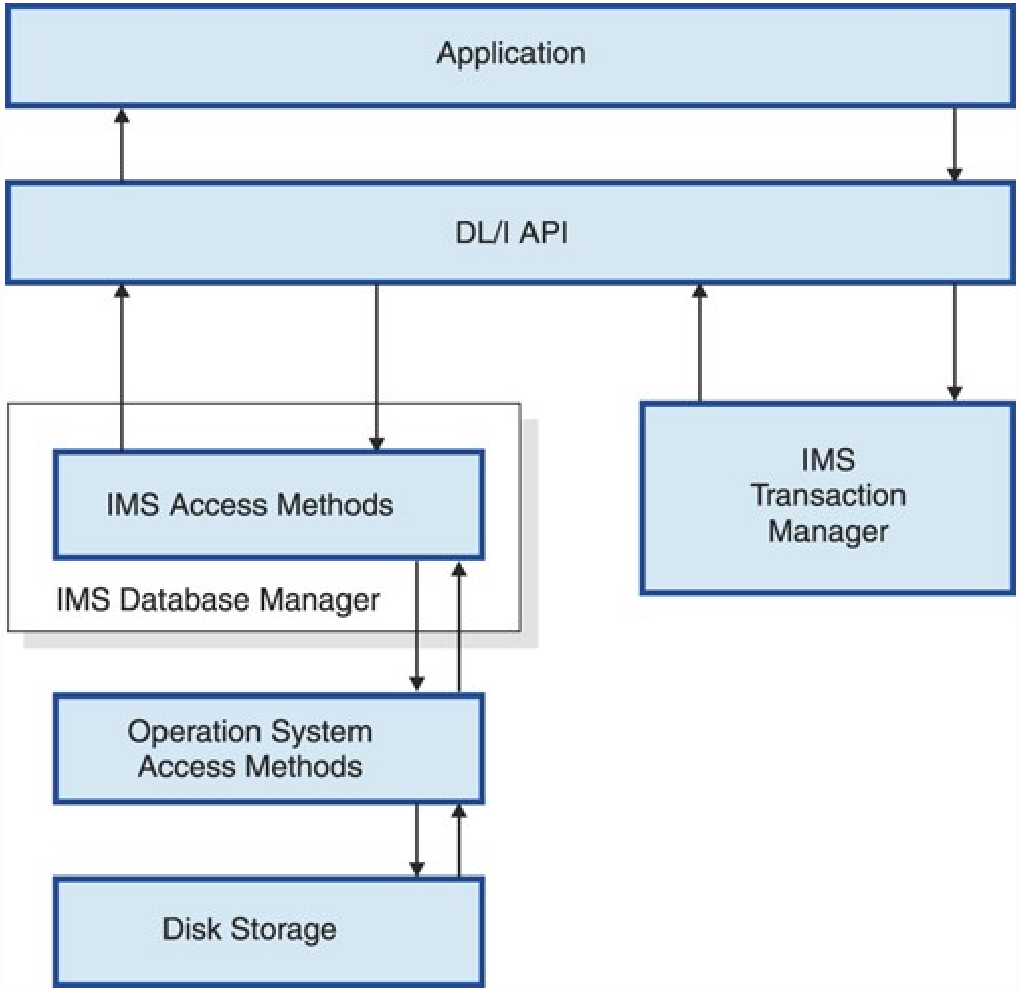
- 1947: Manchester Mark I Williams-Kilburn tube
- Used a cathode ray tube and stored bits as dots

- 2024: Synology DiskStation DS1522+ 5-Bay Network-attached Storage (NAS)
- From 2009: cloud-based network-attached storage solutions appear for online backup

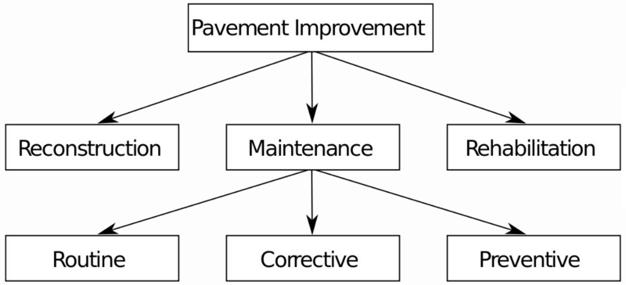
Hierarchical Database Model

- A model where data is organised into a **tree-like** structure, implying a single upward link in each child except the root.
- Rigid structure, efficient for **read**, **search**, and **write** operations within predictable, large datasets.

IBM Information Management System (IMS) is extensively used in banking and manufacturing for large-scale transaction processing (https://www.oreilly.com/api/v2/epubs/9780132886987/files/graphics/08fig01.jpg).



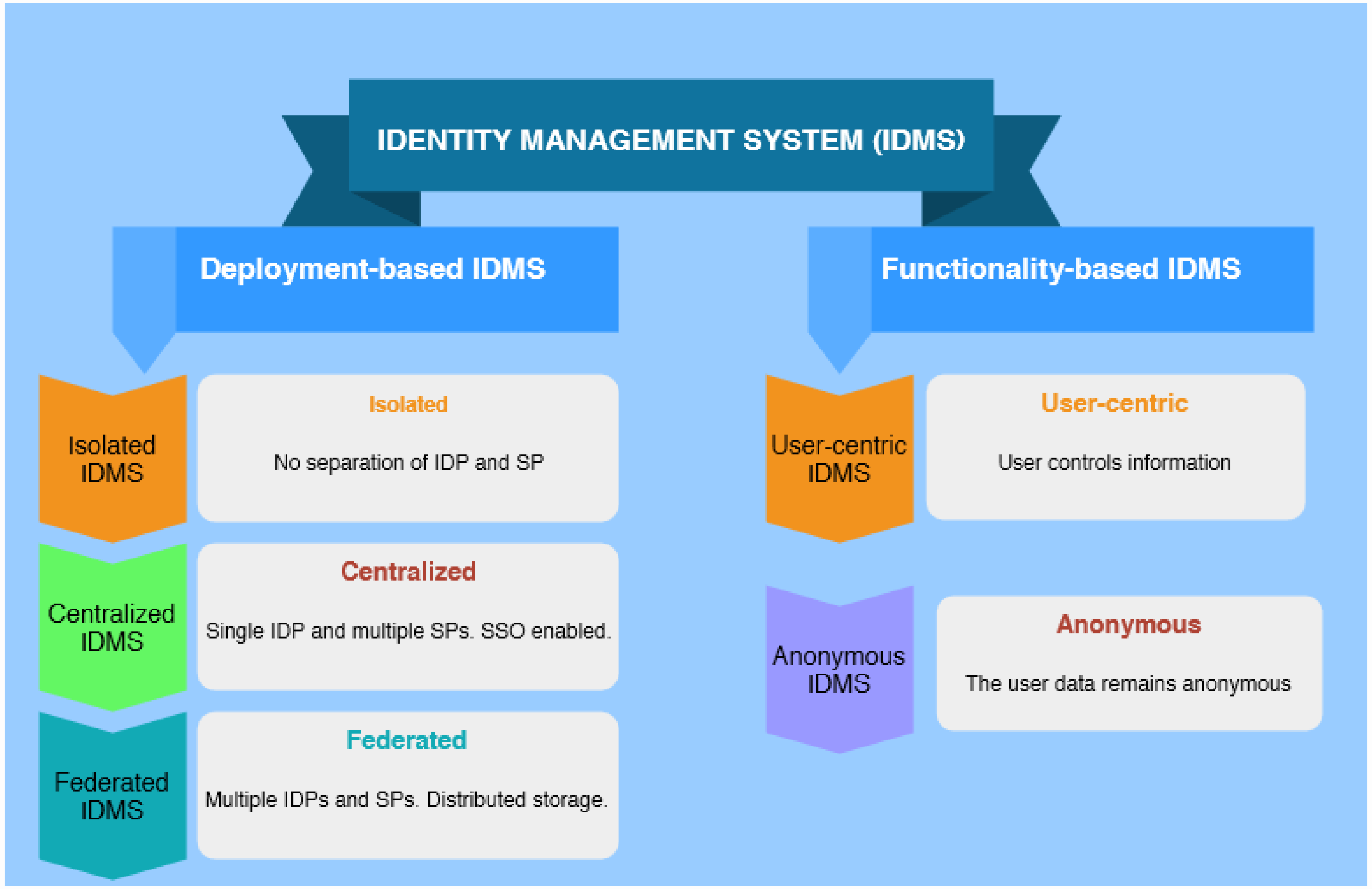
Hierarchical Model



*IMS: Information Management System
*DL/I: Data Language One

Network Database Model

- A flexible model allowing each record to have multiple **parent** and **child** records, forming a graph structure.
- More complex than hierarchical, it allows **many-to-many relationships** and is suited for large databases handling complex applications.
- (IDMS) is used in sectors with complex data relationships like telecommunications and utilities.



IDENTITY MANAGEMENT SYSTEM (IDMS)

Deployment-based IDMS

Isolated IDMS — Isolated: No separation of IDP and SP

Centralized IDMS — Centralized: Single IDP and multiple SPs. SSO enabled.

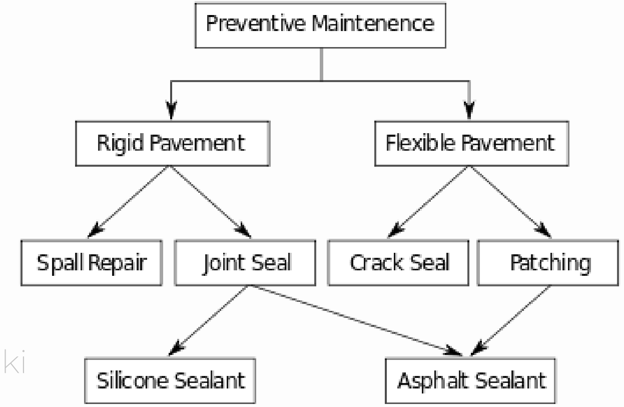
Federated IDMS — Federated: Multiple IDPs and SPs. Distributed storage.

Functionality-based IDMS

User-centric IDMS — User-centric: User controls information

Anonymous IDMS — Anonymous: The user data remains anonymous

Network Model

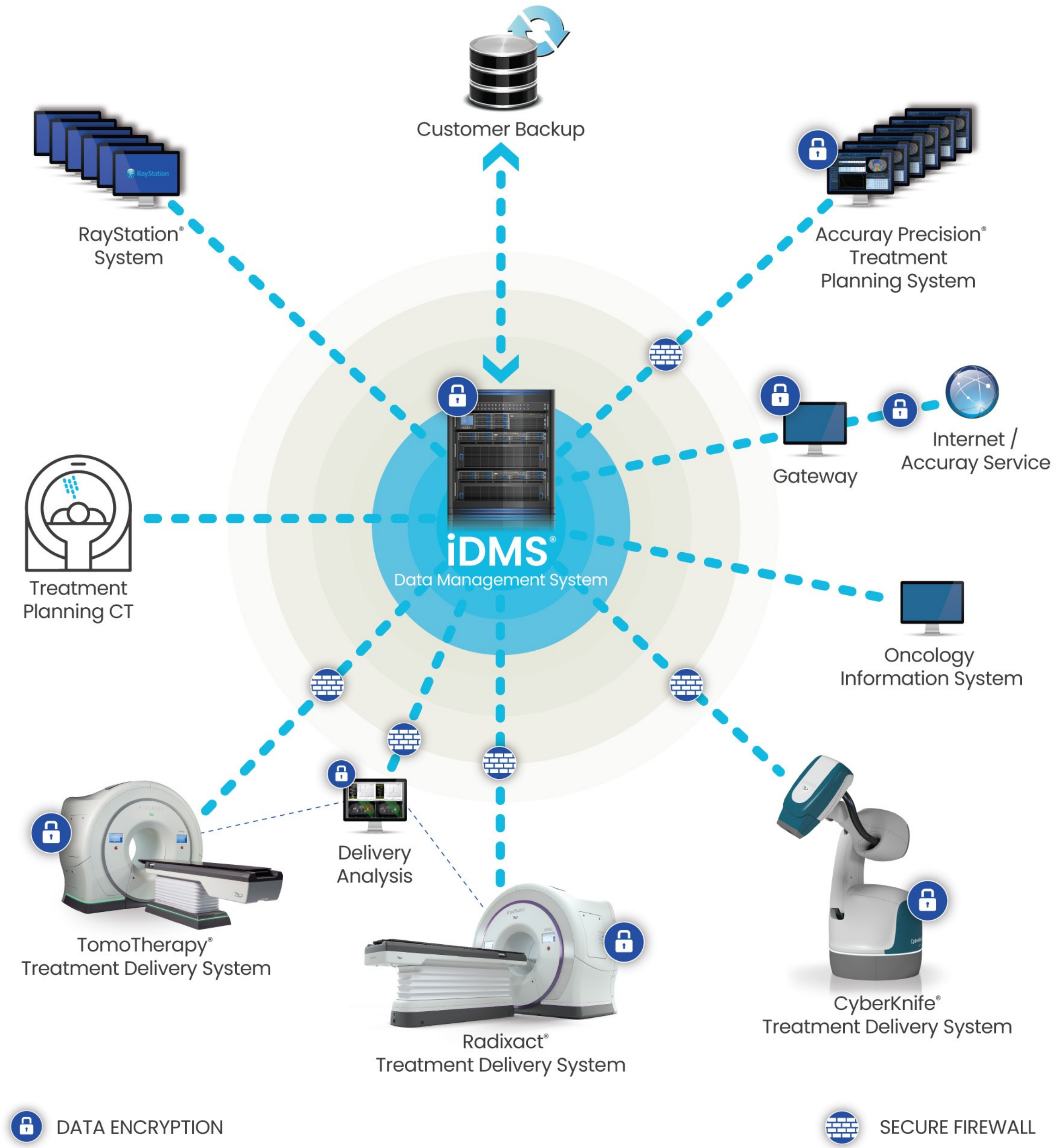


*IDP: Identity Provider
*SP: Service Provider
*SSO: Single Sign On

IDMS example

Accuray IDMS:

- It integrates **patient data** which guide planning and delivery of specified treatments.
- This framework **stores, process** and **manages** all data produced during the process and the delivery phases of each treatment.


Customer Backup
RayStation® System
Accuray Precision® Treatment Planning System
Gateway
Internet / Accuray Service
Treatment Planning CT
iDMS®
Data Management System
Oncology Information System
Delivery Analysis
TomoTherapy® Treatment Delivery System
Radixact® Treatment Delivery System
CyberKnife® Treatment Delivery System
DATA ENCRYPTION
SECURE FIREWALL

Relational Database Model

- A model based on **storing data in tables** that can be linked—or related—based on data common to each.
- Highly flexible, uses **SQL for querying**, well-suited for handling vast amounts of data that require robust transaction processing.
- Oracle Database and MySQL are widely used in almost every industry for web back-ends, customer relationship management (CRM) systems, and more.

Oracle Database

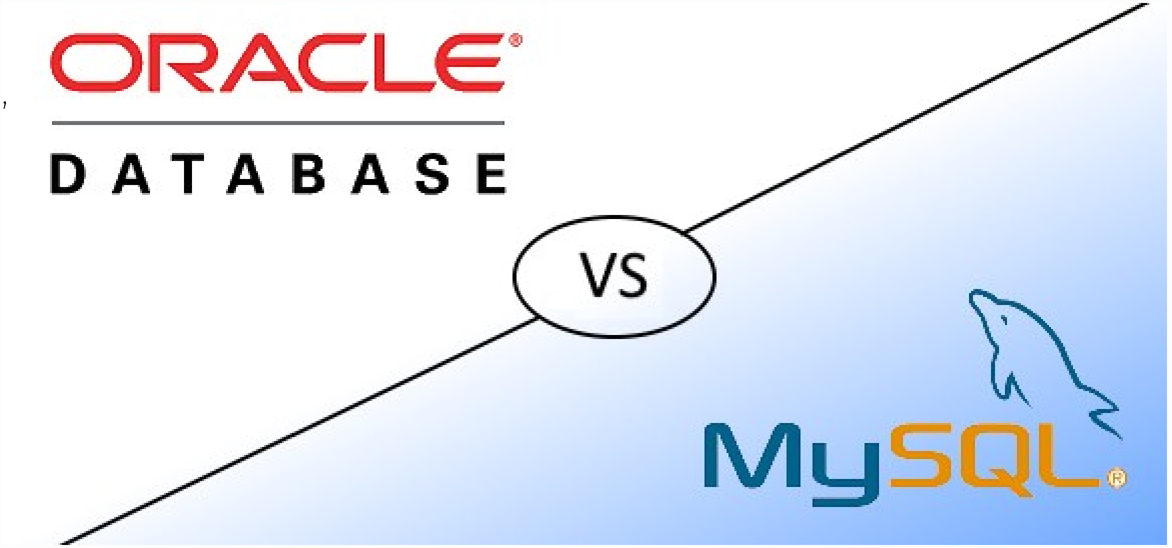
Pros:

-Comprehensive commercial support from Oracle Corporation ensures prompt updates, bug fixes, and security patches.

-Active developer and user community

Cons:

-The extensive support and advanced features come with a significant cost.



MySQL

Pros:

-Open-source community actively enhances development and support, backed by extensive documentation and forums.

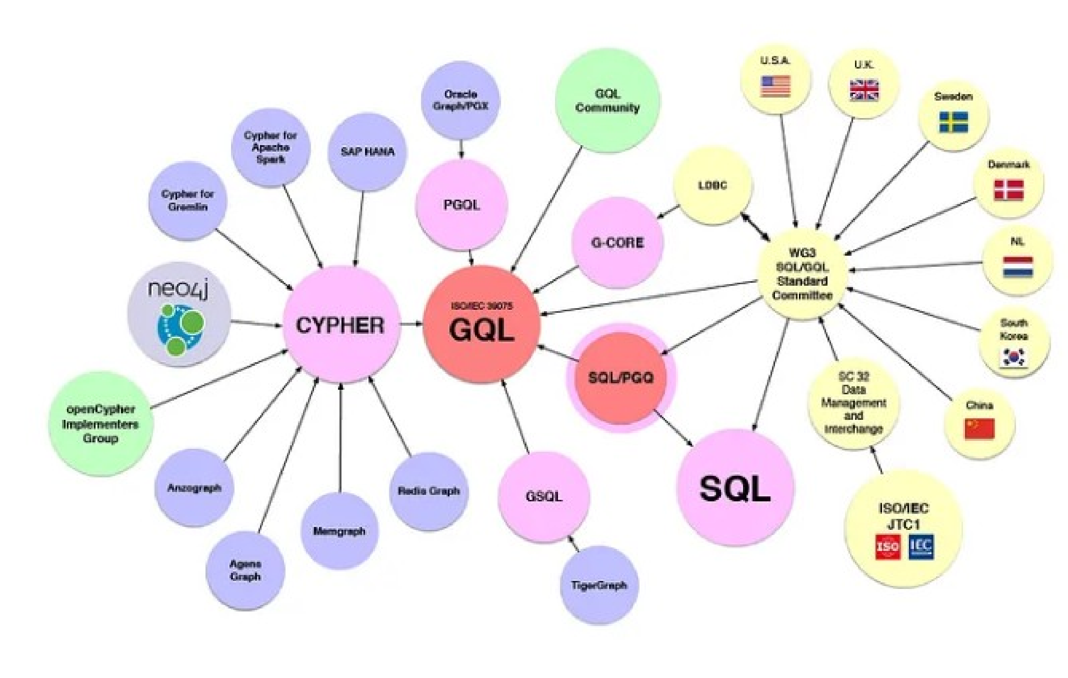
-Oracle provides commercial support for MySQL Enterprise Edition, including 24/7 technical support, training, and consulting.

Cons:

-MySQL lacks some of the advanced features and optimisations.

Graph Database Model

- Stores data in **nodes** and **edges**, where nodes typically represent entities, and edges specify the relationships between these entities.
- Extremely **efficient** for applications that require **frequent navigation across relationships**, like social networks, recommendation engines, and fraud detection systems.
- Examples like **Neo4j** and **ArangoDB**, are used extensively in networking sites, logistics, and complex transaction systems where relationships need to be queried frequently and rapidly.

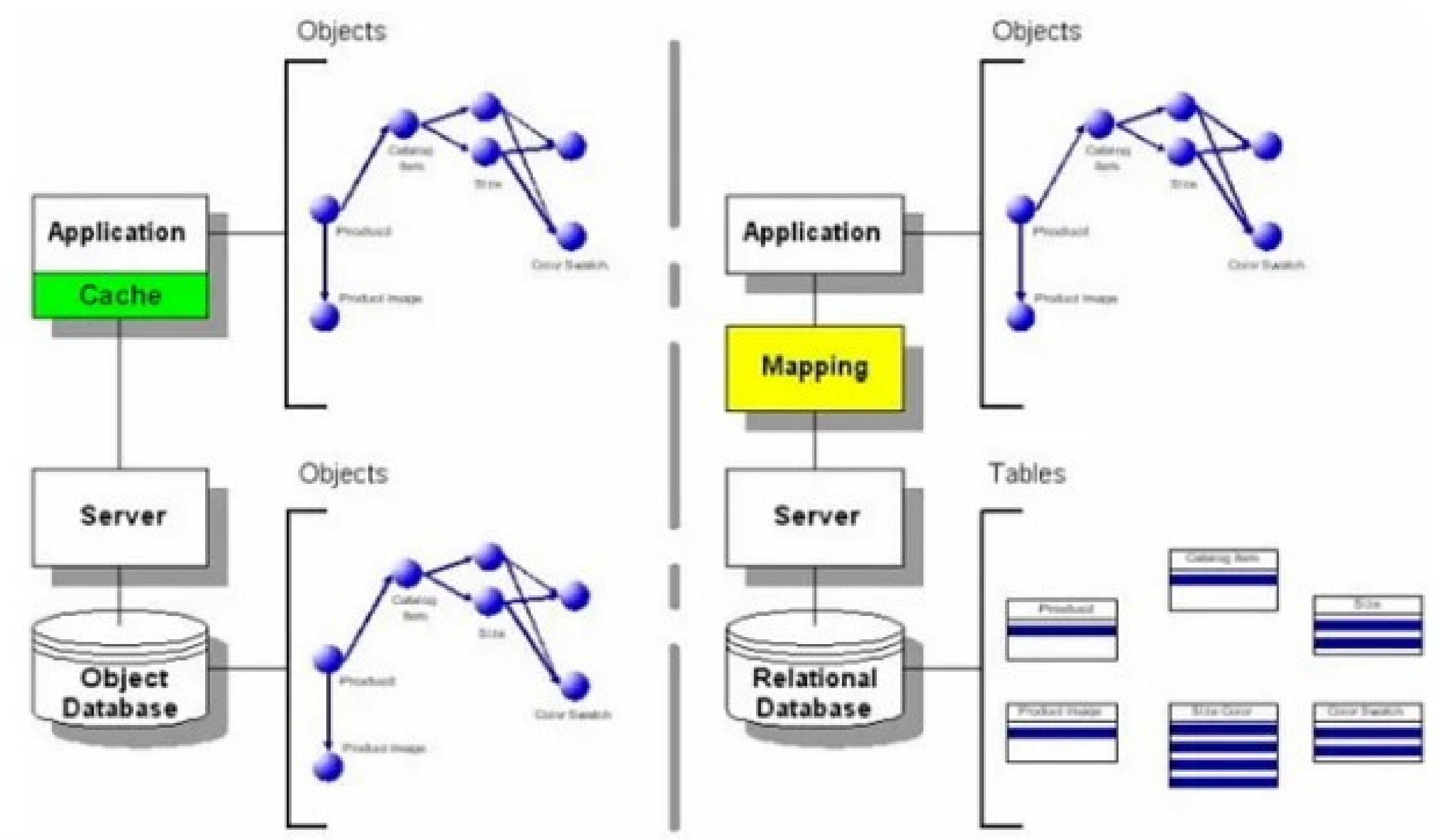


Source: neo4j.com

Source: arangoDB.com

Object-Oriented Database Model

- A Database Management System (DBMS) in which information is represented in the form of objects, as used in object-oriented programming.
- Seamless integration with object-oriented programming languages, supporting complex data types and structures like **encapsulation, inheritance,** and **polymorphism**.
- db4o and ObjectDB, used in engineering applications, complex data applications, and systems where database operations are heavily intertwined with application logic.

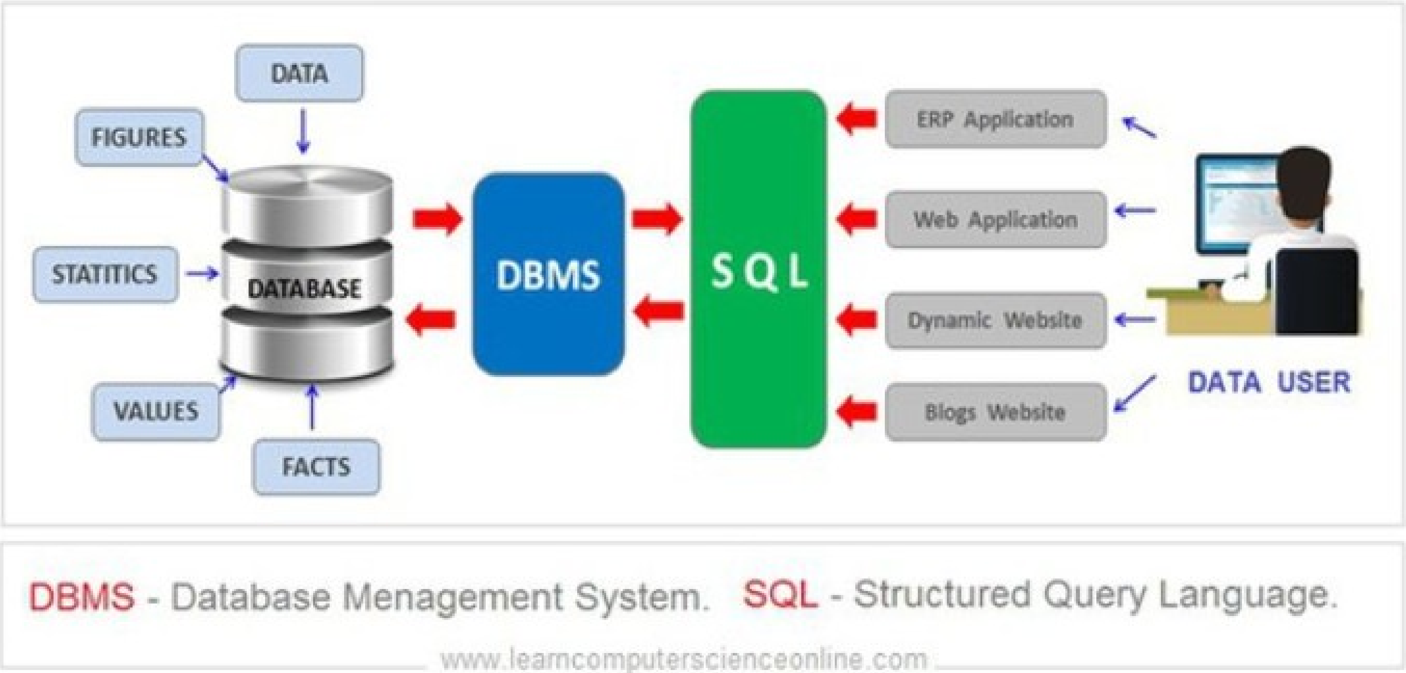


SQL Database Systems

- A domain-specific language used in programming and designed for managing data held in a relational database management system.
- Robust, widely adopted, ideal for complex queries, and highly scalable for transaction-based storage.

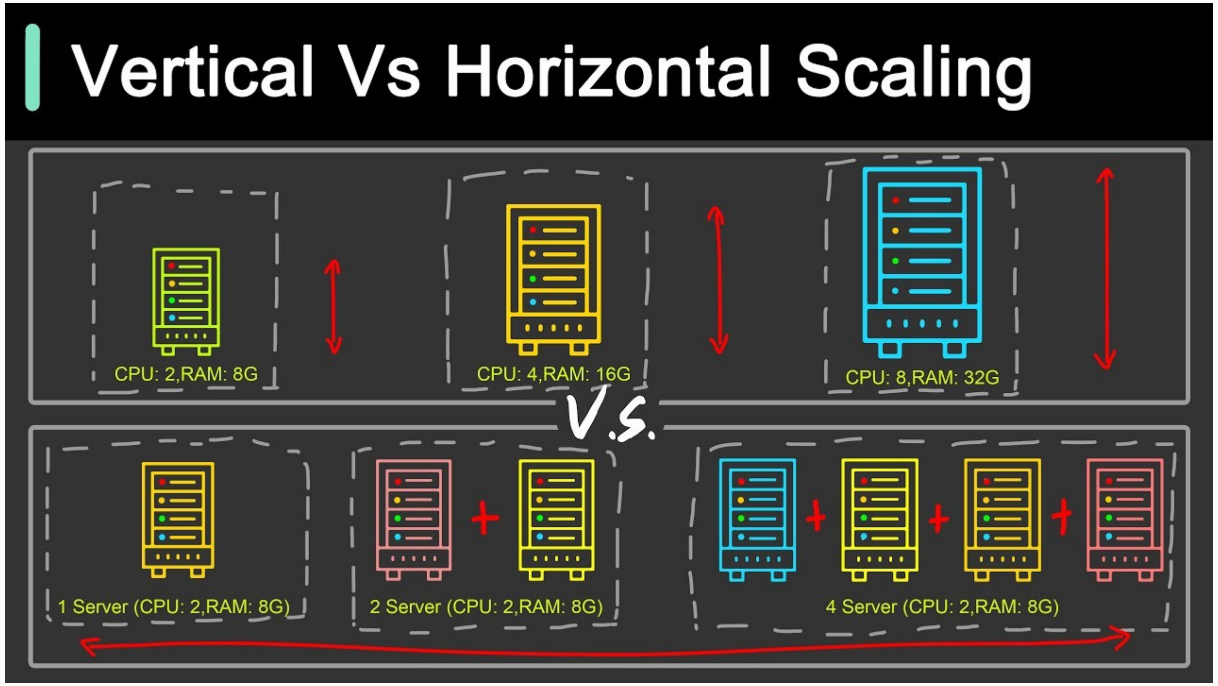
BUT

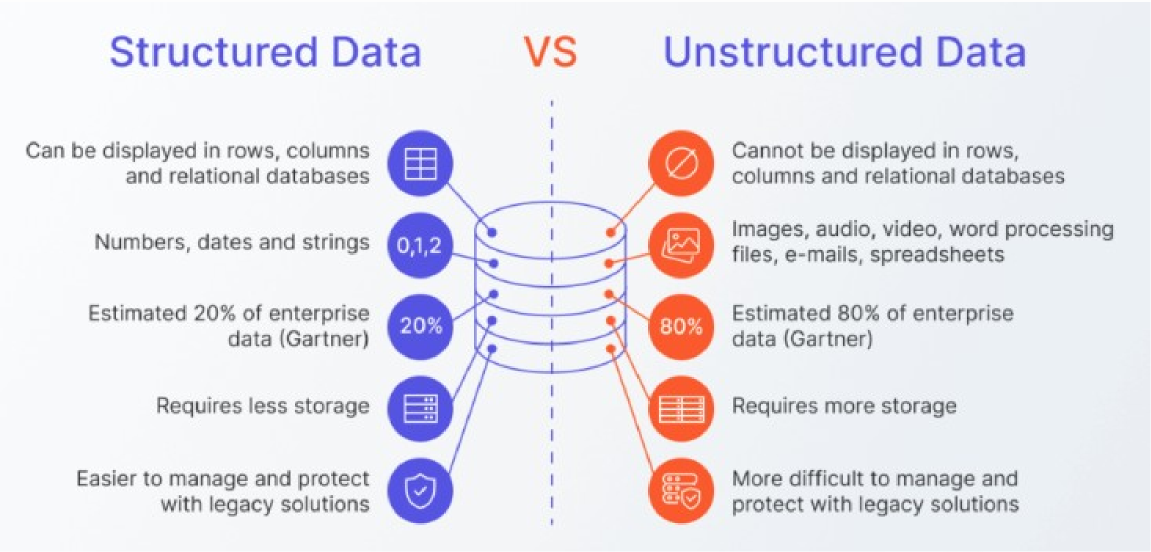
Less efficient for unstructured data, can be complex to scale horizontally.



DBMS - Database Menagement System. SQL - Structured Query Language.

www.learncomputerscienceonline.com


Vertical Vs Horizontal Scaling
CPU: 2,RAM: 8G
CPU: 4,RAM: 16G
CPU: 8,RAM: 32G
V.S.
1 Server (CPU: 2,RAM: 8G)
2 Server (CPU: 2,RAM: 8G)
4 Server (CPU: 2,RAM: 8G)

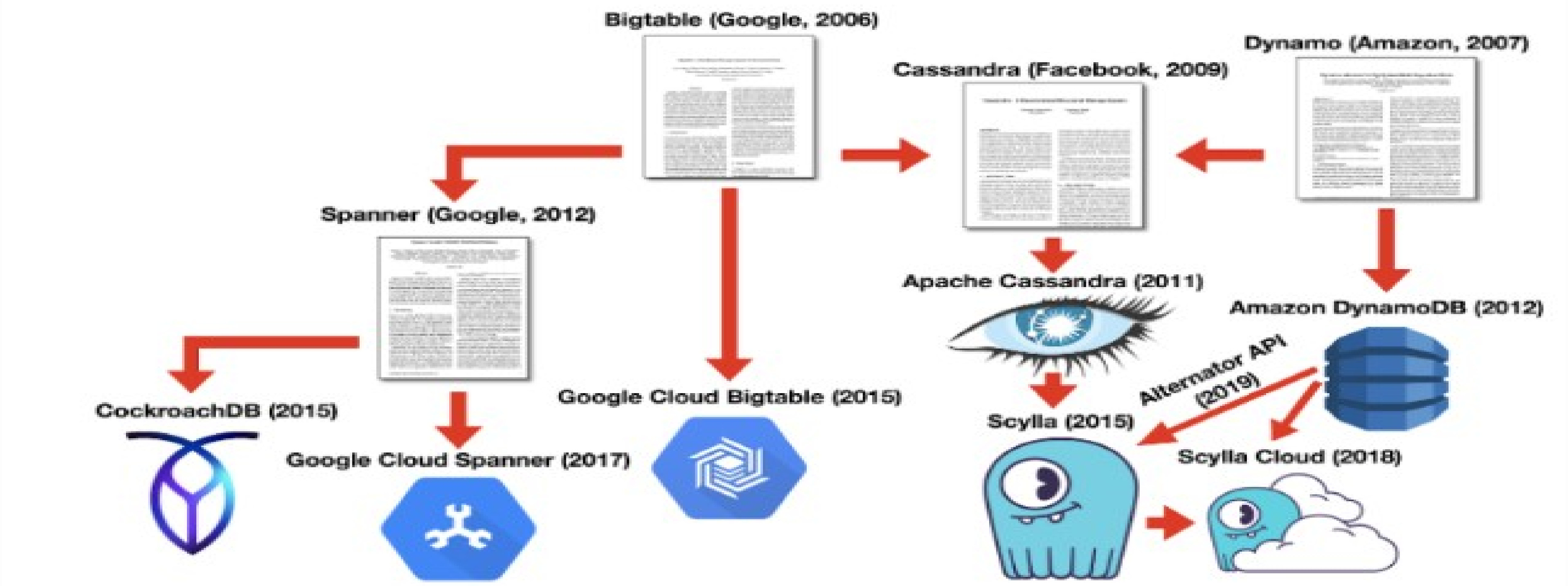


Structured Data VS Unstructured Data

Structured Data	Unstructured Data
Can be displayed in rows, columns and relational databases	Cannot be displayed in rows, columns and relational databases
Numbers, dates and strings	Images, audio, video, word processing files, e-mails, spreadsheets
Estimated 20% of enterprise data (Gartner)	Estimated 80% of enterprise data (Gartner)
Requires less storage	Requires more storage
Easier to manage and protect with legacy solutions	More difficult to manage and protect with legacy solutions

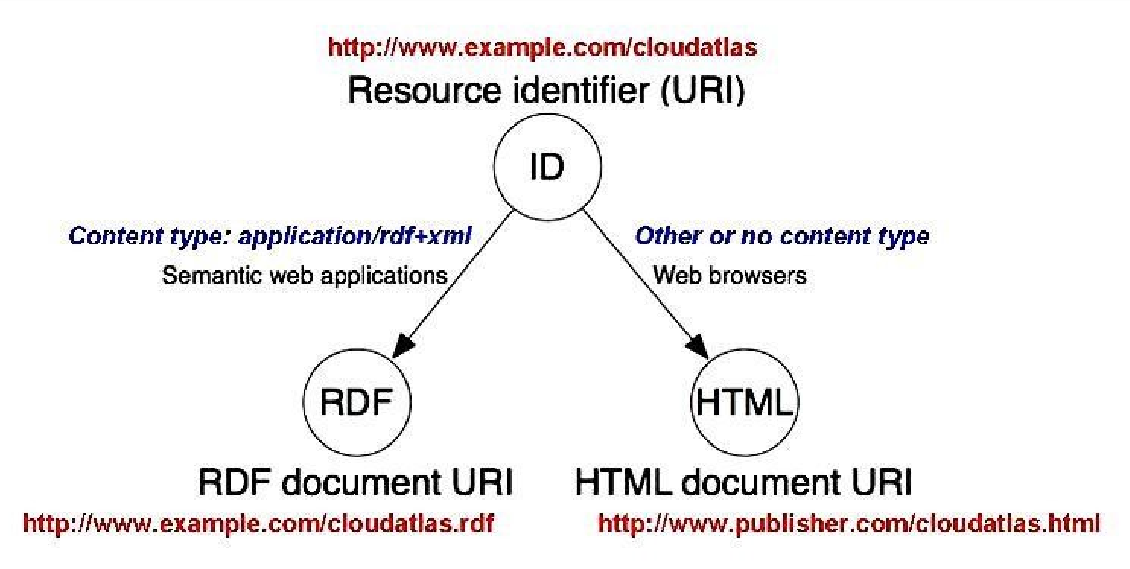
NoSQL Database Systems

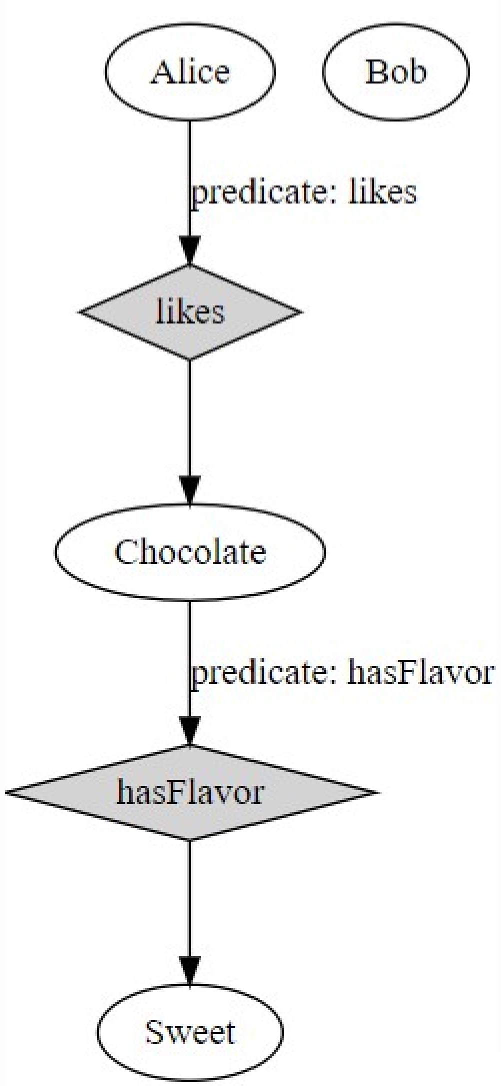
- NoSQL databases represent a range of database technologies that are developed to handle a variety of data models, including key-value, document, column, and graph databases.
- Designed for distributed data stores with large data set sizes, high scalability, and flexible schema designs.
- MongoDB (document-oriented), Cassandra (column-oriented), Redis (key-value store), and Neo4j (graph database).



RDF (Resource Description Framework)

- A standard model for data interchange on the web, RDF extends the linking structure of the web to use Uniform Resource Identifiers (URIs) to name the relationship between things as well as the **two ends of the link** (usually referred to as a “triple”).
- Commonly used in semantic web applications, personal information management, and knowledge management systems.
- **Example:** Data integration across different **web data sources** and **semantic query** applications.


http://www.example.com/cloudatlas
Resource identifier (URI)
ID
Content type: application/rdf+xml
Semantic web applications
Other or no content type
Web browsers
RDF
HTML
RDF document URI
HTML document URI
http://www.example.com/cloudatlas.rdf
http://www.publisher.com/cloudatlas.html

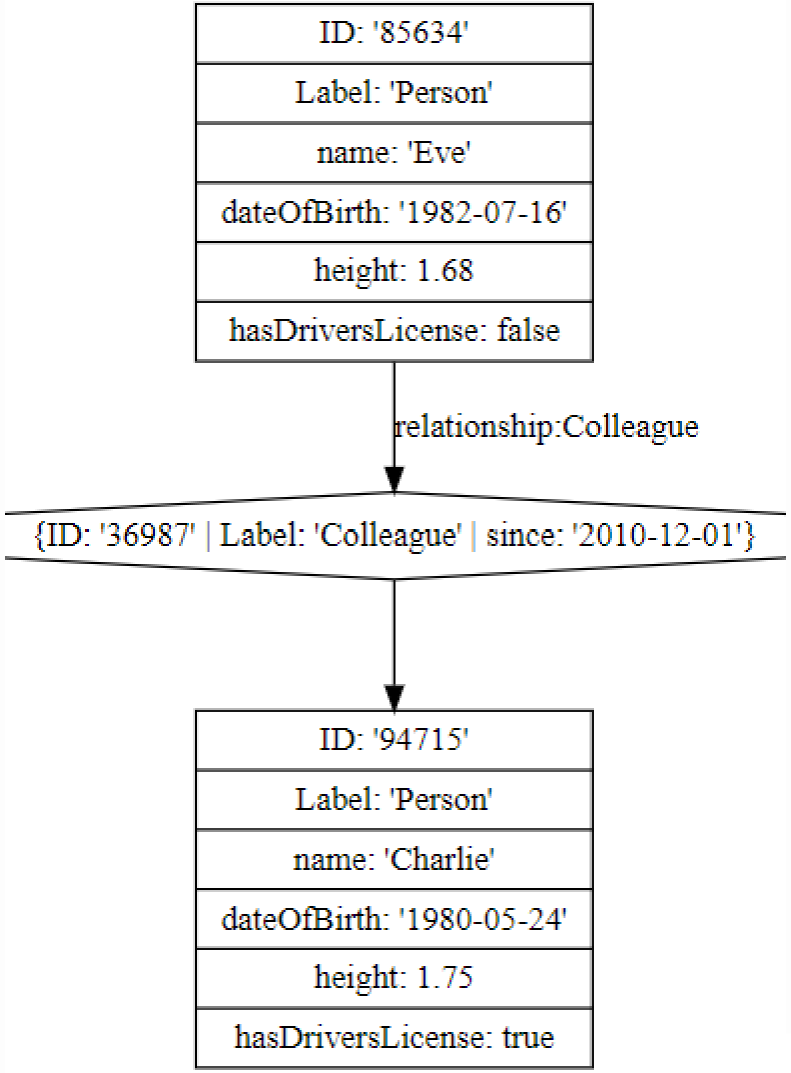


Labeled Property Graph (LPG)

- A type of graph database where each **node** and **edge** has an **associated label** and **properties** that store data about them, allowing for rich, complex data models.
- Highly expressive, suitable for modeling intricate relationships with additional context via properties.
- **Neo4j, utilizes LPG** to enhance queries and analytics in network and IT operations, fraud detection, and recommendation engines.
- There is **no standardised query language** for all LPG-style databases, but Cypher is the most widely adopted one.

Cypher Query example:

```
MATCH (person1:Person {name: "Eve"})-[relationship:Colleague]-(person2:Person {name: "Charlie"})
RETURN person1, relationship, person2
```



ID: '85634'
Label: 'Person'
name: 'Eve'
dateOfBirth: '1982-07-16'
height: 1.68
hasDriversLicense: false

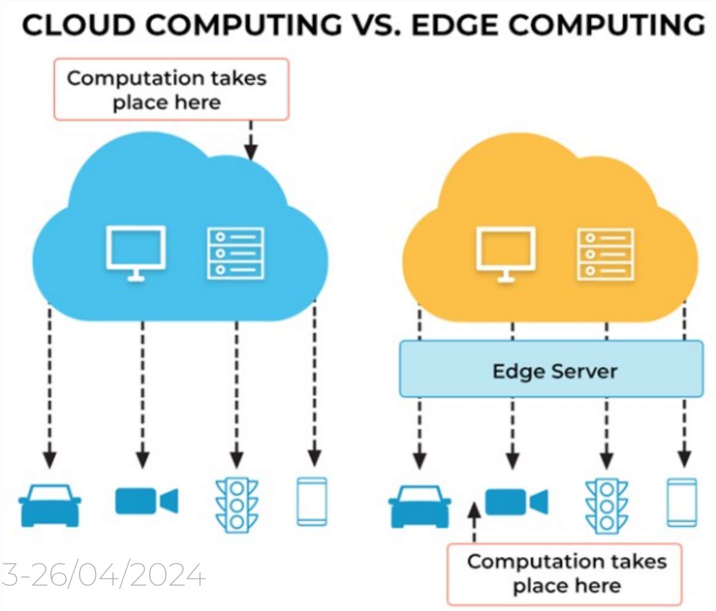
relationship:Colleague

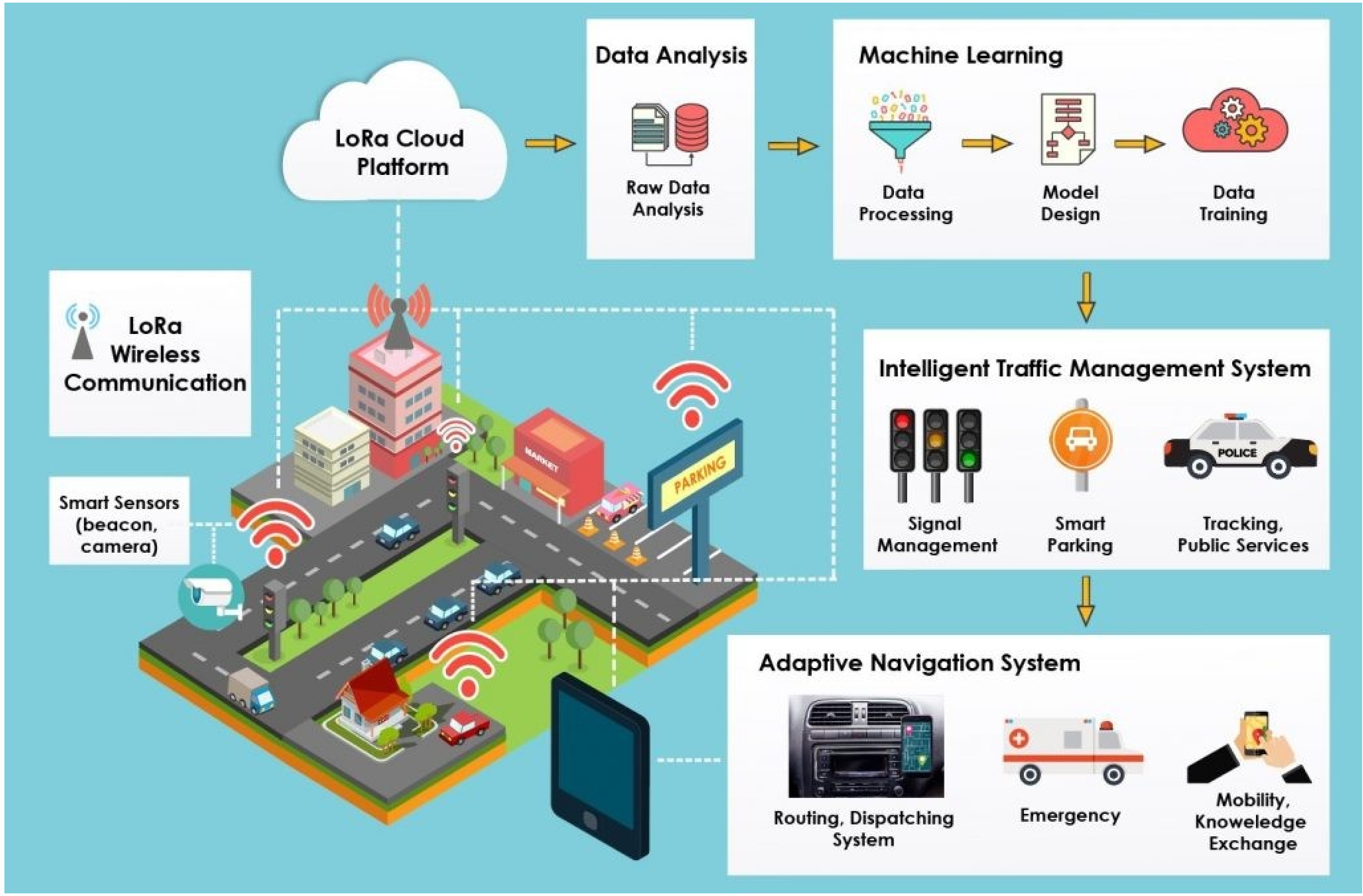
{ID: '36987' | Label: 'Colleague' | since: '2010-12-01'}

ID: '94715'
Label: 'Person'
name: 'Charlie'
dateOfBirth: '1980-05-24'
height: 1.75
hasDriversLicense: true

IoT Data Storage Integration

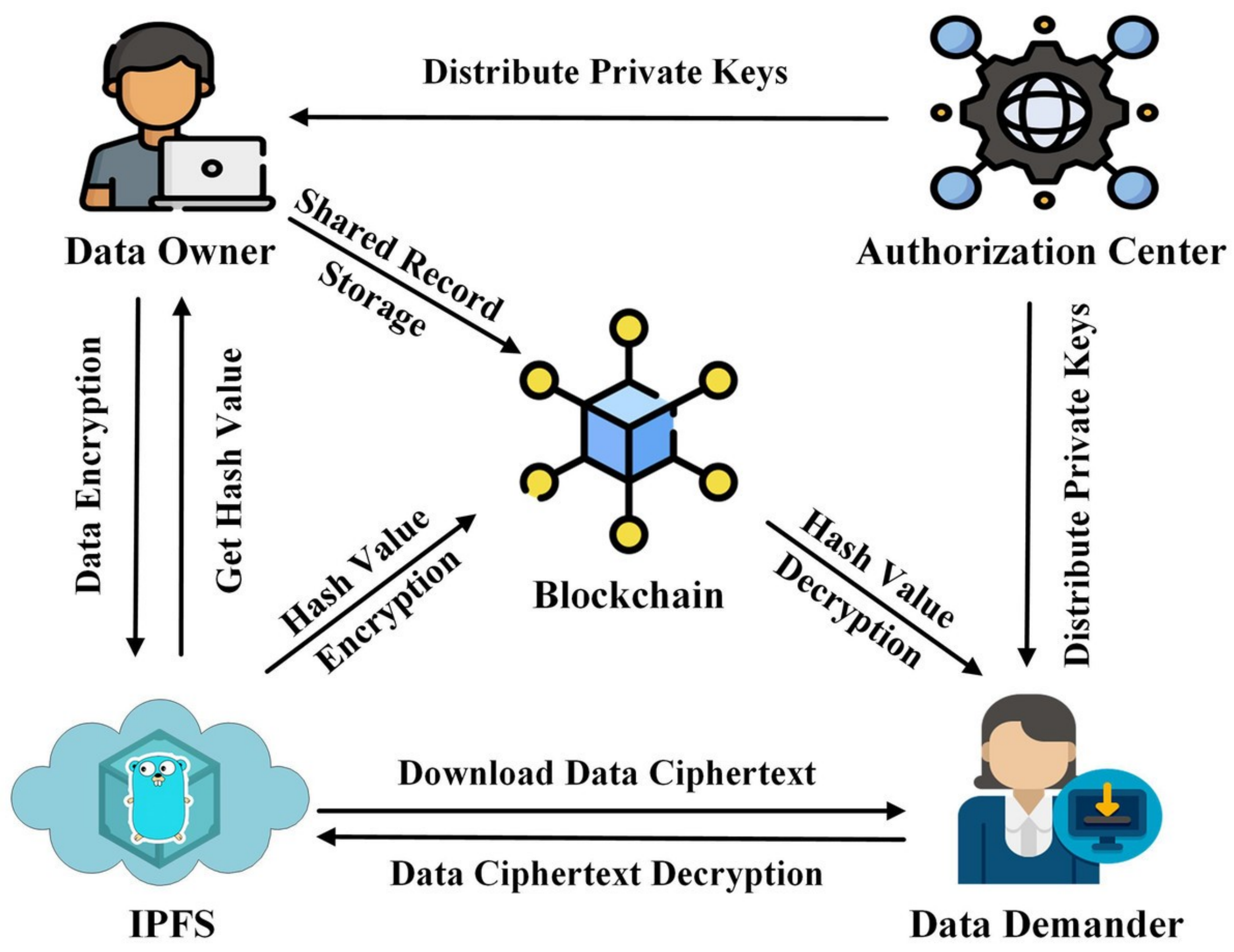
- Handling **real-time**, **voluminous** IoT data streams
- **Scalability** challenges in data ingestion and processing
- **Security protocols** for data in transit and at rest
- Time-series databases like InfluxDB for **efficient storage**
- Edge computing for **low-latency**, **localised** data handling
- Smart city traffic management systems


CLOUD COMPUTING VS. EDGE COMPUTING
Computation takes place here
Edge Server
Computation takes place here


LoRa Cloud Platform
Data Analysis
Raw Data Analysis
Machine Learning
Data Processing
Model Design
Data Training
LoRa Wireless Communication
Smart Sensors (beacon, camera)
MARKET
PARKING
Intelligent Traffic Management System
Signal Management
Smart Parking
POLICE
Tracking, Public Services
Adaptive Navigation System
Routing, Dispatching System
Emergency
Mobility, Knowledge Exchange

Blockchain and Data Security

- Blockchain's decentralized ledger for enhanced **data integrity**
- Encryption and consensus mechanisms prevent unauthorized alterations
- **Applications:** Securing health records, legal documents, transaction logs
- **Trade-offs:** Implementation complexity vs. data immutability benefits

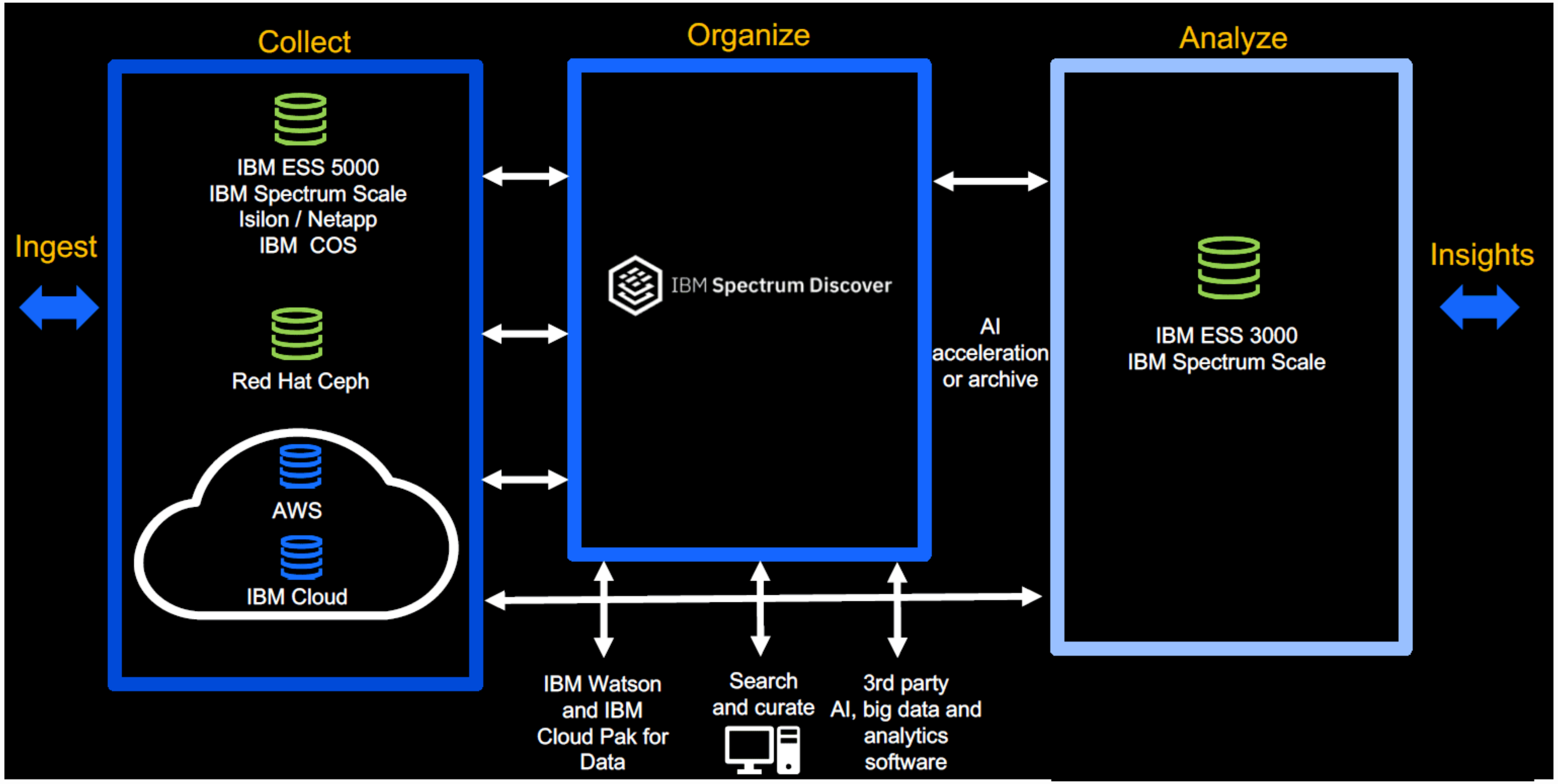


Source: https://dfzljdn9uc3pi.cloudfront.net/2023/cs-1337/1/fig-1-2x.jpg

* Interplanetary File System (IPFS): a file system that allows you to store files and track versions over time, much like Git, keeping track of them on a distributed network, somewhat like BitTorrent.

AI for Data Storage Efficiency

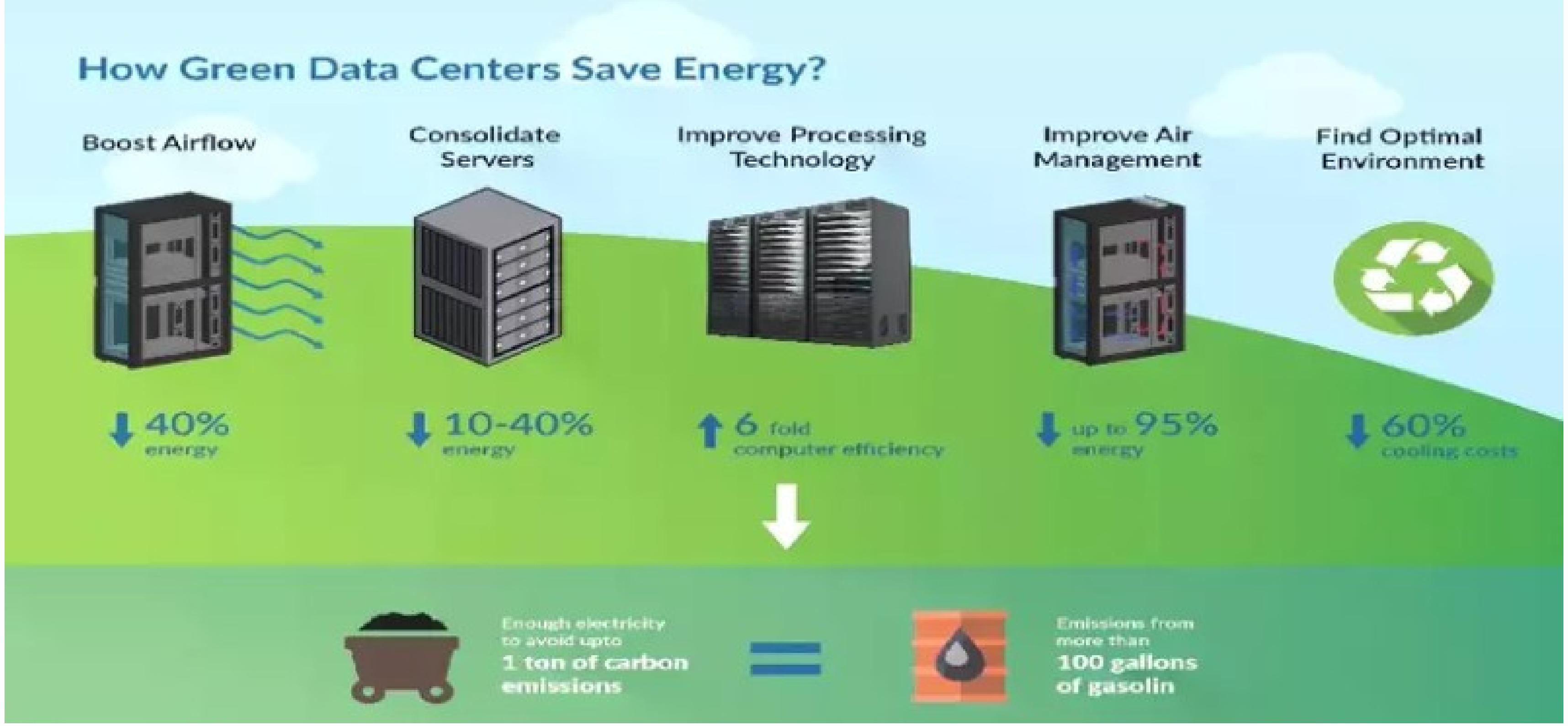
- AI algorithms predictively manage **data life cycles** and **workloads**
- Automated data **deduplication** and **anomaly detection** minimise waste
- AI in storage adapts to user access patterns, **optimises resource allocation**



https://www.nextplatform.com/2020/07/13/an-architecture-for-artificial-intelligence-storage/

Eco-Friendly Data Storage Practices

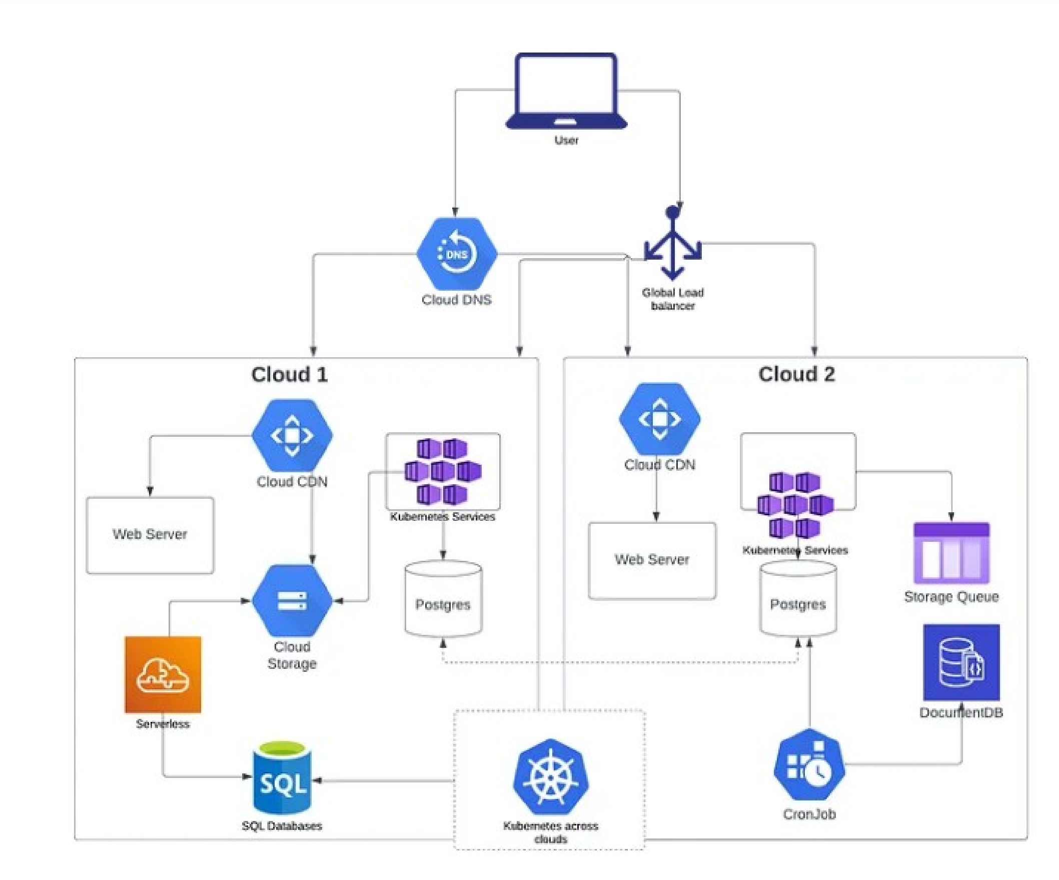
- Assessing the ecological footprint of **data proliferation**
- Innovation in **energy-efficient** servers and cooling systems
- Green data centers reduce **emissions** and **operating costs**
- Policies and practices driving the shift to renewable energy sources


How Green Data Centers Save Energy?
Boost Airflow
Consolidate Servers
Improve Processing Technology
Improve Air Management
Find Optimal Environment
40% energy
10-40% energy
6 fold computer efficiency
up to 95% energy
60% cooling costs
Enough electricity to avoid upto 1 ton of carbon emissions
Emissions from more than 100 gallons of gasolin

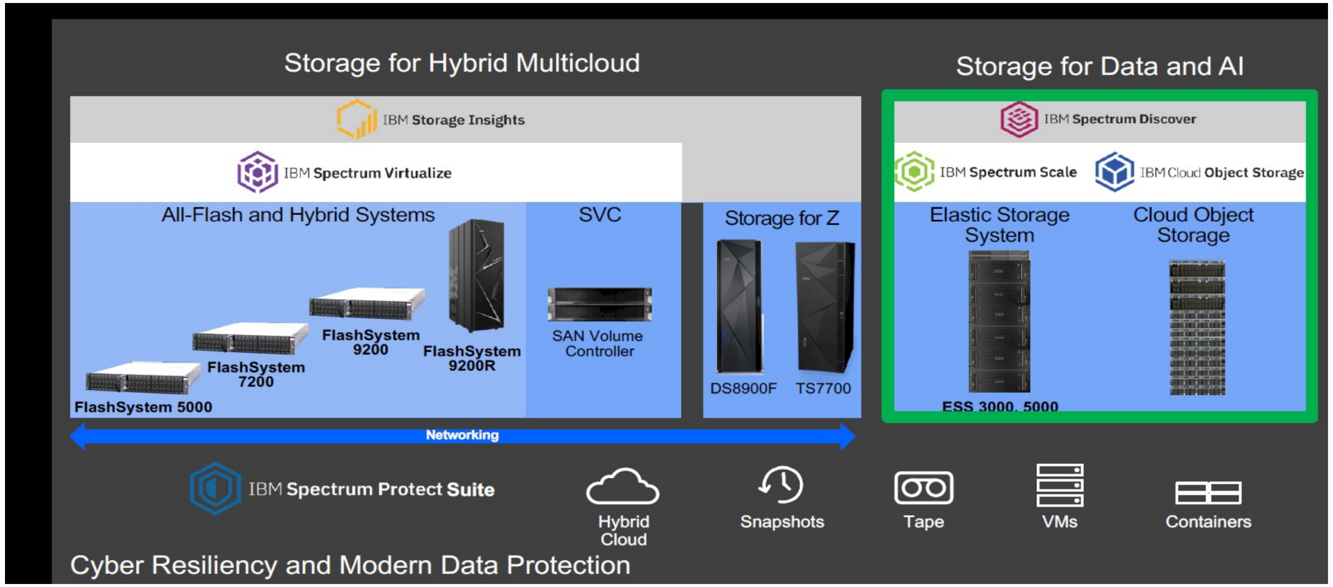
https://rahi.io/green-data-centers/

Strategic Multi-Cloud Deployments

- Multi-cloud approaches mitigate **risks** and ensure service **continuity**
- **Geographical data distribution** aids compliance with local data laws
- Orchestrating **storage solutions** across AWS, Azure, Google Cloud, Apache OpenStack, Apache OpenStack
- **Kubernetes** is the industry-standard tool for deploying microservices and cloud-native applications
- Platforms like OpenShift provide **a unified control plane**, enabling efficient management and monitoring of applications across multiple clouds
- Databases such as NuoDB, FaunaDB, and CockroachDB cater to the **requirements of multi-cloud environments**

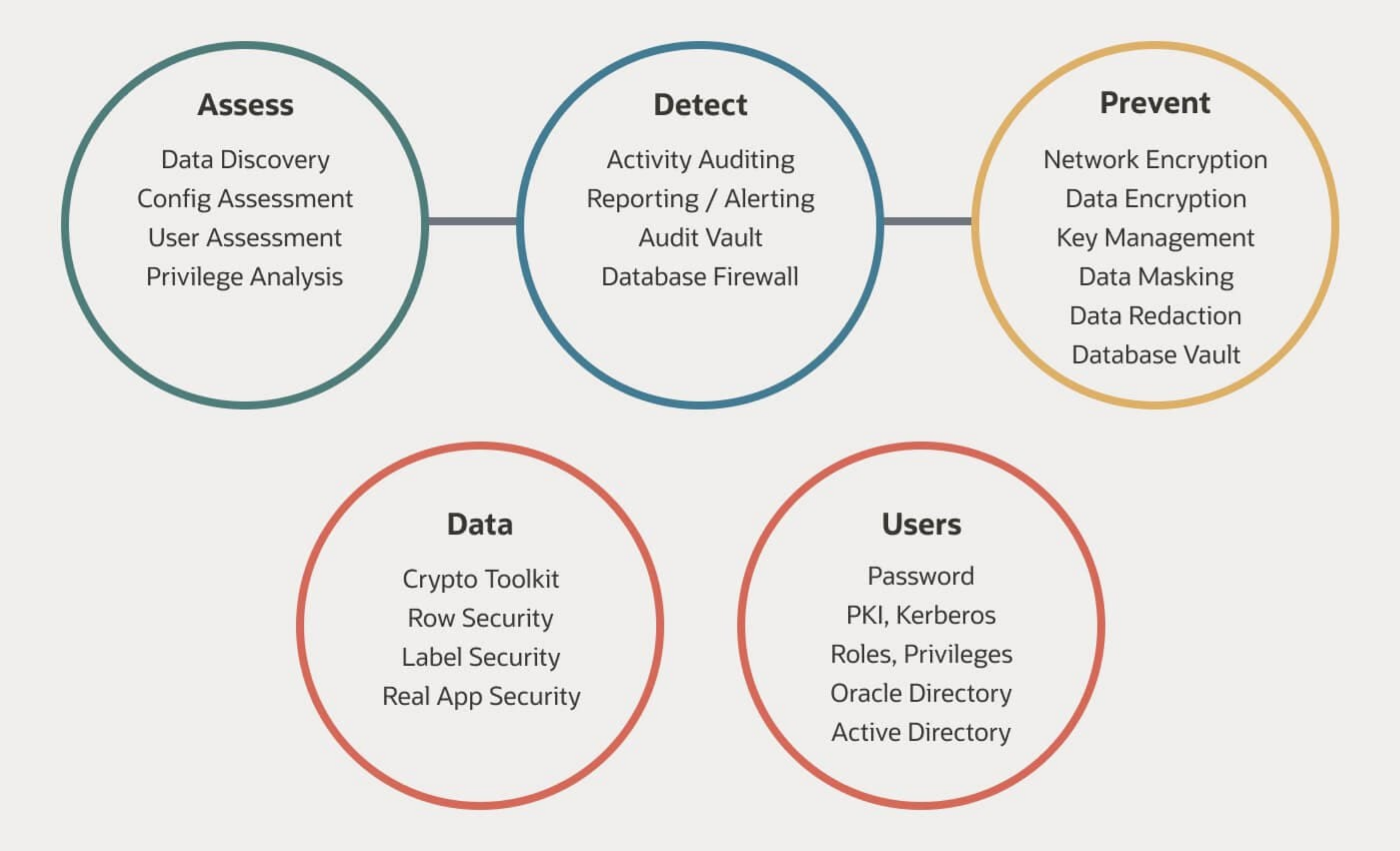


https://miro.medium.com/v2/resize:fit:720/format:webp/1*6d6VaRWpYsJTfzsoGHeS0w.png



https://www.nextplatform.com/2020/07/13/an-architecture-for-artificial-intelligence-storage/

Data Security in Database Systems



Compliance in Data Storage

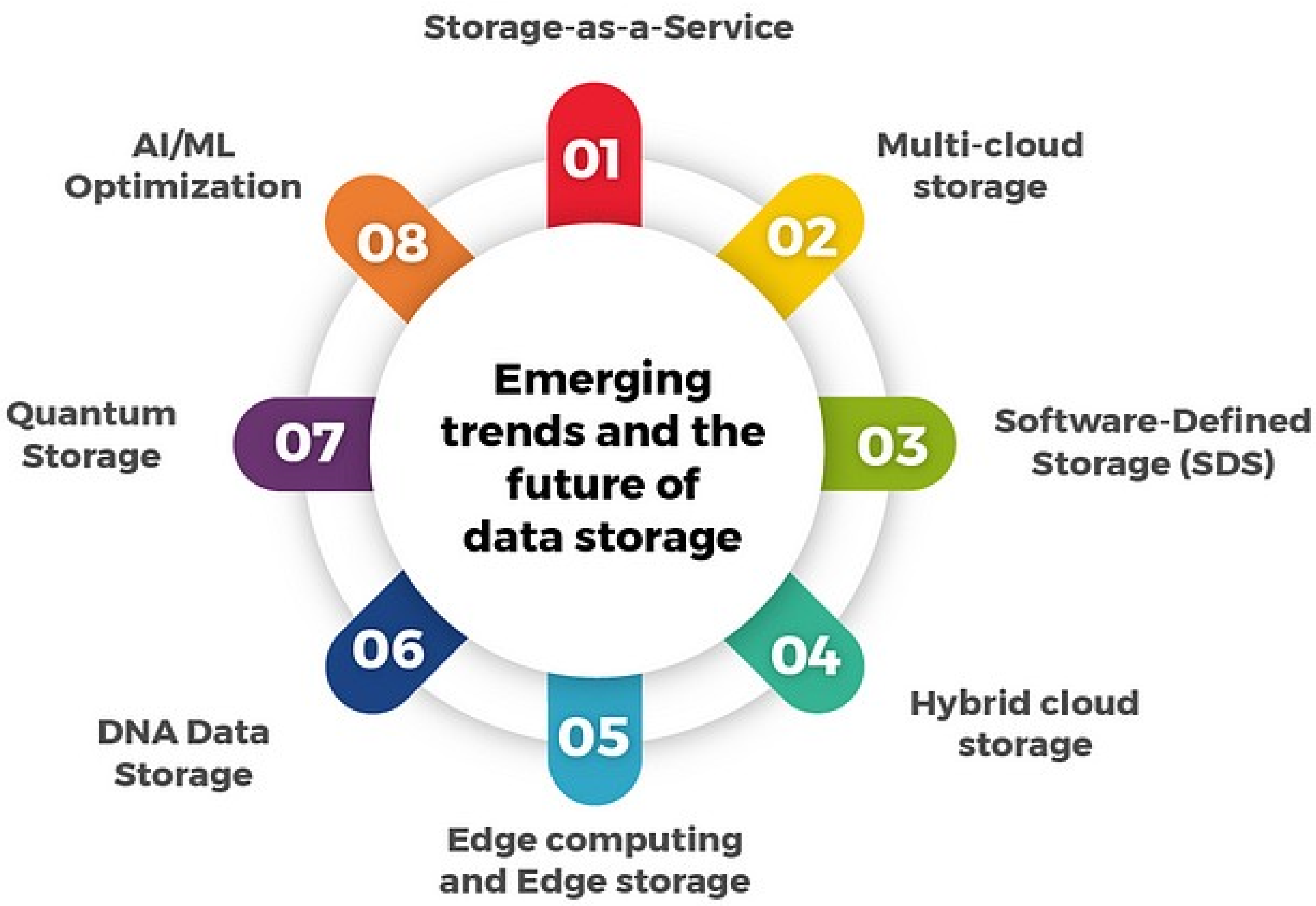
- Aligning database architectures with GDPR, HIPAA, and CCPA
- Implementing compliance through encryption, access controls, and audit trails

Success stories/examples:

Financial Sector: A multinational bank can streamline data governance, achieving GDPR compliance while enhancing customer data protection and transparency.

Healthcare: A hospital network can adopt HIPAA-compliant storage solutions, securing patient records and improving data interoperability across clinics.

Future Data Storage Technologies



Boosting Research for a Smart and Carbon Neutral Built Environment with Digital Twins – **SmartWins**

Project Partners

ktu
1922
kaunas
university of
technology

CERTH
CENTRE FOR
RESEARCH & TECHNOLOGY
HELLAS

POLITECNICO
MILANO 1863

Contecht

INNOTROPE

Funded by
the European Union

This project has received funding from the European Union's Horizon research and innovation programme under grant agreement No 101078997

SmartWins