Database and Information Management

part2

Non-relational cloud databases

- **Non-relational cloud databases** store and manage unstructured data, such as email and mobile message text, documents, surveys, rich media files, and sensor data.
- They don’t follow a clearly-defined schema like relational databases and allow you to save and organize information regardless of its format.

Non-relational database management systems

- Use a more flexible data model
- Designed for managing large data sets across many distributed machines and for easily scaling up or down.
- Useful for accelerating simple queries against large volumes of structured and unstructured data, including web, social media ect.

Need of non-relational database

- Cloud computing
- unprecedented data volumes
- massive workloads for web services
- Need to store new types of data require database

Cloud database

- A cloud database is a database built to run in a public or hybrid cloud environment to help organize, store, and manage data within an organization. Cloud databases can be offered as a managed database-as-a-service (DBaaS)

NoSQL database

NoSQL databases are non-tabular databases and store data differently than relational tables. NoSQL databases come in a variety of types based on their data model.

- Oracle NoSQL database
- MongoDB, Redis, Cassandra, Hbase, and Cloud Bigtable
- Amazon's SimpleDB

Data set using MongoDB

```
{
   "_id": 1,
   "first_name": "Leslie",
   "last_name": "Yepp",
   "cell": "8125552344",
   "city": "Pawnee",
   "hobbies": ["scrapbooking", "eating waffles", "working"]
}
```

News article

MariaDB Launches New Database-as-a-Service, Partners with Google Cloud

By John K. Waters ■ April 2, 2020

MariaDB has launched a new database-as-a-service (DBaaS) this week called SkySQL, which it's billing as the first to provide a "MariaDB in the cloud" experience.

The company behind the popular open source relational database management system (DBMS) also announced a technical partnership with Google Cloud to make SkySQL available on the Google Cloud Platform (GCP).

Amazon Cloud Databases

- Amazon and other cloud computing vendors provide relational database services as well.
- Amazon Relational Database Service (Amazon RDS) offers MySQL, SQL Server, Oracle Database, PostgreSQL, MariaDB, or Amazon Aurora DB (compatible with MySQL) as database engines.
- Pricing is based on usage.

- Oracle has its own Database Cloud Services using its relational Oracle Database
- Microsoft Windows SQL Azure Database is a cloud-based relational database service based on Microsoft's SQL Server DBMS.

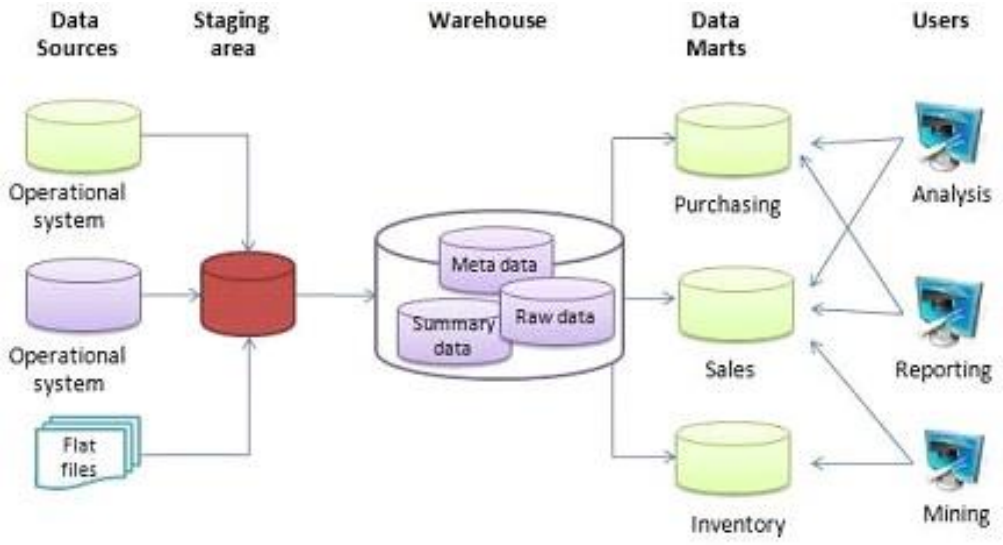
Advantages of Cloud Database

- Special appeal for web-focused start-ups or small to medium-sized businesses seeking database capabilities at a lower price than in-house database products.

What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?

Data Warehouses

- A data warehouse is a database that stores current and historical data of potential interest to decision makers throughout the company.

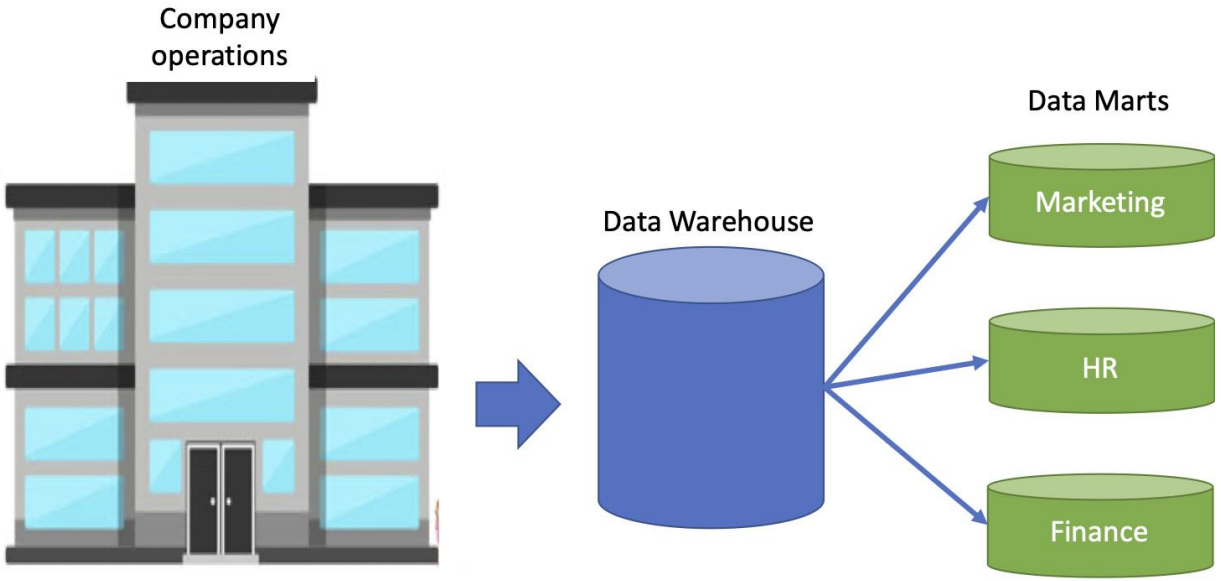


Data warehouse..

- The data warehouse makes the data available for anyone to access as needed, but the data cannot be altered.
- A data warehouse system also provides a range of ad hoc and standardized query tools, analytical tools, and graphical reporting facilities.

Data Marts

- A data mart is a subset of a data warehouse in which a summarized or highly focused portion of the organization's data is placed in a separate database for a specific population of users.



Hadoop

- Hadoop is an open source software framework managed by the Apache Software Foundation that enables distributed parallel processing of huge amounts of data across inexpensive computers.

Hadoop…

- For handling unstructured and semi-structured data in vast quantities, as well as structured data, organizations are using Hadoop.
- It breaks a big data problem down into sub-problems, distributes them among up to thousands of inexpensive computer processing nodes, and then combines the result into a smaller data set that is easier to analyze.

- Hadoop can process large quantities of any kind of data(Facebook and Twitter feeds, complex data such as web server log files, and unstructured audio and video data)
- Companies use Hadoop for analyzing very large volumes of data as well as for a staging before they are loaded into a data warehouse.
- Yahoo uses Hadoop to track users' behavior so it can modify its home page to fit their interests

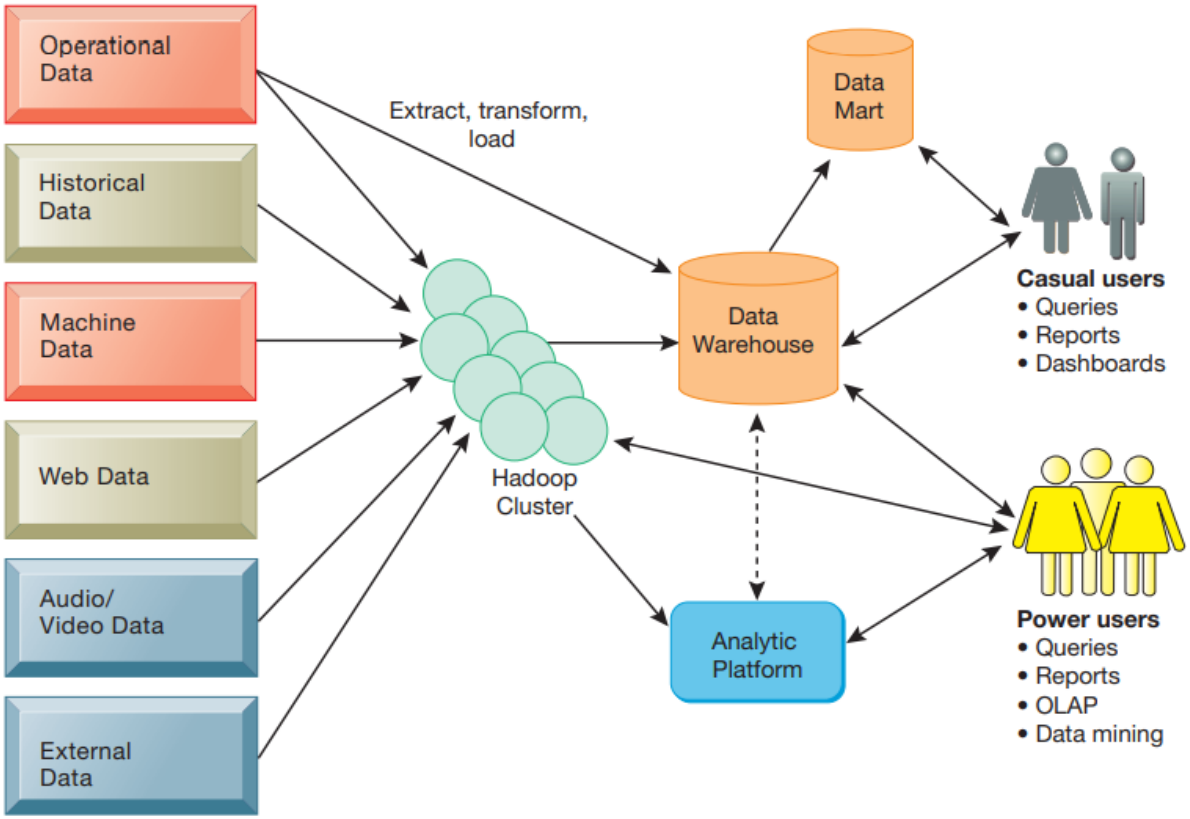
In-Memory Computing

- Facilitating big data analysis is to use in-memory computing ,which relies primarily on a computer's main memory (RAM) for data storage.
- Users access data stored in system primary memory,.
- Dramatically shortening query response times.
- In-memory processing makes it possible for very large sets of data, amounting to the size of a data mart or small data warehouse, to reside entirely in memory.

Analytic Platforms

- Analytic platforms such as IBM PureData System for Analytics, feature preconfigured hardware-software systems that are specifically designed for query processing and analytics.
  - For example, IBM PureData System for Analytics features tightly integrated database, server, and storage components that handle complex analytic queries 10 to 100 times faster than traditional systems.
- Analytic platforms include in-memory systems and NoSQL non-relational database management systems.
- Analytic platforms are now available as cloud services.

FIGURE 6.12 CONTEMPORARY BUSINESS INTELLIGENCE INFRASTRUCTURE



A contemporary business intelligence infrastructure features capabilities and tools to manage and analyze large quantities and different types of data from multiple sources. Easy-to-use query and reporting tools for casual business users and more sophisticated analytical toolsets for power users are included.

Analytical Tools: Relationships, Patterns, Trends

- Once data have been captured and organized using the business intelligence technologies we have just described, they are available for further analysis using software for database querying and reporting, multidimensional data analysis (Online Analytical Processing - OLAP), and data mining.

Online Analytical Processing (OLAP)

- Suppose your company sells four different products—nuts, bolts, washers, and screws—in the East, West, and Central regions.
- If you wanted to ask a fairly straightforward question, such as how many washers sold during the past quarter, you could easily find the answer by querying your sales database.
- But what if you wanted to know how many washers sold in each of your sales regions and compare actual results with projected sales? To obtain the answer, you would need online analytical processing (OLAP).
- OLAP supports multidimensional data analysis, enabling users to view the same data in different ways using multiple dimensions.
- OLAP enables users to obtain online answers to ad hoc questions.

Data mining

- **Data Mining** is a process of finding potentially useful patterns from huge data sets.
- The patterns and rules are used to guide decision making and forecast the effect of those decisions.

The types of information obtainable from data mining

- associations
- sequences
- classifications
- Clusters
- forecasts.

1. Associations

Associations are occurrences linked to a single event.

Eg: A study of supermarket purchasing patterns might reveal that,

- when corn chips are purchased, a cola drink is purchased 65 % of the time, but when there is a promotion, cola is purchased 85% of the time.
- This information helps managers make better decisions because they have learned the profitability of a promotion.

2. sequences

- In sequences events are linked over time.
  - If a house is purchased, a new refrigerator will be purchased within two weeks 65% of the time, and an oven will be bought within one month of the home purchase 45 % of the time.

Classification

- **Classification** recognizes patterns that describe the group to which an item belongs by examining existing items that have been classified and by supposing a set of rules.
  - For example, businesses such as credit card or telephone companies worry about the loss of steady customers.
  - Classification helps discover the characteristics of customers who are likely to leave
  - can provide a model to help managers predict who those customers are so that the managers can devise special campaigns to retain such customers.

Clustering

- **Clustering** works in a manner similar to classification when no groups have yet been defined.
- A data mining tool can discover different groupings within data,
  - Eg: finding affinity groups for bank cards or partitioning a database into groups of customers based on demographics and types of personal investments.

END