Lecture 02: Introduction

1. ***The quality concept***
2. ***Communication***
3. ***Different types of data***
4. ***Different types of data storage***
5. ***Database***
6. ***RDBMS***
7. ***Cloud services***
8. ***Tables***
9. ***ANSI-SPARC***
10. ***SQL***
11. ***Oracle Architecture***

Pär Douhan, pdo@du.se

Quality?

Data quality?

Information quality?

The concept of quality

Quality is defined in the ISO 8402 - 1986 as:

- "The totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs"

According to *Bo Sundgren*, the term quality is used in two different ways:

1. **In a neutral way** - in the neutral mode, quality has about the same meaning as property. Some data can be provided with a number of properties that can be regarded as quality descriptions for the data.
2. **In a normative way** - in the normative mode it makes sense to talk about good quality, better quality, inferior quality and poor quality.

Good quality is related to the user and the use. What in one situation may be good quality, may in another situation be unacceptably poor quality.

Service quality- characteristics

Properties that characterize service quality (*Bergman och Klefsjö*):

1. **Environment -** the environment in which the service is presented.
2. **Reliability -** how reliable the service is performed by the supplier.
3. **Availability -** how easy it is for the customer to get in touch with the supplier.
4. **Willingness -** the willingness of the supplier to help the customer.
5. **Courtesy -** the behavior of the supplier in the form of courtesy and kindness.
6. **Communication ability -** ability to talk to the customer in a way that the customer understands.

Test yourself to evaluate e.g. purchase and installation of a kitchen, or a complete renovation of the bathroom using the above quality criteria!

Product quality- characteristics

Properties that characterize product quality (*Bergman och Klefsjö*):

1. **Performance** - how the product meets customer expectations for performance.
2. **Dependable** - how often there are errors in the product and how serious the errors are.
3. **Maintenance** - how easy it is to maintain and repair the product.
4. **Safety** - how safe the product is to use.

Quote by unknown 90-year-old: "If I still knew that life consists of 90% maintenance"

Usability - characteristics

Properties that characterize usability (*Stefan Cronholm*):

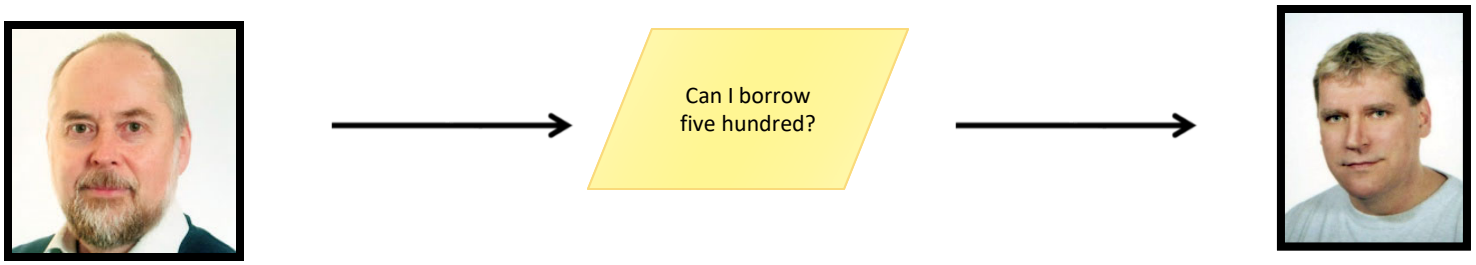
(Usability is a key research topic in the HCI, Human Computer Interaction, area)

1. **Relevance** - how well IS meets the user's needs.
2. **Efficiency** - how efficiently or productively the tasks can be performed.
3. **User attitude** - the user's subjective feelings towards the system.
4. **Learnability** - how easy the system is to learn.
5. **Flexibility** - how IS copes with adaptations to business changes.
6. **Concurrency** – how many concurrent users can the IS handle.

Analyze Learn based on the above criteria.

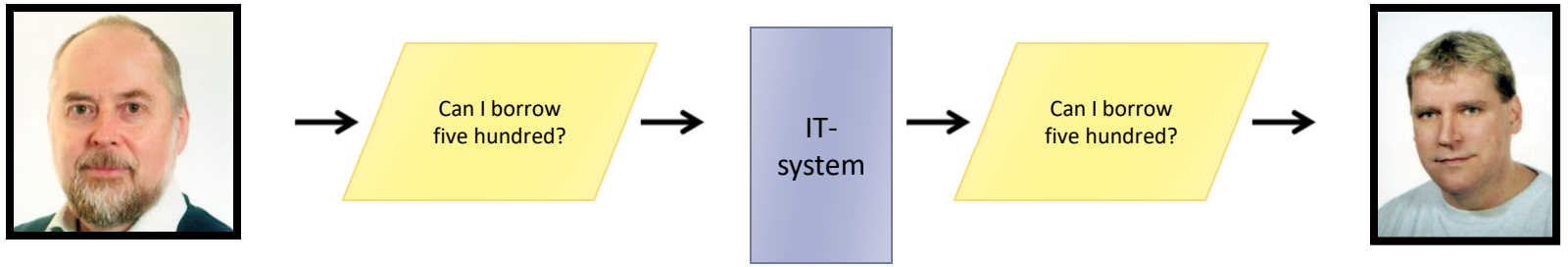
Communication

Direct communication: Hans asks Pär with a submissive tone:



1. Pär perceives the message with his sences,
2. Interprets the message,
3. "Asks" his *Frame of Reference,*
4. Answer: Yes!

Communication via IT system: Hans writes a message:



Communication quality

Generic criteria for communication quality according to (*Owen Eriksson*):

1. **The information must be understandable and relevant to the user.**
2. **Relevant and understandable action aspect** - the verb in the speech act must be comprehensible. *Is it an offer? An order? A confirmation? A wish? An agreement? A question?*
3. **Trustworthy information** – the user must be able to trust that the information is credible.

Credible information

Often it is the lack of routines that makes users begin to believe that the information is not credible. If this happens, it does not matter if the data quality is high or not, it is enough to believe that it is poor.

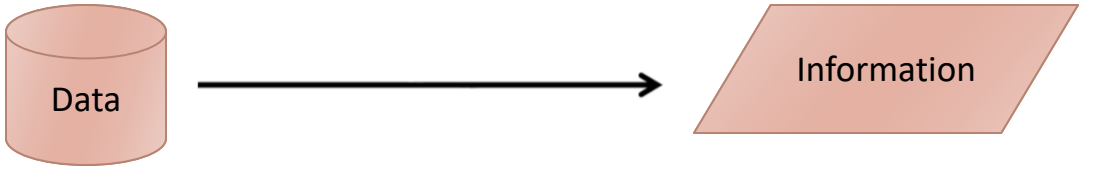
One can increase credibility by storing information about **who** has updated certain information and **when** it was done.

PRICE_CHANGES

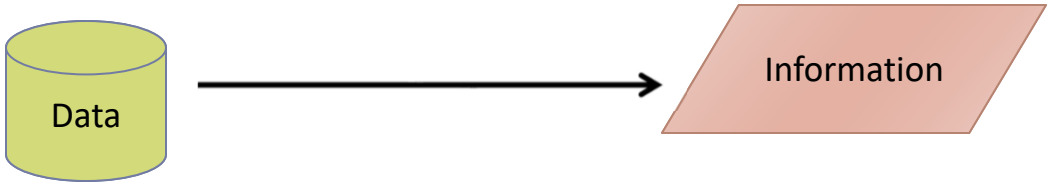
prch_id	product_id	price	**time**	**who**
1234	A-3524-B12	995	2019-08-25:14:55	Ola Nääs
1235	A-3524-B12	1048	2019-08-25:17:21	Kenneth Carling

Routines for data collection?

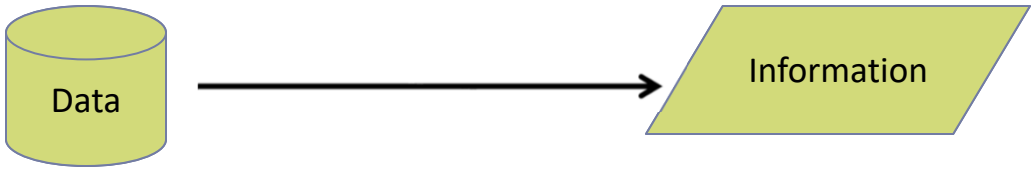
Data and information quality



Low data quality results in low information quality.



High data quality can result in low information quality. Insufficient metadata, low credibility, insufficient frame of reference of the user.



High data quality can result in high information quality. Enough metadata, high credibility, sufficient frame of reference of the user.

Consequences

Information

low information quality

Decisions are made using information. If the information (decision basis) is insufficient, there is a high risk of bad decisions.

Some examples of consequences may be:

1. Cables can break when digging. If there is insufficient location accuracy on geographical coordinates.
2. In war situations, refugee centers and hospitals can be bombed instead of military targets, if the location accuracy is too low or if it is incorrect.
3. Patients may be ill-treated in care.
4. Innocent people can be convicted of crimes.

Different types of data (Oracle)

Data Realm	Structured	Volume	Description	Examples
Master Data	Structured	Low	Enterprise-level data entities	Customer, products
Transaction Data	Structured & semi-structured	Medium - high	Business transactions	Payments, orders
Reference Data	Structured & semi-structured	Low - medium	Internally managed or externally sourced facts to support an organization's ability to effectively process transactions, manage master data, and provide decision support capabilities.	Geo data and market data
Meta Data	Structured	Low	"data about the data"	Units, definitions
Analytical Data	Structured	Medium - high	Derivations of the business operation and transaction data used to satisfy reporting and analytical needs.	Data that reside in data warehouses
Documents	Unstructured	Medium - high	Documents, digital images, geo-spatial data, and multi-media files.	Images, video files
Big Data	Structured, semi-structured, & unstructured	High	Large datasets that are challenging to store, search, share, visualize, and analyze.	User and machine-generated content through social media, web and software logs, cameras, information-sensing mobile devices,

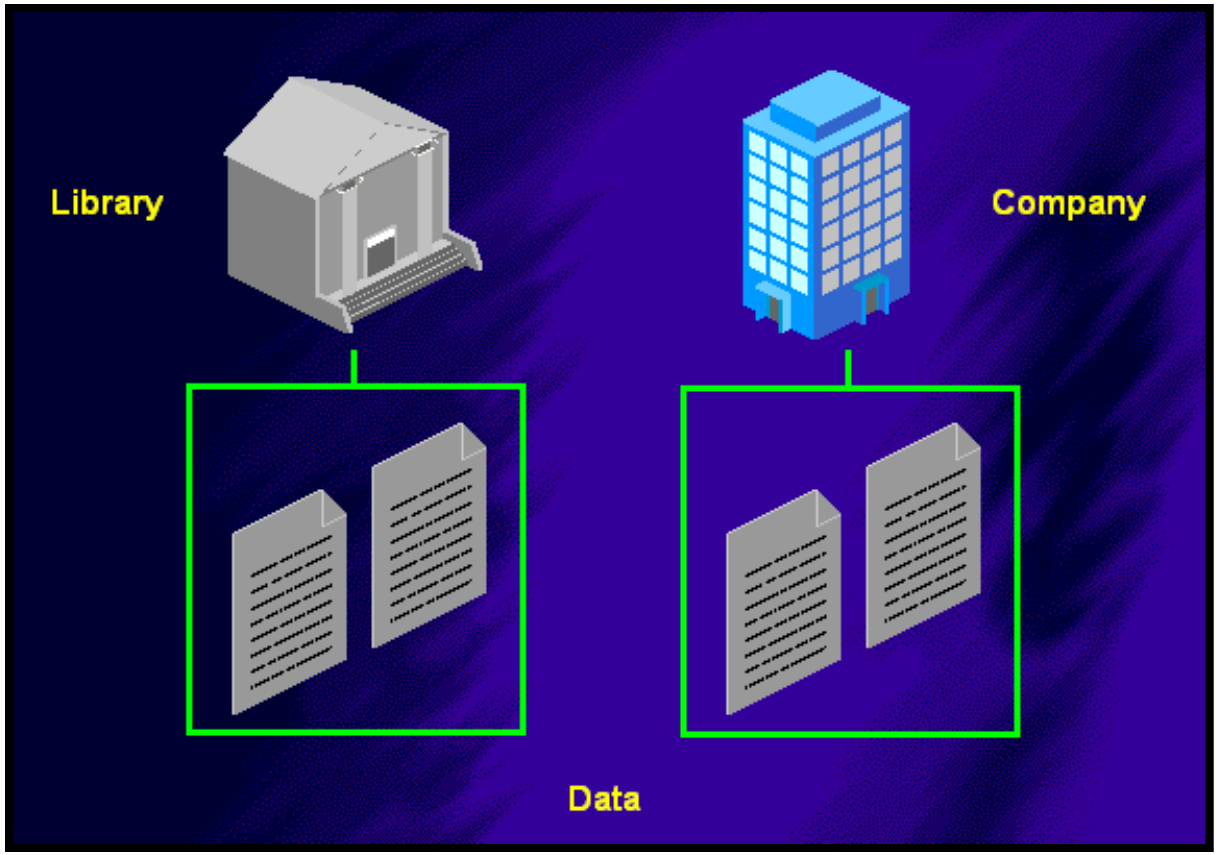
Different types of data storage

Data Realms	Security	Storage & Retrieval	Modeling	Processing & Integration	Consumption
Master data Transactions Analytical data Metadata	Database, app, & user access	RDBMS / SQL	Pre-defined relational or dimensional modeling	ETL/ELT, CDC, Replication, Message	BI & Statistical Tools, Operational Applications
Reference data	Platform security	XML / xQuery	Flexible & Extensible	ETL/ELT, Message	System-based data consumption
Documents and Content	File system based	File System / Search	Free Form	OS-level file movement	Content Mgmt
Big Data - Weblogs - Sensors - Social Media	File system & database	Distributed FS / noSQL	Flexible (Key Value)	Hadoop, MapReduce, ETL/ELT, Message	BI & Statistical Tools

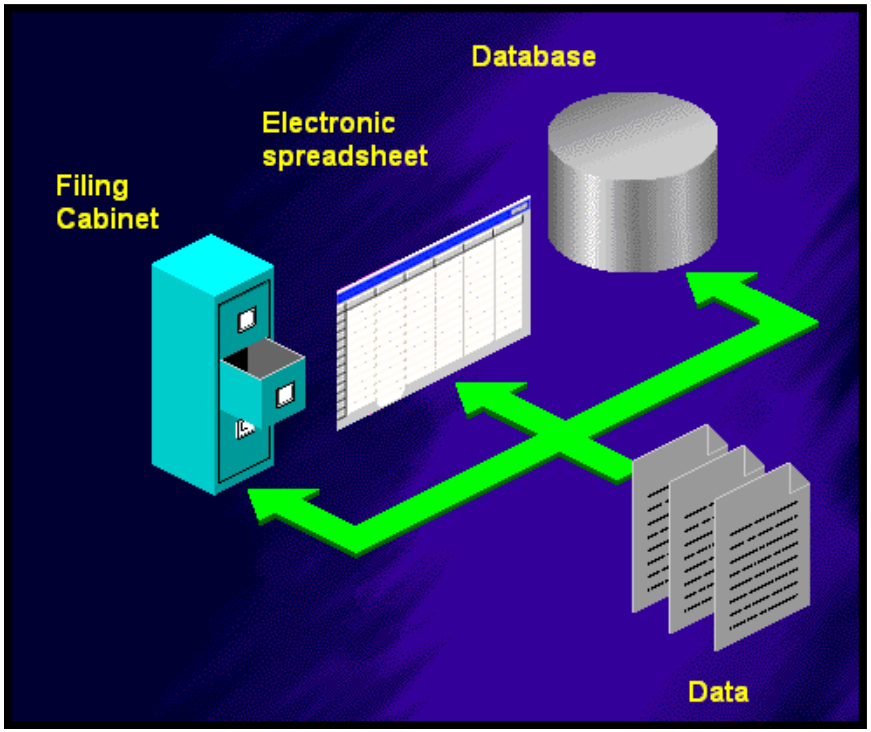
Table 2: Data Realm Characteristics (Oracle Information Architecture Framework)

Information needs

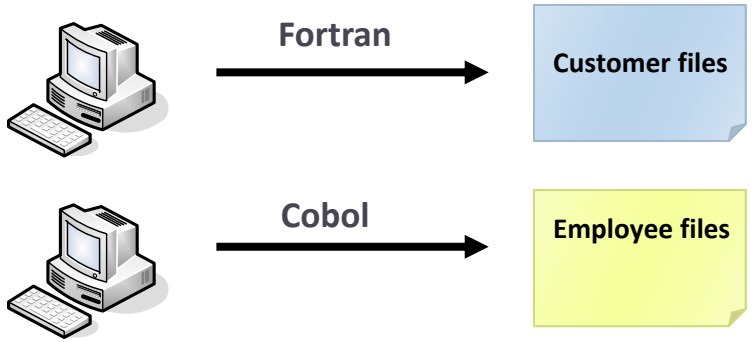
Every company, organization or business has a need for **information**



Data storage



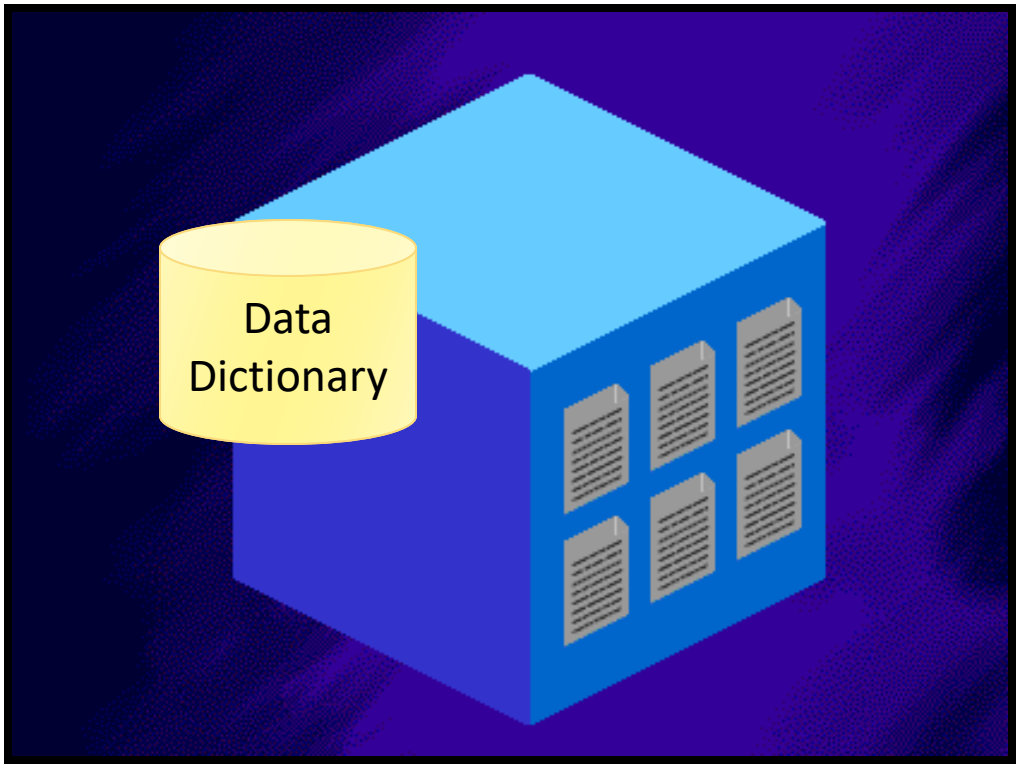
In the 1960's, data storage was dominated by so-called simple file systems.



Each program processed its own data files.

High level of dependency between programs and data files.

Database

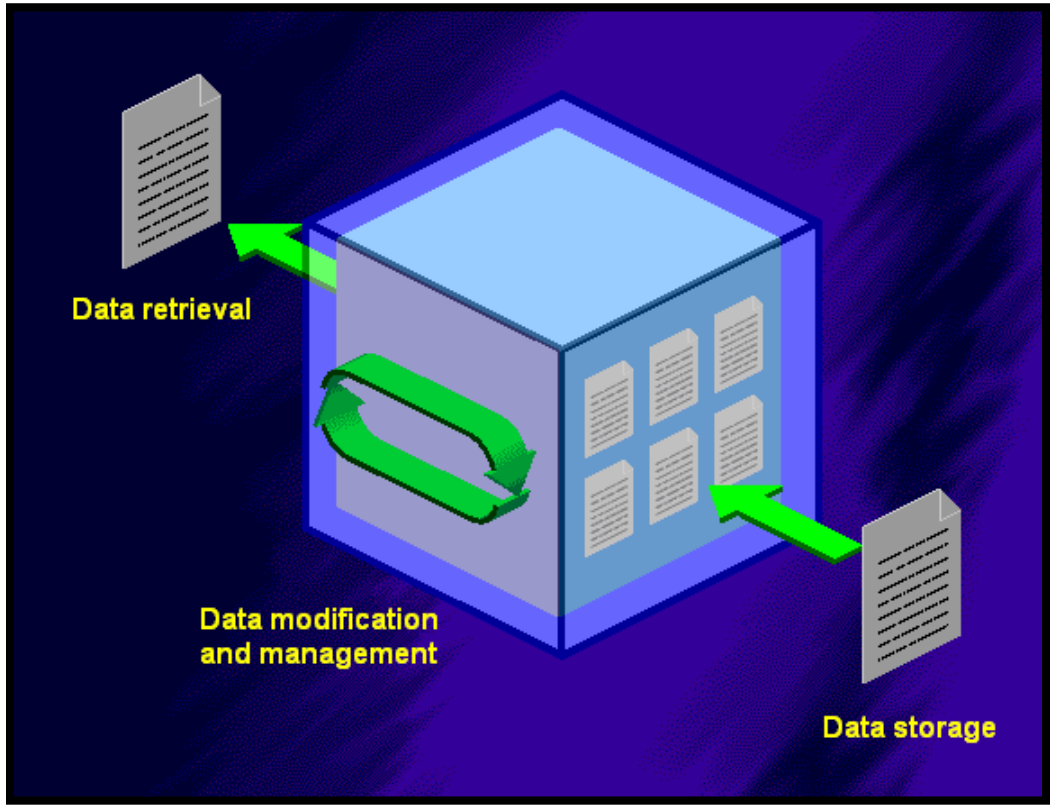


A database is an organized collection of information.

A database always contains a description of stored information. This is usually called the Data Dictionary. It contains metadata about all the objects in the database

It should be **easy** to **read** and **manipulate** information in the database.

DBMS



DBMS = Database Management System

DBMS is a system that stores, modifies and retrieves data in the database.

Different terms are used: ORDBMS, RDBMS and DBMS**.**

Different types of databases

Relational databases, came in the 1970s. Stores master data, transaction data and analytical data.

NoSQL databases, come well into the 2000's. Stores Big Data, data from social media, sensors, web logs etc.

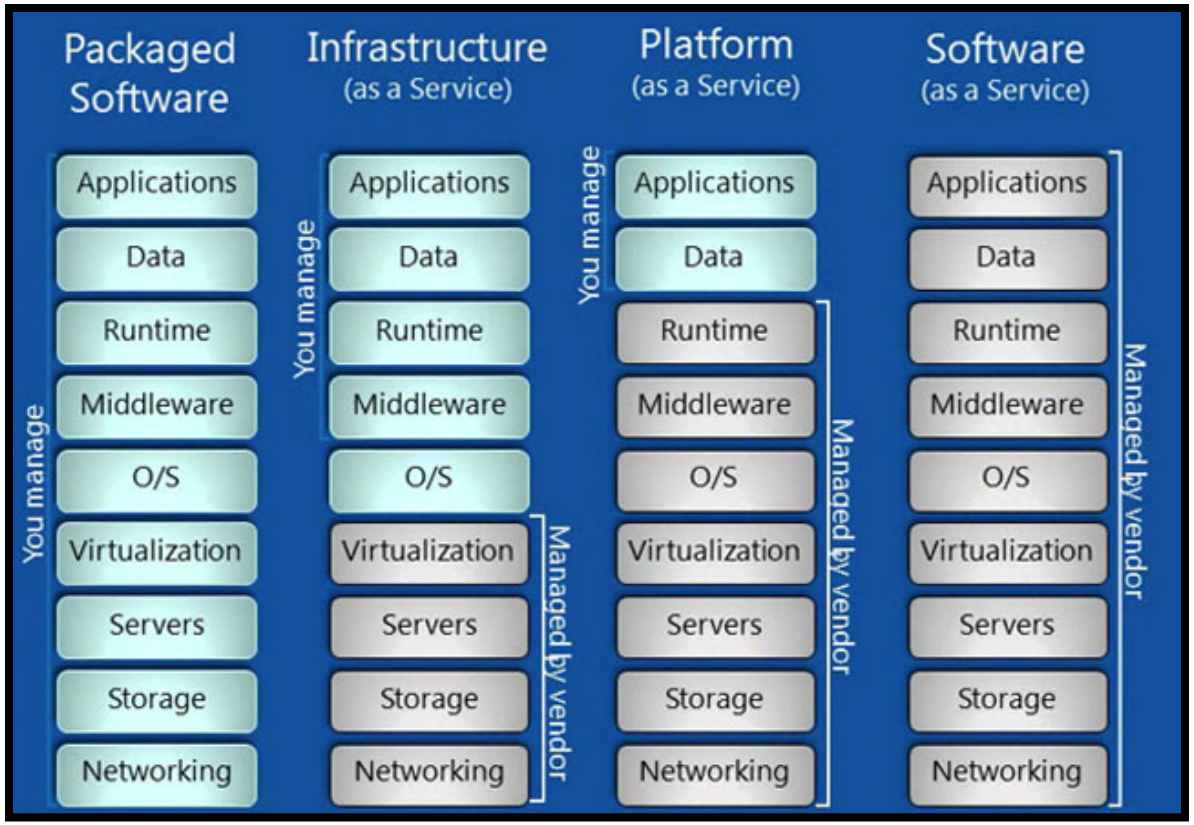
Different suppliers

There are various suppliers of databases. Examples of some common databases:

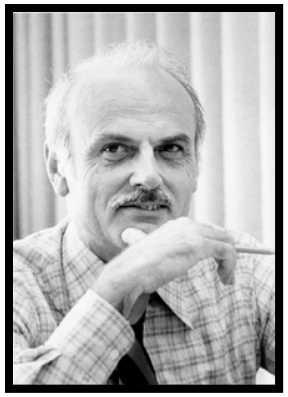
Various cloud services

Data as a service (DaaS) http://www.gapminder.org *Hans Rosling*

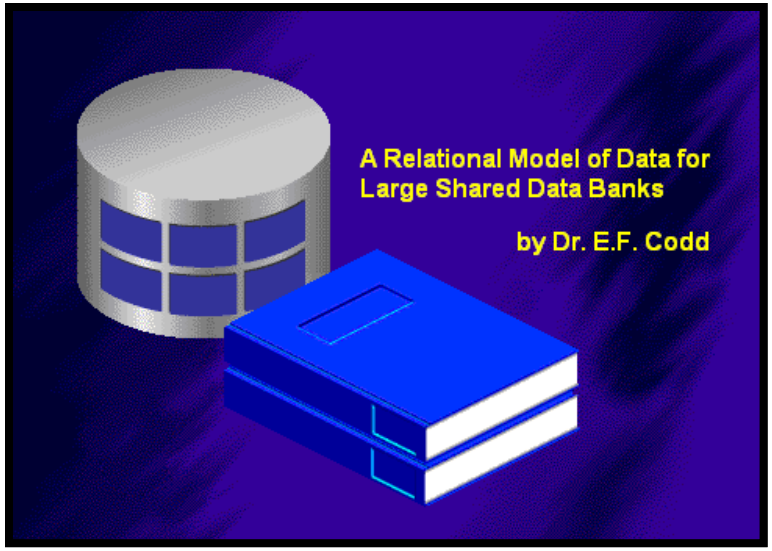
Database as a service (Cloud Database)



Edgar "Ted" Codd



Edgar Frank "Ted" Codd (August 19, 1923 - April 18, 2003) was an English computer scientist who, when working for IBM, in 1965 invented the relational model for database management, the theoretical basis for relational databases.

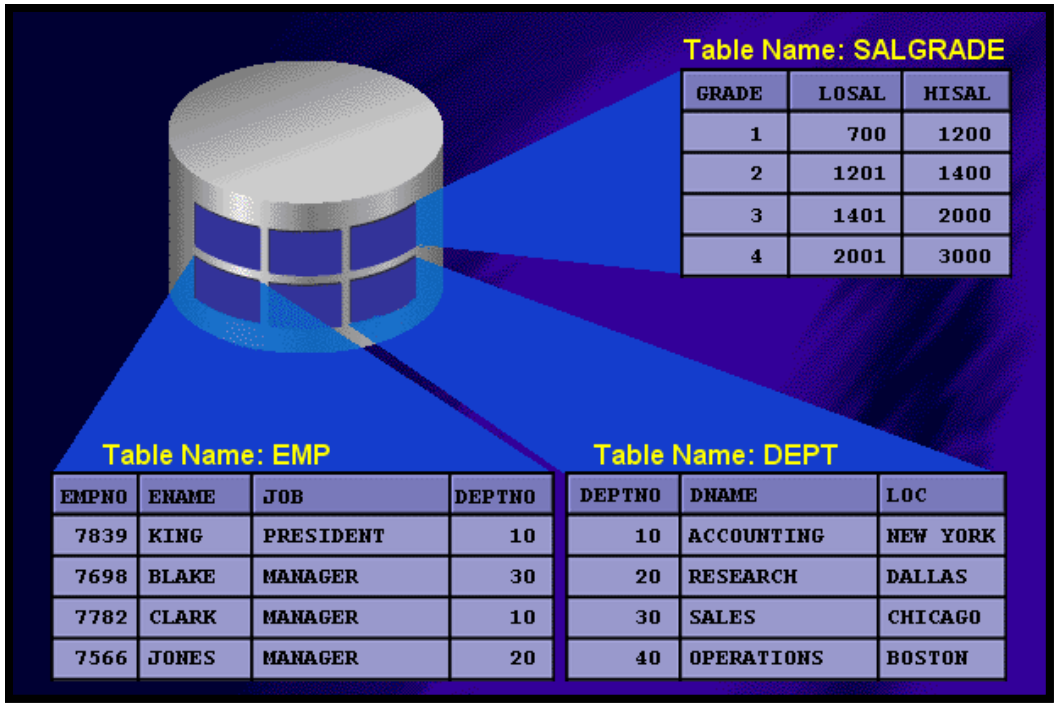


Data is organized in tables at the logical levels.

High level of program / data independence. Client applications do not need to know how data is physically stored in the database.

This is realized by a three-layer (schema) architecture according to ANSI / SPARC, with two logical and one physical level.

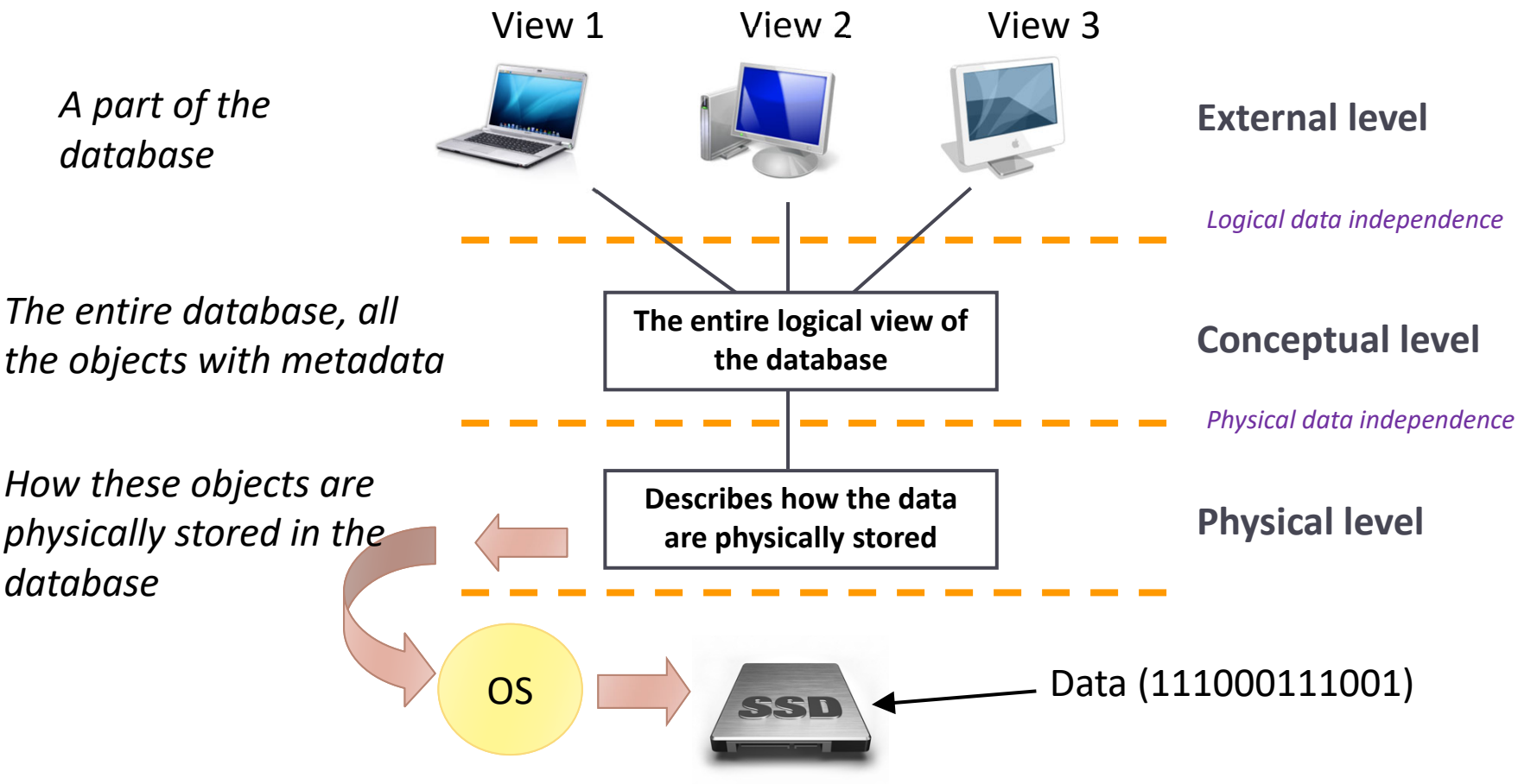
Tables



A relational database is a collection of relations or two-dimensional tables.

A relational database uses these two-dimensional tables to store data.

The ANSI-SPARC architecture



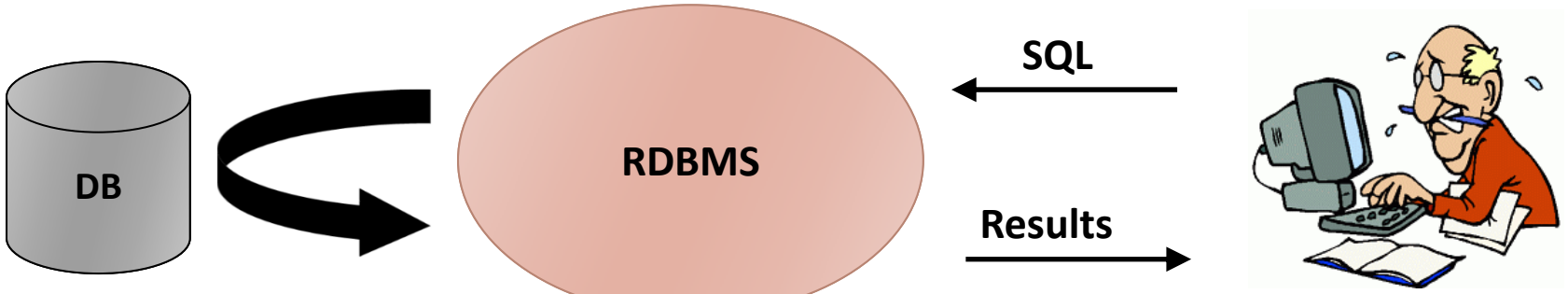
Common for all databases

All databases consist of:

1. **Memory** (physical RAM memory)
2. **Processes** (background processes or tasks)
3. **Storage** (files on disk)

SQL

SQL, Structured Query Language



Database: A collection of files; data files, control files and redo log files stored on disk.

RDBMS: Software, for example Oracle or MySQL. Provides access to the files in the database. Consists of different memory areas and background processes.
In Oracle's architecture the RDBMS is called the *Oracle Instance*.

Oracle

Oracle Server

client

User process

unique session

SQL

Results

Server process

Background processes
(Tasks)

Oracle Instance

SGA

Shared pool

Library cache

Dictionary cache

database buffer cache

Redo log buffer

PMON SMON DBWR LGWR CKPT

RDBMS

Data files

Control files

redo logg 1a

redo logg 1b

Data files

Control files

DB

Different roles

DBA, Database administrator, responsible for security, operation and maintenance of the database.

Developers, those who develop applications.

End users, those who are using the database in their daily work.

SQL: five parts

The SQL language consists of five different parts:

1. **Data Retrieval Language**
   - select
2. **DML, Data Manipulation Language**
   - insert, update och delete
3. **DDL, Data Definition Language**
   - create, alter och drop
4. **DCL, Data Control Language**
   - grant och revoke
5. **TCL, Transaction Control Language**
   - commit och rollback