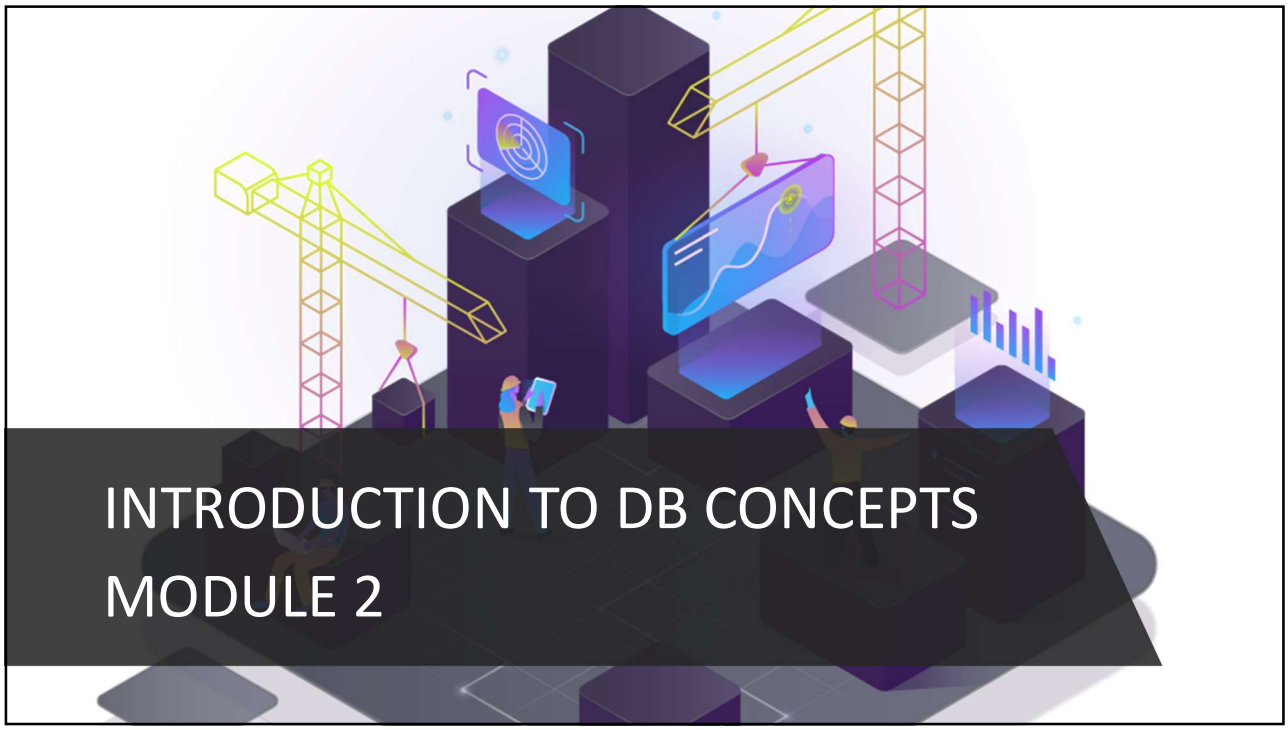


INTRODUCTION TO DB CONCEPTS
MODULE 2

Objectives

After completing this module, you will be able to:

- Define the difference between data and information
- Describe what a database is, the various types of databases, and why they are valuable assets for decision making
- Explain the importance of database design
- Outline the main components of the database system
- Describe the main functions of a database management system (DBMS)

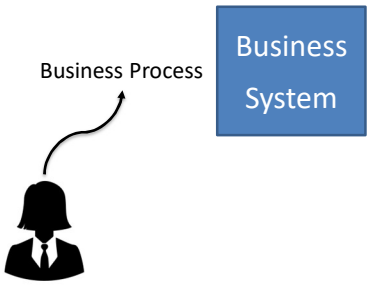
INTRODUCTION TO IT SYSTEMS

Tale of Two Systems

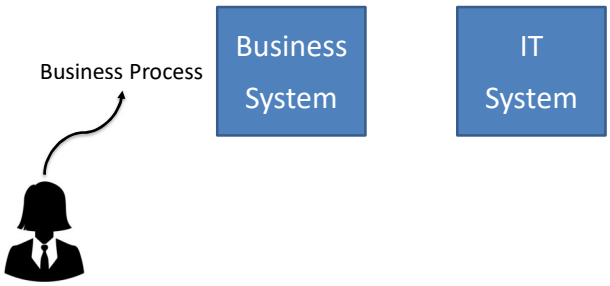
Tale of Two Systems

Business System

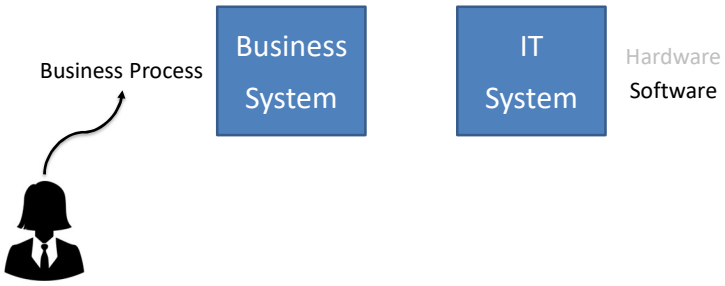
Tale of Two Systems



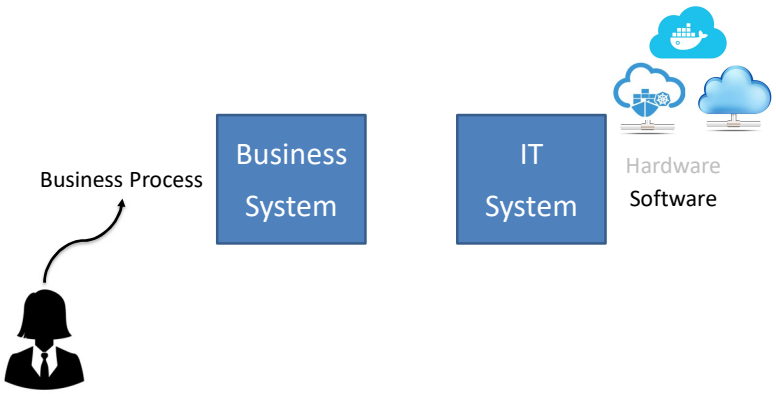
Tale of Two Systems



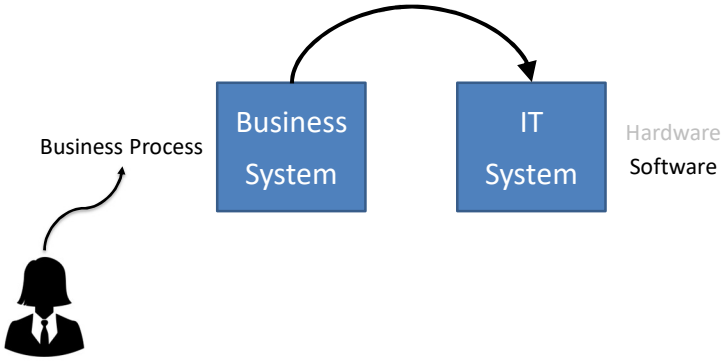
Tale of Two Systems



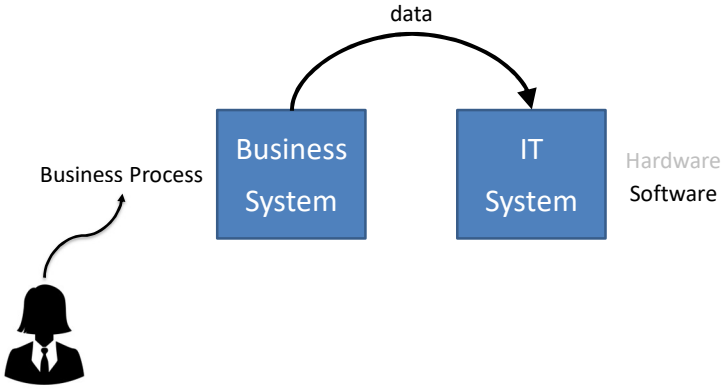
Tale of Two Systems



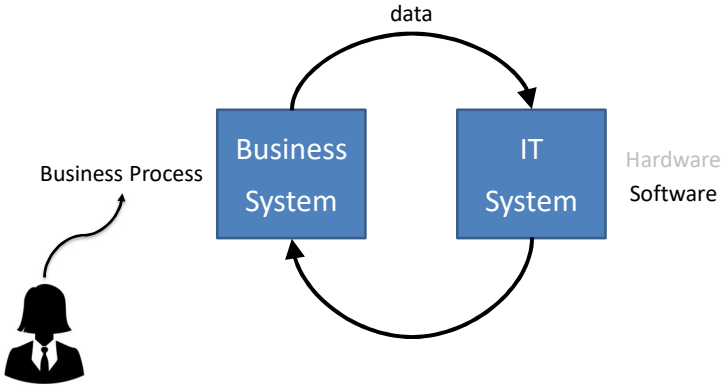
Tale of Two Systems



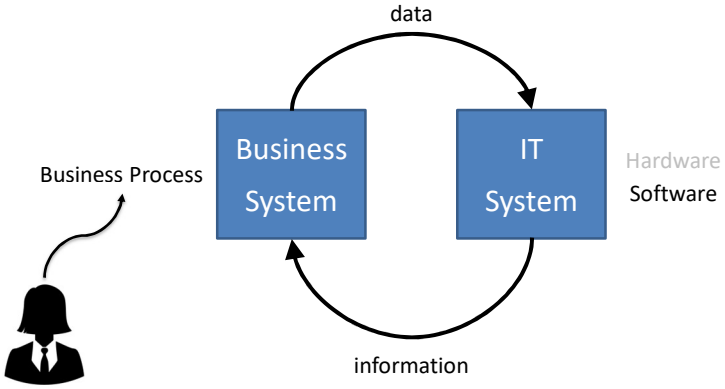
Tale of Two Systems



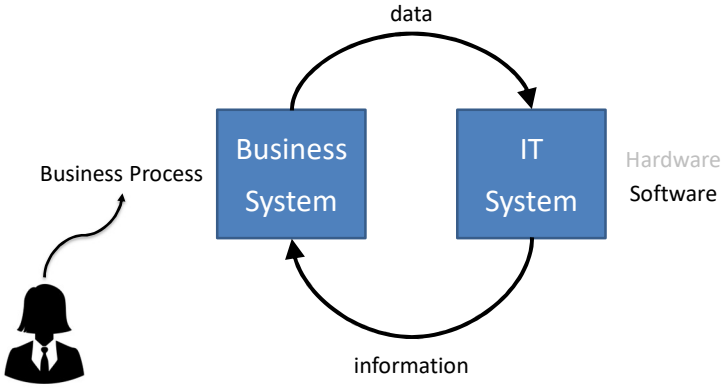
Tale of Two Systems



Tale of Two Systems

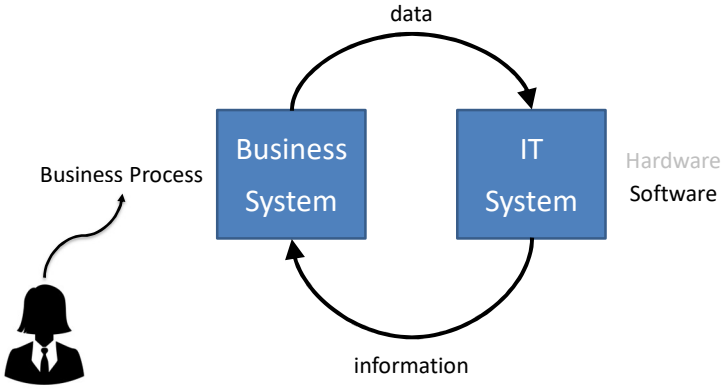


Tale of Two Systems



Alignment Problem

Tale of Two Systems



Alignment Problem

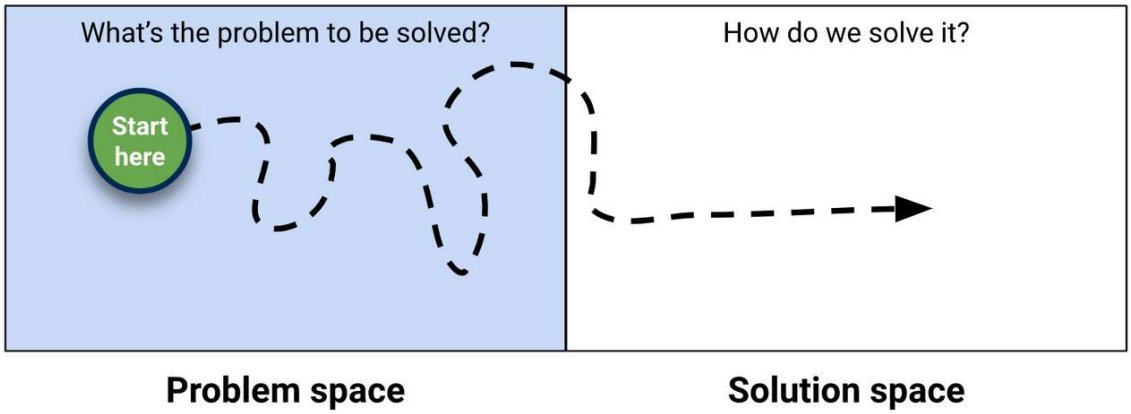
Cloud-native Applications
μ-Service Architecture
Serverless Architecture

Problem Space and Domain

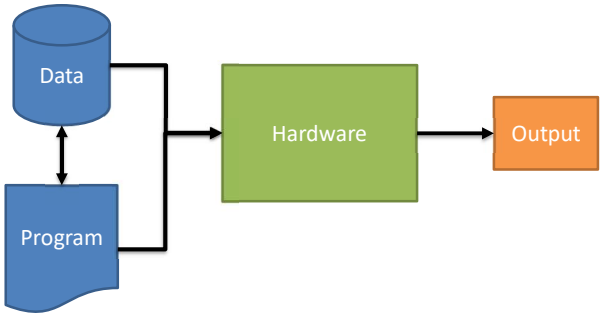
> Domain is the space where the problem is defined
  - Banking
  - Insurance
  - Telecommunication
  - E-commerce

The reality

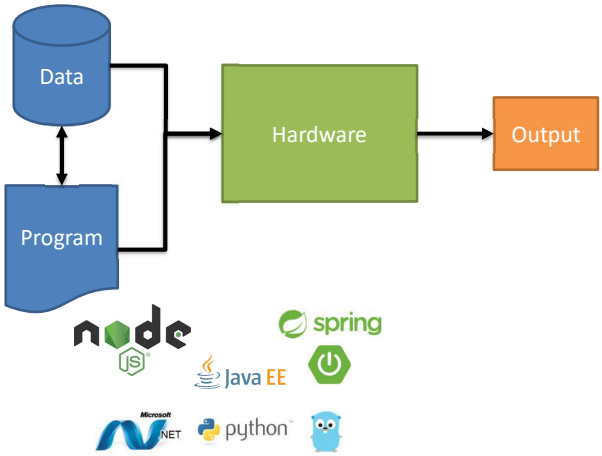
Solution Space



Traditional Programming Paradigm

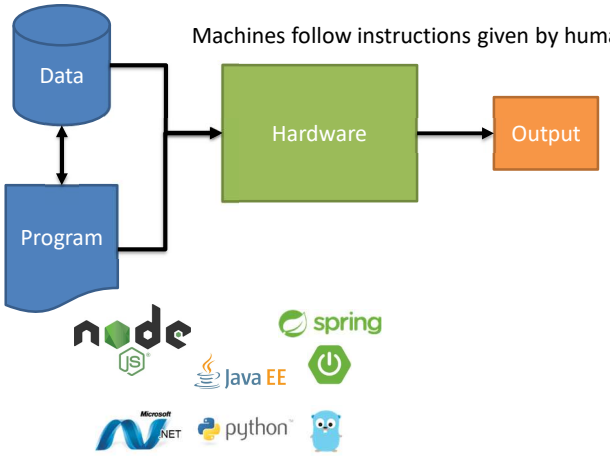


Traditional Programming Paradigm



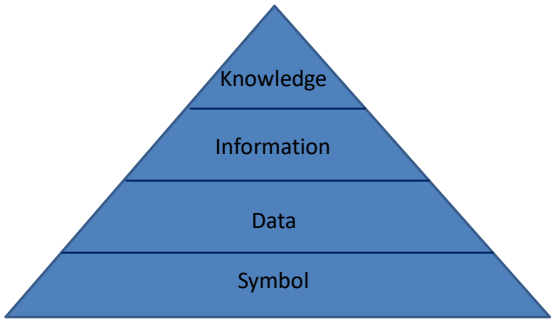
Traditional Programming Paradigm

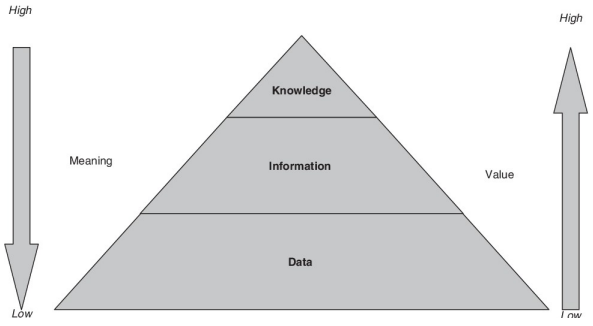
Machines follow instructions given by humans



DATA
VS
INFORMATION
VS
KNOWLEDGE

Symbol → Data → Information → Knowledge





> Symbols (e.g. 0,1,...,9,A,B,...,Z,!,+,-,...)

> Data are facts, numbers, or individual entities without context or purpose.

> 000101020305080D1522375990

Credit Card Number? Insurance Number? Lottery?

What are these symbols?

> F3F4FFFFFFFFFFFFFFFFFFFF7E5212FFFF2542532492412292121921801561581621681762382292011911781691651631621871871871841781691651561551561651741811811791401431431401 35…

Data → Information → Knowledge

> Information is data organized into a meaningful context to aid decision-making.

> F3F4FFFFFFFFFFFFFFFFFFFF7E5212FFFF2542532492412292121921801561581621681762382292011911781691651631621871871871841781691651561551561651741811811791401431431401 35

F3 Hexadecimal number (base:16)

243 Decimal number (base:10)

Data → Information → Knowledge

243	244	255	255	255	255	255	255	255	255	255	255	247	229	212	…
255	255	254	253	249	241	229	212	192	180	156	158	162	168	176	…
238	229	201	191	178	169	165	163	162	162	187	187	187	184	178	…
169	165	156	155	156	165	174	181	181	179	140	143	143	140	135	…
168	178	188	187	189	190	186	171	149	133	151	153	151	145	139	…
174	172	179	168	156	149	144	139	130	121	126	125	122	120	120	…
146	145	141	137	133	129	126	123	123	123	131	126	127	135	138	…
123	128	126	124	123	123	123	126	128	130	138	132	128	132	133	…
123	131	123	124	123	123	125	129	133	135	138	130	128	131	132	…
139	144	138	136	134	132	133	133	135	136	129	125	127	135	137	…
140	140	143	142	140	137	136	136	139	140	127	126	133	143	145	…
133	134	137	137	135	135	137	140	144	146	138	136	141	149	150	…
134	137	133	133	133	134	138	142	147	150	147	144	144	149	149	…
133	141	138	137	136	137	140	143	148	150	150	144	142	147	148	…
125	139	132	133	134	134	137	146	153	155	148	148	148	149	151	…
121	146	151	151	148	142	138	140	144	147	155	155	155	156	157	…
131	153	127	131	137	141	147	154	166	178	164	164	164	163	163	…
128	122	148	147	148	151	149	148	156	167	173	173	172	170	168	…
108	123	166	159	156	162	163	161	165	175	184	184	182	179	175	…
136	159	178	164	163	176	188	189	194	202	195	194	192	188	182	…
201	165	227	200	186	194	201	192	184	184	198	197	195	189	182	…
233	178	221	193	178	195	210	207	199	199	193	192	190	184	176	…
194	172	209	200	207	200	199	203	191	195	209	193	198	183	176	…
177	187	216	209	205	189	181	194	208	228	211	205	216	191	170	…
160	208	233	212	194	200	221	232	224	208	228	198	191	174	177	…

……

Information

An Image with size 317x350

Data → Information → Knowledge

Information

Knowledge

"Red Apple"

Pattern Recognition

Another Example

> 000101020305080D1522375990 (Data)

> 0 1 1 2 3 5 8 13 21 34 55 89 144 (Information)

Another Example

> 000101020305080D1522375990 (Data)

> 0 1 1 2 3 5 8 13 21 34 55 89 144 (Information)

$$a_n = a_{n-1} + a_{n-2}$$
$$a_0 = 0$$
$$a_1 = 1$$

Data – Information – Knowledge

> Knowledge is clear perception/understanding of truth,

$$a_n = a_{n-1} + a_{n-2}$$
$$a_0 = 0$$
$$a_1 = 1$$

$$a_n = \frac{2}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{2}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n$$

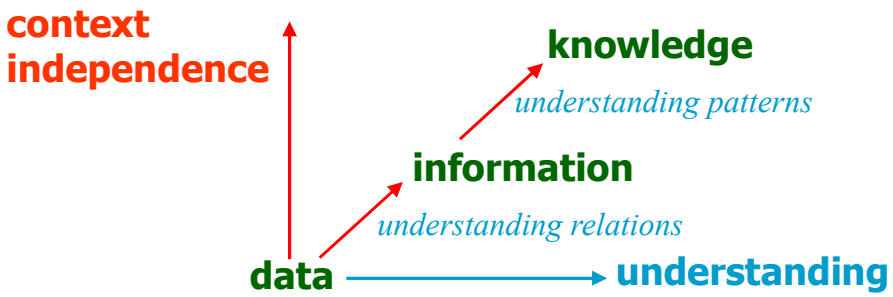
Knowledge

What is the difference between them?

> At the root of information is, "to inform."

> Data don't become information until we have successfully linked meaning to them.

> If we fail to build common meaning and understanding, data remain just a bunch of unconnected events.



Information and Entropy

> How much information does data contain?

> Can we measure it?

> Fortunately, yes:

$$E = -\sum_{\text{each event}} p_i \log(p_i)$$

> Example: Tossing a coin

– $P_H=P_T=0.5$

– E=log2

Information and Entropy

> Toss a coin three times

– H H H

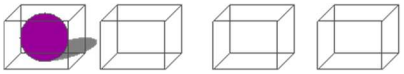
– Probability of three successive H $\frac{1}{8}$

$$\frac{3}{8}\log 2$$

– Less probable events contain more information

Uncertainty

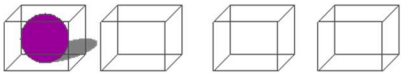
> 4 Boxes, 1 Ball



> You ask yes/no questions to decide on in which box the ball is
> Initially you have no idea, hence the uncertainty is maximum
> As you ask, you get more information, hence the uncertainty decreases
> Finally, you learn the answer in which case the uncertainty is 0
> Information is always a measure of the decrease in uncertainty

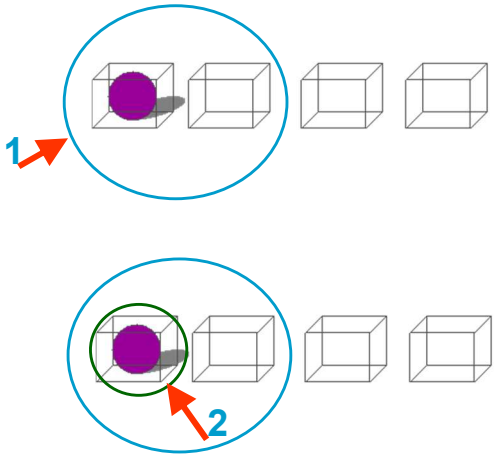
Uncertainty

> 4 Boxes, 1 Ball



> How many questions are enough to learn the box that the ball is in?
  - 4?
  - 3?
  - 2?
  - 1!?

Uncertainty



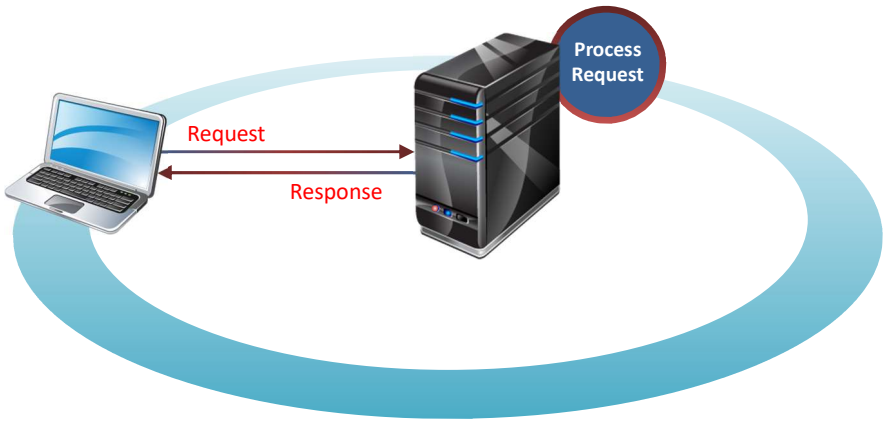
Why Databases?

> Why do we need databases?

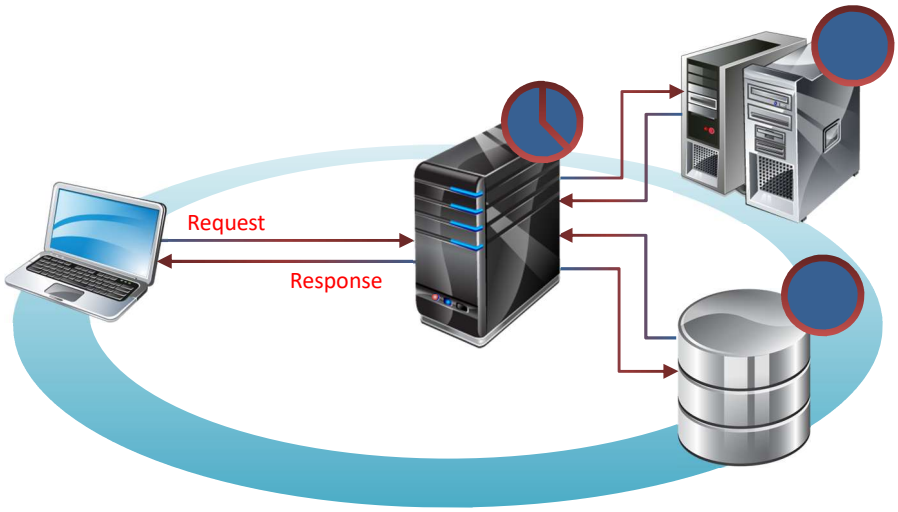
Why Databases?

> Why do we need databases?
  – Applications need to store/persist their state

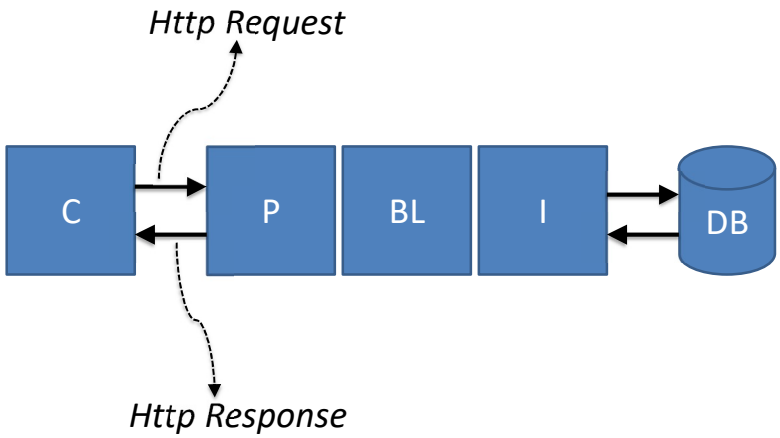
Need for persistence



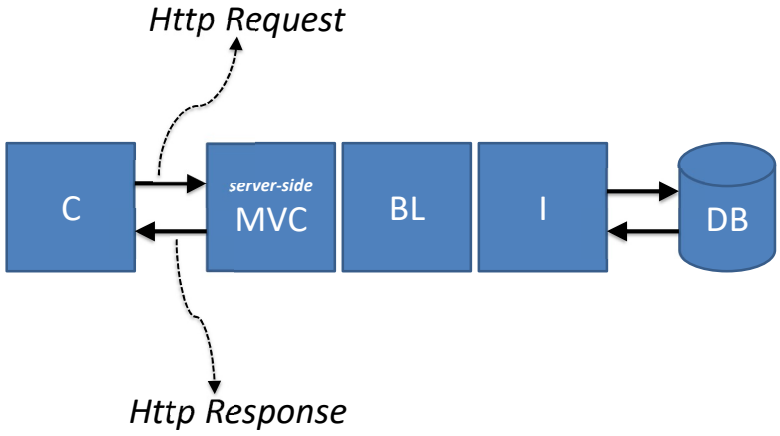
Need for persistence



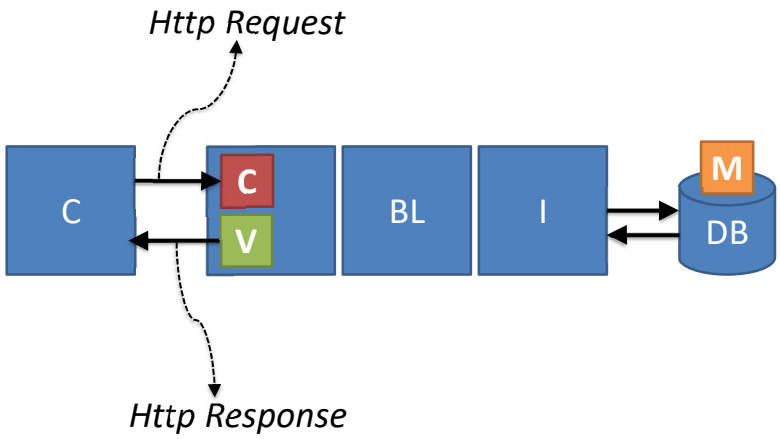
Need for persistence



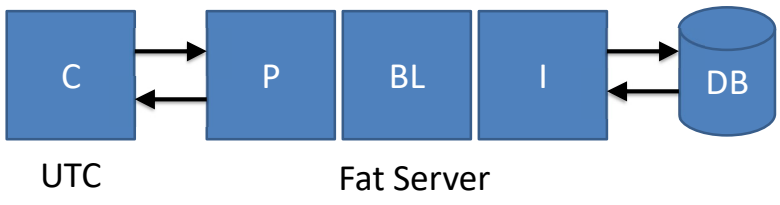
Request-Response



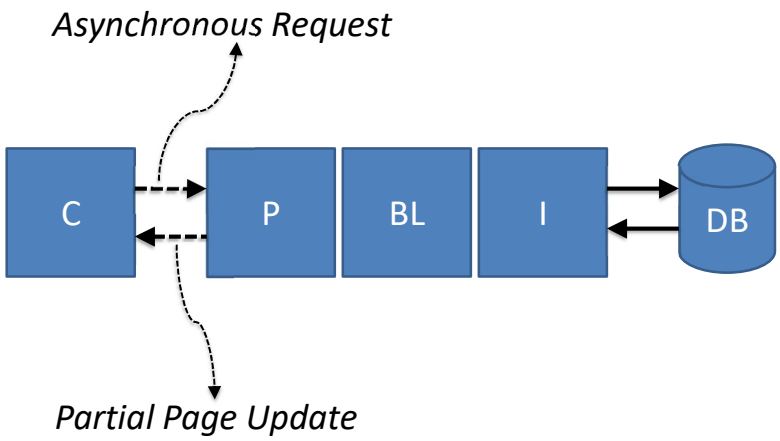
Request-Response

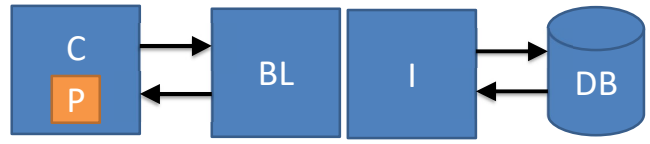


Request-Response

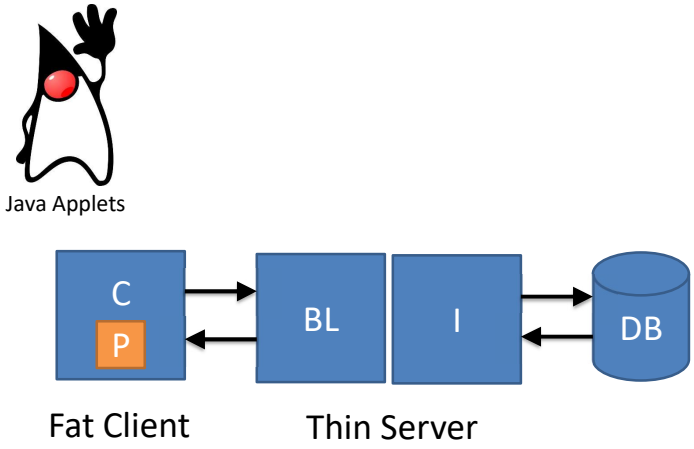


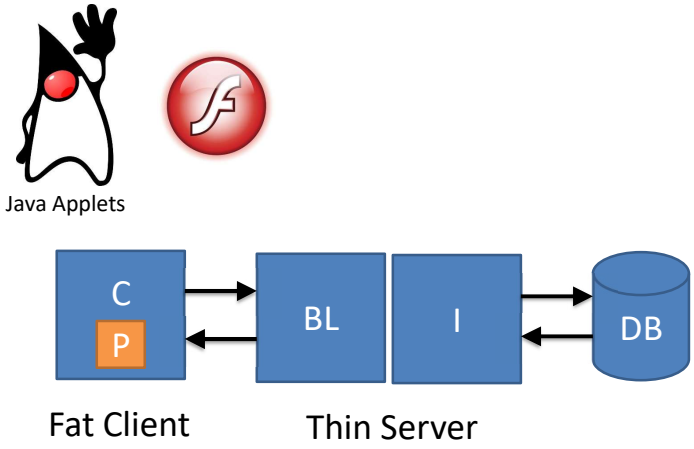
Ajax

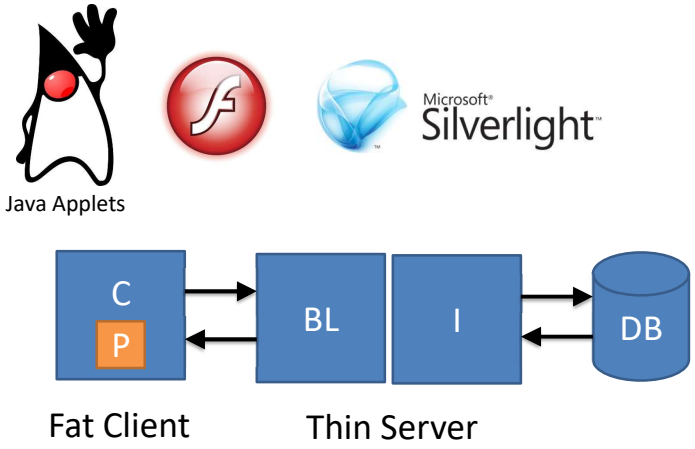


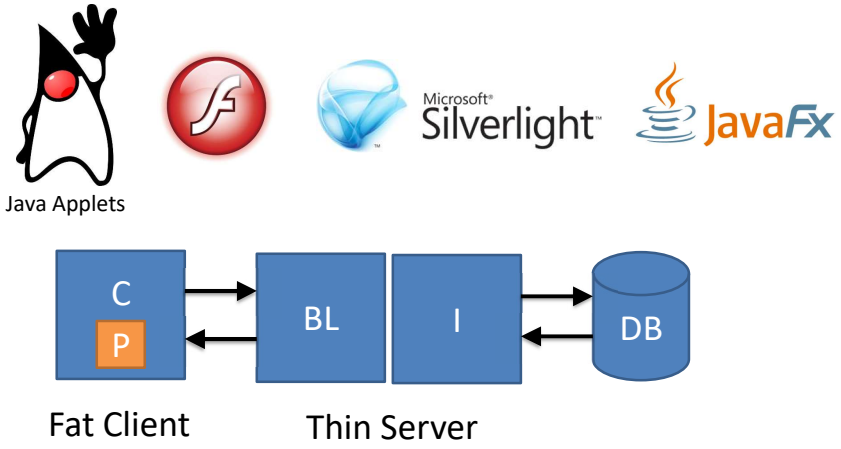


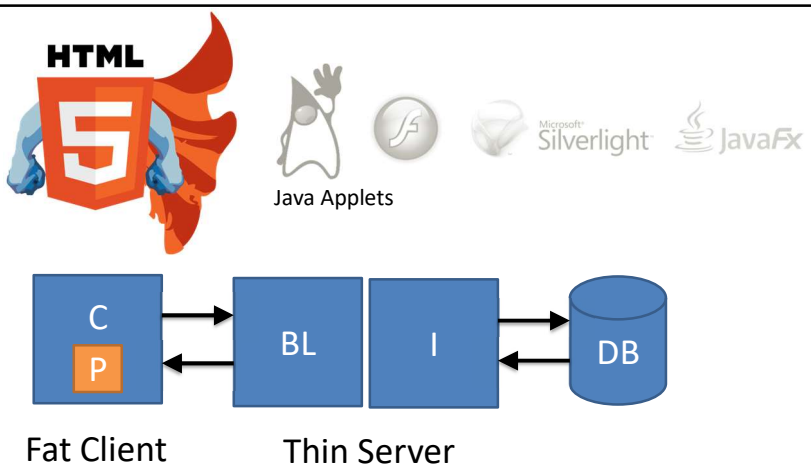
Fat Client Thin Server


Java Applets
C
P
BL
I
DB
Fat Client
Thin Server


Java Applets
C
P
BL
I
DB
Fat Client
Thin Server


Java Applets
Microsoft
Silverlight
C
P
BL
I
DB
Fat Client
Thin Server


Java Applets
Microsoft
Silverlight
JavaFx
C
P
BL
I
DB
Fat Client
Thin Server

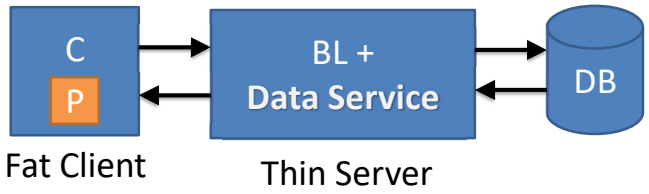

HTML
Java Applets
Microsoft Silverlight
JavaFx
C
P
BL
I
DB
Fat Client
Thin Server

HTML5 APIs

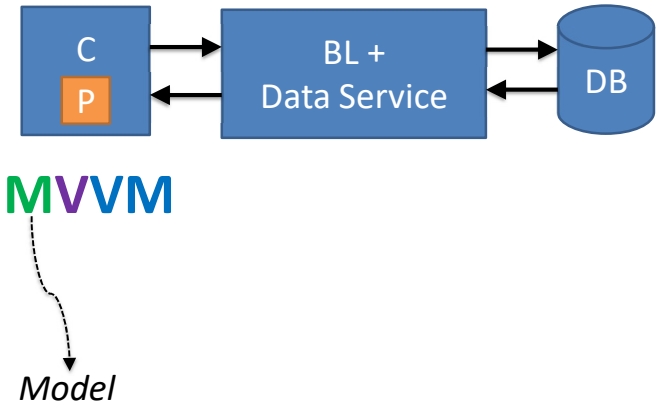

Web Sockets and Messaging, WebRTC
Canvas, SVG, WebGL
File API, File System API, Indexed DB, Offline, Web Storage
DOM
HTML
Drag and Drop, Fullscreen
Animation Timing, Media, Pointer Lock, Web Audio
Browser, Shadow DOM, Typed Arrays, Web Workers
CSS Object Model, Selectors


Semantics
CSS3
Multimedia
Graphics & 3D
Device Access
Performance
Offline & Storage
Connectivity

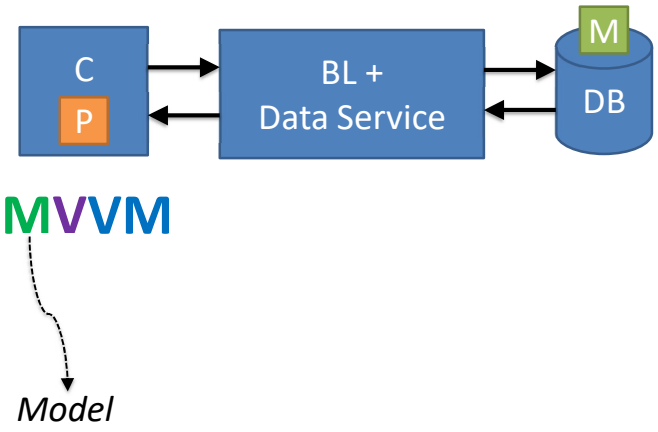
Data Service



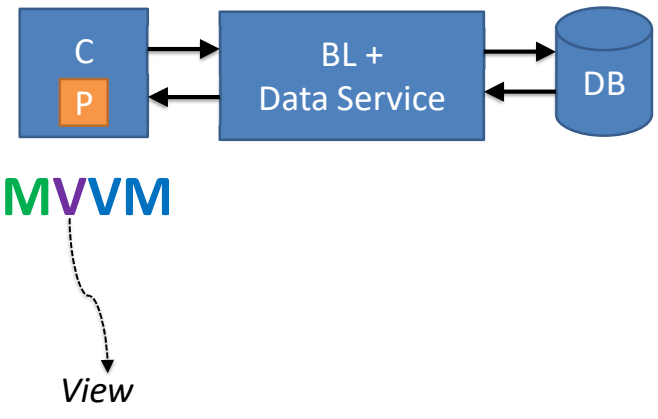
UI Logic



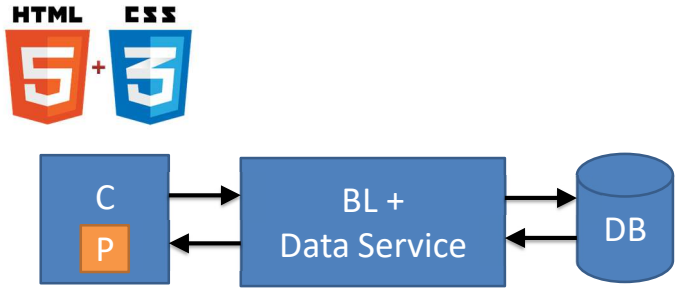
UI Logic



UI Logic



UI Logic



MVVM

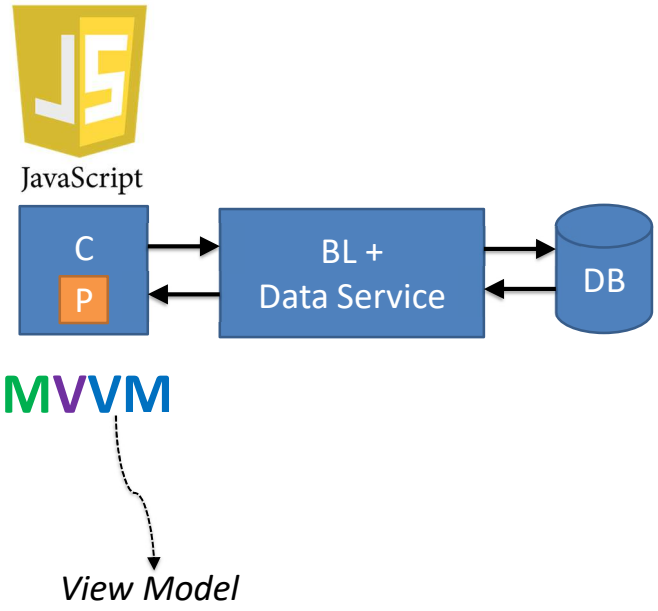
View

UI Logic

MVVM

View Model

UI Logic

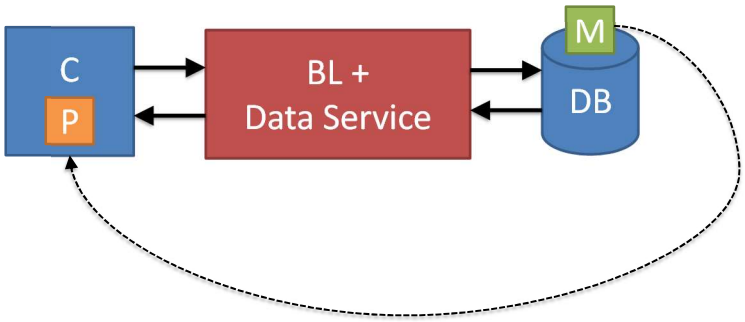


MVVM

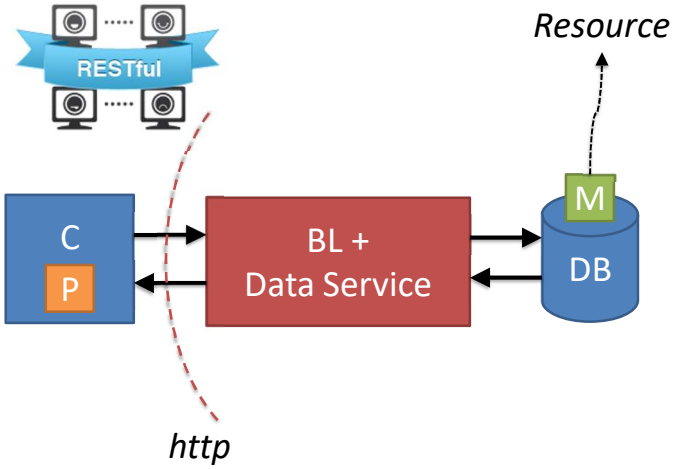
View Model

Data Service

Data Service



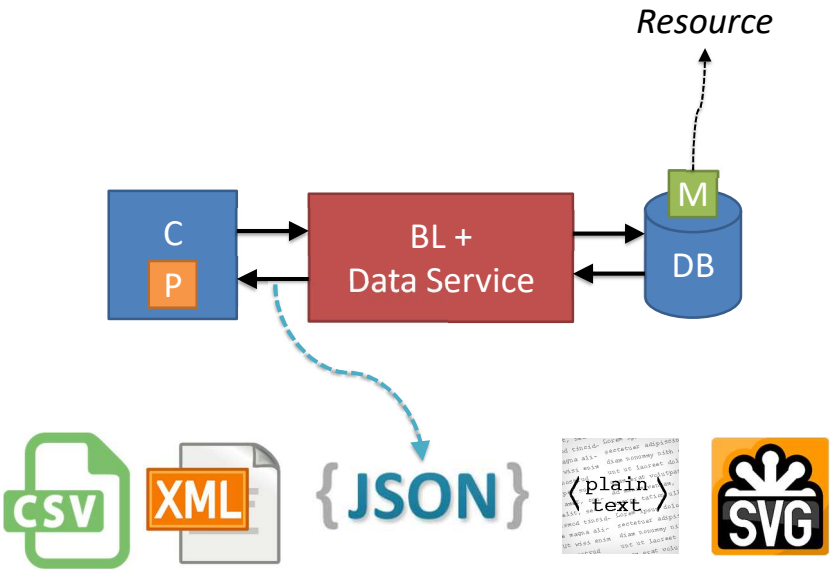
Data Service



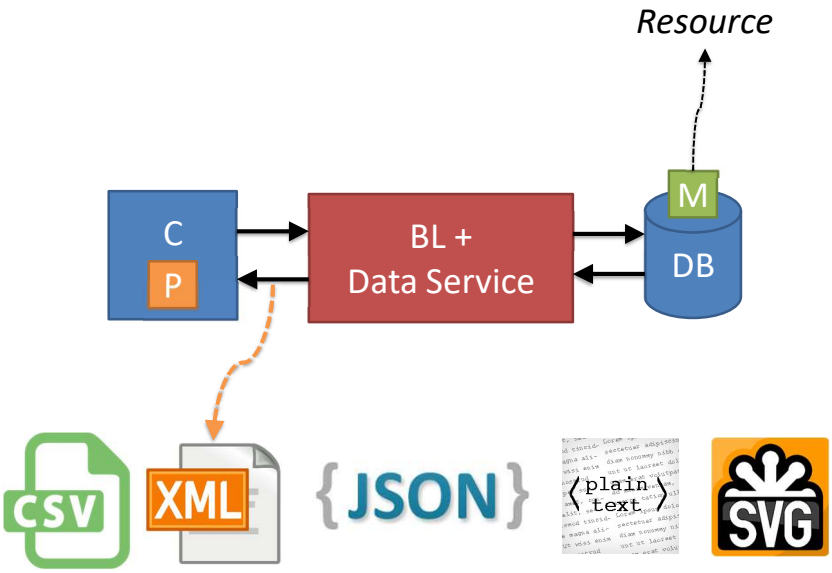
Data Service

HTTP	SQL
GET	SELECT
POST/PUT	INSERT
PUT/POST	UPDATE
DELETE	DELETE

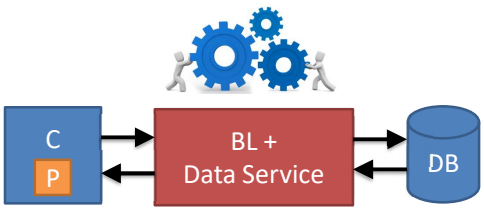
RESTful Data Service



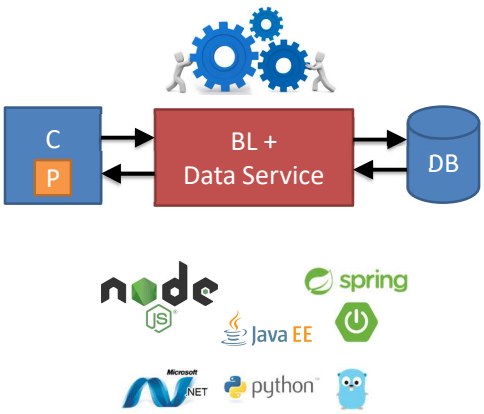
RESTful Data Service



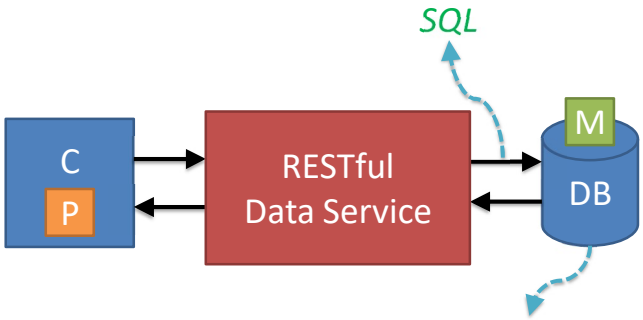
RESTful Data Service



RESTful Data Service



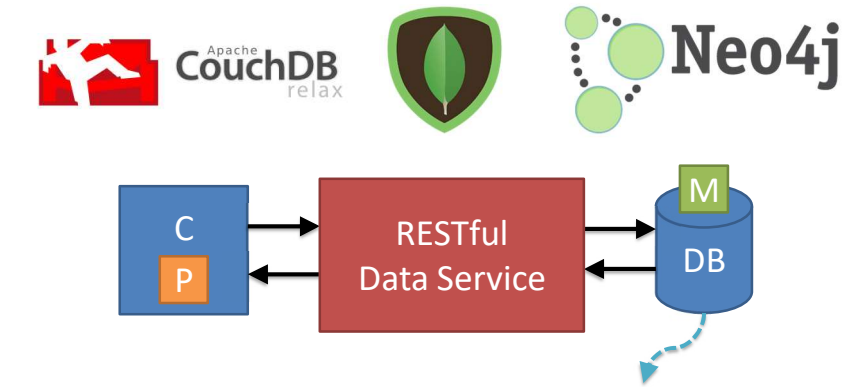
Data Service



JOIN

NOT CLUSTER FRIENDLY

Data Service



redis

cassandra

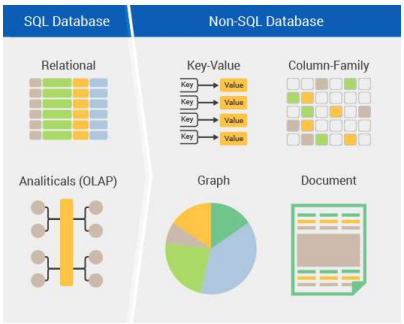
NO-SQL DBs

SCHEMA-LESS

Scalability vs. Consistency

Cluster Friendly

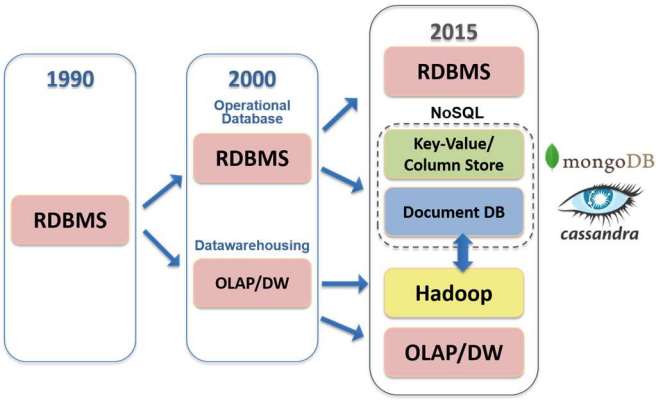
NoSQL Databases



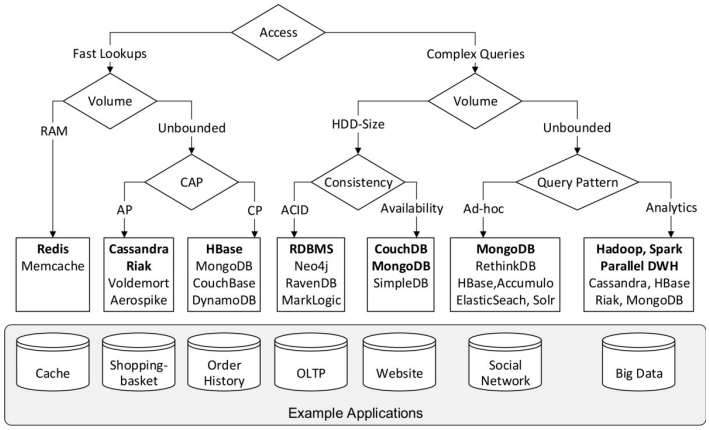
NoSQL Databases

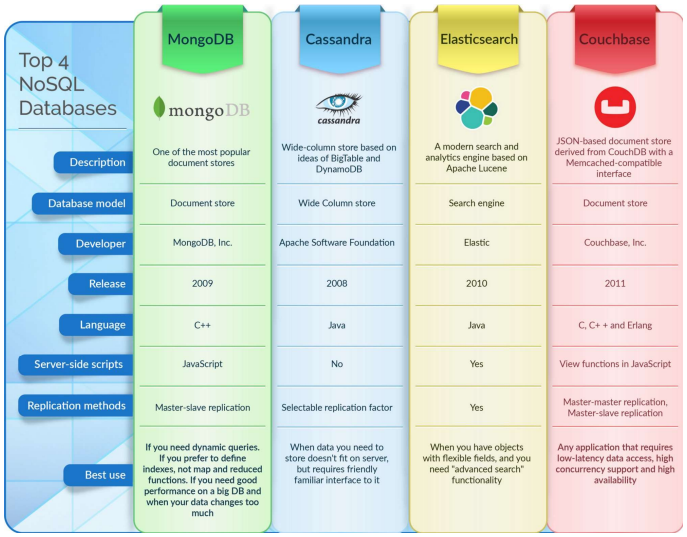
	Data Model	Query API
Cassandra	Columnfamily	Thrift
CouchDB	Document	map/reduce views
HBase	Columnfamily	Thrift, REST
MongoDB	Document	Cursor
Neo4J	Graph	Graph
Redis	Collection	Collection
Riak	Key/value	REST
Scalaris	Key/value	get/put
Tokyo Cabinet	Key/value	get/put
Voldemort	Key/value	get/put

NoSQL Databases

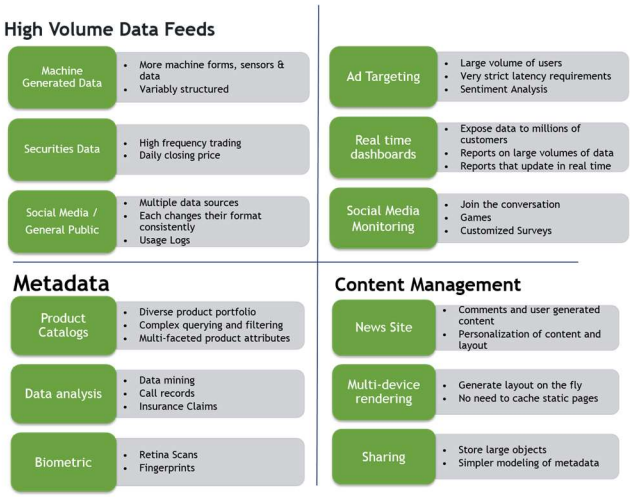


Decision Tree



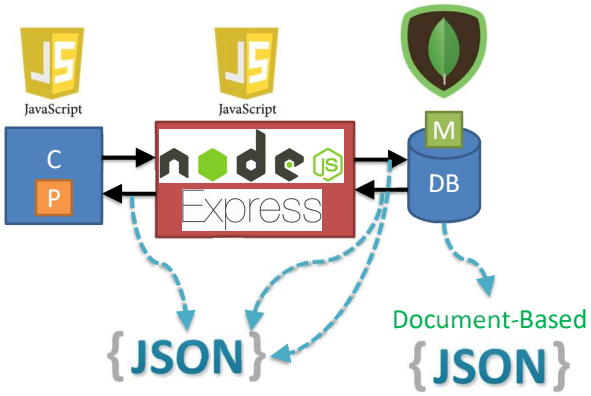


What are the Right Use Cases for NoSQL?



NoSQL Database Features

Example Technology Stack



Introducing the Database

> Data management

- A process that focuses on data collection, storage, and retrieval.
- Common data management functions include addition, deletion, modification, and listing.

Introducing the Database

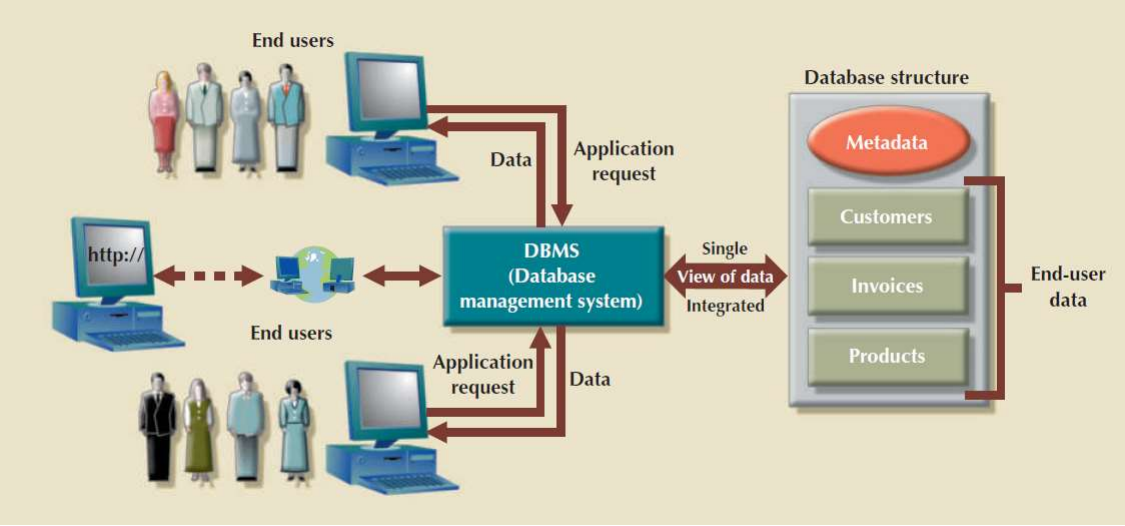
> Database

- A shared, integrated computer structure that houses a collection of related data.
- A database contains two types of data:
  - end-user data (Raw facts)
  - Metadata (Data about data)

Introducing the Database

- A database management system (DBMS) is a collection of programs that manages the database structure and controls access to the data stored in the database.
- In a sense, a database resembles a well-organized electronic filing cabinet in which powerful software (the DBMS) helps manage the cabinet's contents.

Introducing the Database



Role and Advantages of the DBMS

- The DBMS serves as the intermediary between the user and the database.
- The database structure itself is stored as a collection of files, and the only way to access the data in those files is through the DBMS
- The DBMS presents the end user (or application program) with a single, integrated view of the data in the database.
- The DBMS receives all application requests and translates them into the complex operations required to fulfill those requests.
- The DBMS hides much of the database's internal complexity from the application programs and users.

Role and Advantages of the DBMS

- A DBMS provides the following advantages:
  - Improved data sharing
  - Improved data security
  - Better data integration
  - Minimized data inconsistency
  - Improved data access
  - Improved decision making
  - Increased end-user productivity

Types of Databases

- single-user database
  - A database that supports only one user at a time
- desktop database
  - A single-user database that runs on a personal computer
- multiuser database
  - A database that supports multiple concurrent users.
- workgroup database
  - A multiuser database usually supports fewer than 50 users or is used for a specific department in an organization.

Types of Databases

- enterprise database
  - The overall company data representation, which provides support for present and expected future needs.
- centralized database
  - A database located at a single site.
- distributed database
  - A logically related database that is stored in two or more physically independent sites.
- cloud database
  - A database that is created and maintained using cloud services, such as Microsoft Azure or Amazon AWS.

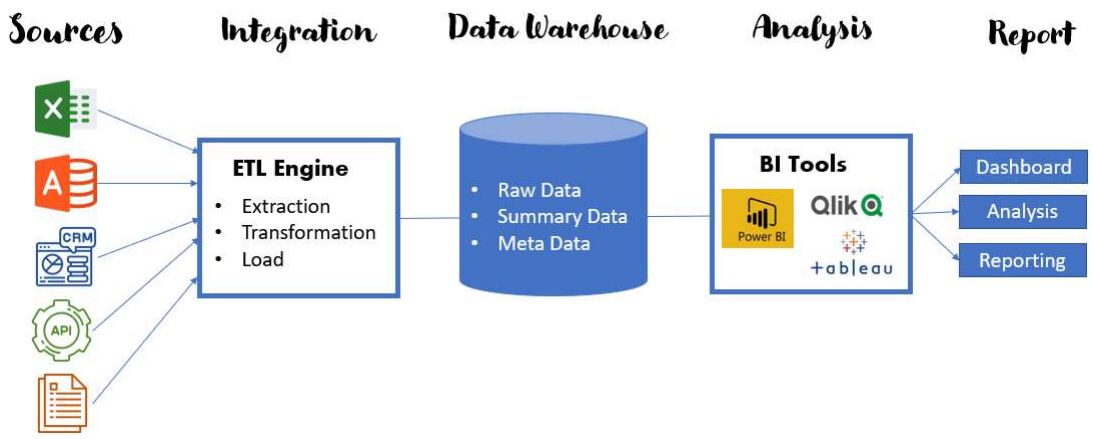
Types of Databases

- General-purpose database
  - A database that contains a wide variety of data used in multiple disciplines.
- Discipline-specific database
  - A database that contains data focused on specific subject areas.
- Operational/Production/OLTP database
  - A database designed primarily to support a company's day-to-day operations. Also known as a transactional database,
- Analytical database
  - A database focused primarily on storing historical data and business metrics used for tactical or strategic decision-making.

Types of Databases

- data warehouse
  - A specialized database that stores historical and aggregated data in a format optimized for decision support.
- online analytical processing (OLAP)
  - A set of tools that provide advanced data analysis for retrieving, processing, and modeling data from the data warehouse.
- business intelligence
  - A set of tools and processes used to capture, collect, integrate, store, and analyze data to support business decision-making.

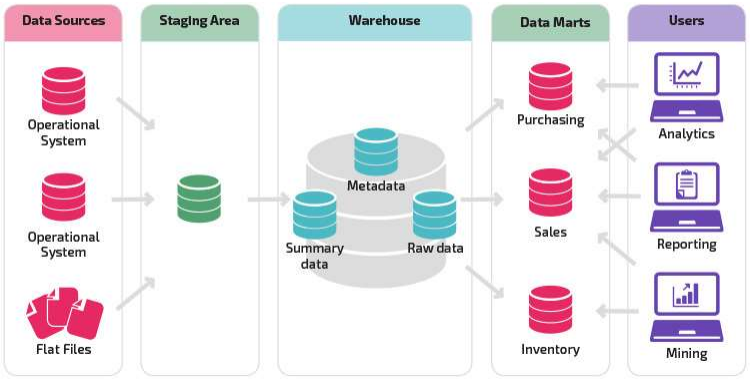
Types of Databases



Data mart and Data warehouse

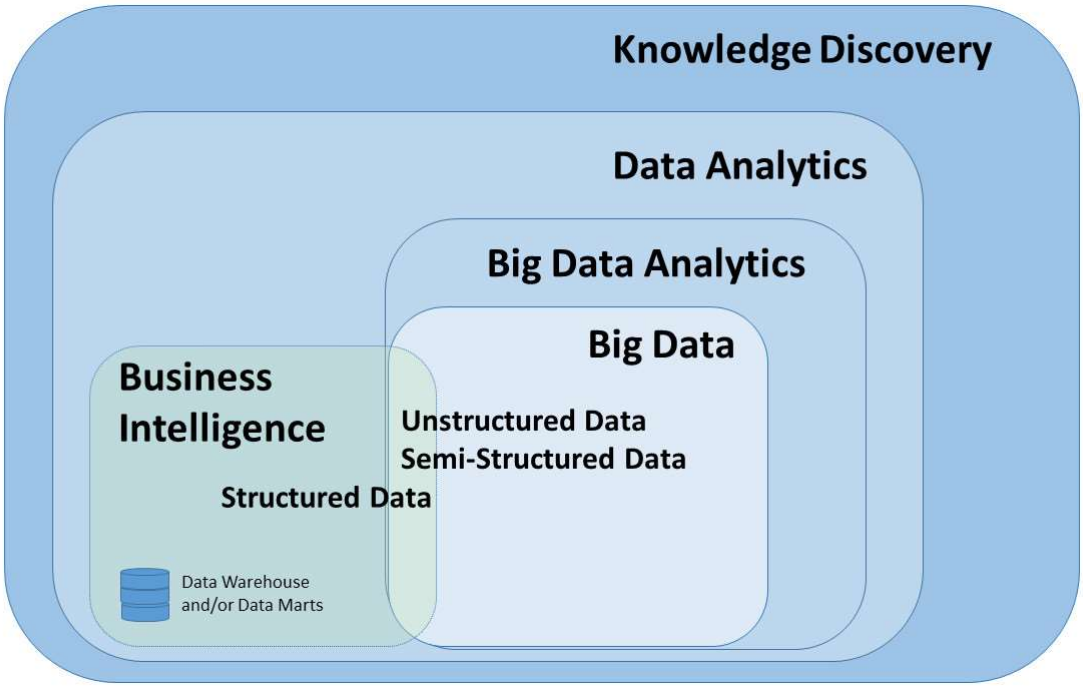
- A data mart is a subset of a data warehouse oriented to a specific business line.
  - Data marts contain repositories of summarized data collected for analysis on a specific section or unit within an organization, for example, the sales department.
- A data warehouse is a large centralized repository of data that contains information from many sources within an organization.
  - The collated data guides business decisions through analysis, reporting, and data mining tools.

Types of Databases



Types of Databases

	Data warehouses	Data lakes	Data marts
Usage	The data analysis and reporting needs of an entire organization	The reporting needs of different kinds and difficulty, predictive analytics	The reporting needs of a specific operational department or subject
Data stored (typically)	Larger volumes of structured data; processed	Huge volumes of structured and unstructured data; raw	A limited amount of structured data; processed
Data sources	An array of external and internal sources, covering different areas of business	Any external or internal sources	Few sources linked to one business area
Size	Larger than 100 GB	Larger than 100 GB	Smaller than 100 GB
Ease of creation	Difficult to set up	Difficult to set up	Easy to set up



Types of Databases

TYPES OF DATABASES

PRODUCT	NUMBER OF USERS			DATA LOCATION		DATA USAGE		XML
	SINGLE USER	MULTIUSER						
		WORKGROUP	ENTERPRISE	CENTRALIZED	DISTRIBUTED	OPERATIONAL	ANALYTICAL	
MS Access	X	X		X		X		
MS SQL Server	X*	X	X	X	X	X	X	X
IBM DB2	X*	X	X	X	X	X	X	X
MySQL	X	X	X	X	X	X	X	X
Oracle RDBMS	X*	X	X	X	X	X	X	X

Data Types

- unstructured data
  - Data exists in its original, raw state; that is, in the format in which it was collected.
- structured data
  - Data formatted to facilitate storage, use, and information generation.
- semi structured data
  - Data that has already been processed to some extent.