SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

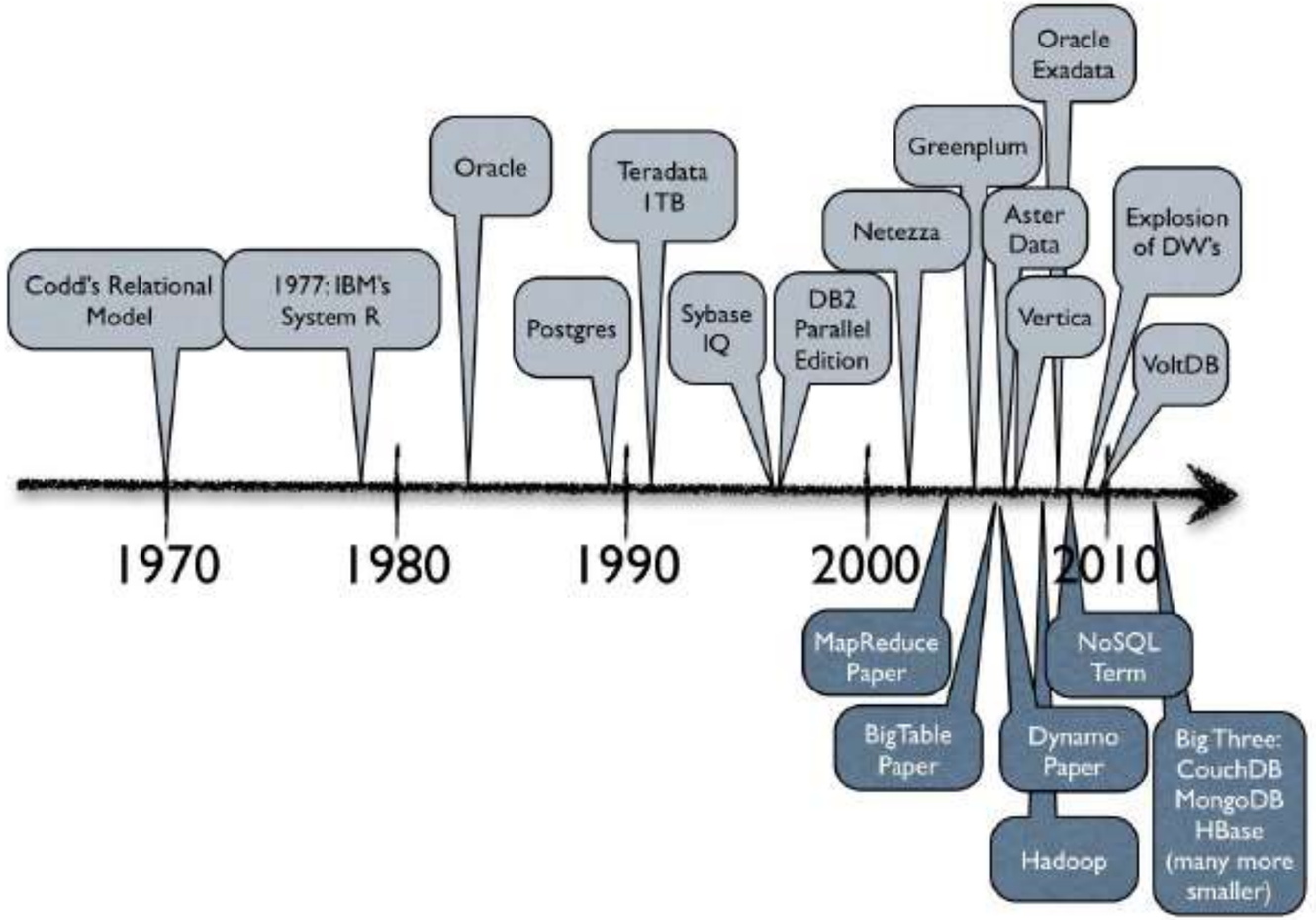
CS8091 Big Data Analytics

III YEAR / VI SEMESTER

Unit 5 – NoSQL and Visualization

Topic 1: NoSQL

DBMS


Oracle Exadata
Greenplum
Oracle
Teradata 1TB
Netezza
Aster Data
Explosion of DW's
Codd's Relational Model
1977: IBM's System R
Postgres
Sybase IQ
DB2 Parallel Edition
Vertica
VoltDB
1970
1980
1990
2000
2010
MapReduce Paper
NoSQL Term
BigTable Paper
Dynamo Paper
Big Three: CouchDB MongoDB HBase (many more smaller)
Hadoop

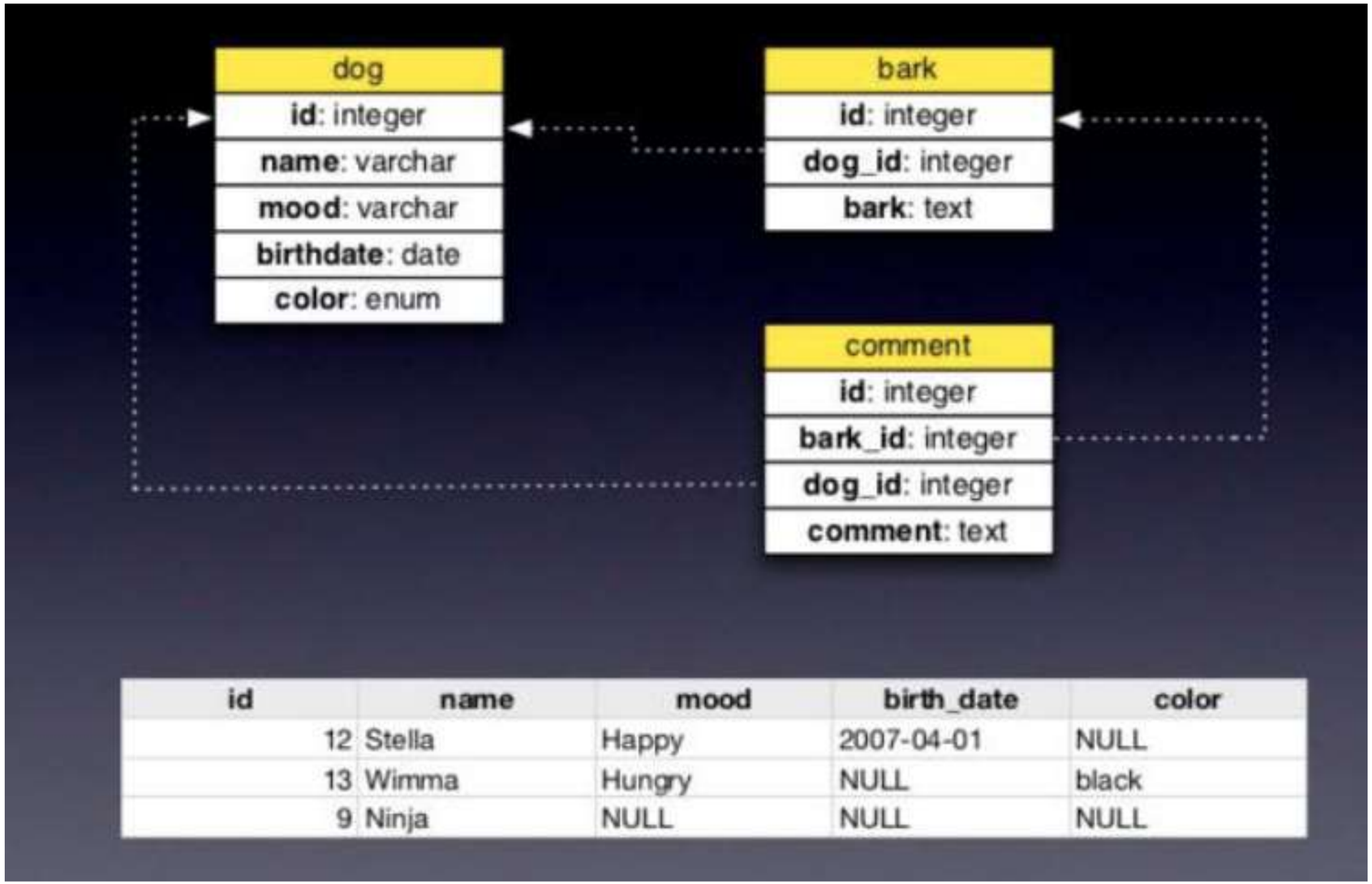
DBMS

Benefits of Relational databases:

- Designed for all purposes
- ACID
- Strong consistancy, concurrency, recovery
- Mathematical background
- Standard Query language (SQL)
- Lots of tools to use with i.e: Reporting services, entity frameworks, ...

SQL Databases

SQL Databases



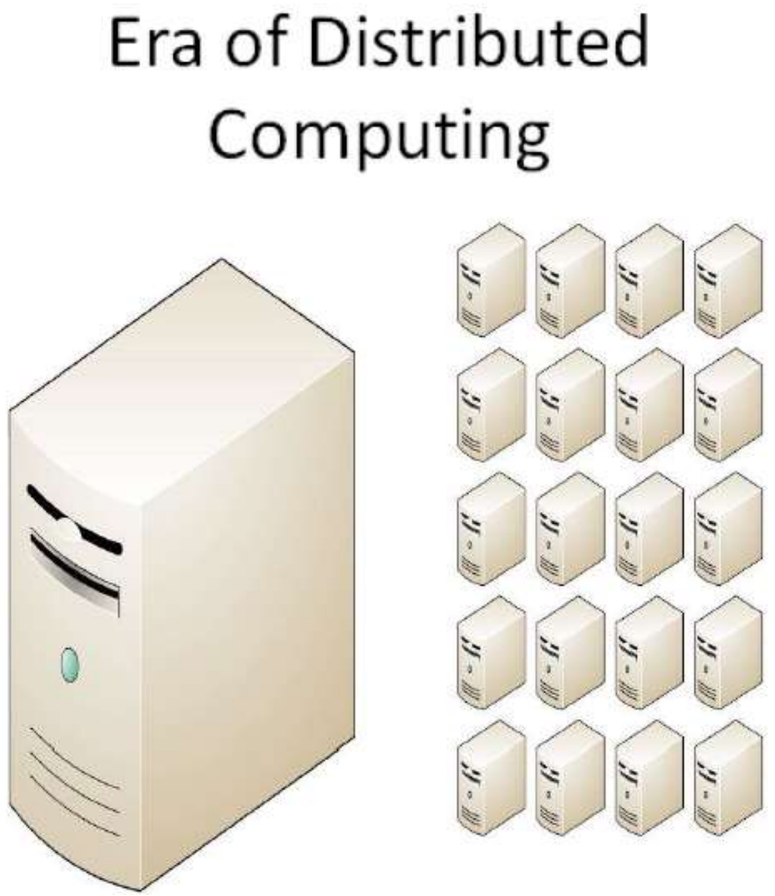
id	name	mood	birth_date	color
12	Stella	Happy	2007-04-01	NULL
13	Wimma	Hungry	NULL	black
9	Ninja	NULL	NULL	NULL

Why NoSQL?

Distributed Computing

Relational databases were not built for distributed applications.
Because...

- Joins are expensive
- Hard to scale horizontally
- Impedance mismatch occurs
- Expensive (product cost, hardware, Maintenance

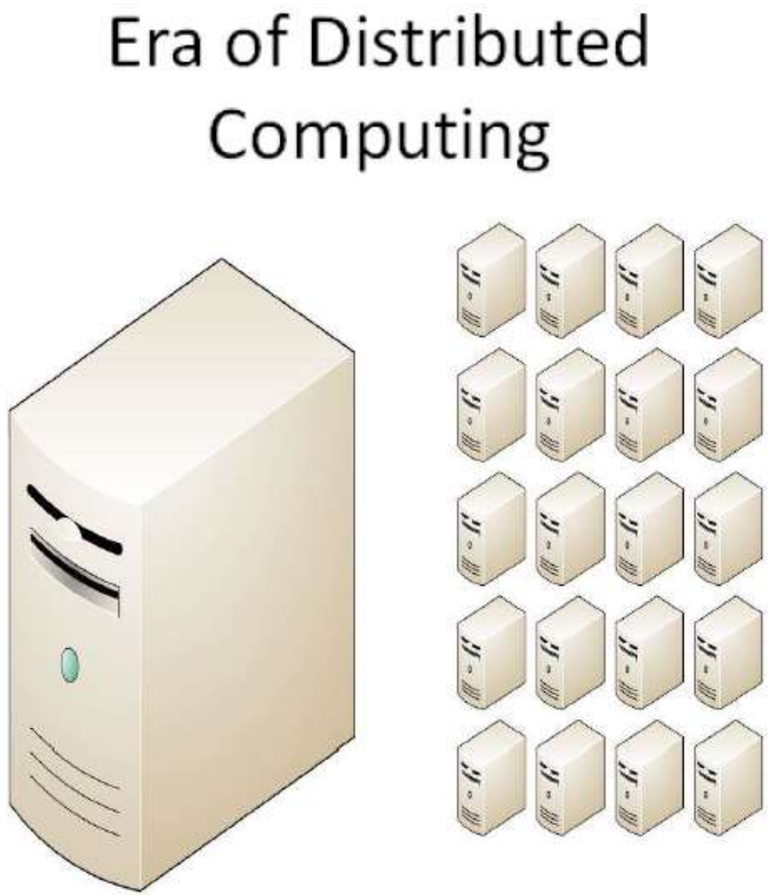


Why NoSQL?

And....

It's weak in:

- Speed (performance)
- High availability
- Partition tolerance



Why NoSQL?


New Trends
2002
2004
2006
2008
2010
2012
Big data
Connectivity
P2P Knowledge
Concurrency
Diversity
Cloud-Grid

What is NoSQL?

A No SQL database provides a mechanism for storage and retrieval of data that employs less constrained consistency models than traditional relational database.

No SQL systems are also referred to as "NotonlySQL“ to emphasize that they do in fact allow SQL-like query languages to be used

What is NoSQL?

NoSQL avoids:

- ▶Overhead of ACID transactions
- ▶Complexity of SQL query
- ▶Burden of up-front schema design
- ▶DBA presence
- ▶Transactions (It should be handled at application layer)

Provides:

- ▶Easy and frequent changes to DB
- ▶Fast development
- ▶Large data volumes(eg.Google)
- ▶Schema less

What is NoSQL?

When and when not to use it?

WHEN / WHY ?

- When traditional RDBMS model is too restrictive (flexible schema)
- When ACID support is not "really" needed
- Object-to-Relational (O/R) impedance
- Because RDBMS is neither distributed nor scalable by nature
- Logging data from distributed sources
- Storing Events / temporal data
- Temporary Data (Shopping Carts / Wish lists / Session Data)
- Data which requires flexible schema
- **Polyglot Persistence** i.e. best data store depending on nature of data.

WHEN NOT ?

- Financial Data
- Data requiring strict ACID compliance
- Business Critical Data

Who uses NoSQL?

What is Schemaless?

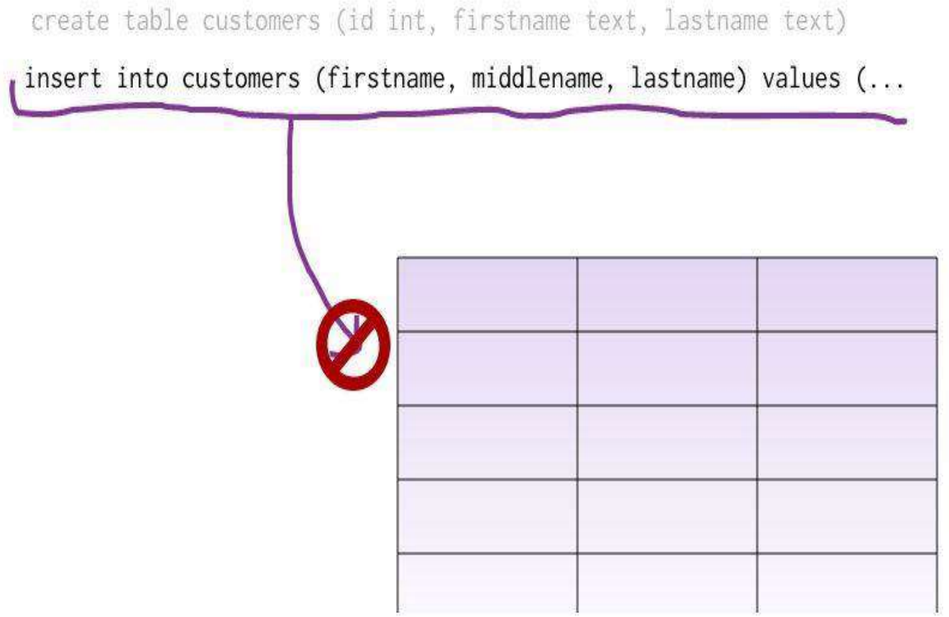
In relational Databases:

▶You can’t add a record which does not fit the schema

▶You need to add NULLs to unused items in a row

▶We should consider the datatypes. i.e : you can’t add a stirn to an interger field

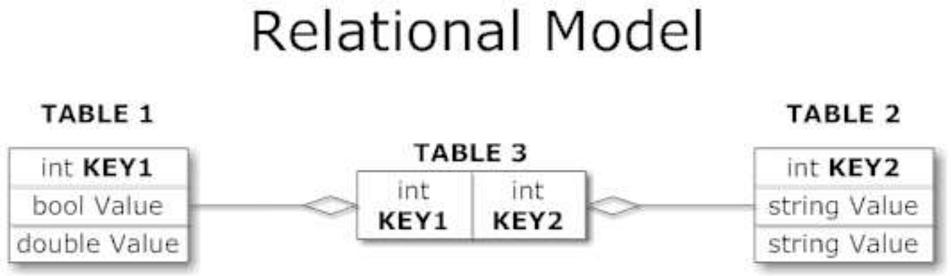
▶You can’t add multiple items in a field (You should create another table: primary-key, foreign key, joins, normalization, .. !!!)

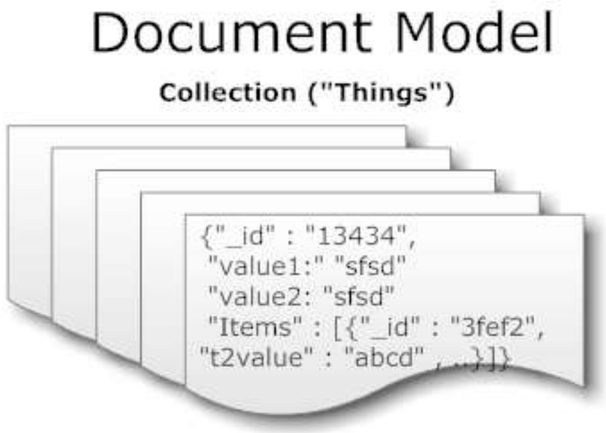


What is Schemaless Datamodel?

In NoSQL Databases:

- There is no schema to consider
- There is no unused cell
- There is no datatype (implicit)
- Most of considerations are done in application layer
- We gather all items in an aggregate (document)

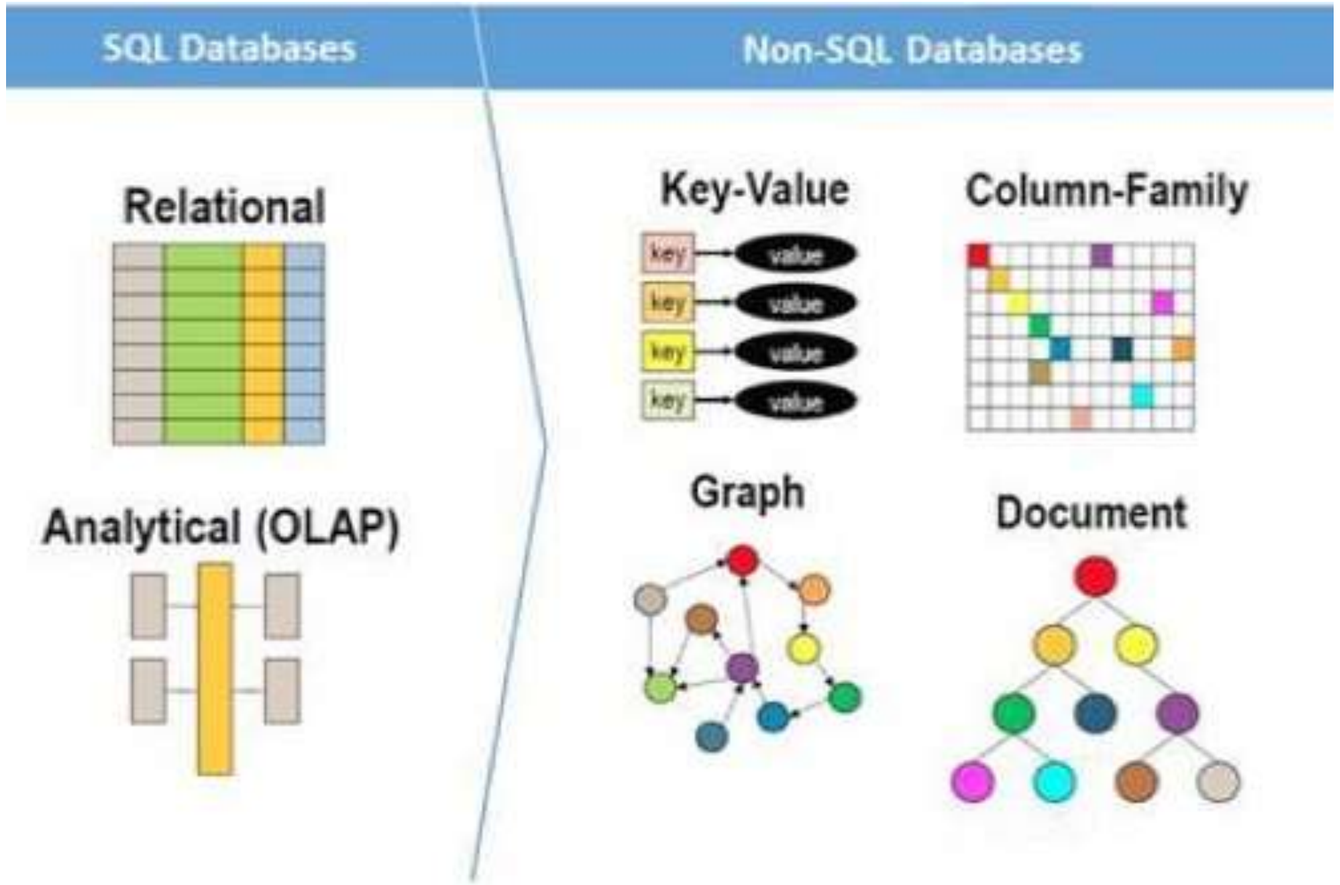




NoSQL Types

NoSQL databases are classified in four major datamodels:

- Key-value
- Document
- Column family
- Graph



NoSQL Types

NoSQL databases are classified in four major datamodels:

- Key-value
- Document
- Column family
- Graph

Type	Example
Key-Value Store	redis, riak
Wide Column Store	HBASE, cassandra
Document Store	mongoDB, CouchDB relax
Graph Store	Neo4j, InfiniteGraph The Distributed Graph Database

NoSQL Types

Key Value Store
Simplest NOSQL databases
- The main idea is the use of a hash table

- Access data (values) by strings called keys
- Data has no required format data may have any format
- Data model: (key, value) pairs
- Basic Operations:

Insert(key,value),
Fetch(key),
Update(key),
Delete(key)

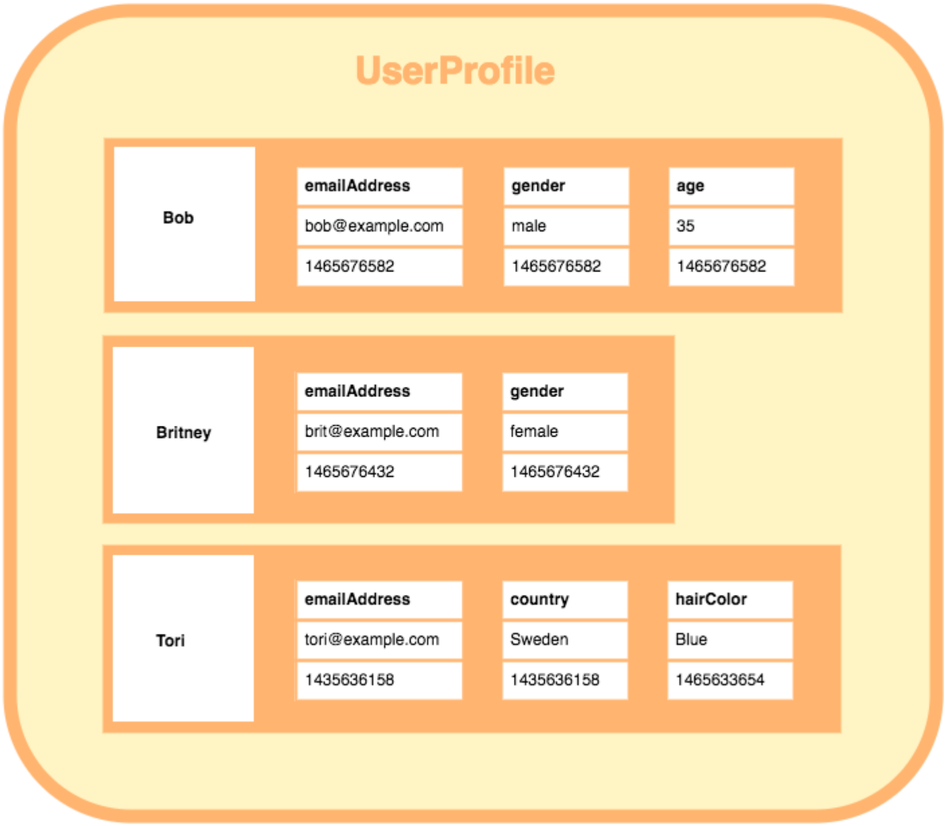
Car	
Key	Attributes
1	Make: Nissan Model: Pathfinder Color: Green Year: 2003
2	Make: Nissan Model: Pathfinder Color: Blue Color: Green Year:2005 Transmission: Auto

NoSQL Types

Column family

The column is lowest/smallest instance of data.

➢It is a tuple that contains a name, a value and a timestamp



Row-oriented

ID	Name	Grade	GPA
001	John	Senior	4.00
002	Karen	Freshman	3.67
003	Bill	Junior	3.33

Column-oriented

Name	ID
John	001
Karen	002
Bill	003

Grade	ID
Senior	001
Freshman	002
Junior	003

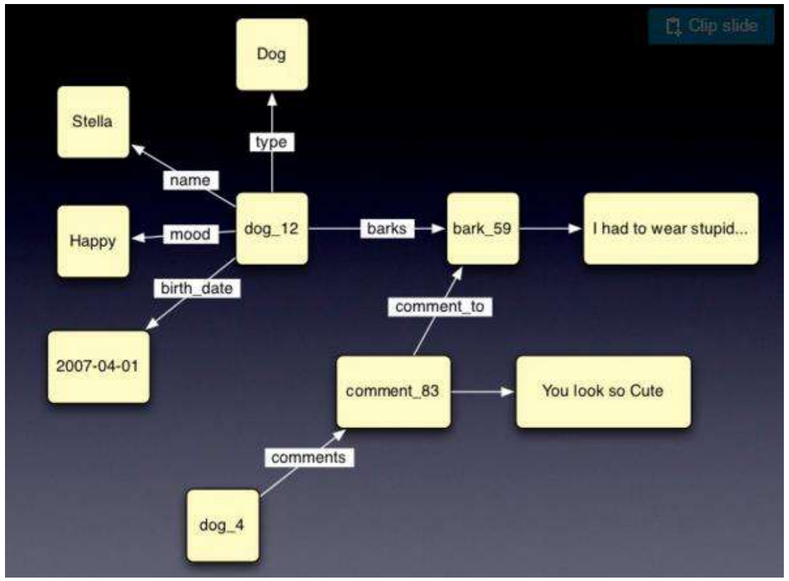
GPA	ID
4.00	001
3.67	002
3.33	003

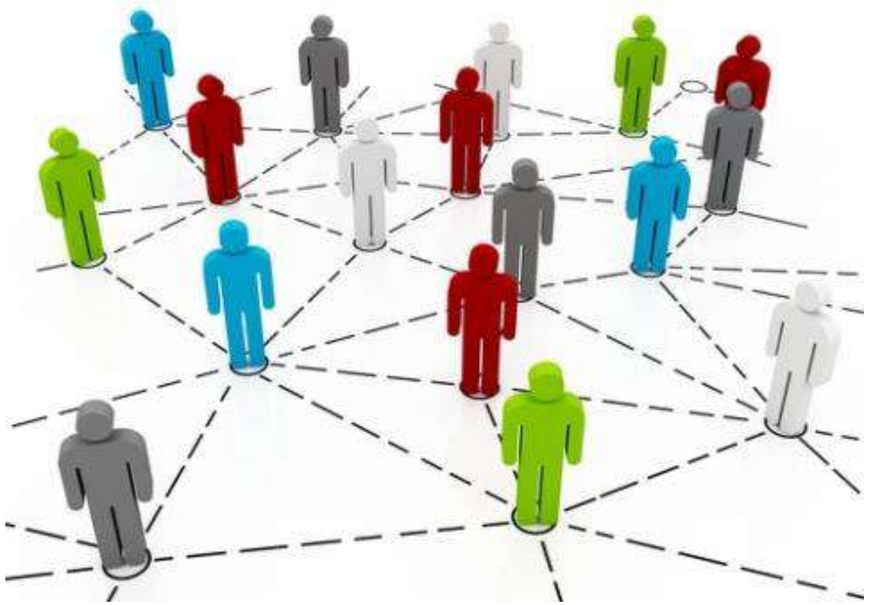
NoSQL Types

Graph Store

Based on Graph Theory.

- Scale vertically, no clustering.
- You can use graph algorithms easily
- Transactions

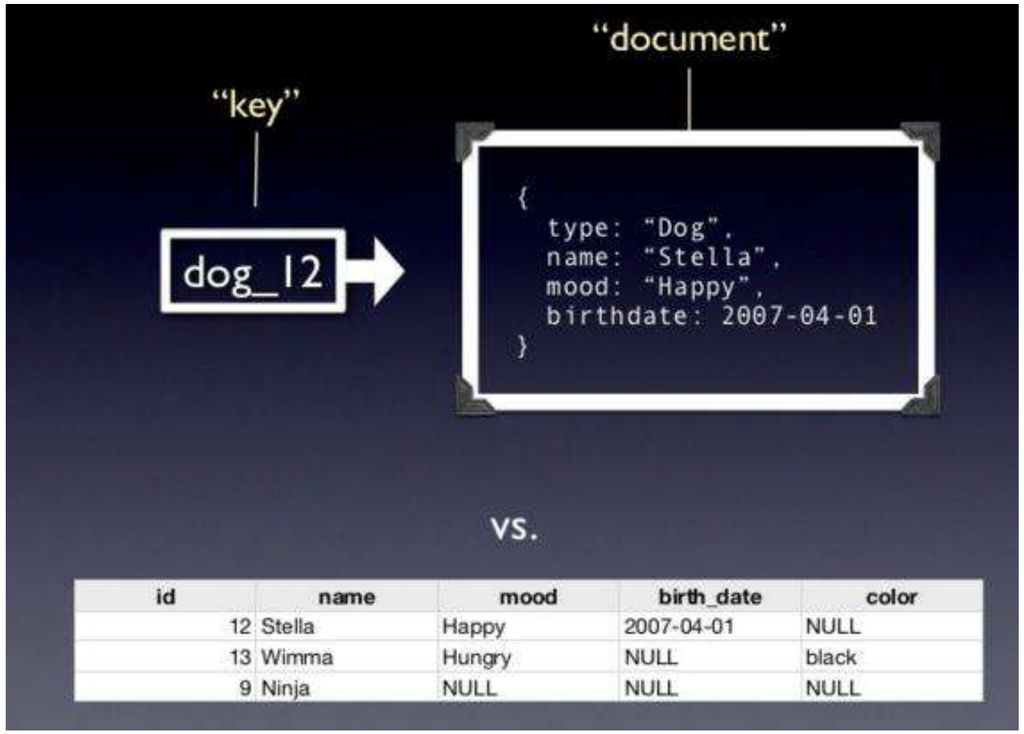




NoSQL Types

Document store
Pair each key with complex data structure known as data structure.

- Indexes are done via B-Trees.
- Documents can contain many different key-value pairs, or key-array pairs, or even nested documents.



THANK YOU