Tools for Data Mining & Analytics

- NoSQL
- Features
- Types

What is No SQl data base

- It’s Not No SQL it’s NOT ONLY SQL.
- It’s not even a replacement to RDBMS.
- As compared to the good olden days we are saving more and more data.
- Connection between the data is growing in which we require an architecture that takes advantage of these two key issues.

What is NoSQL?

- Non Relational data storage system
- No Fixed table schema
- Distributed
- Relaxes one or more ACID Properties

Types of No SQl data base *

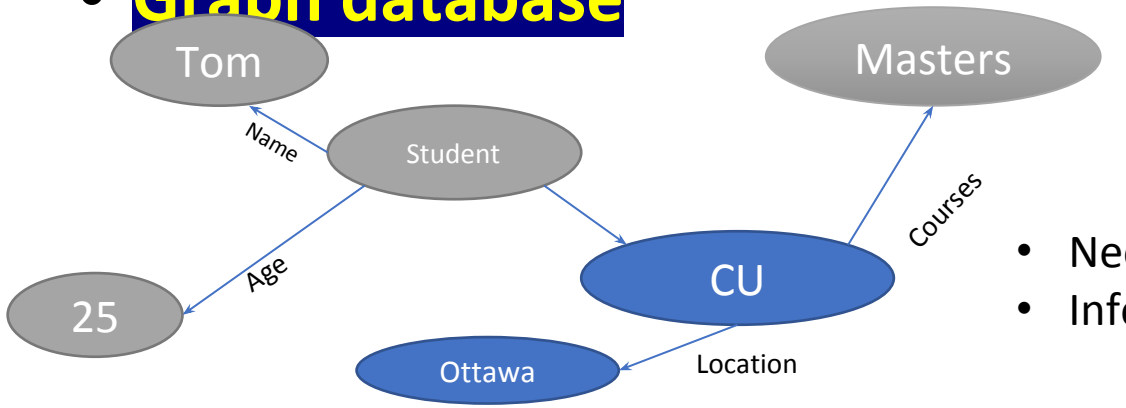
- **Key Value pair**

Dynamo DB

Azure Table Storage

```
(#key,#value)
(Name, Tom)
(Age,25)
(Role, Student)
(University, CU)
```

- **Graph database**



- Neo4j
- Infogrid

- **Document Based**

```
[
 {
  "Name": "Tom",
  "Age": 30,
  "Role": "Student",
  "University": "CU",

 }
]
```

Mongo Db

AmazonSimple DB

Couch DB

Column Oriented database

Row Id	Columns	
1	Name	Tom
	Age	25
	Role	Student

Bigtable(Google)

Hbase, Maria DB, Cra

Motivations

- **Problems with SQL**
  - Rigid schema
  - Not easily scalable (designed for 90's technology or worse)
  - Requires unintuitive joins
- **Perks of mongoDB**
  - Easy interface with common languages (Java, Javascript, PHP, etc.)
  - DB tech should run anywhere (VM's, cloud, etc.)
  - Keeps essential features of RDBMS's while learning from key-value noSQL systems

http://www.slideshare.net/spf13/mongodb-9794741?v=qf1&b=&from_search=13

Advantages of NoSQL

- Can easily scale up
- Doesn’t require a pre-define schema
- Cheap, easy to implement
- Relaxes data consistency requirement
- Data can be replicated to multiple nodes and can be partitioned

SQL vs NoSQL

- Relational Database
- Relational model
- Pre-defined schema

- Table-based schema

- Vertically scalable
- Not preferred for larger datasets
- Emphasis on ACID properties
- Eg Oracle, DB2,MySQL etc.

- Non-relational, distributed database
- Model-less approach
- Dynamic schema for unstructured data
- Document-based or Graph based or wide column store or key-value pairs database
- Horizontally Scalable
- largely preferred for large datasets
- Emphasis CAP theorem
- MongoDB,Hbase,Cassandra,Redis,Neo4j,CouchDBetc.