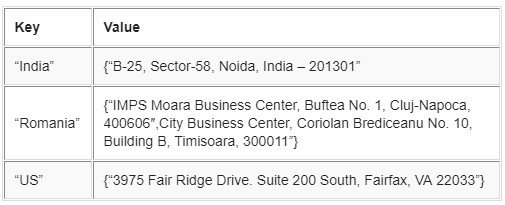
# Big Data Session 9 Assignment 3

NoSQL Databases:

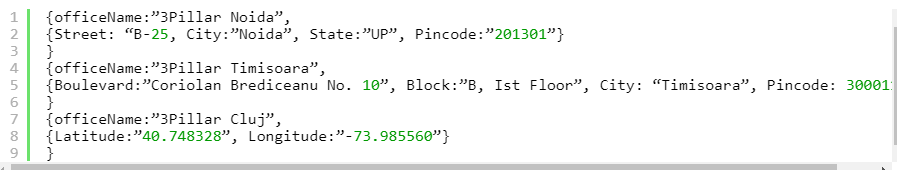
NoSQL database is a way to store and retrieve data from database not by the way traditional databases used to store and retrieve data. Traditional databases mainly store data in tabular format i.e. in rows and columns forms. So, NoSQL databases mainly work on non-relational data.

Types of NoSQL Databases

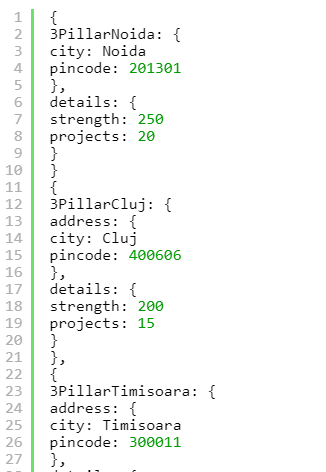
1. **Key-Value Store** – It has a Big Hash Table of keys & values {Example- Riak, Amazon S3 (Dynamo)}



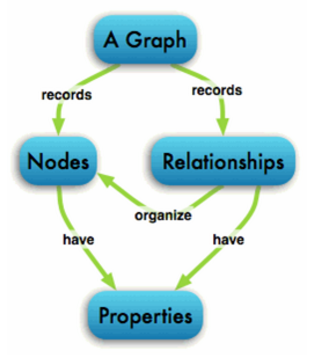
1. **Document-based** **Store- It**stores documents made up of tagged elements. {Example- CouchDB}



1. **Column-based Store-**Each storage block contains data from only one column, {Example- HBase, Cassandra}



1. **Graph-based**-A network database that uses edges and nodes to represent and store data. {Example- Neo4J}



CAP Theorem

**Consistency** - This means that the data in the database remains consistent after the

execution of an operation. For example after an update operation, all clients see the

same data.

**Availability** - This means that the system is always on (service guarantee availability),

no downtime.

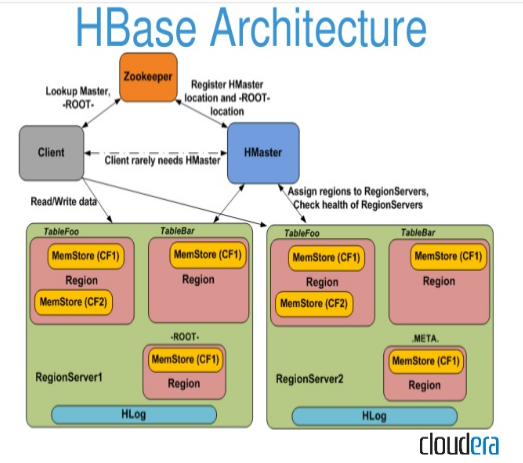
**Partition** **Tolerance** - This means that the system continues to function even if the

communication among the servers is unreliable, i.e. the servers may be partitioned into

multiple groups that cannot communicate with one another.

HBase Architecture

HBase provides low-latency random reads and writes on top of HDFS. In HBase, tables are dynamically distributed by the system whenever they become too large to handle (Auto Sharding). The simplest and foundational unit of horizontal scalability in HBase is a Region. A continuous, sorted set of rows that are stored together is referred to as a region (subset of table data).  HBase architecture has a single HBase master node (HMaster) and several slaves i.e. region servers. Each region server (slave) serves a set of regions, and a region can be served only by a single region server. Whenever a client sends a write request, HMaster receives the request and forwards it to the corresponding region server.



HBase vs RBMS

|  |  |  |
| --- | --- | --- |
| **Feature** | **NoSQL** | **RDBMS** |
| **Data Volume** | Handles Huge Data Volumes | Handles Limited Data Volumes |
| **Data Validity** | Highly Guaranteed | Less Guaranteed |
| **Scalability** | Horizontally | Horizontally & Vertically |
| **Query Language** | No declarative query language | Structured Query Language (SQL) |
| **Schema** | No predefined schema or less rigid schemas | Predefined Schema (Data Definition Laguage & Data Manipulation Language) |
| **Data Type** | Supports unstructured and unpredictable data | Supports relational data and its relationships are stored in separate tables |
| **ACID/BASE** | Based on BASE principle (Basically, Available, Soft State, Eventually Consistent) | Based on ACID principle (Atomicity, Consistency, Isolation and Durability) |
| **Transaction Management** | Weaker transactional guarantee | Strong transactional guarantees |
| **Data Storage Technique** | Schema-free collections are utilized to store different typesand document structures, such as {“color”, “blue”} and {“price”, “23.5”} can be stored within a single collection. | No collections are used for data storage; instead use DML for it. |