

## NoSQL Systems

### - Key-Value Store

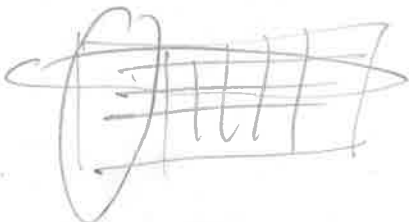
- Map (associative array)
- Simple, non-trivial
- Keys may be sorted to support range requests
- Ex. Riak, Redis, etc.

### - Document Store

- store data as a collection of documents
- XML, JSON
- Documents may have different fields
- API to query the data in documents
- Can support nested structures
- Each document has a key associated with it
- ex. CouchDB

### Column Store

- good for OLAP workload
- store data in columns instead of rows



- Related to timeseries DB
- Ex. C-Store
- kdb

## Large Scale Distributed Data Processing

- 3 Vs: Volume, Velocity, Variety

### - Hadoop

- talk about Map Reduce
- 2 steps on disk

### - Storm

- sprouts and bolts to define sources and operators
- batch processing of stream data
- DAG topology
- Streaming - running indefinitely

### - Spark

- streaming model
- RDDs - resilient data sets
- in memory, can be faster than Hadoop

### - Streaming database

### - Graph Databases

- see slides