Lecture 2

Dr. R. Karthi

What is NoSQL?

- Non-relational data storage systems
- No fixed table schema
- No Joins
- No multi-document transactions
- Relaxes one or more ACID properties

Advantages of NoSQL

- Cheap, Easy to implement
- Easy to distribute
- Can easily scale up & down
- Relaxes the data consistency requirement
- Doesn’t require a pre-defined schema
- Data can be replicated to multiple nodes and can be partitioned

Types of NoSQL

Key value data store	Column-oriented data store	Document data store	Graph data store
• Riak • Redis • Membase	• Cassandra • HBase • HyperTable	• MongoDB • CouchDB • RavenDB	• InfiniteGraph • Neo4 • Allegro Graph

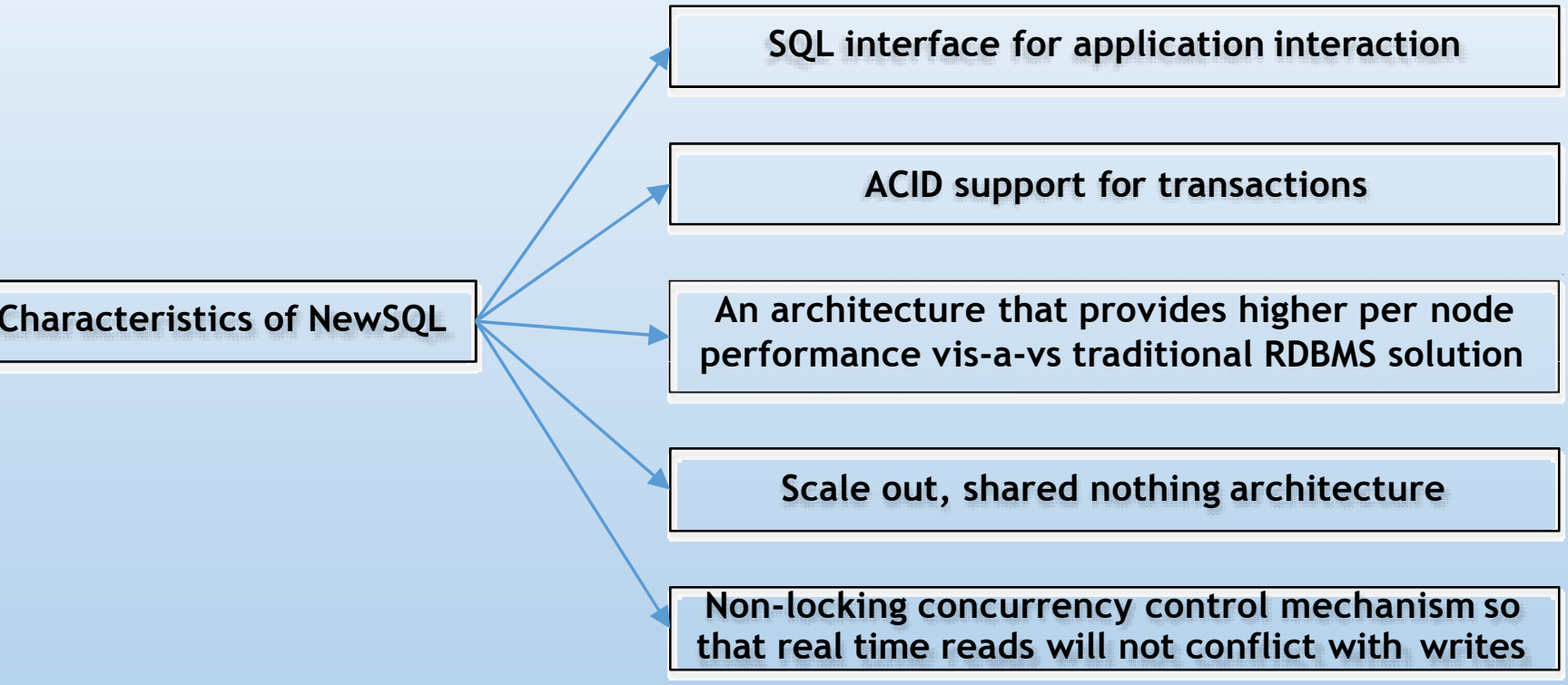
NoSQL Vendors

Company	Product	Most widely used by
Amazon	DynamoDB	LinkedIn, Mozilla
Facebook	Cassandra	Netflix, Twitter, eBay
Google	BigTable	Adobe Photoshop

SQL Vs. NOSQL databases

SQL	NoSQL
elational database	Non-relational, distributed database
elational model	Model-less approach
re-defined schema	Dynamic schema for unstructured data
able based databases	Document-based or graph-based or wide column store or key-value pairs databases
ertically scalable (by increasing system esources)	Horizontally scalable (by creating a cluster of commodity machines)
ses SQL	Uses UnSQL (Unstructured Query Language)
ot preferred for large datasets	Largely preferred for large datasets
ot a best fit for hierarchical data	Best fit for hierarchical storage as it follows the key-value pair of storing data similar to JSON (Java Script Object Notation)
mphasis on ACID properties	Follows Brewer's CAP theorem
xcellent support from vendors	Relies heavily on community support
upports complex querying and data eeping needs	Does not have good support for complex querying
an be configured for strong consistency	Few support strong consistency (e.g., MongoDB), few others can be configured for eventual consistency (e.g., Cassandra)
xamples: Oracle, DB2, MySQL, MS SQL, ostgreSQL, etc.	MongoDB, HBase, Cassandra, Redis, Neo4j, CouchDB, Couchbase, Riak, etc.

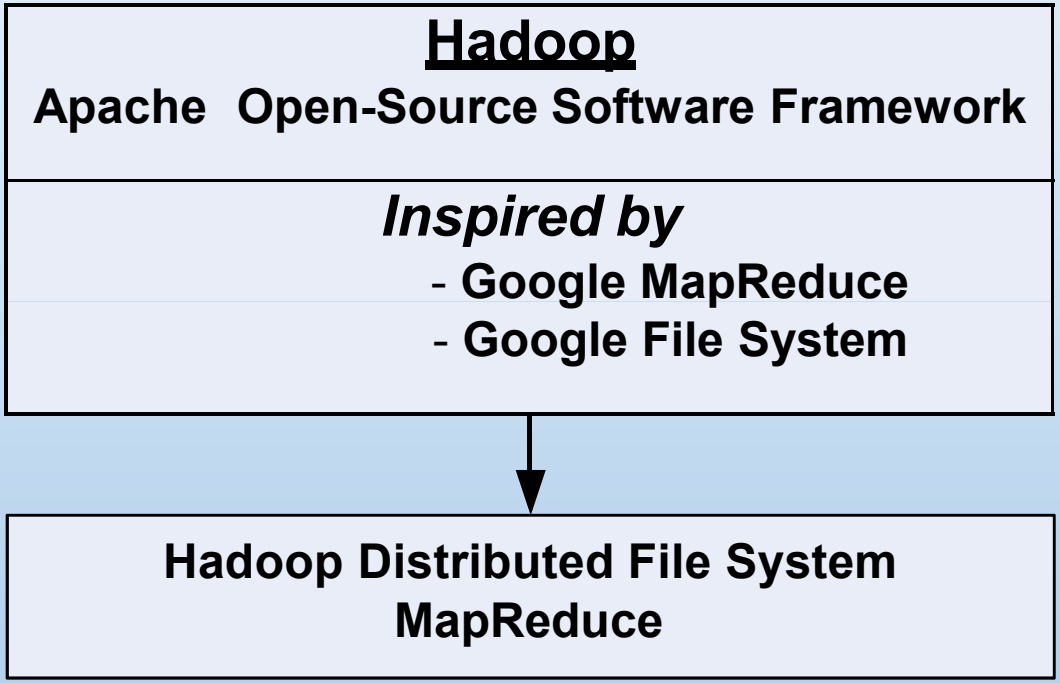
NewSQL


Characteristics of NewSQL
SQL interface for application interaction
ACID support for transactions
An architecture that provides higher per node performance vis-a-vs traditional RDBMS solution
Scale out, shared nothing architecture
Non-locking concurrency control mechanism so that real time reads will not conflict with writes

SQL Vs. NoSQLVs. NewSQL

	SQL	NoSQL	NewSQL
Adherence to ACID properties	Yes	No	Yes
OLTP/OLAP	Yes	No	Yes
Schema rigidity Adherence to data model	Yes Adherence to relational model	No	Maybe
Data Format Flexibility	No	Yes	Maybe
Scalability	Scale up Vertical Scaling	Scale out Horizontal Scaling	Scale out
Distributed Computing	Yes	Yes	Yes
Community Support	Huge	Growing	Slowly growing

Hadoop



Key Advantages of Hadoop

- Stores data in its native format
- Scalable
- Cost-effective
- Resilient to failure
- Flexibility
- Fast

Versions of Hadoop