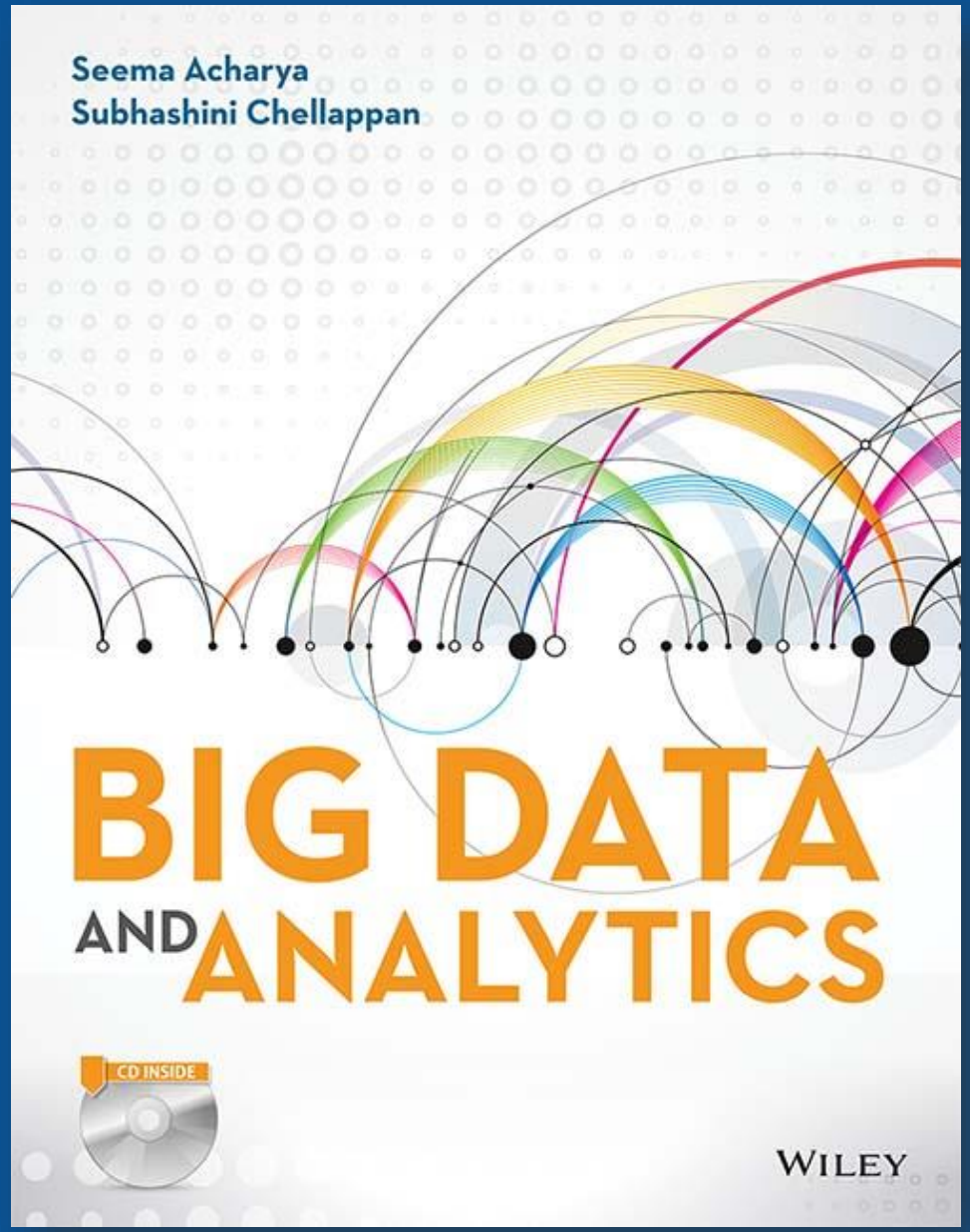


Big Data And Analytics

Seema Acharya
Subhashini Chellappan

Chapter 4

The Big Data Technology Landscape

Learning Objectives and Learning Outcomes

Learning Objectives	Learning Outcomes
The big data technology landscape 1. What is NoSQL databases? 2. Why NoSQL? 3. Key advantages of NoSQL. 4. What is NewSQL? 5. SQL Vs. NoSQL. 6. Getting familiar with Hadoop.	a) To understand the significance of NoSQL databases. b) To understand the need for NewSQL. c) To understand the Hadoop platform and be able to appreciate the difference between Hadoop 1.0 and Hadoop 2.0.

Session Plan

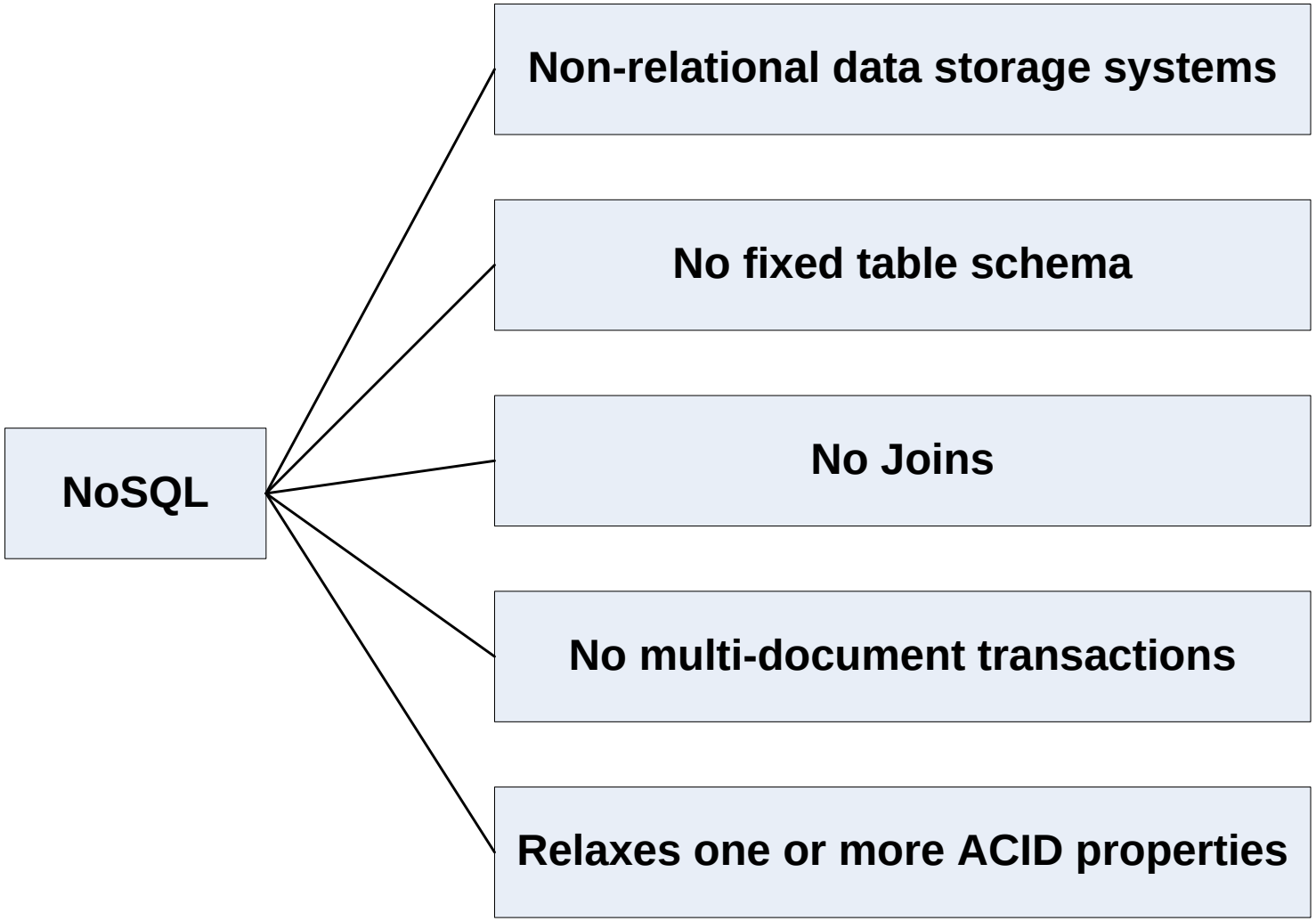
Lecture time	45 to 60 minutes
Q/A	15 minutes

Agenda

- NoSQL
  - What is it?
  - Types of NoSQL Databases
  - Why NoSQL?
  - Advantages of NoSQL
  - NoSQL Vendors
  - SQL versus NoSQL
  - NewSQL
  - Comparison of SQL, NoSQL and NewSQL
- Hadoop
  - Features of Hadoop
  - Key Advantages of Hadoop
  - Versions of Hadoop

What is NoSQL?

What is NoSQL?



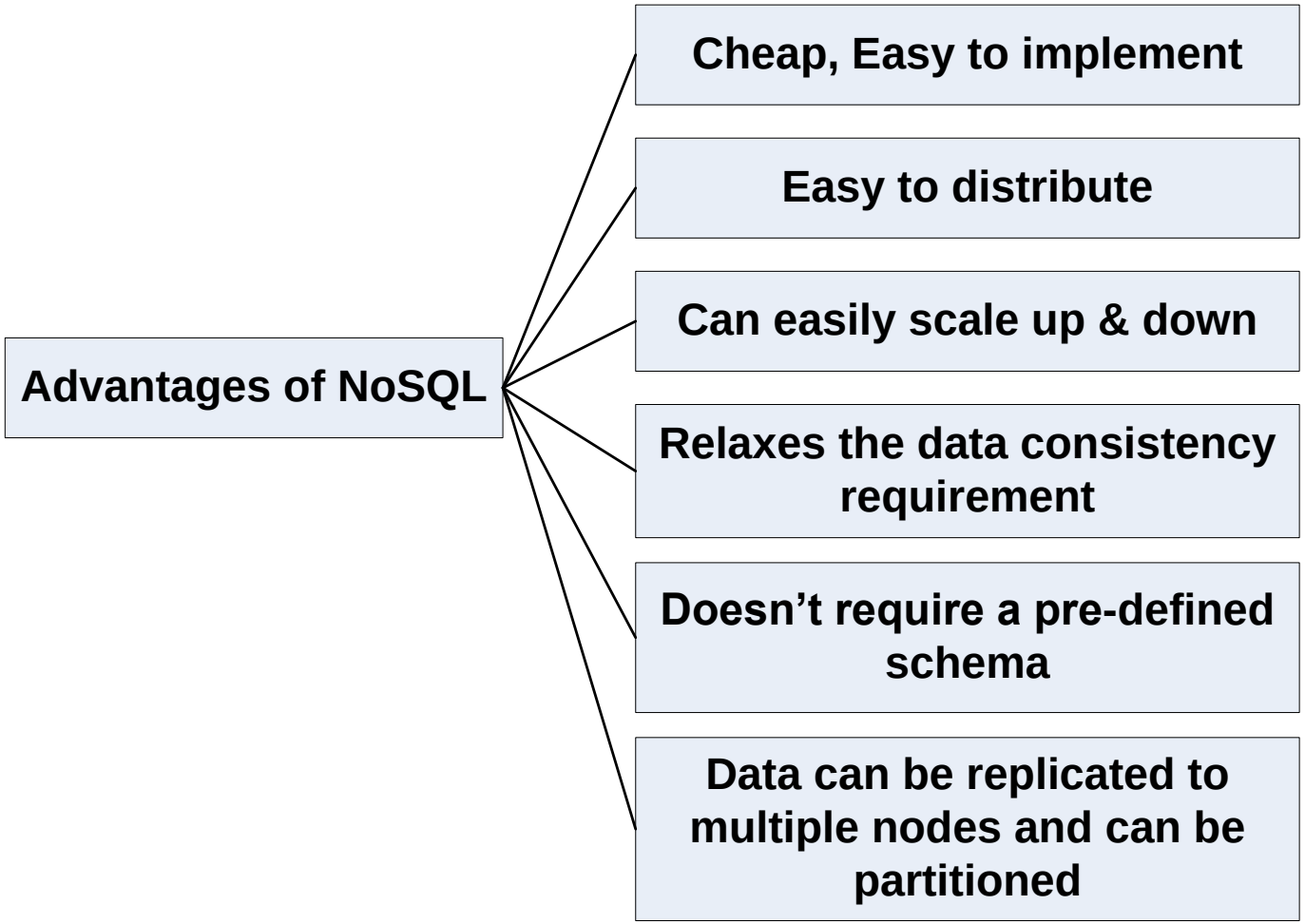
Types of NoSQL

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

Advantages of NoSQL

Advantages of NoSQL



NoSQL Vendors

NoSQL Vendors

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

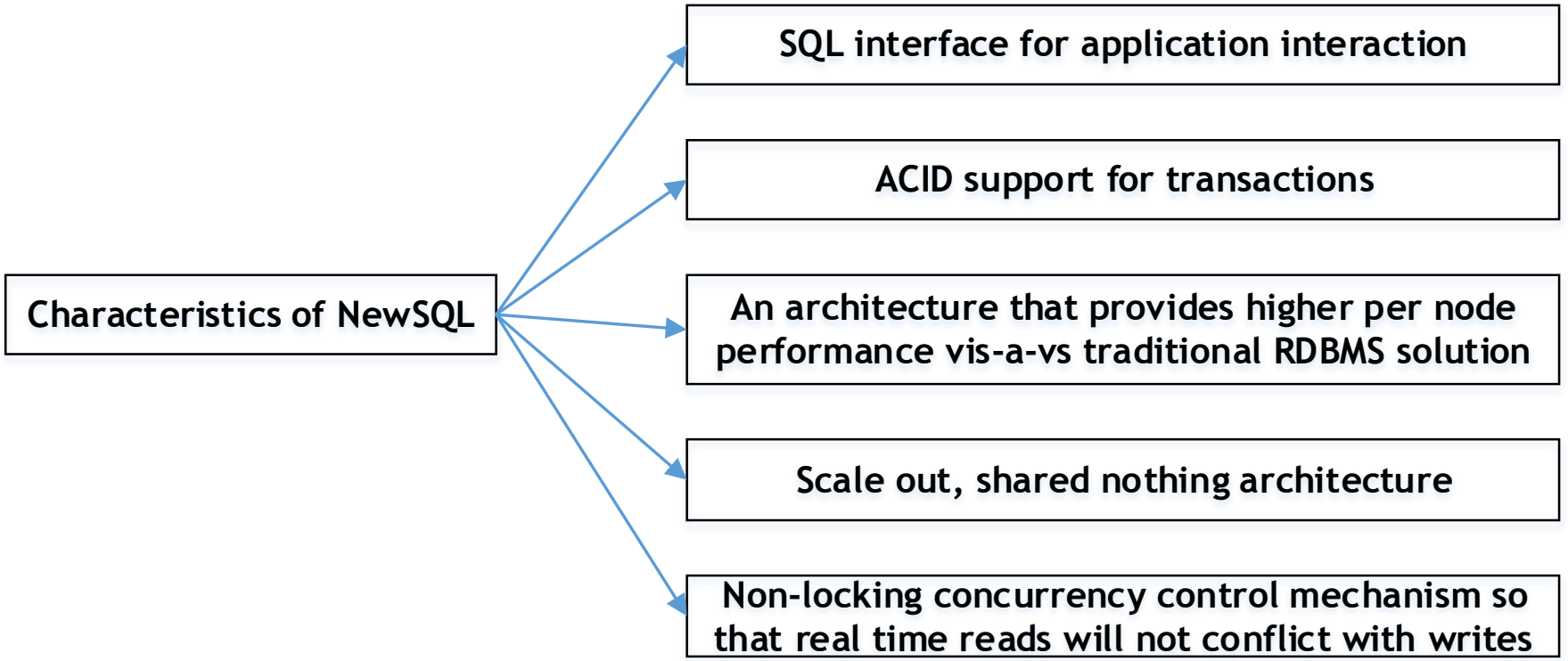
SQL Vs. NoSQL

SQL Vs. NoSQL

SQL	NoSQL
Relational database	Non-relational, distributed database
Relational model	Model-less approach
Pre-defined schema	Dynamic schema for unstructured data
Table based databases	Document-based or graph-based or wide column store or key-value pairs databases
Vertically scalable (by increasing system resources)	Horizontally scalable (by creating a cluster of commodity machines)
Uses SQL	Uses UnQL (Unstructured Query Language)
Not preferred for large datasets	Largely preferred for large datasets
Not 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)
Emphasis on ACID properties	Follows Brewer's CAP theorem
Excellent support from vendors	Relies heavily on community support
Supports complex querying and data keeping needs	Does not have good support for complex querying
Can be configured for strong consistency	Few support strong consistency (e.g., MongoDB), few others can be configured for eventual consistency (e.g., Cassandra)
Examples: Oracle, DB2, MySQL, MS SQL, PostgreSQL, etc.	MongoDB, HBase, Cassandra, Redis, Neo4j, CouchDB, Couchbase, Riak, etc.

NewSQL

NewSQL



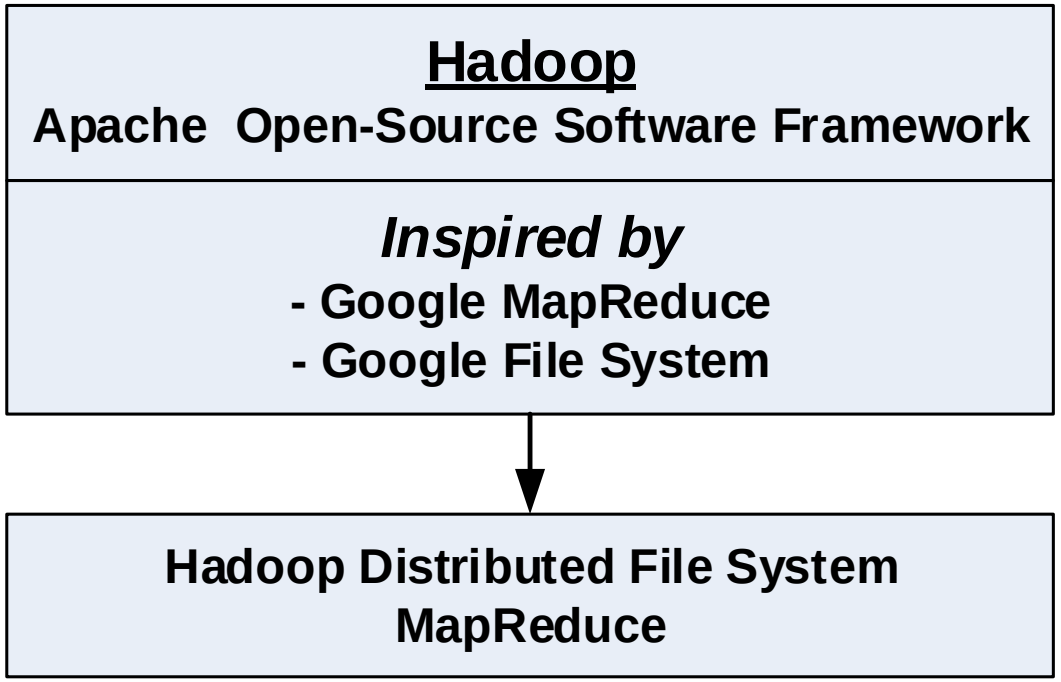
SQL Vs. NoSQL Vs. NewSQL

SQL Vs. NoSQL Vs. 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

Hadoop

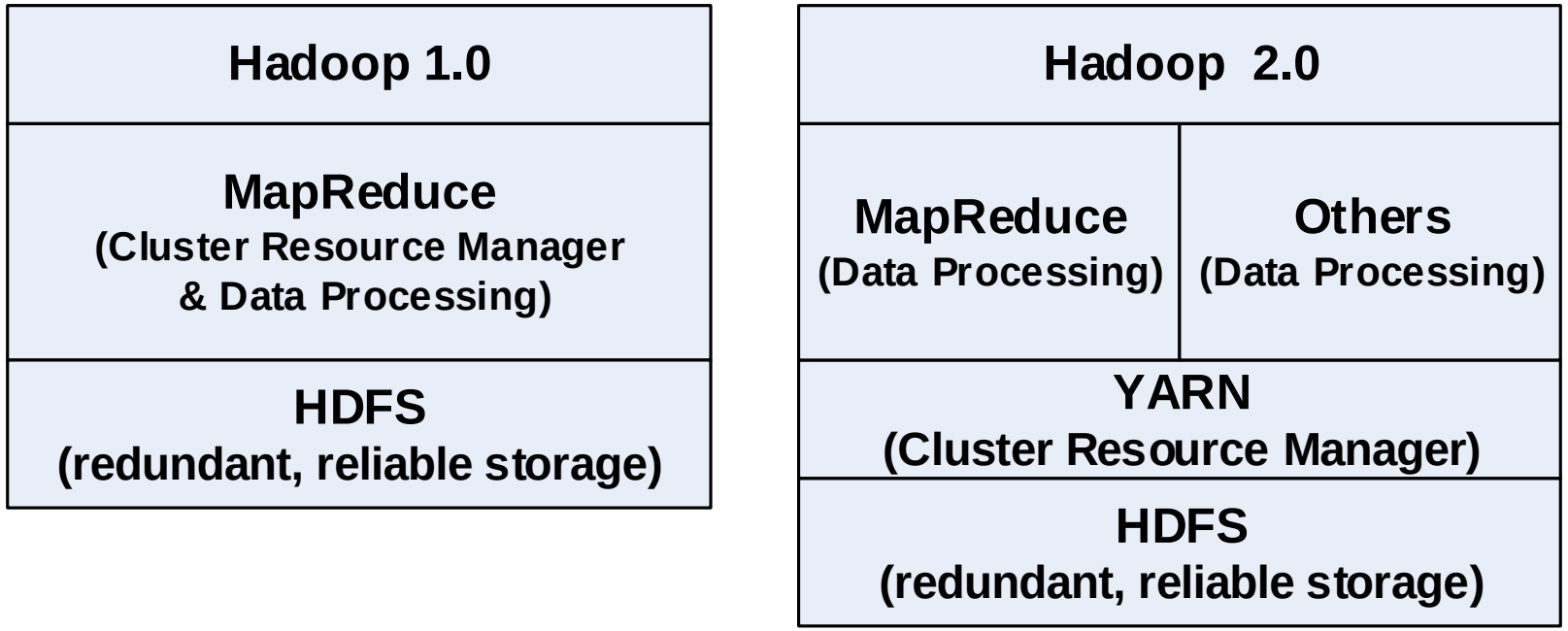


Key Advantages of Hadoop

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

Versions of Hadoop

Versions of Hadoop



Answer a few quick questions …

Fill in the blanks

1. The expansion for CAP is _____, _____ and _____.
2. The expansion of BASE is _____.
3. MongoDB is _____ and _____.
4. Cassandra is _____ and _____.
5. _____ has no support for ACID properties of transactions.
6. _____ is a robust database that supports ACID properties of transactions and has the scalability of NoSQL.

Answer Me

- Cite the difference between Hadoop 1.0 and Hadoop 2.0.
- Compare and contrast SQL, NoSQL and NewSQL.

Summary please…

Ask a few participants of the learning program to summarize the lecture.

References …

Further Readings

- http://www.mongodb.com/nosql-explained
- http://nosql-database.org/
- http://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduce_Compatibility_Hadoop1_Hadoop2.html
- http://hadoop.apache.org/

Thank you