Lesson 1
Select Key Terms

MapReduce programming model

- Refers to a programming paradigm for processing Big Data sets with a parallel and distributed environment using map and reduce tasks.

"Big Data Analytics ", Ch.04 L01: MapReduce, Hive and Pig, Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

Yarn

- A component of Hadoop, providing:

- Resources management using multiple machines (data nodes)

- Running and scheduling of the parallel programs for map and reduce tasks

- Allocating parallel processing resources for computing subtasks

NoSQL data model

- Refers to model offering relaxation in one or more of the ACID properties (Atomicity, consistence, isolation, and durability) of the database

- Follows CAP (consistency, availability and partitions) theorem

- Relies upon a model known as the BASE model.

CAP Theorem

- CAP (consistency, availability and partitions) theorem states that out of three properties, two must at least present for the application/service/ process.

-

"Big Data Analytics ", Ch.04 L01: MapReduce, Hive and Pig,
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

BASE Model

- Model has three principles:

- Basic Availability (even in the presence of multiple failures)

- Soft State (data consistency is the developer's problem and should not be handled by the database)

- Eventual Consistency (return to the last updated value).

"Big Data Analytics ", Ch.04 L01: MapReduce, Hive and Pig, Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

Key-Value Data Store

- A simplest way to implement a schema-less database.

- A string called, key maps to values in a large data string or BLOB (basic large object)

- Use primary-key access; therefore, the storage easily scales up and data-retrievals are fast

"Big Data Analytics ", Ch.04 L01: MapReduce, Hive and Pig,
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

Tabular Data Store

- Refers to table, column-family, or BigTable like Data Store

Column family Big-Data

- *A* storage in logical groups of column families

- May be similar to columns of sparse matrix

- Use a pair of row and column keys to access fields in the columns

Object Data Store

- A repository for:

(i) objects (such as files, ….,

(ii) system metadata which provides information such as filename, creation_date, last_modified, language_used, and

Object Data Store

(iii) Custom metadata which provides information such as subject, category and sharing permission.

Tuples and Collections

- *Tuple* means an ordered list of elements. An n-tuple relates to set theory, a collection (sequence) of "n" elements. Tuples implement the records.

- *Collection* means a well-defined collection of distinct objects in a set, the objects of a set are the elements.

"Big Data Analytics ", Ch.04 L01: MapReduce, Hive and Pig,
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

Projection

- A unary operation (single input or operand) written as $\Pi_{attr1, attr2,\dots, attrn}$ where($attr1, attr2,\dots, attrn$) is a set of n attribute-names

- Projection returns a set obtained by selecting only the n-attributes

Natural Join

- Two tables join based on all common columns

- Both the tables must have same column name and the data types

- .

"Big Data Analytics ", Ch.04 L01: MapReduce, Hive and Pig,
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

Inner Join

- A default natural join

- Inner Join; Two tables join

- Based on common columns, mentioned using the ON clause.

- INNER JOIN returns all rows from both tables if the columns match

"Big Data Analytics ", Ch.04 L01: MapReduce, Hive and Pig, Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

Indexing to a field

- Providing reference to a field in a document of collections that support the queries and operations using that index

- A DB creates an index on the _id field of every collection

-

Summary

We learnt meanings of:

- ACID Property

- CAP

- BASE

- NoSQL Data Stores

- Join

"Big Data Analytics ", Ch.04 L01: MapReduce, Hive and Pig,
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

End of Lesson 1 on Meanings of Key Terms

"Big Data Analytics ", Ch.04 L01: MapReduce, Hive and Pig,
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India