Cloud computing

* + Software as a service (SaaS)
  + Utility computing

Slice 1 18/19

26-29

32

34-

* Infrastructure as a service
* Platform as a service
* Software as a service

Slice 2

5 Problems (Example)

Key Idea

Big Concepts

How does Hadoop Work?

Goals of Hadoop Distributed File System HDFS

Hadoop Distributed File System HDFS

Cluster, Job (Task)

HQL -> JOB -> Cluster

Submit Job….XXX

Select avg(score) from table where age >10

Map n Node (10W)

Task1 Filter Age > 10

Task 2 Avg ()

Job finished xx

Hive

Pig and Hive Similar/ Difference

Schema

HQL

SELECT content FROM table1 WHERE condition JOIN table2 ON table2.prop = table1.prop GROUP BY table1.prop2 HAVING condition2 ORDER BY table1.prop LIMIT n

UNION -

JOIN HAVING

HAVING function()– Groupby

ID SCROE COURSE

A 40 YW

A 60 SX

B 30 YW

ID NAME

A HUANG MEIJING

B MOHANMODE

GPA

SELECT NAME , AVG (SCORE) AS avgScore FROM ScoreTable GROUP BY NAME

A 50

B 30

SELECT \* FROM (SELECT NAME , AVG (SCORE) FROM ScoreTable GROUP BY NAME) WHERE avgScore >= 50

**SELECT ID , AVG (SCORE) AS avgScore FROM ScoreTable GROUP BY NAME HAVING avgScore>=50**

SELECT ID , AVG (SCORE) AS avgScore FROM ScoreTable JOIN T2 ON T1.id = T2.id

GROUP BY NAME HAVING avgScore>=50

A 50

B 30

A HUANG MEIJING

B MOHANMODE

SELECT ID , AVG (SCORE) AS avgScore FROM ScoreTable GROUP BY NAME HAVING avgScore>=50

JOIN T2 ON T1.id = T2.id

A 50

A HUANG MEIJING

B MOHANMODE

JOIN T2 ON T1.id = T2.id

JOIN

SELECT \* FROM T1 JOIN T2 ON T1.id = T2.id

YARN

NOSQL

**SPARK**

**&&, ||, !**

**RDD**

**Val xxx = fsddsd**

**Xxx type RDD**

**Map, filter, ~~join~~ , (reduceByKey), groupByKey, aggregateByKey**

**SQL (DATAFRAME)**

Val Df = spark.sql(  
  """SELECT \*  
     FROM people  
     JOIN json ...""")

**DATAFRAME df**

**Select() filter() where() groupby() join() agg()**

**Agg() limit() union()**