



Recap

2

► Spark

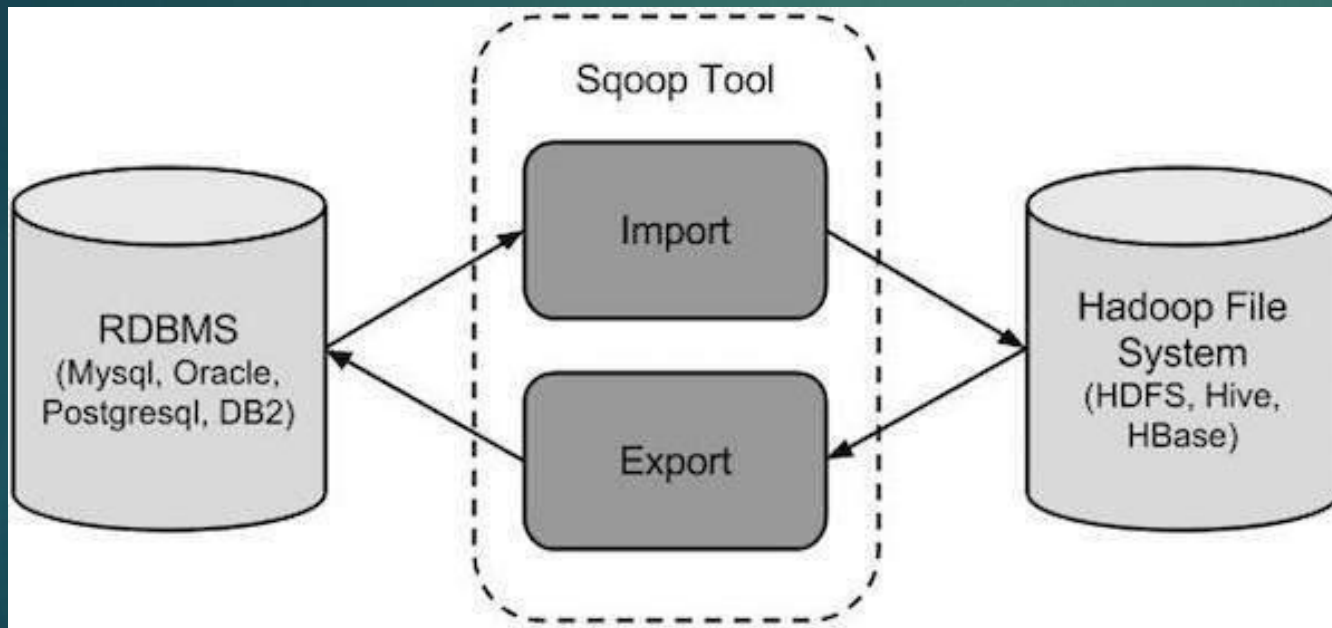
Agenda for today

3

- ▶ Spark

- ▶ Sqoop

Introduction



Export

5

Parameter	Description
--table	Target table name
--export-dir	HDFS source dir name
--fields-terminated-by	Field delimiter
-m, --num-mappers	#mappers to launch
--staging-table	Staging table for temp storage
--jar-file	Use mentioned jar file to export
--update-key	Update data in RDBMS based on mentioned key

- ▶ `sqoop export --connect jdbc:mysql://localhost :3306/retail_db --username retail_dba --password cloudera --table test --fields-terminated-by ',' --export-dir <HDFS DIRECTORY NAME>`

Import

6

Parameter	Description
--table	Source table name
--target-dir	HDFS target dir name
--fields-terminated-by	Field delimiter
-m, --num-mappers	#mappers to launch
--split-by	Unique column name
--delete-target-dir	Delete target HDFS dir if exists
--where	Condition to apply while fetching data from RDBMS

- ▶ `sqoop import --connect jdbc:mysql://localhost :3306/<DATABASE NAME> --username root -p --table <TABLE NAME> --m 1 --target-dir <HDFS DIRECTORY NAME>`

Jobs

7

- ▶ Compile sqoop jobs for regular execution

- ▶ Create Job

```
sqoop job --create myjob -- import --connect jdbc:mysql:// localhost  
:3306/retail_db --username retail_dba --password cloudera --table departments  
--target-dir <HDFS DIRECTORY NAME>
```

- ▶ List all created jobs

```
sqoop job --list
```

- ▶ Show details of one specific job

```
sqoop job --show myjob
```

- ▶ Execute created job

```
sqoop job --exec myjob
```


- ▶ Evaluate a single command on RDBMS
`sqoop eval --connect jdbc:mysql://localhost:3306/retail_db --username retail_dba --password cloudera -e "INSERT INTO Test VALUES(999, 'name999')"`
- ▶ What could be the use case of eval tool?

References

- ▶ <https://sqoop.apache.org/docs/1.4.0-incubating/SqoopUserGuide.html>
- ▶ Hadoop: the definitive guide 4th edition