Prepared By
M.Sasidhar
Cloudera Certified Hadoop Professional
Sasis937@gmail.com

Sqoop

Apache Sqoop is a tool designed for transferring bulk data between Apache Hadoop and structured datastores such as relational databases

Sqoop imports data from external structured datastores into HDFS

Sqoop can also be used to extract data from Hadoop and export it to external structured datastores such as relational databases

slide 3 --- download and unarchive

download sqoop from below url

https://archive.apache.org/dist/sqoop/1.4.0-incubating/

Give permissions

Chmod 777 sqoop….

unarchive by using

tar -xzvf sqoop...

slide4 – sqoop releases

Sqoop is an open source software product of the Apache Software Foundation.

Prepared By
M.Sasidhar
Cloudera Certified Hadoop Professional
Sasis937@gmail.com

Software development for Sqoop occurs at http://svn.apache.org/repos/asf/incubator/sqoop/trunk

slide – usage

Sqoop will read the table row-by-row into HDFS

The output of this import process is a set of files containing a copy of the imported table

The import process is performed in parallel

For this reason, the output will be in multiple files

slide – sqoop commands

```
$ sqoop help
usage: sqoop COMMAND [ARGS]

Available commands:
  codegen            Generate code to interact with database records
  create-hive-table  Import a table definition into Hive
  eval               Evaluate a SQL statement and display the results
  export             Export an HDFS directory to a database table
  help               List available commands
  import             Import a table from a database to HDFS
  import-all-tables  Import tables from a database to HDFS
  list-databases     List available databases on a server
  list-tables        List available tables in a database
  version            Display version information
```

slide – sqoop arguments

```
Common arguments:
   --connect <jdbc-uri>     Specify JDBC connect string
   --connect-manager <jdbc-uri>     Specify connection manager class to use
   --driver <class-name>    Manually specify JDBC driver class to use
   --hadoop-home <dir>      Override $HADOOP_HOME
```

Prepared By
M.Sasidhar
Cloudera Certified Hadoop Professional
Sasis937@gmail.com

```
   --help                    Print usage instructions
-P                           Read password from console
   --password <password>     Set authentication password
   --username <username>     Set authentication username
```

slide – import

```
$ sqoop import --connect jdbc:mysql://localhost/db --username foo --table
TEST
$ sqoop --options-file /users/homer/work/import.txt --table TEST
import.txt file contains
import
--connect
jdbc:mysql://localhost/db
--username
foo
```

slide –

```
#
# Options file for Sqoop import
#

# Specifies the tool being invoked
import

# Connect parameter and value
--connect
jdbc:mysql://localhost/db

# Username parameter and value
--username
foo

#
# Remaining options should be specified in the command line.
#
```


connect

sqoop import –connect jdbc:oracle:thin:@192.168.137.23:1521:india --username CORE_80_CLEANUP –paasword core_80_cleanup –table AC_AMOUNT

Prepared By
M.Sasidhar
Cloudera Certified Hadoop Professional
Sasis937@gmail.com