

Slide # 67

```
[root@sandbox-hdp ~]# cd /etc/hive/conf
[root@sandbox-hdp conf]# ls
atlas-application.properties          hive-exec-log4j.properties
beeline-log4j.properties.template    hive-log4j.properties
conf.server                           hive-site.xml
hive-default.xml.template             ivysettings.xml
hive-env.sh                           mapred-site.xml
hive-env.sh.template                 parquet-logging.properties
[root@sandbox-hdp conf]# cat hive-site.xml
<configuration>

  <property>
    <name>ambari.hive.db.schema.name</name>
    <value>hive</value>
  </property>
```

Slide # 68

```
[root@sandbox-hdp ~]# mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 51
Server version: 5.7.22 MySQL Community Server (GPL)

Copyright (c) 2000, 2018, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| druid |
| hive |
| mysql |
| performance_schema |
| ranger |
| superset |
| sys |
+-----+
8 rows in set (0.18 sec)

mysql> use hive;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
```

Slide # 69

```
mysql> select * from ROLES;
```

ROLE_ID	CREATE_TIME	OWNER_NAME	ROLE_NAME
1	1529334605	admin	admin
2	1529334605	public	public

```
2 rows in set (0.00 sec)
```

Slide # 74

```
[root@sandbox-hdp ~]# sqoop-import --connect jdbc:mysql://sandbox-
hdp.hortonworks.com/hive --username root --P --table TBLs --target
-dir /user/anup/sqoop --fields-terminated-by '|' --num-mappers 7
Warning: /usr/hdp/2.6.5.0-292/accumulo does not exist! Accumulo im
ports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installatio
n.
20/05/23 17:06:13 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6.2.6.5.0-292
Enter password:
20/05/23 17:06:20 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
20/05/23 17:06:20 INFO tool.CodeGenTool: Beginning code generation
```

Slide # 76

```
AttemptID:attempt_1590192907351_0005_m_000002_0 Timed out after 300 secs
20/05/23 17:19:08 INFO mapreduce.Job: map 86% reduce 0%
20/05/23 17:19:12 INFO mapreduce.Job: map 100% reduce 0%
20/05/23 17:19:24 INFO mapreduce.Job: Job job_1590192907351_0005 completed successfully
20/05/23 17:19:25 INFO mapreduce.Job: Counters: 31
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=1198463
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=754
    HDFS: Number of bytes written=2544
    HDFS: Number of read operations=28
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=14
  Job Counters
    Failed map tasks=7
    Launched map tasks=14
    Other local map tasks=14
    Total time spent by all maps in occupied slots (ms)=3170093
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=3170093
    Total vcore-milliseconds taken by all map tasks=3170093
    Total megabyte-milliseconds taken by all map tasks=792523250
  Map-Reduce Framework
    Map input records=37
    Map output records=37
    Input split bytes=754
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=27651
    CPU time spent (ms)=48980
    Physical memory (bytes) snapshot=1184595968
    Virtual memory (bytes) snapshot=13667975168
    Total committed heap usage (bytes)=309329920
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=2544
20/05/23 17:19:25 INFO mapreduce.ImportJobBase: Transferred 2.4844 KB in 723.3548 seconds (3.5169 bytes/sec)
20/05/23 17:19:25 INFO mapreduce.ImportJobBase: Retrieved 37 records.
```

Slide # 77

```
[root@sandbox-hdp ~]# hdfs dfs -ls /user/anup/
Found 1 items
drwxr-xr-x  - root hdfs          0 2020-05-23 17:19 /user/anup/sqoop
[root@sandbox-hdp ~]# hdfs dfs -ls /user/anup/sqoop
Found 8 items
-rw-r--r--  1 root hdfs          0 2020-05-23 17:19 /user/anup/sqoop/_SUCCESS
-rw-r--r--  1 root hdfs        424 2020-05-23 17:19 /user/anup/sqoop/part-m-00000
-rw-r--r--  1 root hdfs        385 2020-05-23 17:19 /user/anup/sqoop/part-m-00001
-rw-r--r--  1 root hdfs        403 2020-05-23 17:19 /user/anup/sqoop/part-m-00002
-rw-r--r--  1 root hdfs        305 2020-05-23 17:19 /user/anup/sqoop/part-m-00003
-rw-r--r--  1 root hdfs        324 2020-05-23 17:19 /user/anup/sqoop/part-m-00004
-rw-r--r--  1 root hdfs        321 2020-05-23 17:19 /user/anup/sqoop/part-m-00005
-rw-r--r--  1 root hdfs        382 2020-05-23 17:19 /user/anup/sqoop/part-m-00006
```

Slide # 78

```
[root@sandbox-hdp ~]# hdfs dfs -cat /user/anup/sqoop/part-m-00000
1|1529335142|2|0|hive|0|1|sales_fact_1997|EXTERNAL_TABLE|null|null
2|1529335145|2|0|hive|0|2|sales_fact_1998|EXTERNAL_TABLE|null|null
3|1529335147|2|0|hive|0|3|sales_fact_dec_1998|EXTERNAL_TABLE|null|null
4|1529335150|2|0|hive|0|4|inventory_fact_1997|EXTERNAL_TABLE|null|null
5|1529335152|2|0|hive|0|5|inventory_fact_1998|EXTERNAL_TABLE|null|null
6|1529335154|2|0|hive|0|6|agg_pl_01_sales_fact_1997|EXTERNAL_TABLE|null|null
```

Slide # 79

```
Database changed
mysql> select * from TBLs LIMIT 10;
```

TBL_ID	CREATE_TIME	DB_ID	LAST_ACCESS_TIME	OWNER	RETENTION	SD_ID	TBL_NAME	TBL_TYPE	VIEW_EXPANDED_TEXT	VIEW_ORIGINAL_TEXT
1	1529335142	2	0	hive	0	1	sales_fact_1997	EXTERNAL_TABLE	NULL	NULL
2	1529335145	2	0	hive	0	2	sales_fact_1998	EXTERNAL_TABLE	NULL	NULL
3	1529335147	2	0	hive	0	3	sales_fact_dec_1998	EXTERNAL_TABLE	NULL	NULL
4	1529335150	2	0	hive	0	4	inventory_fact_1997	EXTERNAL_TABLE	NULL	NULL
5	1529335152	2	0	hive	0	5	inventory_fact_1998	EXTERNAL_TABLE	NULL	NULL
6	1529335154	2	0	hive	0	6	agg_pl_01_sales_fact_1997	EXTERNAL_TABLE	NULL	NULL
7	1529335157	2	0	hive	0	7	agg_ll_01_sales_fact_1997	EXTERNAL_TABLE	NULL	NULL
8	1529335159	2	0	hive	0	8	agg_l_03_sales_fact_1997	EXTERNAL_TABLE	NULL	NULL
9	1529335161	2	0	hive	0	9	agg_l_04_sales_fact_1997	EXTERNAL_TABLE	NULL	NULL
10	1529335164	2	0	hive	0	10	agg_l_05_sales_fact_1997	EXTERNAL_TABLE	NULL	NULL

```
10 rows in set (0.00 sec)
```

Slide # 80

```
mysql> select * from ROLES;
```

ROLE_ID	CREATE_TIME	OWNER_NAME	ROLE_NAME
1	1529334605	admin	admin
2	1529334605	public	public

2 rows in set (0.00 sec)

Slide # 80

Try --help for usage instructions.

```
[root@sandbox-hdp ~]# sqoop-import --connect jdbc:mysql://sandbox-hdp.hortonworks.com/hive --username root --P --table ROLES --target-dir /user/anup/sqoop/hive_roles --columns "ROLE_ID, OWNER_NAME, ROLE_NAME" --fields-terminated-by '|'
```

Warning: /usr/hdp/2.6.5.0-292/accumulo does not exist! Accumulo imports will fail.

Please set \$ACCUMULO_HOME to the root of your Accumulo installation.

20/05/23 17:57:07 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6.2.6.5.0-292

Enter password:

20/05/23 17:57:12 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.

20/05/23 17:57:12 INFO tool.CodeGenTool: Beginning code generation

Slide # 80

20/05/23 17:57:10 INFO db.DBInputFormat: Using Read Committed transaction isolation

20/05/23 17:57:40 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN(`ROLE_ID`), MAX(`ROLE_ID`) FROM `ROLES`

20/05/23 17:57:40 INFO db.IntegerSplitter: Split size: 0; Num splits: 4 from: 1 to: 2

20/05/23 17:57:41 INFO mapreduce.JobSubmitter: number of splits:2

20/05/23 17:57:42 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1590192907351_0006

Slide # 81

```
[root@sandbox-hdp ~]# hdfs dfs -ls /user/anup/sqoop/hive_roles
```

Found 3 items

```
-rw-r--r-- 1 root hdfs 0 2020-05-23 17:58 /user/anup/sqoop/hive_roles/_SUCCESS
```

```
-rw-r--r-- 1 root hdfs 14 2020-05-23 17:58 /user/anup/sqoop/hive_roles/part-m-00000
```

```
-rw-r--r-- 1 root hdfs 16 2020-05-23 17:58 /user/anup/sqoop/hive_roles/part-m-00001
```

```
[root@sandbox-hdp ~]# hdfs dfs -cat /user/anup/sqoop/hive_roles/part-m-00000
```

```
1|admin|admin
```

```
[root@sandbox-hdp ~]# hdfs dfs -cat /user/anup/sqoop/hive_roles/part-m-00001
```

```
2|public|public
```

Slide # 84

```
[root@sandbox-hdp ~]# sqoop-import --connect jdbc:mysql://sandbox-hdp.hortonworks.com/hive --username root --P --table TBLs --target-dir /user/anup/sqoop/tbls --fields-terminated-by '|' --num-mappers 8
Warning: /usr/hdp/2.6.5.0-292/accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
20/05/23 18:19:40 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6.2.6.5.0-292
Enter password:
```

```
20/05/23 18:20:13 INFO db.IntegerSplitter: Split size: 4; Num splits: 8 from: 1 to: 37
20/05/23 18:20:13 INFO mapreduce.JobSubmitter: number of splits:8
20/05/23 18:20:13 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1590192907351_0007
20/05/23 18:20:14 INFO impl.YarnClientImpl: Submitted application application_1590192907351_0007
```

Slide # 85

```
[root@sandbox-hdp ~]# hdfs dfs -ls /user/anup/sqoop/
Found 10 items
-rw-r--r-- 1 root hdfs 0 2020-05-23 17:19 /user/anup/sqoop/_SUCCESS
drwxr-xr-x - root hdfs 0 2020-05-23 17:58 /user/anup/sqoop/hive_roles
-rw-r--r-- 1 root hdfs 424 2020-05-23 17:19 /user/anup/sqoop/part-m-00000
-rw-r--r-- 1 root hdfs 385 2020-05-23 17:19 /user/anup/sqoop/part-m-00001
-rw-r--r-- 1 root hdfs 403 2020-05-23 17:19 /user/anup/sqoop/part-m-00002
-rw-r--r-- 1 root hdfs 305 2020-05-23 17:19 /user/anup/sqoop/part-m-00003
-rw-r--r-- 1 root hdfs 324 2020-05-23 17:19 /user/anup/sqoop/part-m-00004
-rw-r--r-- 1 root hdfs 321 2020-05-23 17:19 /user/anup/sqoop/part-m-00005
-rw-r--r-- 1 root hdfs 382 2020-05-23 17:19 /user/anup/sqoop/part-m-00006
drwxr-xr-x - root hdfs 0 2020-05-23 18:21 /user/anup/sqoop/tbls
[root@sandbox-hdp ~]# hdfs dfs -ls /user/anup/sqoop/tbls/
Found 9 items
-rw-r--r-- 1 root hdfs 0 2020-05-23 18:21 /user/anup/sqoop/tbls/_SUCCESS
-rw-r--r-- 1 root hdfs 347 2020-05-23 18:21 /user/anup/sqoop/tbls/part-m-00000
-rw-r--r-- 1 root hdfs 384 2020-05-23 18:21 /user/anup/sqoop/tbls/part-m-00001
-rw-r--r-- 1 root hdfs 402 2020-05-23 18:21 /user/anup/sqoop/tbls/part-m-00002
-rw-r--r-- 1 root hdfs 326 2020-05-23 18:21 /user/anup/sqoop/tbls/part-m-00003
-rw-r--r-- 1 root hdfs 254 2020-05-23 18:21 /user/anup/sqoop/tbls/part-m-00004
-rw-r--r-- 1 root hdfs 256 2020-05-23 18:21 /user/anup/sqoop/tbls/part-m-00005
-rw-r--r-- 1 root hdfs 253 2020-05-23 18:21 /user/anup/sqoop/tbls/part-m-00006
-rw-r--r-- 1 root hdfs 322 2020-05-23 18:21 /user/anup/sqoop/tbls/part-m-00007
```

Slide # 86

```
[root@sandbox-hdp ~]# hdfs dfs -cat /user/anup/sqoop/tbls/part-m-00001
6|1529335154|2|0|hive|0|6|agg_pl_01_sales_fact_1997|EXTERNAL_TABLE|null|null
7|1529335157|2|0|hive|0|7|agg_ll_01_sales_fact_1997|EXTERNAL_TABLE|null|null
8|1529335159|2|0|hive|0|8|agg_l_03_sales_fact_1997|EXTERNAL_TABLE|null|null
9|1529335161|2|0|hive|0|9|agg_l_04_sales_fact_1997|EXTERNAL_TABLE|null|null
10|1529335164|2|0|hive|0|10|agg_l_05_sales_fact_1997|EXTERNAL_TABLE|null|null
```

```
Database changed
mysql> select count(*) from TBLs;
+-----+
| count(*) |
+-----+
|      37 |
+-----+
1 row in set (0.08 sec)
```


Slide # 88

```
[root@sandbox-hdp ~]# sqoop help
Warning: /usr/hdp/2.6.5.0-292/accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
20/05/23 18:37:17 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6.2.6.5.0-292
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
```

Slide # 95

```
[root@sandbox-hdp ~]# sqoop-import --connect jdbc:mysql://sandbox-hdp.hortonworks.com/hive --username root --P --table TBLS --target-dir /user/anup/sqoop/tbls --fields-terminated-by '|' --num-mappers 8 --delete-target-dir
Warning: /usr/hdp/2.6.5.0-292/accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
20/05/23 18:41:39 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6.2.6.5.0-292
Enter password:
20/05/23 18:41:45 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
```

Slide # 97

```
[root@sandbox-hdp ~]# sqoop-import --connect jdbc:mysql://sandbox-hdp.hortonworks.com/hive --username root --P --query 'select * from TBLS where $CONDITIONS limit 75' --target-dir /user/anup/sqoop/tbls --fields-terminated-by '|' --num-mappers 8 --delete-target-dir --split-by TBL_ID
Warning: /usr/hdp/2.6.5.0-292/accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
20/05/23 18:53:12 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6.2.6.5.0-292
Enter password:
20/05/23 18:53:22 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
```

Slide # 98

```
[root@sandbox-hdp ~]# hdfs dfs -ls /user/anup/sqoop/tbls/
Found 9 items
-rw-r--r-- 1 root hdfs 0 2020-05-23 18:54 /user/anup/sqoop/tbls/_SUCCESS
-rw-r--r-- 1 root hdfs 347 2020-05-23 18:54 /user/anup/sqoop/tbls/part-m-00000
-rw-r--r-- 1 root hdfs 384 2020-05-23 18:54 /user/anup/sqoop/tbls/part-m-00001
-rw-r--r-- 1 root hdfs 402 2020-05-23 18:54 /user/anup/sqoop/tbls/part-m-00002
-rw-r--r-- 1 root hdfs 326 2020-05-23 18:54 /user/anup/sqoop/tbls/part-m-00003
-rw-r--r-- 1 root hdfs 254 2020-05-23 18:54 /user/anup/sqoop/tbls/part-m-00004
-rw-r--r-- 1 root hdfs 256 2020-05-23 18:54 /user/anup/sqoop/tbls/part-m-00005
-rw-r--r-- 1 root hdfs 253 2020-05-23 18:54 /user/anup/sqoop/tbls/part-m-00006
-rw-r--r-- 1 root hdfs 322 2020-05-23 18:54 /user/anup/sqoop/tbls/part-m-00007
[root@sandbox-hdp ~]# hdfs dfs -ls /user/anup/sqoop/tbls/part-m-00007
-rw-r--r-- 1 root hdfs 322 2020-05-23 18:54 /user/anup/sqoop/tbls/part-m-00007
[root@sandbox-hdp ~]# hdfs dfs -cat /user/anup/sqoop/tbls/part-m-00007
33|1529335216|2|0|hive|0|33|store|EXTERNAL_TABLE|null|null
34|1529335218|2|0|hive|0|34|store_ragged|EXTERNAL_TABLE|null|null
35|1529335221|2|0|hive|0|35|time_by_day|EXTERNAL_TABLE|null|null
36|1529335223|2|0|hive|0|36|warehouse|EXTERNAL_TABLE|null|null
37|1529335225|2|0|hive|0|37|warehouse_class|EXTERNAL_TABLE|null|null
```

Slide # 99

Database changed
mysql> select * from TBLS limit 10;

TBL_ID	CREATE_TIME	DB_ID	LAST_ACCESS_TIME	OWNER	RETENTION	SD_ID	TBL_NAME	TBL_TYPE	VIEW_EXPANDED_TEXT	VIEW_ORIGIN
1	1529335142	2	0	hive	0	1	sales_fact_1997	EXTERNAL_TABLE	NULL	NULL
2	1529335145	2	0	hive	0	2	sales_fact_1998	EXTERNAL_TABLE	NULL	NULL
3	1529335147	2	0	hive	0	3	sales_fact_dec_1998	EXTERNAL_TABLE	NULL	NULL
4	1529335150	2	0	hive	0	4	inventory_fact_1997	EXTERNAL_TABLE	NULL	NULL
5	1529335152	2	0	hive	0	5	inventory_fact_1998	EXTERNAL_TABLE	NULL	NULL
6	1529335154	2	0	hive	0	6	agg_pl_01_sales_fact_1997	EXTERNAL_TABLE	NULL	NULL
7	1529335157	2	0	hive	0	7	agg_ll_01_sales_fact_1997	EXTERNAL_TABLE	NULL	NULL