1.

Hive的几种常见的数据导入方式

这里介绍四种：

（1）、从本地文件系统中导入数据到Hive表；

（2）、从HDFS上导入数据到Hive表；

（3）、从别的表中查询出相应的数据并导入到Hive表中；

（4）、在创建表的时候通过从别的表中查询出相应的记录并插入到所创建的表中

1.从本地文件系统中导入数据到Hive表

1.1

[hadoop@h91 hive-0.9.0-bin]$ bin/hive

创建ha表

hive> create table ha(id int,name string)

> row format delimited

> fields terminated by '\t'

> stored as textfile;

[ROW FORMAT DELIMITED]关键字，是用来设置创建的表在加载数据的时候，支持的列分隔符。

[STORED AS file\_format]关键字是用来设置加载数据的数据类型,默认是TEXTFILE，如果文件数据是纯文本，就是使用 [STORED AS TEXTFILE]，然后从本地直接拷贝到HDFS上，hive直接可以识别数据。

1.2

操作系统中的文本

[hadoop@h91 ~]$ cat haha.txt

101 zs

102 ls

103 ww

1.3导入数据

hive> load data local inpath '/home/hadoop/haha.txt' into table ha;

hive> select \* from ha;

\*\*\*\*\*

和我们熟悉的关系型数据库不一样，Hive现在还不支持在insert语句里面直接给出一组记录的文字形式，也就是说，Hive并不支持INSERT INTO …. VALUES形式的语句。

\*\*\*\*\*

--------------------------------------------------

2.

从HDFS上导入数据到Hive表；

2.1

[hadoop@h91 hadoop-0.20.2-cdh3u5]$ bin/hadoop fs -mkdir abc

[hadoop@h91 ~]$ cat hehe.txt

1001 aa

1002 bb

1003 cc

[hadoop@h91 hadoop-0.20.2-cdh3u5]$ bin/hadoop fs -put /home/hadoop/hehe.txt abc/.

（上传到 hdfs中）

2.2

hive> create table he(id int,name string)

> row format delimited

> fields terminated by '\t'

> stored as textfile;

导入

hive> load data inpath '/user/hadoop/abc/hehe.txt' into table he;

---------------------------------------------------------

3.从别的表中查询出相应的数据并导入到Hive表中

3.1

hive> select \* from he;

OK

1001 aa

1002 bb

1003 cc

hive> create table heihei(id int,name string)

> row format delimited

> fields terminated by '\t'

> stored as textfile;

3.2

hive> insert into table heihei select \* from he;

或

hive> insert overwrite table heihei select \* from ha;

（insert overwrite 会覆盖数据）

--------------------------------------------------

4.在创建表的时候通过从别的表中查询出相应的记录并插入到所创建的表中

hive> create table ga as select \* from he;

================================================================

导出数据

（1）、导出到本地文件系统；

（2）、导出到HDFS中；

（3）、导出到Hive的另一个表中。

1.导出到本地文件系统；

hive> insert overwrite local directory '/home/hadoop/he1' select \* from he;

[hadoop@h91 ~]$ cd he1（he1为目录，目录下有000000\_0文件 ）

[hadoop@h91 he1]$ cat 000000\_0

（发现 列之间没有分割 ）

可以下面的方式增加分割

hive> insert overwrite local directory '/home/hadoop/he1' select id,concat('\t',name) from he;

\*\*\*\*\*\*

和导入数据到Hive不一样，不能用insert into来将数据导出

\*\*\*\*\*\*

---------------------------------------------------------

2.导出到HDFS中。

hive> insert overwrite directory '/user/hadoop/abc' select \* from he;

（/user/hadoop/abc 为hdfs下目录）

[hadoop@h91 hadoop-0.20.2-cdh3u5]$ bin/hadoop fs -ls abc

[hadoop@h91 hadoop-0.20.2-cdh3u5]$ bin/hadoop fs -cat abc/000000\_0

-------------------------------------------------------------

3.导出到Hive的另一个表中

hive> insert into table he12 select \* from he;