Hive Homework

作业要求**:**

使用附件中包含的 Stock.csv 文件，结合 Hive 然后完成以下要求:

(1) Check how many rows are inserted.

(2) Select all the stock info in 2008.

(3) Check if table is internal or external. If internal, then change it to external. If external, then change it to internal.

(4) Find the maximum historical open price.

步骤提示：

1. Prepare File

将.csv 文件导入 Hadoop 集群服务器。

2. Create a New Database

3. Build Table

Hive 中存在 Date 格式，形式为 YYYY-MM-DD，然而我们的文件中的日期格式为 MM/DD/YYYY。

Here we need to create a temporary table deal with the date format (from MM/DD/YYYY to YYYY-MM-DD), the way to change format is by using:

TO_DATE(from_unixtime(UNIX_TIMESTAMP(stockdate,'MM/dd/yy')))

4. Check how many rows are inserted

By using the following command:

SELECT COUNT(*) FROM temp;

5. Create a Partitioned Table to select all the stock info in 2008

- PARTITIONED BY(year STRING)
- 新建一个 PARTITION (year = "2008")，导入满足 stockDate > '2007-12-31' AND stockDate < '2009-01-01’ 的数据即可。

6. Build internal and external table separately and check the status

Use the following command to check the table status:

DESCRIBE FORMATTED table_name;

- Switch a table from internal to external:

  ALTER TABLE table_name SET TBLPROPERTIES('EXTERNAL'='TRUE');

- Switch a table from external to internal:

  ALTER TABLE table_name SET TBLPROPERTIES('EXTERNAL'='FALSE');

7. Find the maximum historical open price.

Method 1: 按 open price 降序排序找出第一个数值。

Method 2: CAST(MAX())命令去直接寻找最大值。

Method 3: …