cd apache-drill-<version>

z

bin/drill-embedded

出现当天的消息和0: jdbc:drill:zk=local> 提示。

!quit

json:

SELECT \* FROM cp.`employee.json` LIMIT 3;

dfs:

//SELECT \* FROM dfs.`<path-to-installation>/apache-drill-<version>/sample-data/region.parquet`;

SELECT \* FROM dfs.`Users/drilluser/apache-drill/sample-data/region.parquet`;

0: jdbc:drill:zk=local> select state, city, count(\*) totalreviews

from dfs.`/<path-to-yelp-dataset>/yelp/yelp\_academic\_dataset\_business.json`

group by state, city order by count(\*) desc limit 10;

0: jdbc:drill:zk=local> alter system set `store.json.all\_text\_mode` = true;

0: jdbc:drill:zk=local> select attributes from dfs.`/<path-to-yelp-dataset>/yelp/yelp\_academic\_dataset\_business.json` limit 10;

$ ssh mapr@localhost -p 2222

Password:

Last login: Mon Sep 15 13:46:08 2014 from 10.250.0.28

Welcome to your Mapr Demo virtual machine.

[mapr@maprdemo ~]$ sqlline

apache drill 1.1.0

"Does your data know the Drill?"

0: jdbc:drill:>

alter session set `planner.enable\_decimal\_data\_type`=true;

0: jdbc:drill:> show databases;

0: jdbc:drill:> use hive.`default`;

0: jdbc:drill:> describe orders;

0: jdbc:drill:> select \* from orders limit 5;

use maprdb;

Drill支持对多种文件格式的查询，包括文本，CSV，Parquet和JSON。

0: jdbc:drill:> use dfs.clicks;

"location": "/mapr/demo.mapr.com/data/nested"

0: jdbc:drill:> select \* from `clicks/clicks.json` limit 2;

唯一的特殊要求是使用反引号将文件路径括起来。只要文件路径包含Drill保留字或字符，就必须这样做。

0: jdbc:drill:> use dfs.logs;

0: jdbc:drill:> select \* from logs limit 2;

dfs.clicks工作空间位置属性指向包含日志目录的目录，这使得此查询的FROM子句引用非常简单。您不必引用文件系统上的完整目录路径。

0: jdbc:drill:> use hive.`default`;

0: jdbc:drill:> select `month`, sum(order\_total)

from orders group by `month` order by 2 desc;

请注意，仅因为“月”是SQL中的保留字，才需要对“月”列进行反勾。

0: jdbc:drill:> use dfs.clicks;

0: jdbc:drill:> select t.user\_info.device, count(\*) from `clicks/clicks.json` t

group by t.user\_info.device

having count(\*) > 1000;

0: jdbc:drill:> select t.trans\_id transaction, t.user\_info.cust\_id customer from `clicks/clicks.campaign.json` t

union all

select u.trans\_id, u.user\_info.cust\_id from `clicks/clicks.json` u limit 5;

0: jdbc:drill:> select cast(row\_key as int) as cust\_id, cast(t.personal.name as varchar(20)) as name,

cast(t.personal.gender as varchar(10)) as gender, cast(t.personal.age as varchar(10)) as age,

cast(t.address.state as varchar(4)) as state, cast(t.loyalty.agg\_rev as dec(7,2)) as agg\_rev,

cast(t.loyalty.membership as varchar(20)) as membership

from customers t limit 5;

表中的每一列都需要CAST函数。此函数以可读的整数和字符串形式返回MapR-DB / HBase二进制数据。另外，您可以使用CONVERT\_TO / CONVERT\_FROM函数对字符串列进行解码。在大多数情况下，CONVERT\_TO / CONVERT\_FROM比CAST更有效。仅使用CONVERT\_TO将二进制类型转换为VARCHAR以外的任何类型。

row\_key列用作表的主键（在这种情况下为客户ID）。

表别名t是必需的；否则，列族名称将被解析为表名称，并且查询将返回错误。

0: jdbc:drill:> select cast(row\_key as int), regexp\_replace(cast(t.address.state as varchar(10)),'"','')

from customers t limit 1;

| EXPR$0 | EXPR$1 |

0: jdbc:drill:> use dfs.views;

0: jdbc:drill:> create or replace view custview as select cast(row\_key as int) as cust\_id,

cast(t.personal.name as varchar(20)) as name,

cast(t.personal.gender as varchar(10)) as gender,

cast(t.personal.age as varchar(10)) as age,

cast(t.address.state as varchar(4)) as state,

cast(t.loyalty.agg\_rev as dec(7,2)) as agg\_rev,

cast(t.loyalty.membership as varchar(20)) as membership

from maprdb.customers t;

https://drill.apache.org/docs/lesson-2-run-queries-with-ansi-sql/