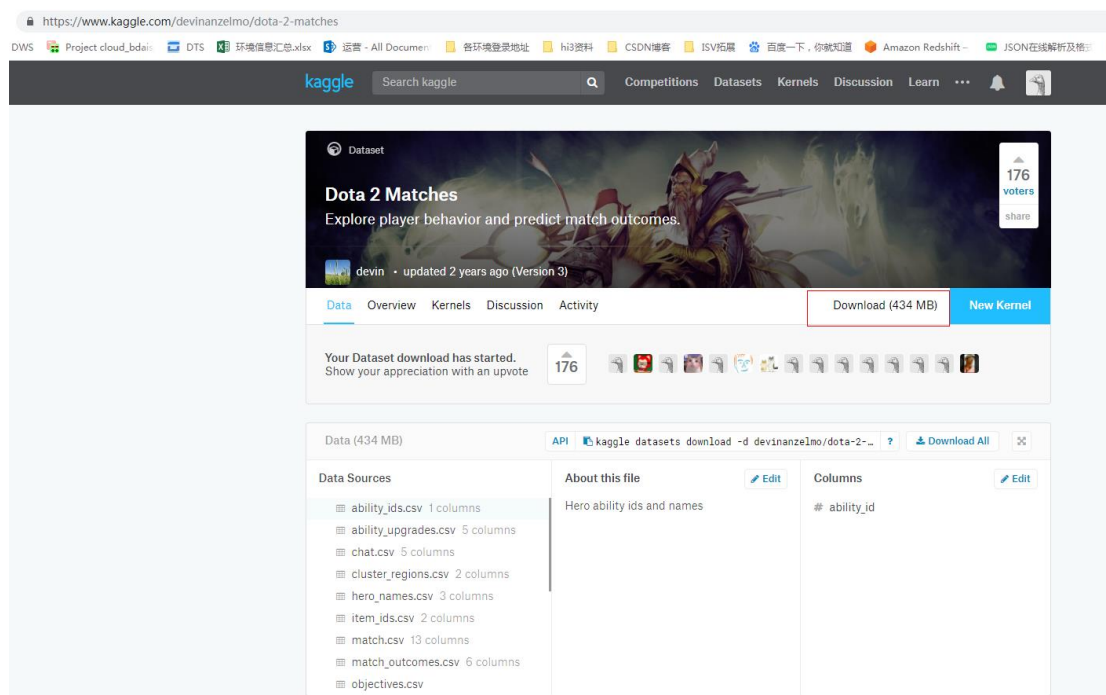


数据仓库服务DWS使用---数据分析

一．主要完成作业二

作业2：根据下面提供的资料与提示，进行Dota2比赛分析

(1) 下载数据 (<https://www.kaggle.com/devinanzelmo/dota-2-matches>)



(2) (https://support.huaweicloud.com/devg-dws/import_from_obs_0001.html) 根据导入指导，将数据导入以创建好的数据仓库

(3) 用户可根据自己的爱好，对dota2的比赛数据进行分析，步骤可以参照第一天的操作文档。不强制要求将所有的表上传并分析，用户可根据自己的实际情况，上传几张表即可。

下面是样例,请参考：

1、创建OBS桶：

区域

华北-北京一

不同区域的云产品之间内网互不相通，请选择靠近您业务的区域，可减少网络时延，提高访问速度。桶创建成功后不支持变更区域，请谨慎选择。

桶名称

obs-demo

命名规则

- 需全局唯一，不能与已有的任何桶名称重复。
- 长度范围为3到63个字符，支持小写字母、数字、中划线（-）、英文句号（.）。
- 禁止两个英文句号（.）或英文句号（.）和中划线（-）相邻，禁止以英文句号（.）和中划线（-）开头或结尾。
- 禁止使用IP地址。
- 如果名称中包含英文句号（.），访问桶或对象时可能会进行安全证书校验。

存储类别

标准存储

低频访问存储

归档存储

适用于有大量热点文件或小文件，且需要频繁访问（平均一个月多次）并快速获取数据的业务场景。

桶策略

私有

公共读

公共读写

任何用户都可以对桶内对象进行读/写/删除操作。

多AZ

所选区域不支持配置多AZ。

2、导入数据到OBS桶中：

对象

已删除对象

碎片

对象是数据存储的基本单位，在OBS中文件和文件夹都是对象。您可以上传任何类型（文本、图片、视频等）的文件，并在桶中对这些文件进行管理。 [了解更多](#)

上传文件

新建文件夹

複製

删除

输入对象名前缀搜索

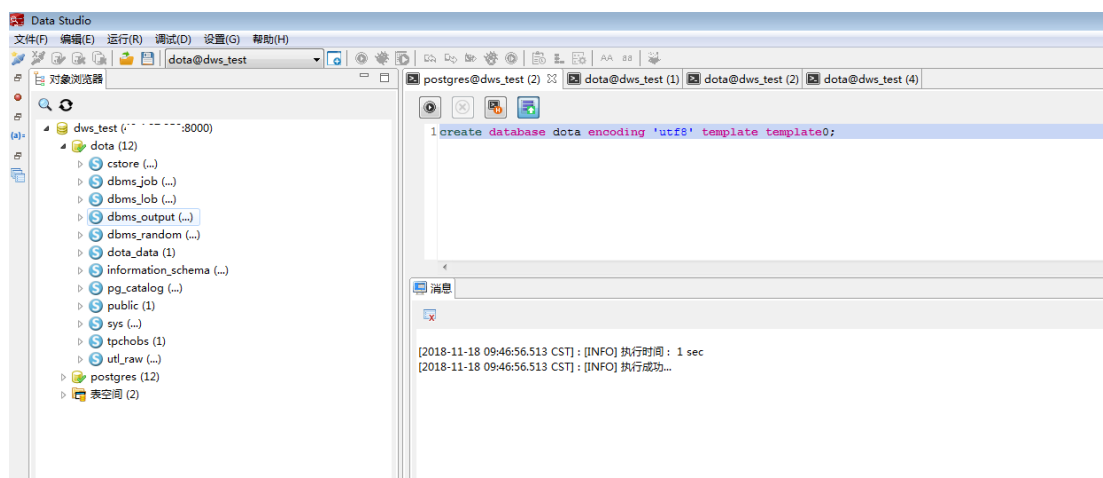
Q

C

<input type="checkbox"/>	名称	存储类别	大小	加密状态	恢复状态	最后修改时间	操作
<input type="checkbox"/>	ability.csv	标准存储	18.807 KB	未加密	-	2018/11/18 10:31:21 GMT+08:00	下载 分享 更多
<input type="checkbox"/>	ability_ids.csv	标准存储	18.807 KB	未加密	-	2018/11/18 10:30:39 GMT+08:00	下载 分享 更多

3、在DWS集群中创建数据库：

```
create database dota encoding 'utf8' template template0;
```



4、数据库中创建表：

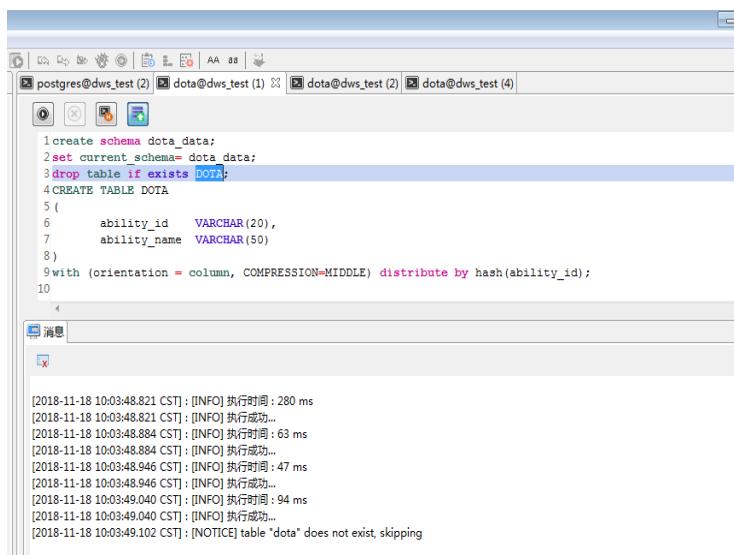
```
create schema traffic_data;

set current_schema= traffic_data;

drop table if exists DOTA;

CREATE TABLE DOTA
(
    ability_id VARCHAR(20),
    ability_name VARCHAR(50)
)

with (orientation = column, COMPRESSION=MIDDLE) distribute by
hash(ability_id );
```



```
1 create schema dota_data;
2 set current_schema= dota_data;
3 drop table if exists DOTA;
4 CREATE TABLE DOTA
5 (
6     ability_id VARCHAR(20),
7     ability_name VARCHAR(50)
8 )
9 with (orientation = column, COMPRESSION=MIDDLE) distribute by hash(ability_id);
10
```

消息

```
[2018-11-18 10:03:48.821 CST]: [INFO] 执行时间: 280 ms
[2018-11-18 10:03:48.821 CST]: [INFO] 执行成功...
[2018-11-18 10:03:48.884 CST]: [INFO] 执行时间: 63 ms
[2018-11-18 10:03:48.884 CST]: [INFO] 执行成功...
[2018-11-18 10:03:48.946 CST]: [INFO] 执行时间: 47 ms
[2018-11-18 10:03:48.946 CST]: [INFO] 执行成功...
[2018-11-18 10:03:49.040 CST]: [INFO] 执行时间: 94 ms
[2018-11-18 10:03:49.040 CST]: [INFO] 执行成功...
[2018-11-18 10:03:49.102 CST]: [NOTICE] table "dota" does not exist, skipping
```

5、创建外表;

```
create schema tpchobs;
```

```
set current_schema = 'tpchobs';

drop FOREIGN table if exists DOTA_OBS;

CREATE FOREIGN TABLE DOTA_OBS
(
    like dota_data.DOTA
)SERVER gsmpp_server
OPTIONS
(
    encoding 'utf8',
    location 'obs://obs-demo/ability_ids/ability.csv',
    format 'text',
    delimiter ',',
    access_key 'AAAAAAAAAAAAAAAAAAAA',
    secret_access_key 'BBBBBBBBBBBBBBBBBB',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on'
);
```

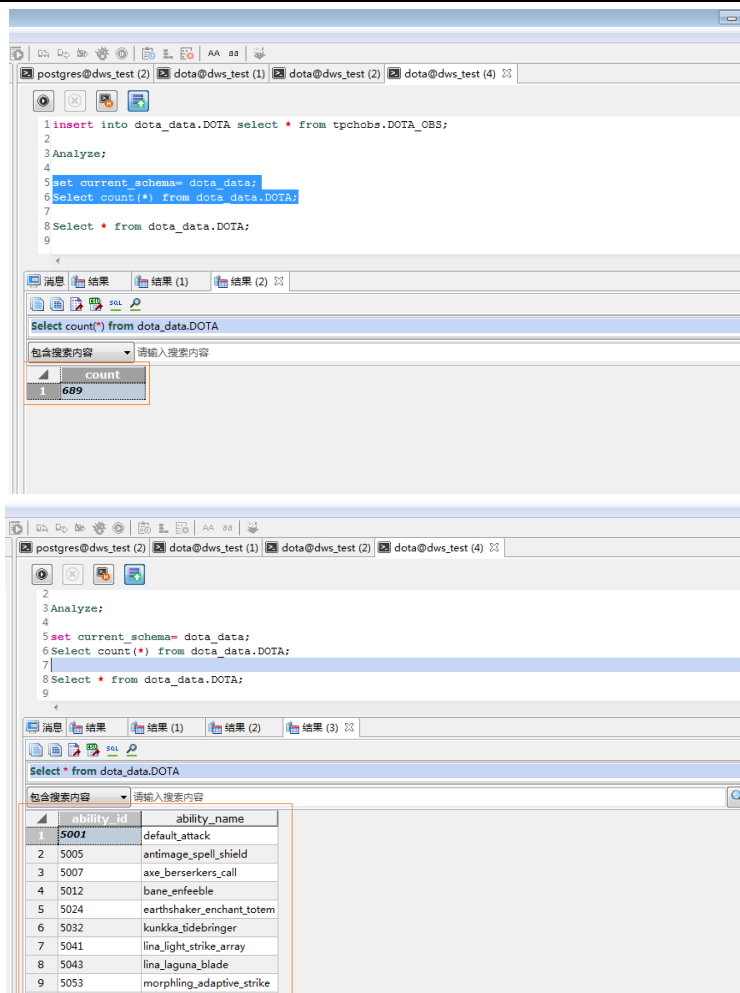
6、查询数据样例：

```
Analyze;

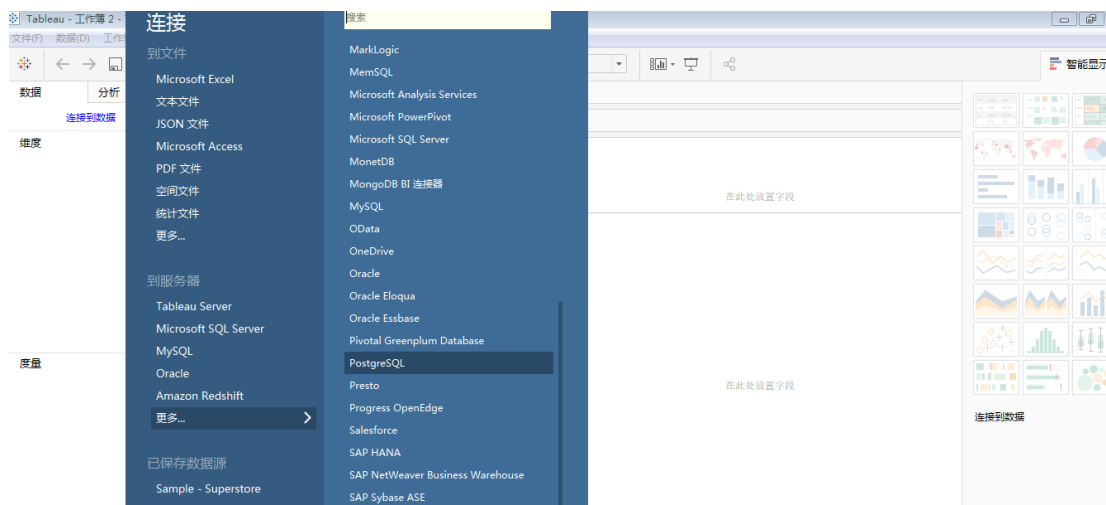
set current_schema = dota_data;
```

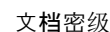
select count(*) from dota_data.DOTA;

select * from dota_data.DOTA;



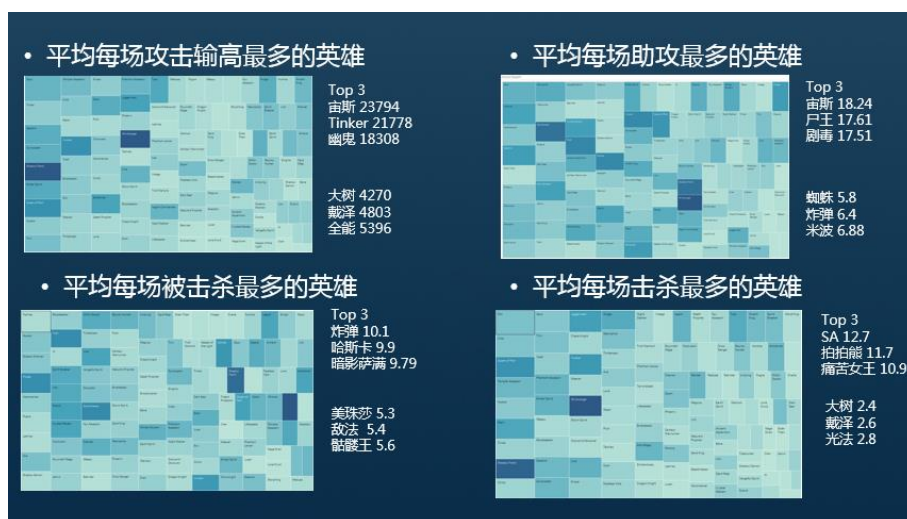
7、使用Tableau工具连接DWS数据仓库：





8、使用Tableau分析数据：

第 6 页, 共 8 页

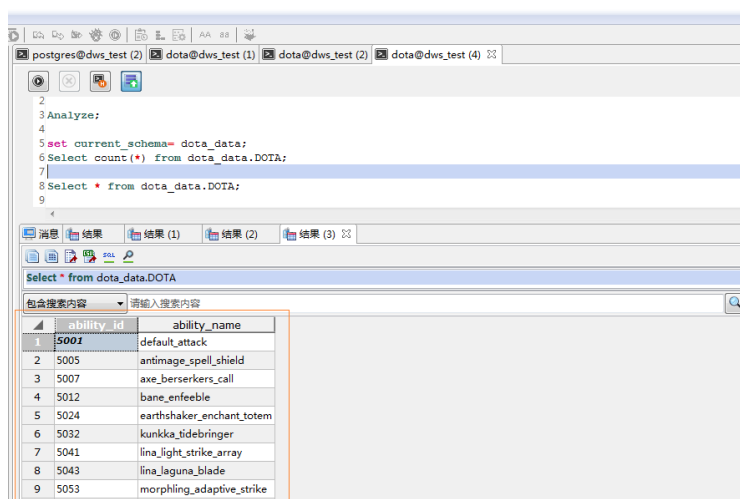


注：

使用Tableau连接DWS集群时要保证网络的连通性，否则会发生连接断开的现象。

二．打卡任务

完成作业二后，可以对分析结果进行截图。可以对DataStudio的结果截图，类似于下图：



也可对Tableau的分析结果截图，类似于下图：

• 平均每场攻击输出最高的英雄



Top 3
宙斯 23794
Tinker 21778
幽鬼 18308

大树 4270
戴泽 4803
全能 5396

• 平均每场助攻最多的英雄



Top 3
宙斯 18.24
尸王 17.61
剧毒 17.51

蜘蛛 5.8
炸弹 6.4
米波 6.88

• 平均每场被击杀最多的英雄



Top 3
炸弹 10.1
哈斯卡 9.9
暗影萨满 9.79

美妹莎 5.3
敌法 5.4
骷髅王 5.6

• 平均每场击杀最多的英雄



Top 3
SA 12.7
拍拍熊 11.7
痛苦女王 10.9

大树 2.4
戴泽 2.6
光法 2.8