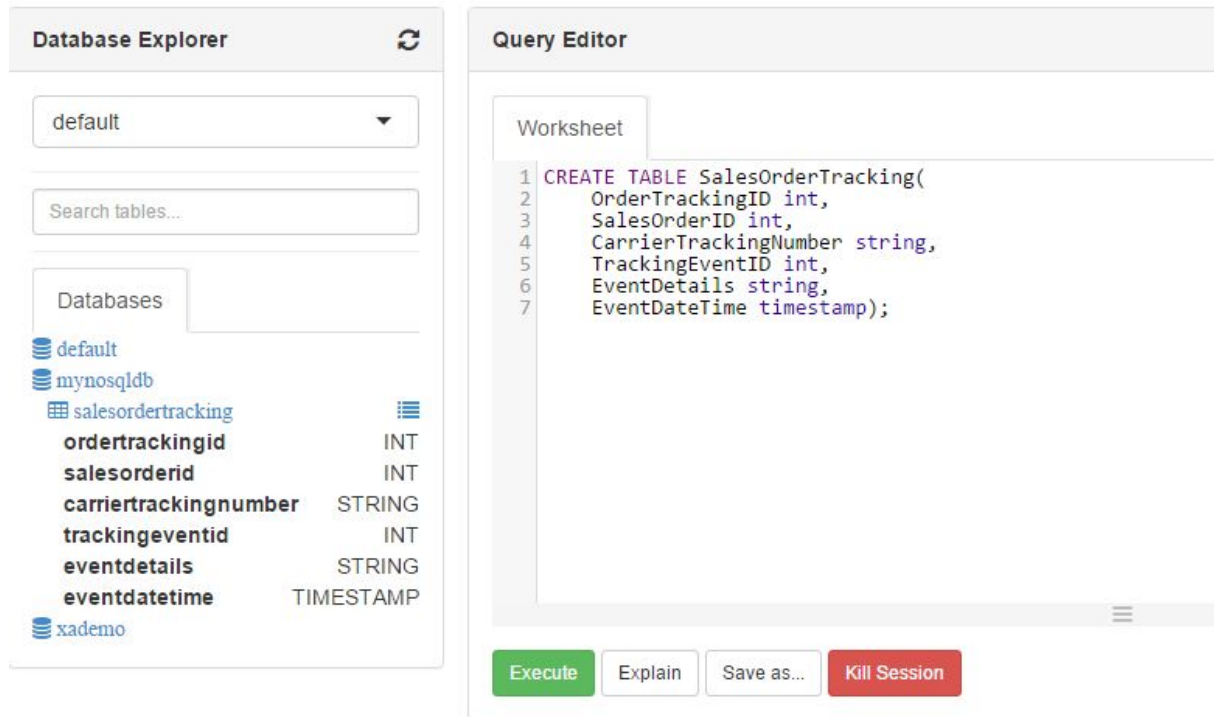


Bu bölümde Apache Hive ile tablo ve veritabanı işlemlerinin nasıl yapıldığını inceleyeceğiz



Hive Veritabanı Nasıl oluşturulur ?

Hive ile oluşturduğumuz tablolar bir veritabanı altında görünür. Bu nedenle öncelikle bir veritabanı oluşturmamız gerekir. Veritabanı 2 şekilde oluşturulur

```
hive> CREATE DATABASE [IF NOT EXISTS] my_db;
```

yada

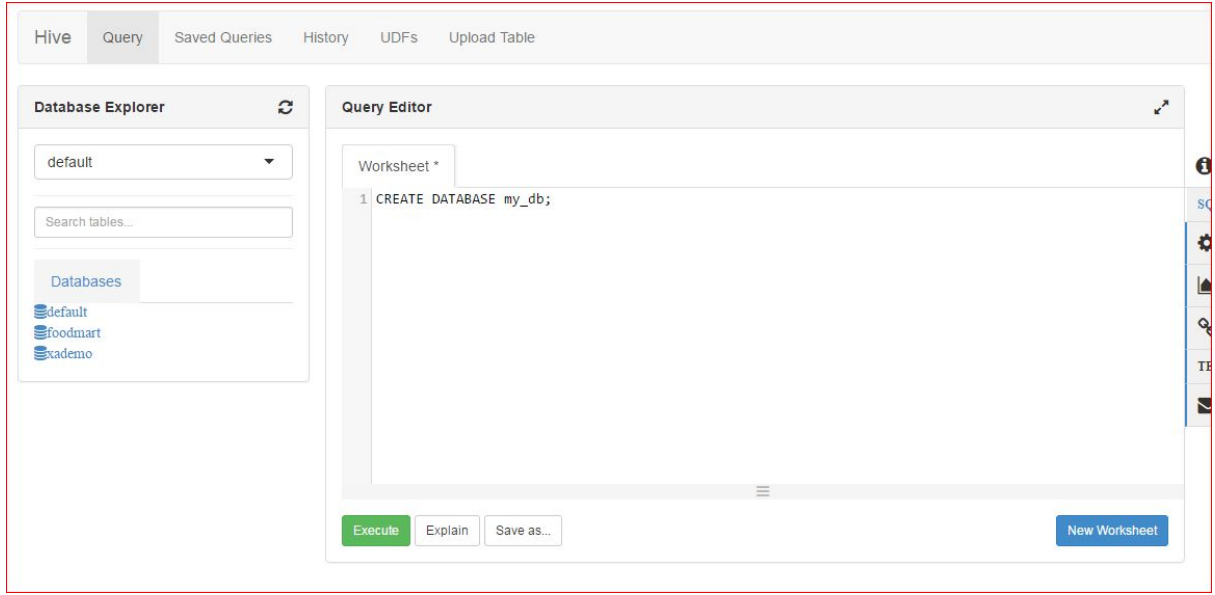
```
hive> CREATE SCHEMA my_db;
```

test için

```
hive> SHOW DATABASES;
```

```
default  
my_db
```

Cloud bir sistem üzerinden rahatlıkla veritabanı oluşturabilirsiniz



hive şema

Hive Tablo Nasıl oluşturulur ?

Hive ile tablo oluşturarak aslında HDFS üzerinde var olan bir veriyi bir şema bilgisi tanımlamış oluruz. Veriler yine hdfs sistemine eklenir yada silinir, fakat hive şema bilgisini saklayarak SQL ile kolayca sorgulama yapmamızı sağlar

HDFS üzerinde alttaki gibi bir veri kümemiz olsun

1,US,5,<https://www.google.com.tr/#q=apache+pig,apache,pig>
2,US,25,<https://www.google.com.tr/#q=apache+hive,apache,hive>
3,TR,20,<https://www.google.com.tr/#q=apache+hadoop,apache,hadoop>
4,EN,10,<https://www.google.com.tr/#q=apache+oozie,apache,oozie>
5,PL,16,<https://www.google.com.tr/#q=apache%20flume,apache,flume>
6,US,24,<https://www.google.com.tr/#q=apache+spark,,spark>
7,US,36,<https://www.google.com.tr/#q=apache+kafka,,kafka>
8,EN,48,<https://www.google.com.tr/#q=+storm,apache storm in hadoop ,storm>

Bu veri üzerinden şu şekilde tablo oluşturabiliriz

```
CREATE EXTERNAL TABLE IF NOT EXISTS weblog (  
  id int,  
  country String,  
  duration int,  
  url String,  
  keyword1 String,  
  keyword2 String)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ','  
LINES TERMINATED BY '\n'  
STORED AS TEXTFILE  
LOCATION '/user/hue/weblog';
```

Bu script ile Hive tablosu oluşturabiliriz. Burada dikkat etmemiz gereken konu **EXTERNAL** anahtar kelimesini eklememiz. Eğer eklemeyizse tabloyu sildiğimiz zaman HDFS altındaki loglarda silinir. Bu yüzden geçici(temporary) bir tablo kullanmıyorsak **EXTERNAL** anahtar kelimesini eklemeliyiz.

Daha sonra işlemleri test etmek için basit SQL sorguları yazalım

Query Editor



```
1 select * from weblog;
```

Execute

Save as...

Explain

or create a

New query

Results Query Log Columns

	weblog.id	weblog.country	weblog.duration	weblog.url	weblog.keyword1	weblog.keywc
0	1	US	5	https://www.google.com.tr/#q=apache+pig	apache	pig
1	2	US	25	https://www.google.com.tr/#q=apache+hive	apache	hive
2	3	TR	20	https://www.google.com.tr/#q=apache+hadoop	apache	hadoop
3	4	EN	10	https://www.google.com.tr/#q=apache+oozie	apache	oozie
4	5	PL	16	https://www.google.com.tr/#q=apache%20flume	apache	flume
5	6	US	24	https://www.google.com.tr/#q=apache+spark		spark
6	7	US	36	https://www.google.com.tr/#q=apache+kafka		kafka
7	8	EN	48	https://www.google.com.tr/#q=+storm	apache storm in hadoop	storm

hive-sorgu-sonucu[/caption]

Tablo Yapısındaki Değişiklikler(Alter)

Hive içerisindeki tablo yapılarında değişiklik için alttaki script'leri kullanabiliriz

ALTER TABLE weblog RENAME TO new_weblog (tablo ismi değiştirme)

ALTER TABLE weblog ADD COLUMNS (access_count int) (yeni kolon ekleme)

ALTER TABLE weblog DROP COLUMN id (kolon silme)

ALTER TABLE weblog CHANGE id new_id (kolon adı değiştirme)

ALTER TABLE weblog CHANGE id String (kolon tipi değiştirme)

ALTER TABLE weblog REPLACE COLUMNS(id int,url STRING,duration int); (burada belirtilmeyen kolonlar silinir)

Hive Partition nedir?

HDFS içerisinde veriler genellikle klasörler altında atılır.Örnek verirsek browser üzerinden logların toplandığı bir sistemde günlük klasörler vardır. 1 Ocak tarihinde gelen log'lar 2016_01_01 klasörünün altına, 2 Ocak tarihinden gelen loglar 2016_01_02 altına atılabilir.Bunun en önemli nedeni ise sorgulama performansını artırmaktır

File Browser

Search for file name							Rename	Move	Copy	Change Permissions	Download	Delete	New	Upload
Home / user / browserlog							Trash							
Type	Name	Size	User	Group	Permissions	Date								
	.		hue	hdfs	drwxr-xr-x	November 29, 2016 11:								
	..		hdfs	hdfs	drwxr-xr-x	November 29, 2016 11:								
	2016_01_01		hue	hdfs	drwxr-xr-x	November 29, 2016 11:								
	2016_01_02		hue	hdfs	drwxr-xr-x	November 29, 2016 11:								
	2016_01_03		hue	hdfs	drwxr-xr-x	November 29, 2016 11:								
	2016_01_04		hue	hdfs	drwxr-xr-x	November 29, 2016 11:								

hdfs-logs[/caption]

Bu noktada alttaki gibi bir hive sorgusu yapmak isteyelim

```
select * from weblog where log_date = '2016_01_01'
```

Böyle bir sorgu yapabilmek için **log_date** olarak isimlendirilmiş bir partition(bölümlendirme) eklememiz gerekmektedir

Partition şu şekilde eklenir

```
ALTER TABLE weblog ADD PARTITION (log_date='2016_01_01') location  
'/user/browserlog/2016_01_01';
```

Adım adım partition oluşturalım

Adım 1) Tablo oluşturulur

```
CREATE EXTERNAL TABLE IF NOT EXISTS weblog1P (  
    id int,  
    country String,  
    duration int,  
    url String,  
    keyword1 String,  
    keyword2 String ) partitioned BY (log_date String)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ','  
LINES TERMINATED BY '\n'  
STORED AS TEXTFILE  
;
```

Adım 2) Partition eklenir

```
ALTER TABLE weblog1P ADD PARTITION (log_date='2016_01_02') location '/user/hive/02';
```

Adım 3) Sorgu ile sonuçlar alınır

```
select * from weblog1P where log_date = '2016_01_02'
```