

Apache Hive Nedir?

Büyük verileri Hadoop üzerinde paralel olarak işlemek için MapReduce programı yazmamız gerekir. MapReduce programı Java, [Apache Pig](#) veya Apache Hive ile geliştirilebilir

```
-- import the file as lines

CREATE EXTERNAL TABLE lines(line string)
LOAD DATA INPATH 'books' OVERWRITE INTO TABLE lines;

-- create a virtual view that splits the lines

SELECT word, count(*) FROM lines
      LATERAL VIEW explode(split(text, ' ')) lTable as word
GROUP BY word;
```

hive-ornek

Bu bölümde Hadoop MapReduce geliştirme yöntemlerinden Apache Hive'ı inceleyeceğiz. Apache Hive, SQL biçiminde sorgular geliştirmemizi sağlayan Apache Hadoop tabanlı veri ambarı(datawarehouse) kütüphanesidir.

Hangi durumlarda Apache Hive tercih edilmeli ?

Büyük veriler üzerinde basit analizler yapmak için Java MapReduce yada Apache Pig yazmak yerine hızlı bir şekilde Apache Hive tabanlı SQL sorguları geliştirilebilir. Mesela 1 milyar log üzerinde şöyle bir analiz yapalım

- Sitemizi hangi ülkeden kaç tane tekil kullanıcı ziyaret etmiş

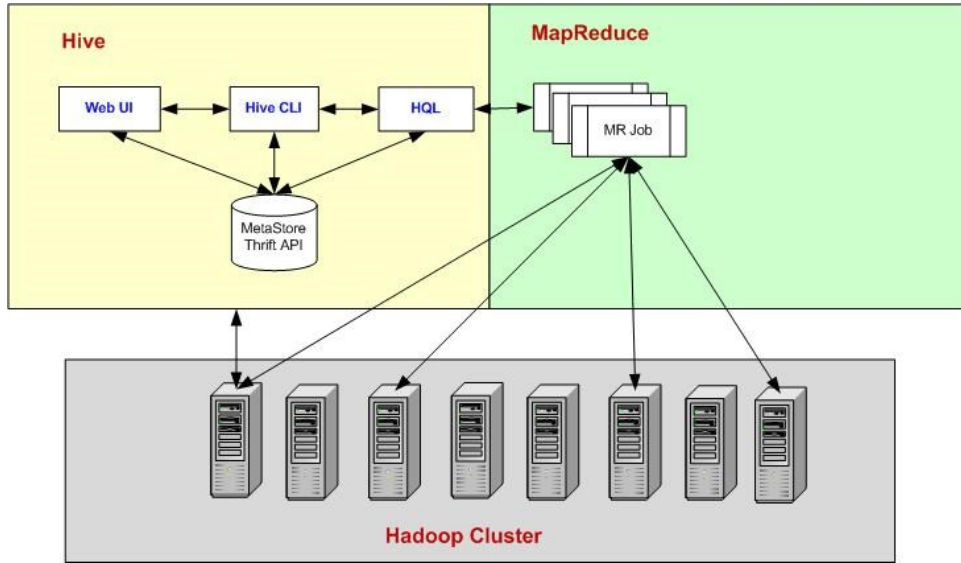
Bu durumda şöyle bir hive kodu işimizi görecektir

```
select country, count(distinct user_id) from data where log_date >= '2015-04-17'
and log_date <= '2015-04-18' group by country;
```

Ayrıca benzer sorguları sürekli big data developer 'a yaptırmak yerine iş birimine yaptırabiliriz.

Hive Mimarisi

Web arayüzünden yada Hive client üzerinden çalıştırdığımız HQL(Hive Query Language) sorguları daha önce girilmiş şema bilgileri kullanılarak arka planda MapReduce job'larına çevrilir.Hive kurulduğu zaman beraberinde Thrift server ile gelir.Bu servisin amacı ise Hive üzerinde belirli portlara erişmeyi sağlamaktır



Hive MetaStore Nedir?

Şema bilgilerinin ve bu şema altında yer alan tablolar,kolonlar,kolon tiplerinin saklandığı bölümdür

	◆	Name	◆	Type
0		id		int
1		lastname		string
2		firstname		string
3		dob		date
4		newsletter		boolean
5		contacts		map<string,string>
6		orders		array<string>
7		site		string

hive metastore