

Apache Pig Nedir ?

MapReduce ile büyük verileri dağıtık sistemlerde analiz edebiliriz . MapReduce programı yazmak için alttaki kodlama yöntemlerinden bir tanesini izlememiz gerekir

- Java MapReduce
- Apache Pig
- Apache Hive

Pig-latin example

- Query : Get the list of pages visited by users whose age is between 20 and 25 years.

```
users = load 'users' as (name, age);
```

```
users_18_to_25 = filter users by age > 20 and age <= 25;
```

```
page_views = load 'pages' as (user, url);
```

```
page_views_u18_to_25 = join users_18_to_25 by name,  
page_views by user;
```

pig-nedir

Aslında Apache Pig ve Apache Hive yazmış olduğumuz kodları arka planda Java MapReduce kodlarına çevirir.

Pig ve Hive kullanmamızın en önemli sebebi Java MapReduce ile kod geliştirmek zahmetlidir , bu yüzden Pig ve Hive ile daha hızlı MapReduce uygulamaları geliştirebiliriz

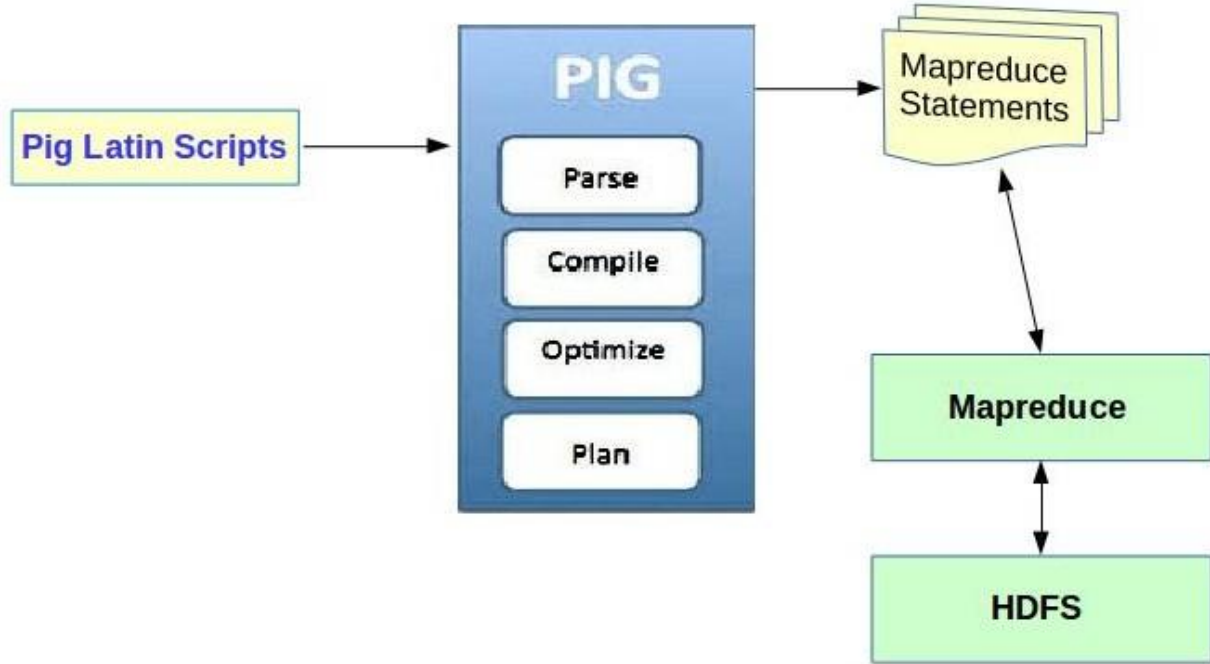
Apache Pig Mimarisi

Geliştirmiş olduğumuz Pig kodları arka planda şu adımları izleyerek MapReduce kodlarına çevrilir

Parse : Bu adımda geliştirmiş olduğumuz Pig kodları syntax olarak kontrol edilir.Yanlış bir yazım varsa hata verilir

Compile : Bu adımda yazmış olduğumuz kodlar MapReduce job larına çevrilir

Optimize ve Plan : Bu adımda yazmış olduğumuz kodları Apache Pig bizim için optimize eder. Çünkü geliştirmiş olduğumuz kodların daha performanslı çalışması için Pig çeşitli optimizasyonlar yapar



pig-mimari

Altteki şekilde Apache Pig ile yazılmış örnek bir kod bloğu bulunmaktadır. Burada örnek bir word-count uygulaması yapılmıştır. Benzer örneği Java MapReduce ile yazmak istersek yüzlerce satır kod geliştirmemiz gerekir

```

1 Data = LOAD '/temp/loaded_data' USING PigStorage() AS
2 (
3     id: chararray,
4     keyword: chararray
5 );
6
7 FilteredData = FILTER DATA BY keyword != '' and keyword IS NOT NULL;
8
9
10 GroupedDataByKeyword = GROUP FilteredData BY (keyword);
11
12 WordCount = FOREACH GroupedDataByKeyword {
13     GENERATE
14         group as groupedKeyword,
15         COUNT(keyword) as countOfKeyword:long;
16 }

```

pig-örnek-kod

Apache Pig'in avantajları

- Öğrenmesi ve geliştirmesi kolaydır
- Büyük verileri üzerinde kolaylıkla analizler yapılabilir
- Geliştirmiş olduğumuz MapReduce job'larını optimize eder
- Veri üzerinde analizler yapabileceğimiz metodlar sunar (filter, group, join, ordering vs..)
- İhtiyaç halinde java,javascript ve python ile kütüphaneler yazıp Apache pig içerisinde kullanabiliriz (udf)
- Hadoop içerisindeki ayarlamalar Apache Pig ile yapılabilir (mapreduce.reduce.memory.mb ..)