

Apache Pig ile Hadoop Programlama

Apache Pig , büyük verileri işlememizi sağlayan açık kaynak kodlu bir kütüphanedir . Geliştirilen kodlar arka planda Hadoop MapReduce job larına çevrilir

- 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;
```

Apache Hive ile Hadoop Programlama

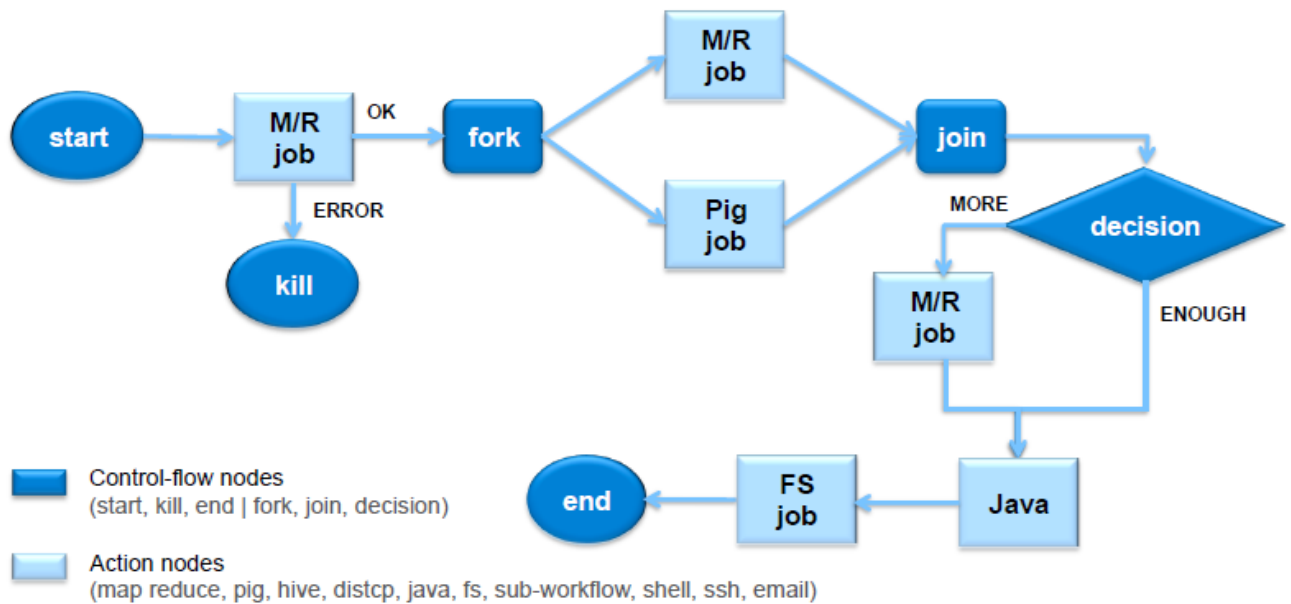
Apache Hive büyük verileri işlememizi sağlayan SQL yapısına benzer açık kaynak kodlu kütüphanedir . Geliştirilen kodlar arka planda Hadoop MapReduce job larına çevrilir

■ Sample Hive query:

```
SELECT stock.product, SUM(orders.purchases)
FROM stock JOIN orders
ON (stock.id = orders.stock_id)
WHERE orders.quarter = 'Q1'
GROUP BY stock.product;
```

Apache Oozie ile zamanlanmış büyük veri uygulamaları

Apache Oozie büyük veri işlemlerini zamanlamak ve sıralandırmak için kullanılır



Mongo

Mongo DB verileri JSON tabanlı kaydeden açık kaynak kodlu ilişkisel olmayan bir veritabanıdır

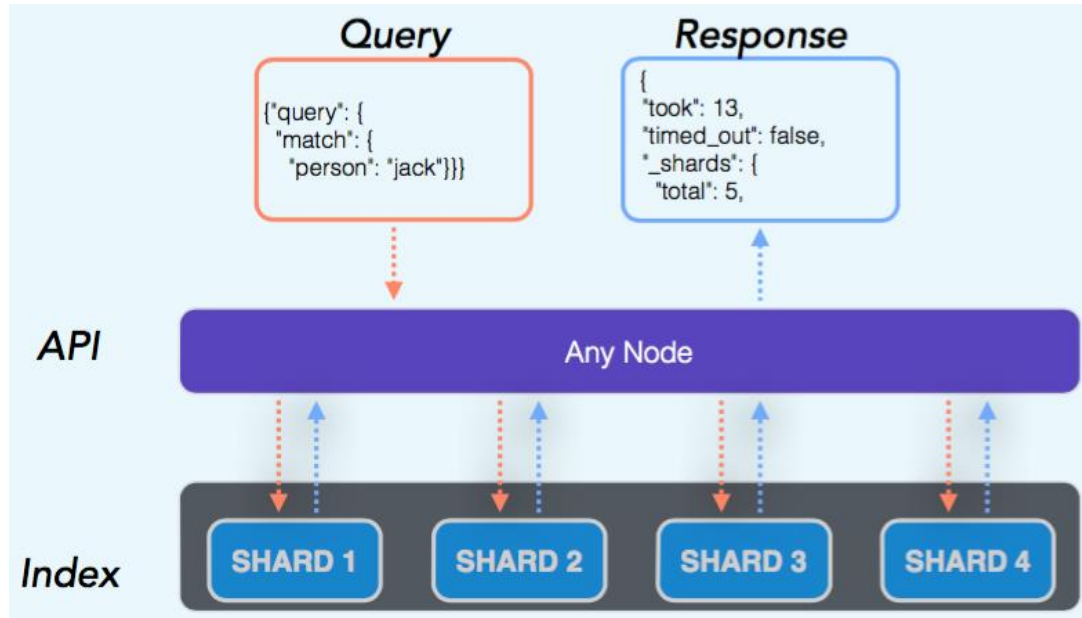
```
db.users.update(           ← collection
  { age: { $gt: 18 } },    ← update criteria
  { $set: { status: "A" } }, ← update action
  { multi: true }          ← update option
)
```

```
db.users.remove(           ← collection
  { status: "D" }           ← remove criteria
)
```

```
db.users.insert (          ← collection
  {
    name: "sue",            ← field: value
    age: 26,                 ← field: value
    status: "A"             ← field: value
  }                          } document
)
```

ElasticSearch ile metin arama uygulamaları

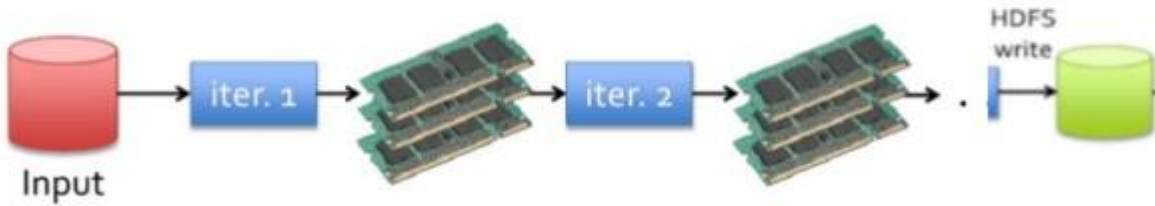
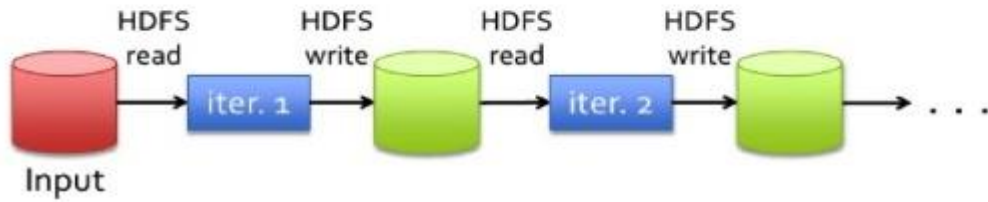
ElasticSearch açık kaynak kodlu java tabanlı arama işlemlerinde kullanılan bir kütüphanedir



Spark Programlama

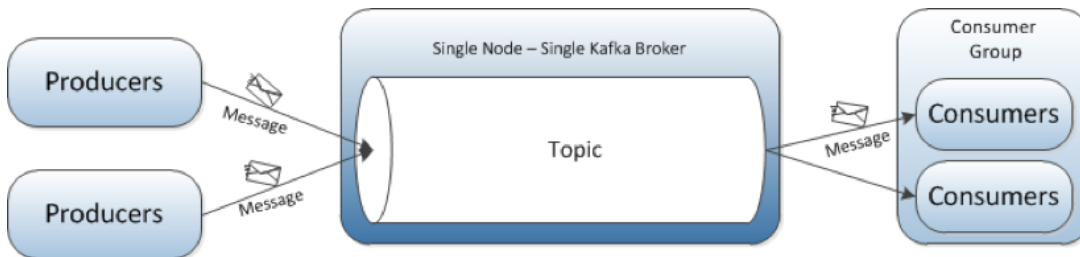
Spark büyük verileri işlemek için kullanılan bir kütüphanedir . Hadoop ' a göre daha az disk I/O işlemleri yapıldığı için Hadoop ' a göre daha hızlıdır

HADOOP MAPREDUCE VS SPARK



Apache Kafka ile mesajlasma sistemleri

Apache Kafka büyük veri akışını düşük bir gecikme zamanı ile sağlayan açık kaynak kodlu bir kütüphanedir



Apache Storm ile gerek zamanli hesaplama sistemleri

Apache Storm akan verileri anlık olarak iřleyebilme zelliđine sahip aık kaynak kodlu bir ktphanedir

