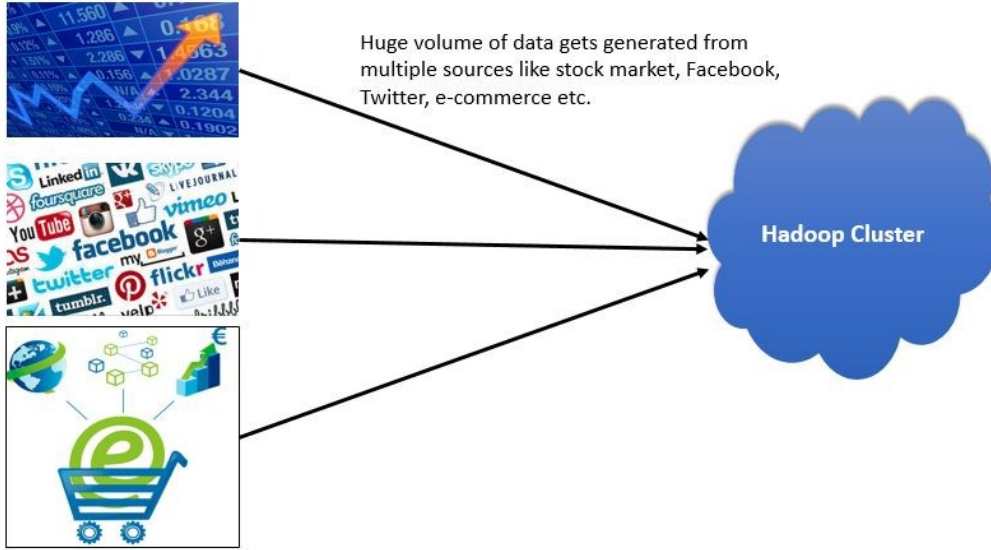


Büyük Verileri Hadoop ' a Nasıl Aktarabiliriz ?

Dış sistemlerden topladığınız verileri Hadoop üzerinde analiz edebilmek için öncelikle verileri Hadoop a taşımamız gerekir . Bu noktada aşağıdaki yöntemleri tercih edebiliriz

- Local bilgisayardan Hadoop a veri kopyalama ,
- Hadoop Sandbox ile ,
- Spark üzerinden ,
- Veritabanı üzerinden ,
- Storm üzerinden ,
- Kafka üzerinden

Şimdi bunları biraz daha detaylı inceleyelim



Local bilgisayardan Hadoop sistemine dosya kopyalama (copyFromLocal)

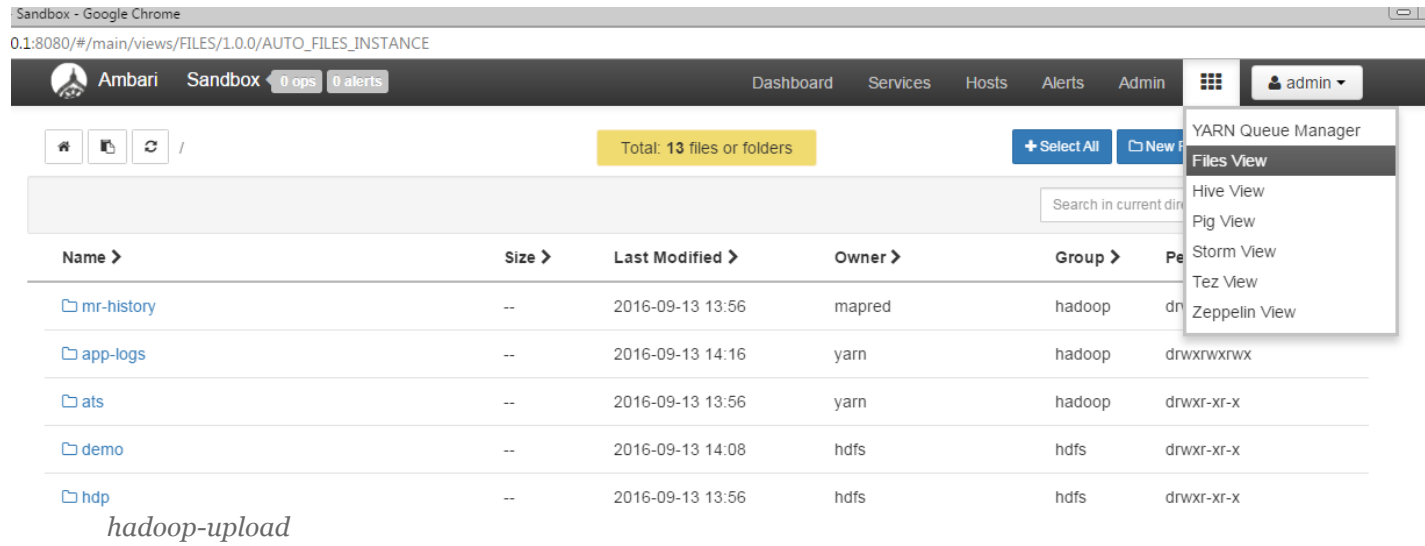
Hadoop içerisindeki **-copyFromLocal** komutuyla local makinadaki logları hadoop sistemine kopyalayabilirsiniz

```
hdfs dfs -copyFromLocal /home/networklogs.txt hdfs://cluster:9000/data/
```

Burada /home dizini altındaki networklogs.txt dosyasını hadoop üzerinde /data/logs/networklogs altına kopyalıyoruz

Hadoop Sandbox ile

Adım 1) Öncelikle 'Files View' (Dosya görünümü) açılır . Burada hdfs dosyalarını görebilirsiniz



Sandbox - Google Chrome

0.1:8080/#/main/views/FILES/1.0.0/AUTO_FILES_INSTANCE

Ambari Sandbox 0 ops 0 alerts

Dashboard Services Hosts Alerts Admin

admin

YARN Queue Manager

Files View

Hive View

Pig View

Storm View

Tez View

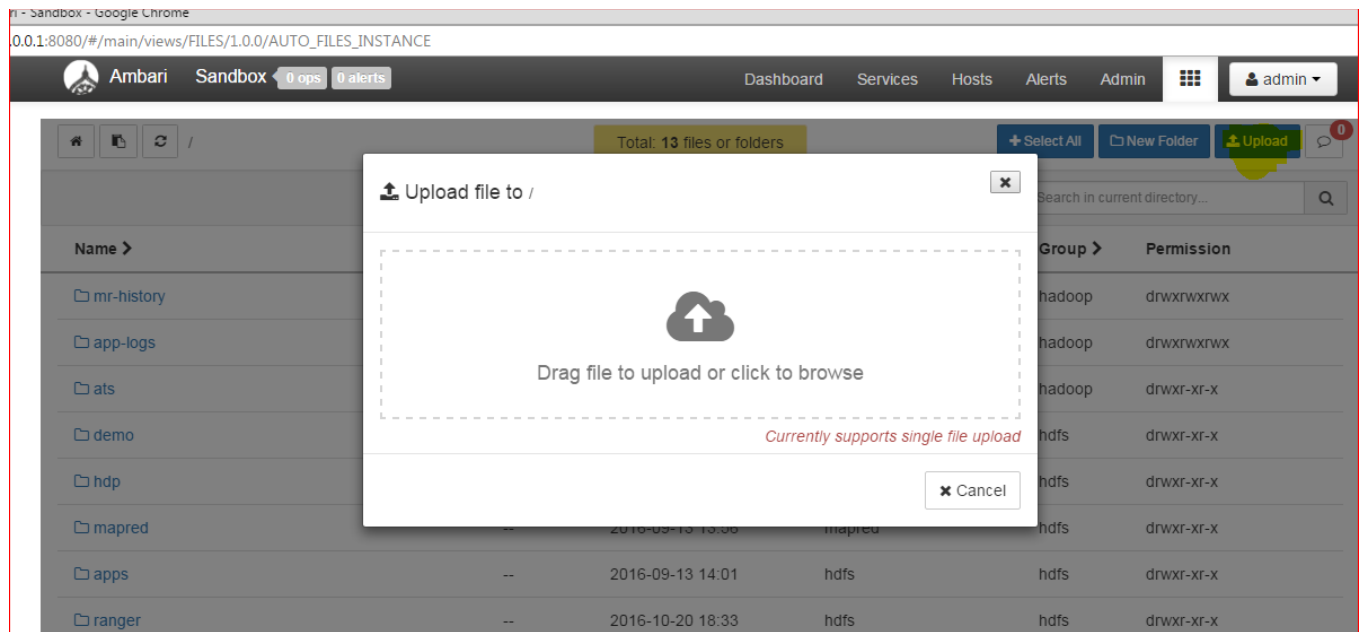
Zeppelin View

Search in current dir

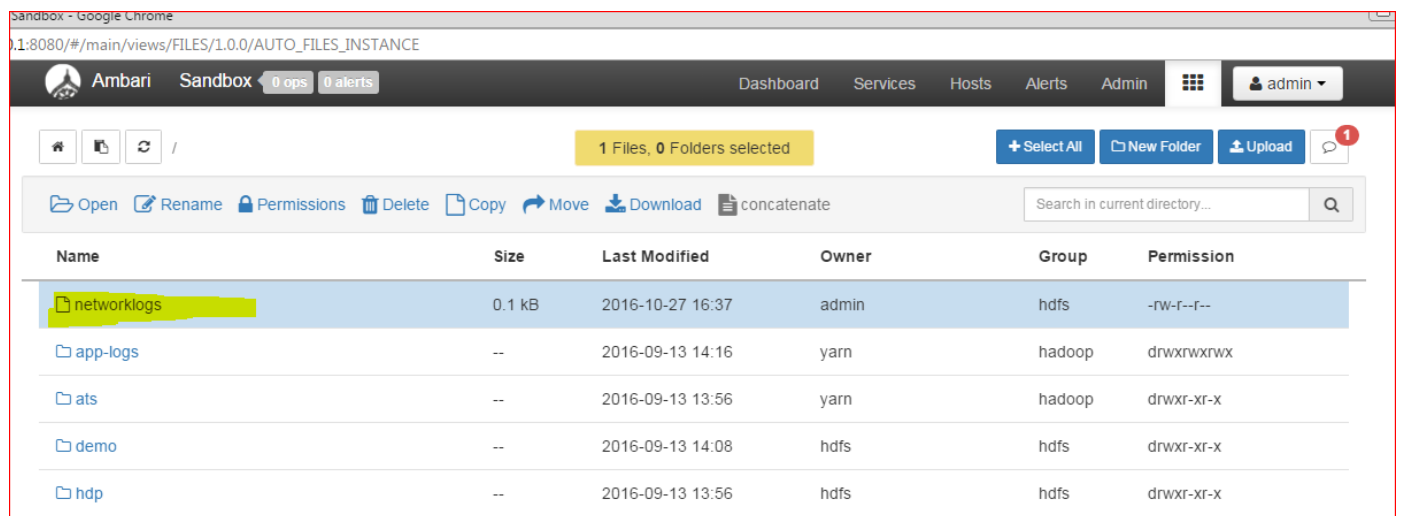
Name >	Size >	Last Modified >	Owner >	Group >	Permissions
mr-history	--	2016-09-13 13:56	mapred	hadoop	drwxr-xr-x
app-logs	--	2016-09-13 14:16	yarn	hadoop	drwxrwxrwx
ats	--	2016-09-13 13:56	yarn	hadoop	drwxr-xr-x
demo	--	2016-09-13 14:08	hdfs	hdfs	drwxr-xr-x
hdp	--	2016-09-13 13:56	hdfs	hdfs	drwxr-xr-x

hadoop-upload

Adım 2) Açılan bölümde 'Upload' seçilerek çıkan pencereden kopyalayacağımız dosyayı seçeriz



hadoop-kopyalama



hadoop-kopyalama

Spark üzerinden kopyalama

Spark üzerinde çeşitli analizler yaptıktan sonra verileri şu şekilde Hadoop sistemine aktarabilirsiniz

DataFrame yapısı üzerinden ;

```
dataFrame.write().parquet("hdfs://hadoopcluster:9000/data/logs/networklogs");
```

Rdd ile ;

```
rdd.saveAsTextFile("hdfs://hadoopcluster:9000/data/logs/networklogs");
```

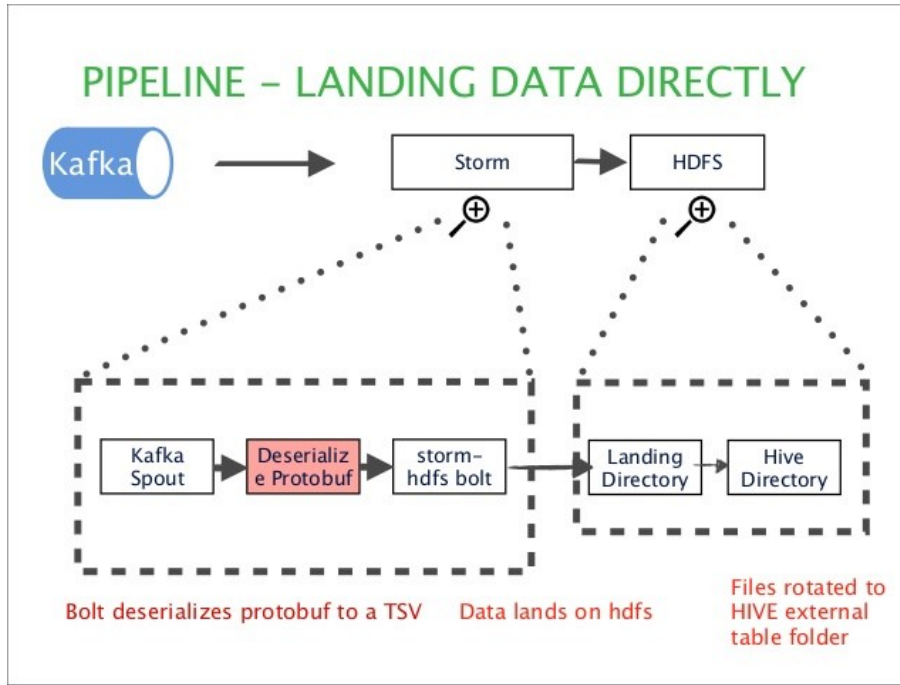
Veritabanı üzerinden Hadoop a dosya kopyalama

Veritabanı tabloları içerisinde bulunan dosyaları [Sqoop](#) ile Hadoop a rahatlıkla atabilirsiniz . Sqoop açık kaynak kodlu ve kullanılması kolay bir uygulamadır

```
$ sqoop import
--connect jdbc:mysql://localhost/db_name
--username root
--password 123
--table customer
--query "SELECT * from customer where status='active'"
--target-dir /data/customer
```

Storm üzerinden Hadoop a dosya kopyalama

[Apache Storm](#) üzerinden Hadoop a dosya kopyalamak için Storm içerisinde HdfsBolt gibi sınıflar aracılığıyla verileri Hadoop a aktarabiliriz

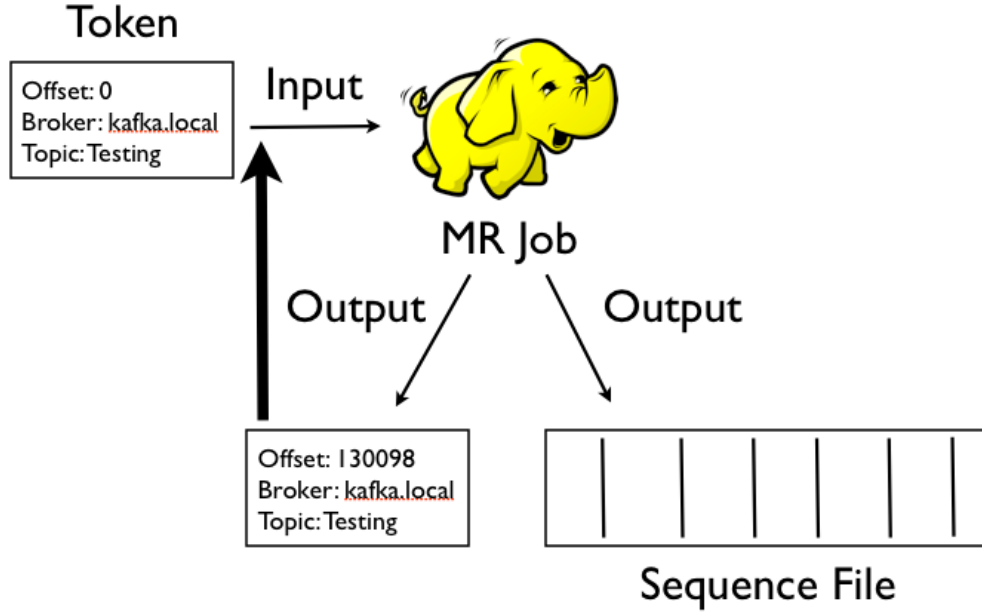


Friday, April 24, 15

Storm

Kafka üzerinden Hadoop a dosya kopyalama

[Apache Kafka](#) üzerinden Hadoop a dosya kopyalamak için hadoop-client kullanarak bir proje geliştirmeniz gereklidir. Arama yaptığınız zaman github üzerinde hazır projeler bulabileceğiniz gibi bu projeyi kendinizde geliştirebilirsiniz



kafka_hadoop