

LE MOIS DE LA DATA



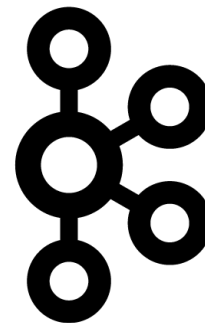
Data, Data Everywhere!



SMACK

Embrassez le Fast Data







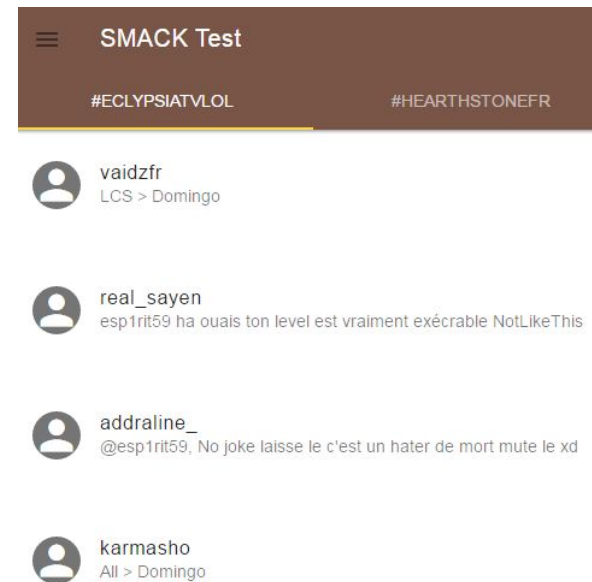
Use case

Yet another Twitter app ?



Twitch Chat

- Live Stream multiple twitch channel
- Metrics on Users



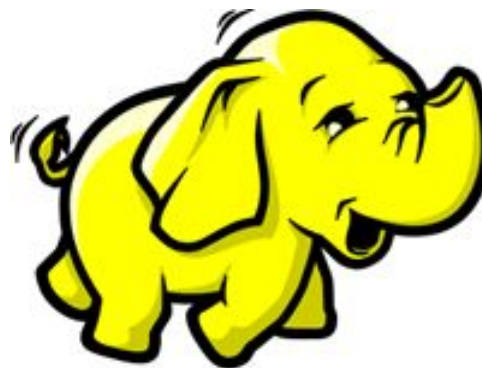


From Big Data to Fast Data



Hadoop Era (2005-2015)

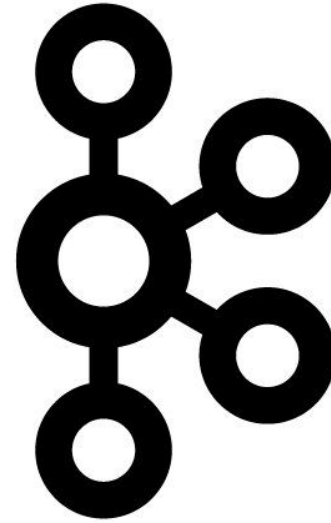
- Large Volume
- Batch Processing
- First Gen MapReduce
- Second Gen Spark

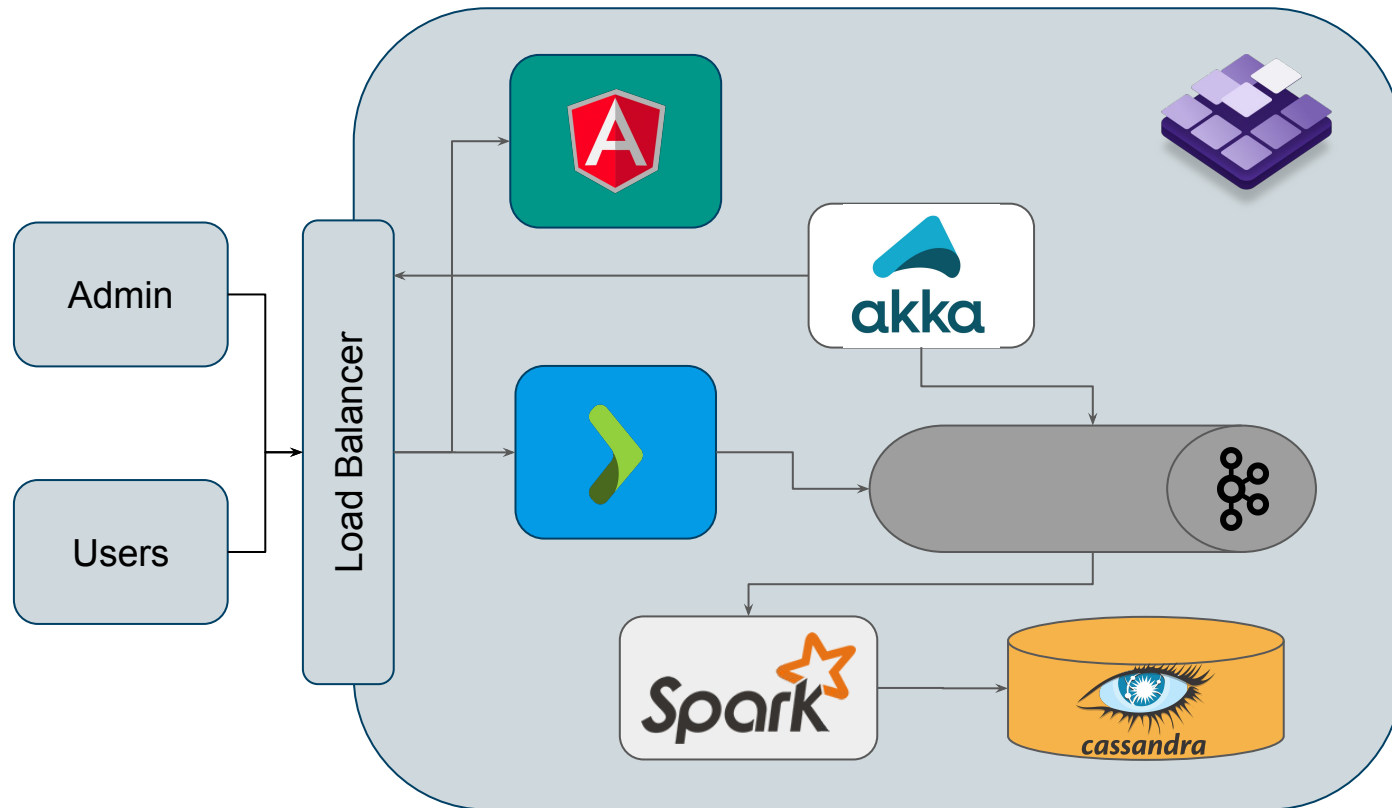




Fast Data Era (2015 - Now)

- Large Volume
- Near Real Time Processing
- Start with Kafka





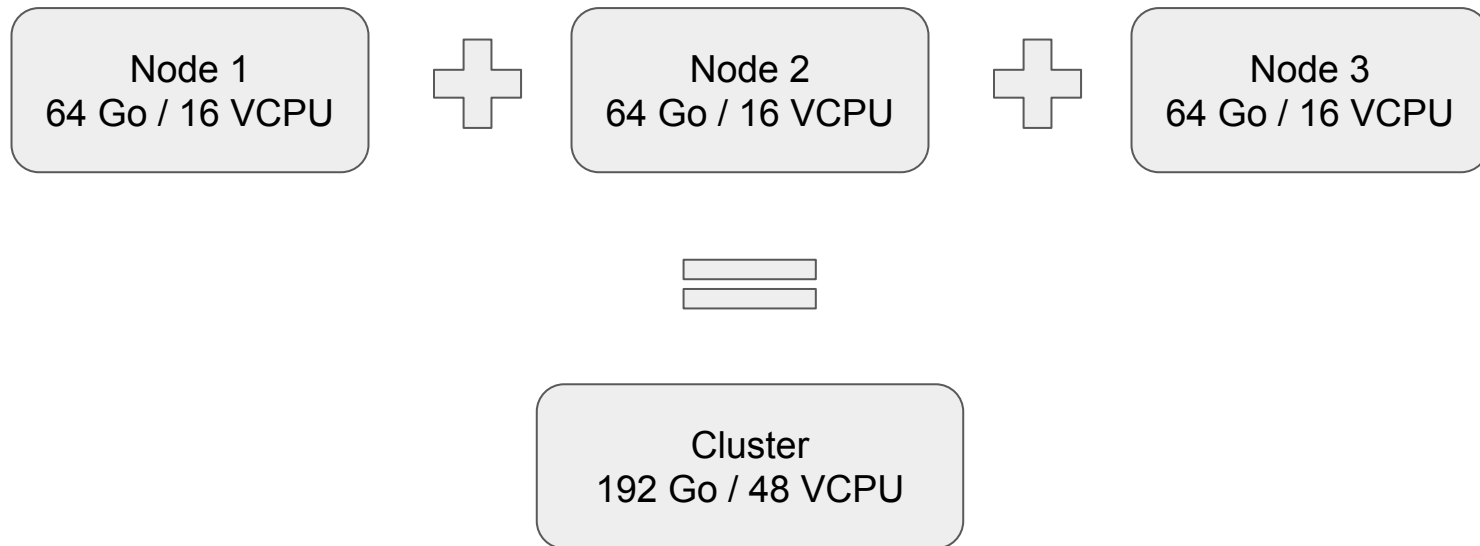


Mesos

sMACK



Sharing resources



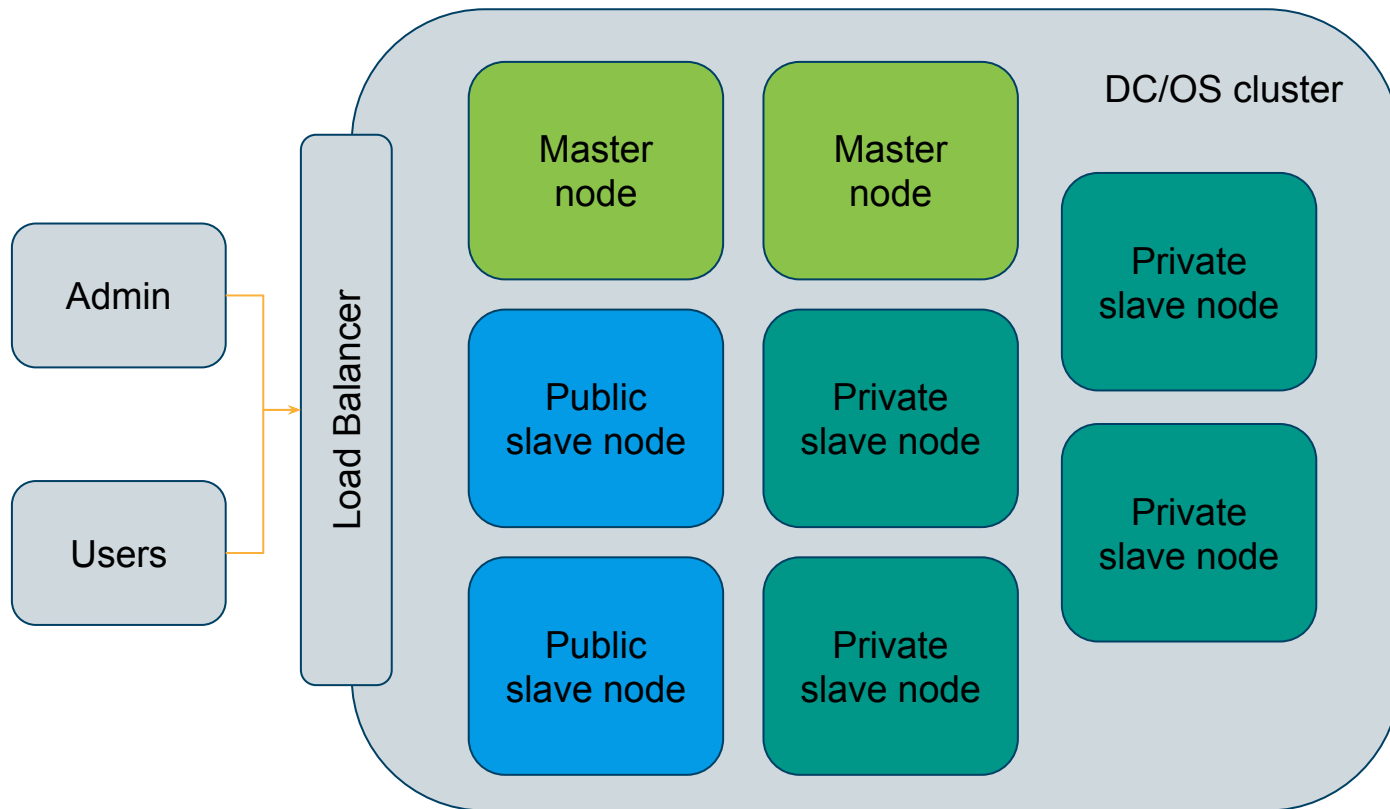
```
{  
  "id": "akka-twitch",  
  "cpus": 1,  
  "mem": 512.0,  
  "container": {  
    "type": "DOCKER",  
    "docker": {  
      "image": "slequeux/akka-twitch:1.2"  
    }  
  },  
}
```



docker



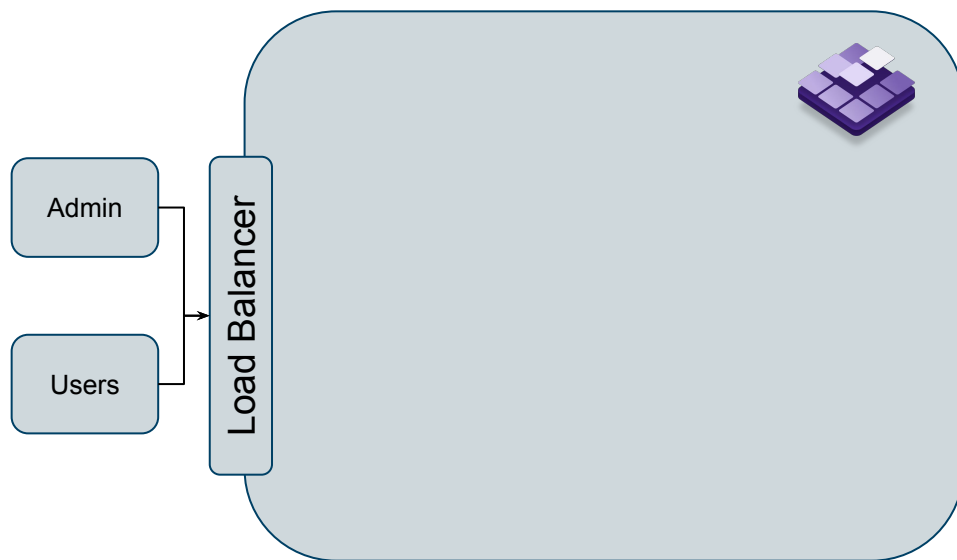
MARATHON



DC/OS Installation

Quick tour of GUI

Marathon-Ib

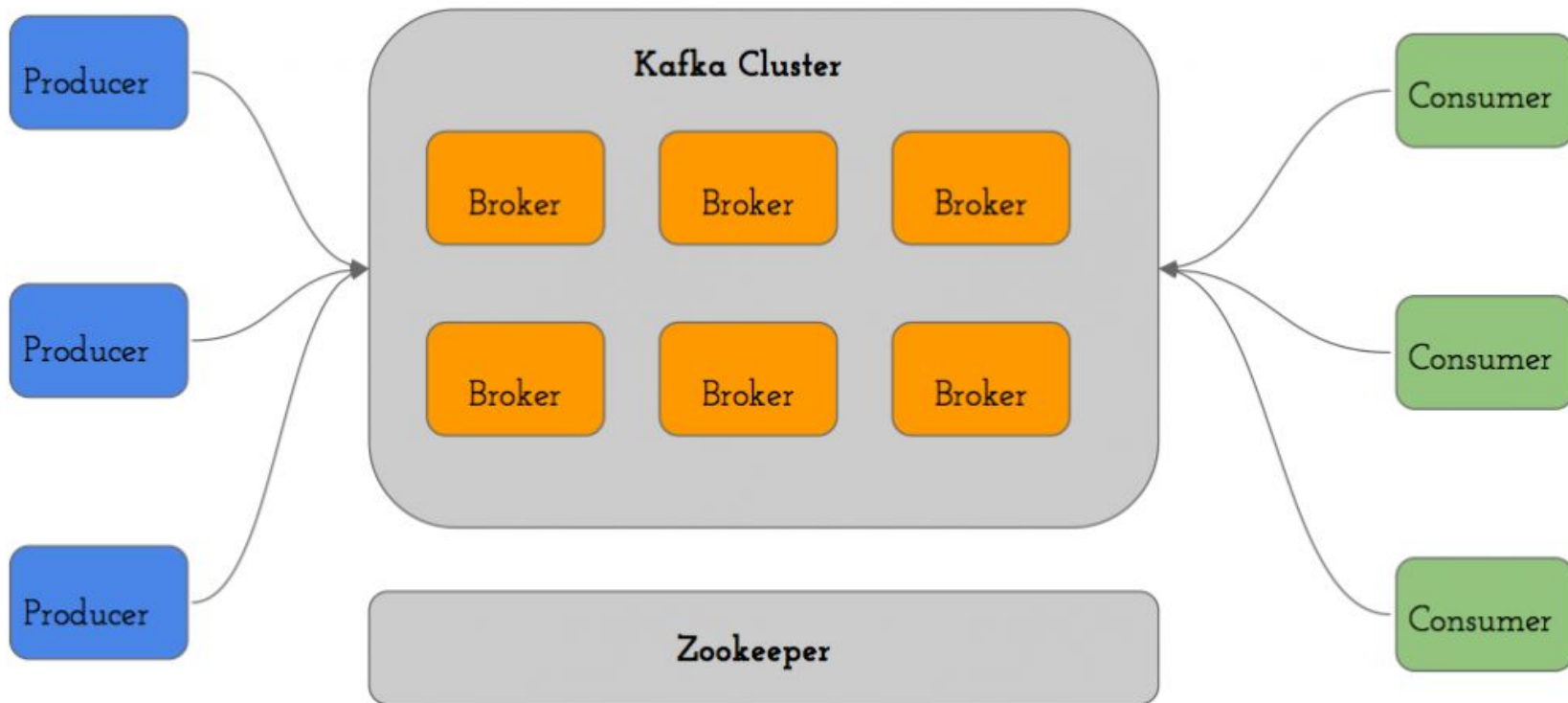




Kafka

SMACK



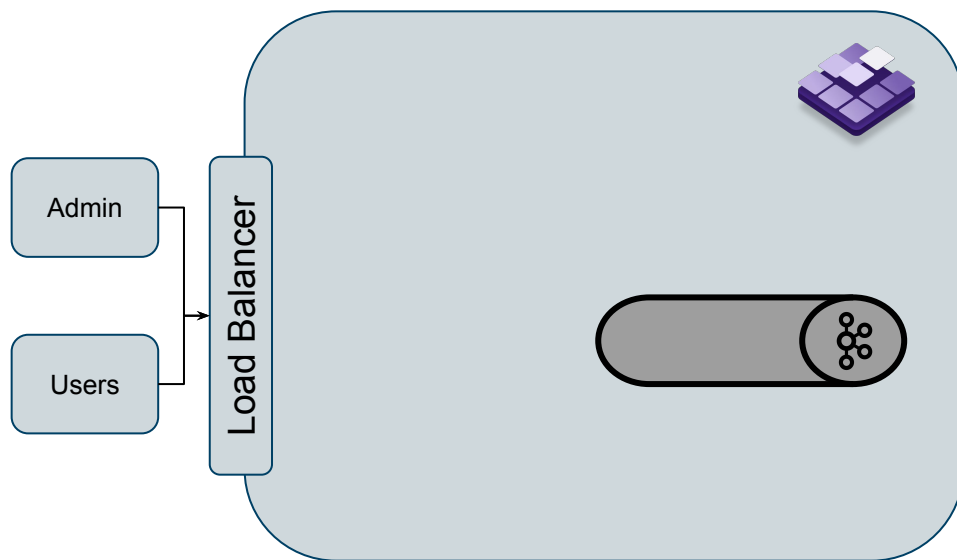


Service installation

Topic creation

Console consumer

Console producer





Akka

SMACK

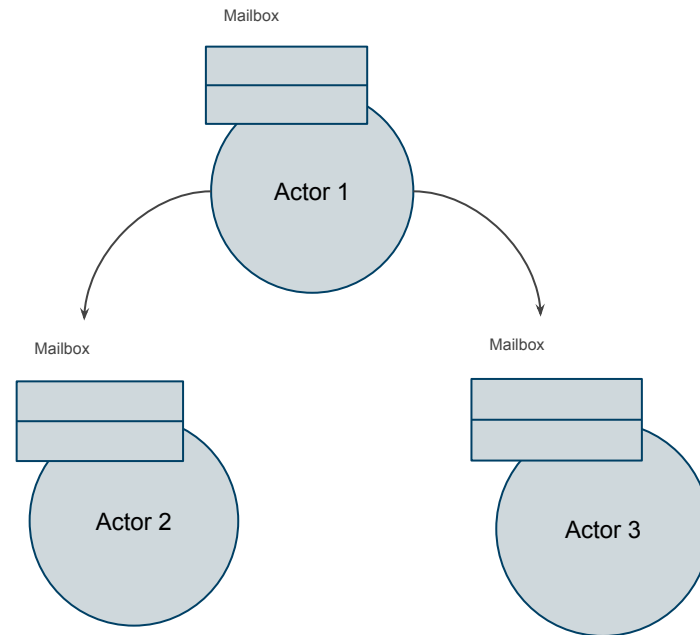


Lightweight

Actor System

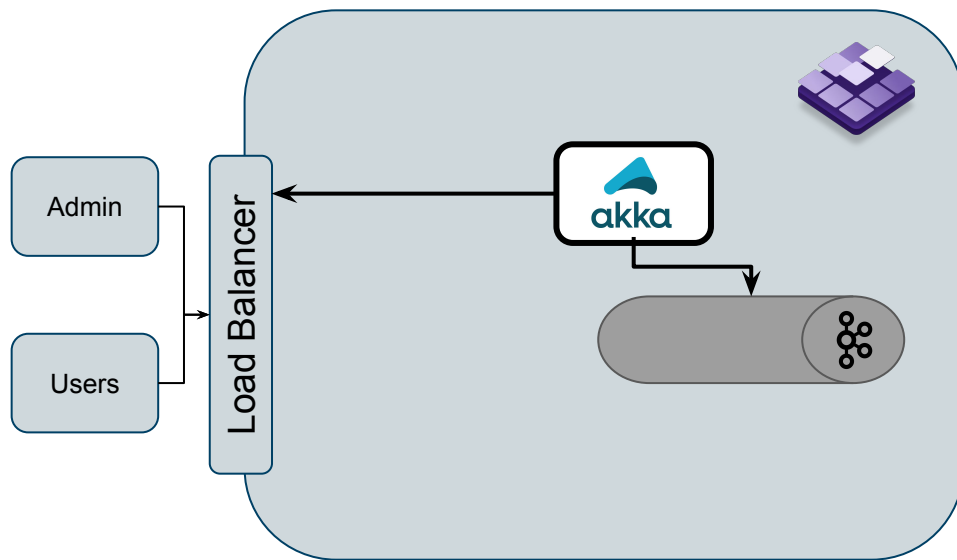
High level Api For Streaming

Native Kafka Source and Sink



Deploying an actor system

Example limitations





Restitution

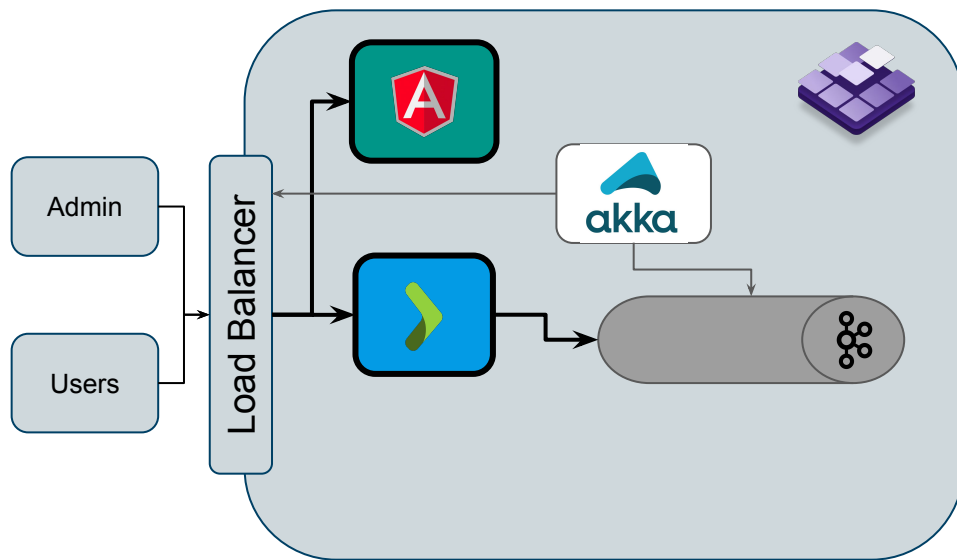
Let's have a break



Backend deployment

Frontend deployment

App Tour





Cassandra

SMACk

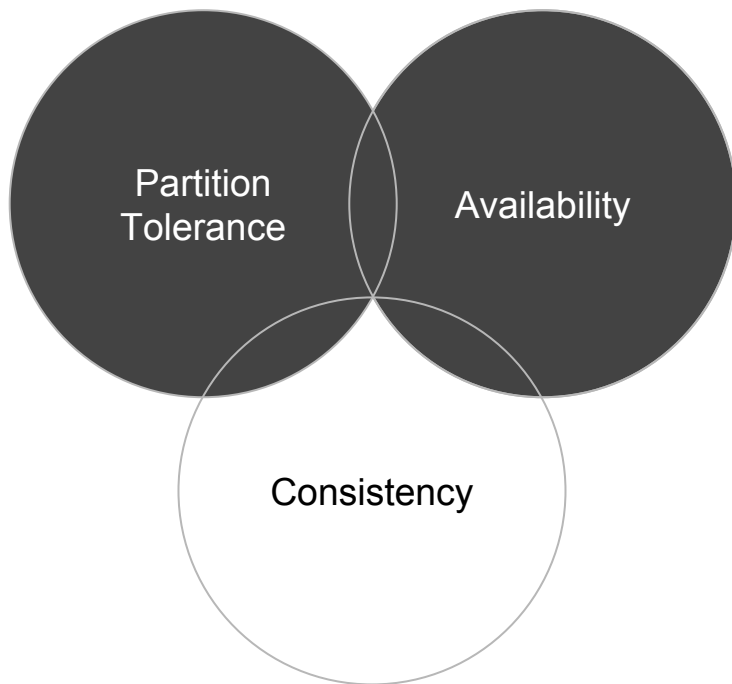


- Highly Available
- Tuneably Consistent
- Row Store
- NoSQL



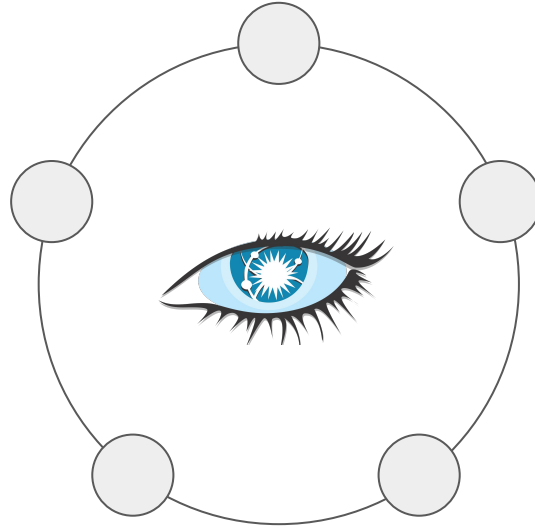
Highly Available

LE MOIS DE LA DATA



#lemoisdeladata

powered by **Xebia**





```
CREATE KEYSPACE IF NOT EXISTS smack
```

```
WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 1 };
```



- Query || Cluster || Data Center
- Read || Write
- Different Level (One, Quorum, ...)



partition key

columns

42	email	name	phone
	foo@bar.fr	foo	12345

14	name	phone
	doe	98765



- CQL \sim SQL
- No Join
- Limited Aggregation
- Filter on Key
- Model Around Query

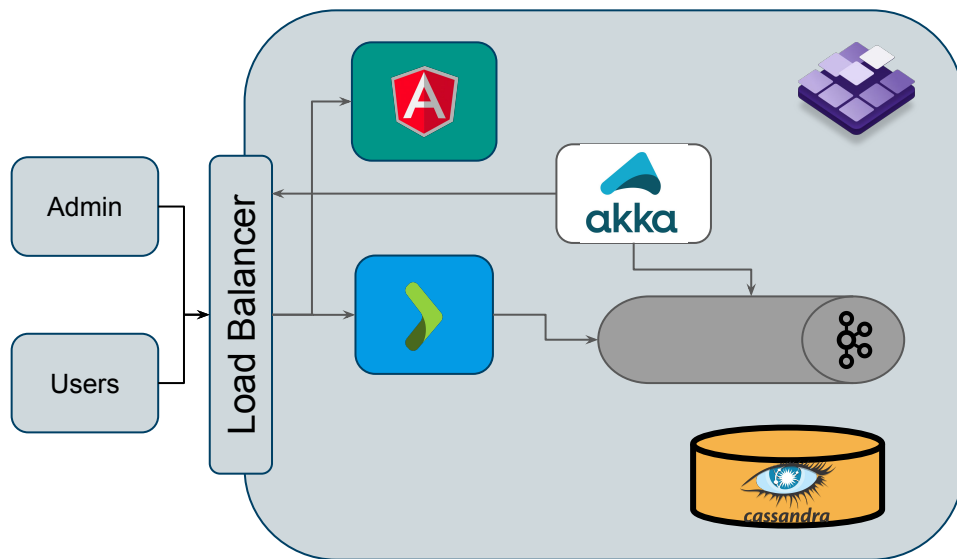


Package installation

Keyspace creation

Schema creation

Query DB

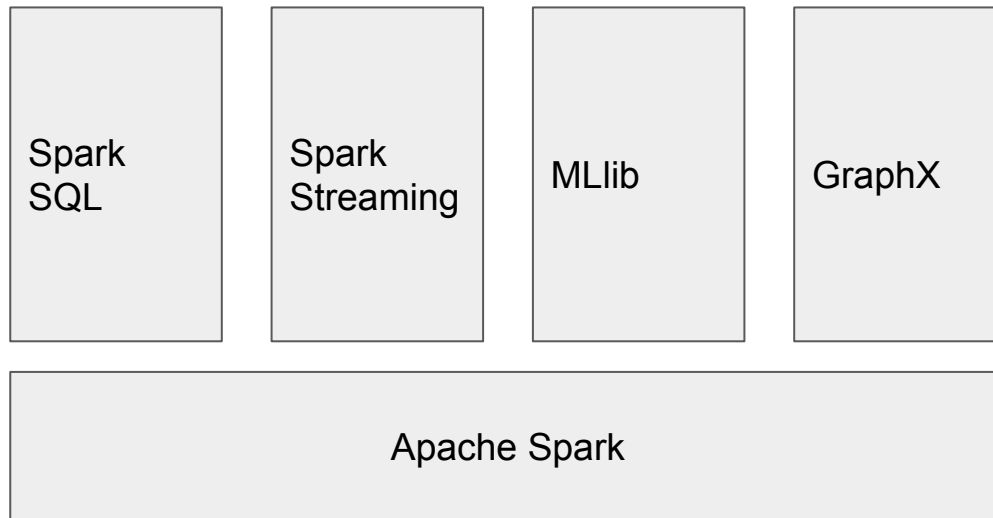


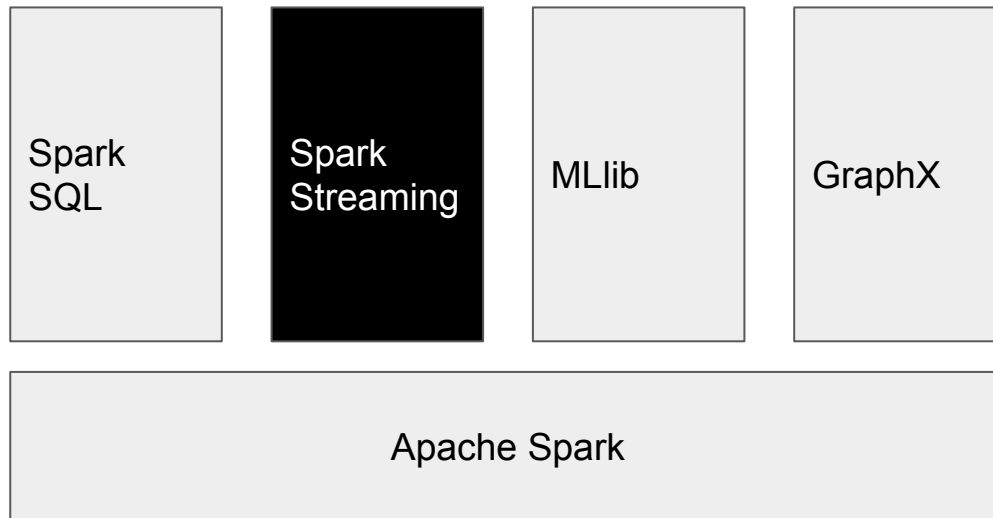


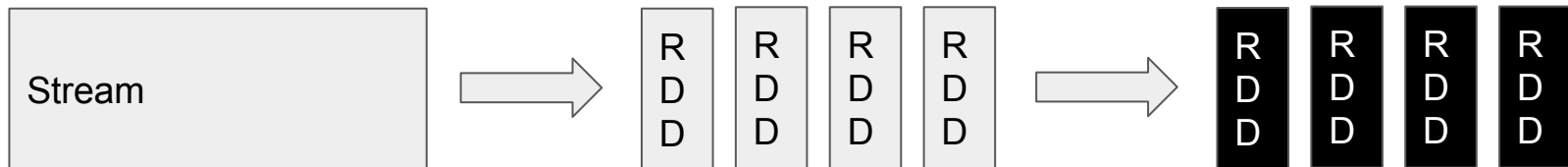
Spark

SMACK



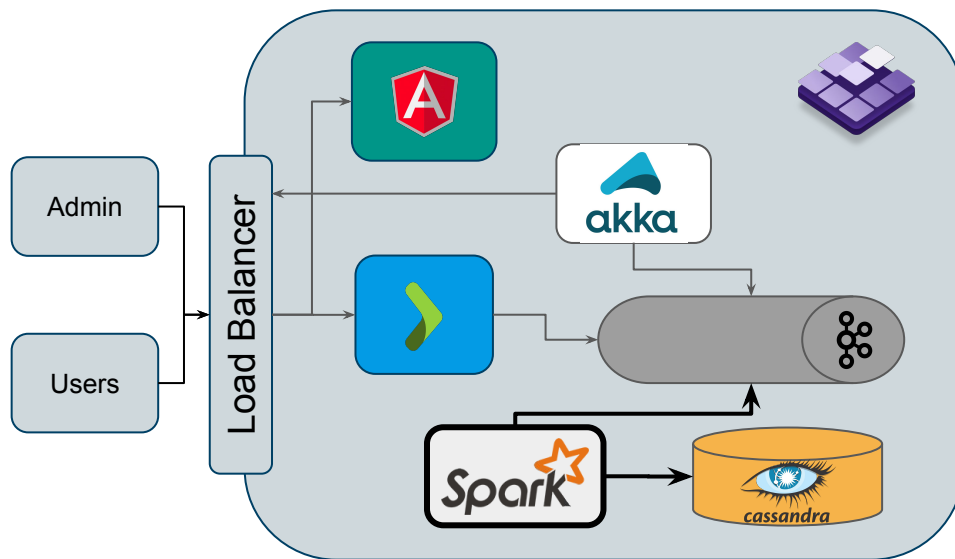


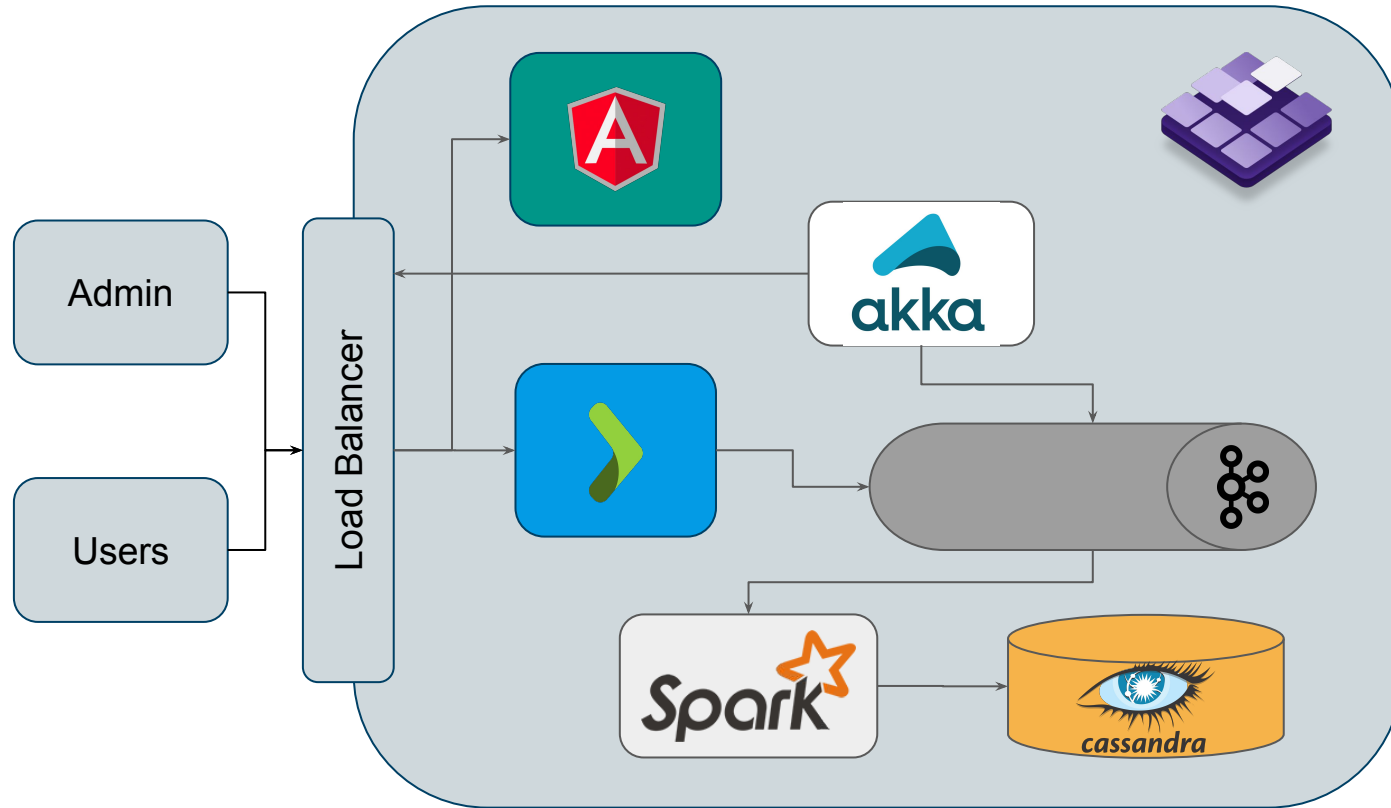




Package installation

Deploying a streaming job







To infinity and beyond



PANCAKE

STACK



Presto



Arrow



NiFi



Cassandra



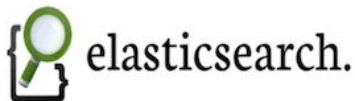
AirFlow



Kafka



ElasticSearch



Spark



TensorFlow

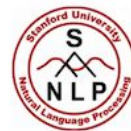


TensorFlow

Algebird



CoreNLP



Kibana





That's all folks

Thank you





#BeXebian
recrutement@xebia.fr

