

BIG DATA EUROPE

Empowering Communities
with Data Technologies



SC4 Pilot 1

Brussels
22.09.2016

L. Selmi - BDE - Societal Challenge on
Transport



The Societal Challenge

BIG DATA EUROPE

Empowering Communities
with Data Technologies



From a data perspective we see endless streams of data collected by sensors and devices immersed in a graph of infrastructures



L. Selmi - BDE - Societal Challenge on
Transport



BIG DATA EUROPE

Empowering Communities
with Data Technologies

Objective of the SC4 Pilots

A scalable, fault-tolerant and flexible platform based on open source frameworks that can process unbounded data sets and graphs.



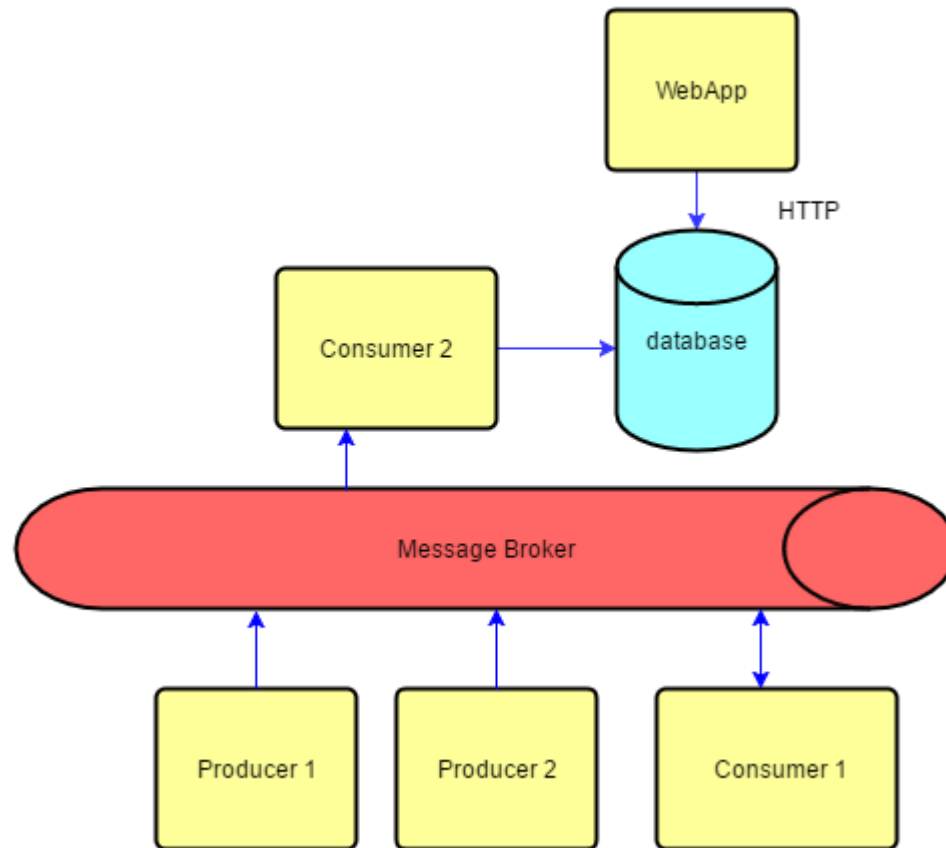
L. Selmi - BDE - Societal Challenge on
Transport



BIG DATA EUROPE

Empowering Communities
with Data Technologies

Microservice Architecture



L. Selmi - BDE - Societal Challenge on
Transport



BIG DATA EUROPE

Empowering Communities
with Data Technologies

Message Broker

Apache Kafka



Apache Kafka is a high-throughput distributed durable messaging system



L. Selmi - BDE - Societal Challenge on
Transport

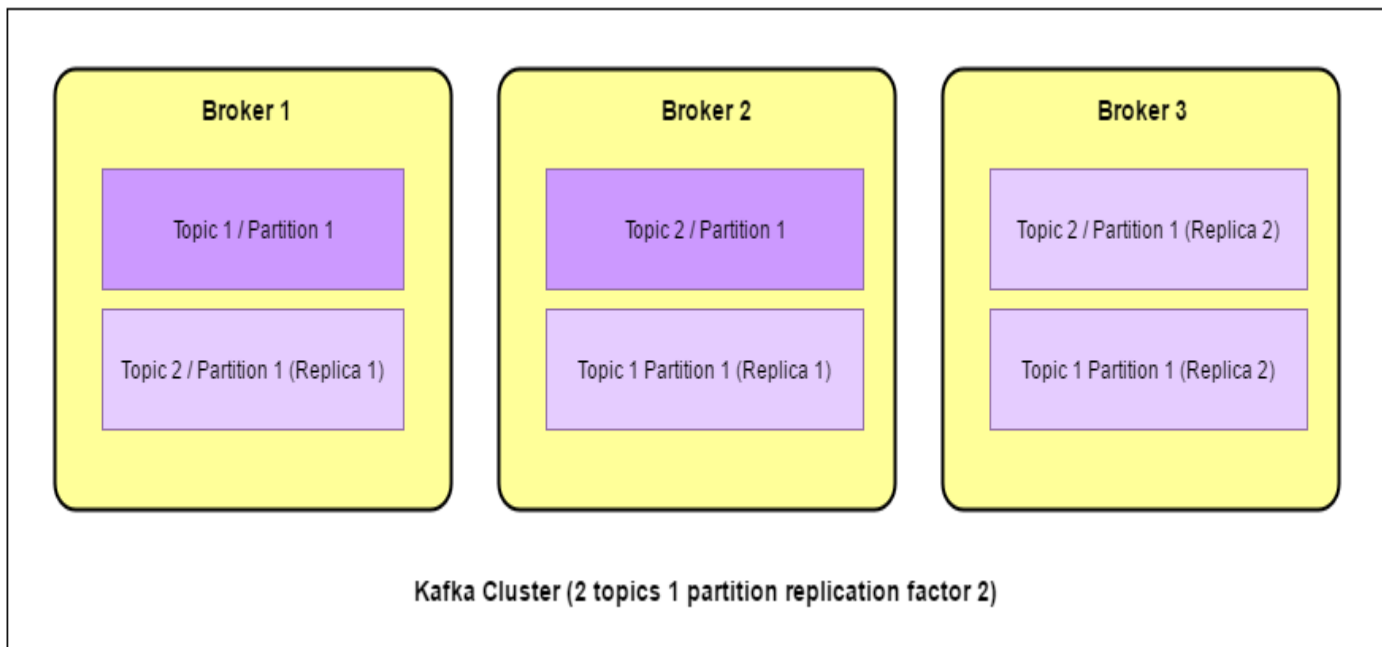


BIG DATA EUROPE

Empowering Communities
with Data Technologies

Kafka Cluster

Apache Kafka



L. Selmi - BDE - Societal Challenge on
Transport



BIG DATA EUROPE

Empowering Communities
with Data Technologies

Stream and Batch Processor

Apache Flink



Apache Flink is an open source platform for distributed stream and batch data processing.



L. Selmi - BDE - Societal Challenge on
Transport

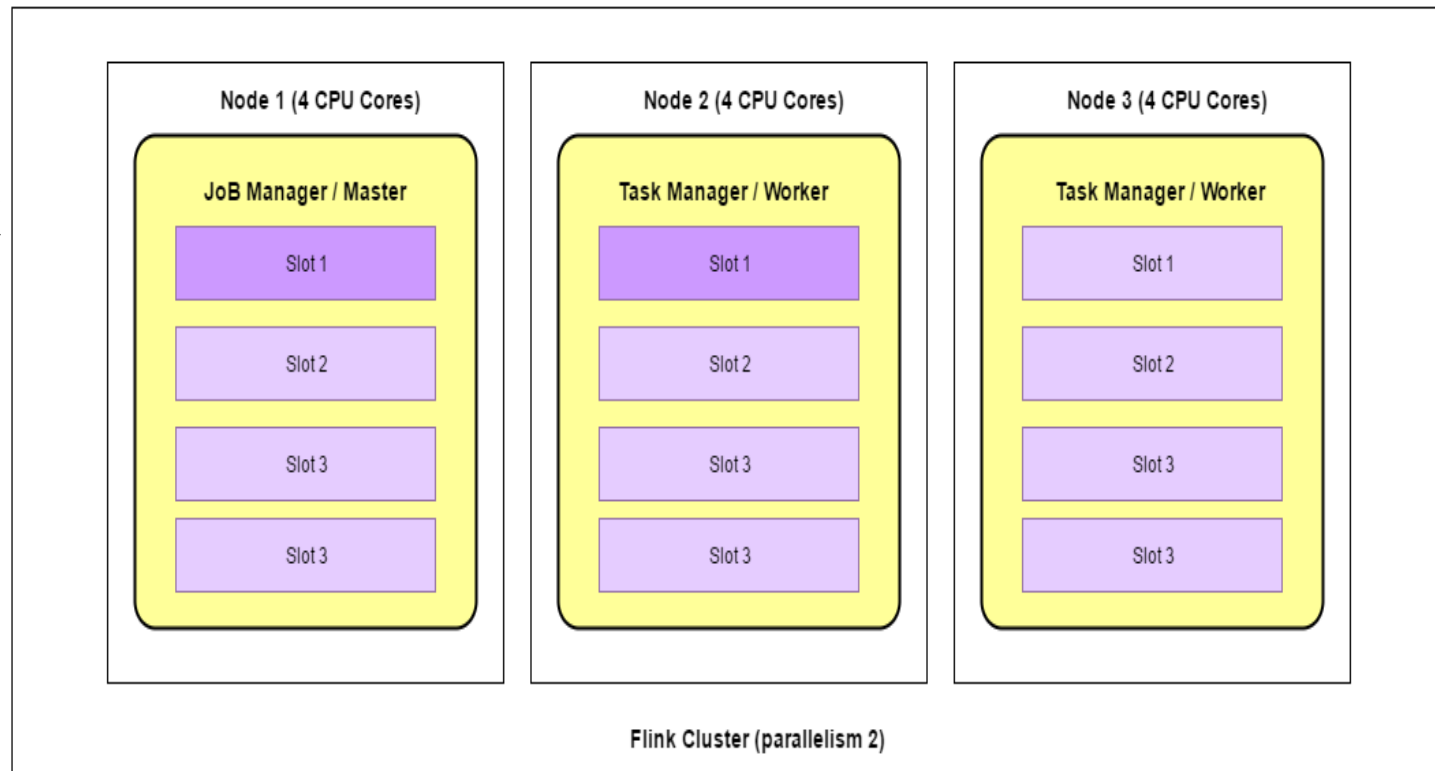


BIG DATA EUROPE

Empowering Communities
with Data Technologies

Flink Cluster

Apache Flink



L. Selmi - BDE - Societal Challenge on
Transport



BIG DATA EUROPE

Empowering Communities
with Data Technologies

Storage and Indexing



elastic

Elasticsearch is a distributed open source document database built on top of Apache Lucene



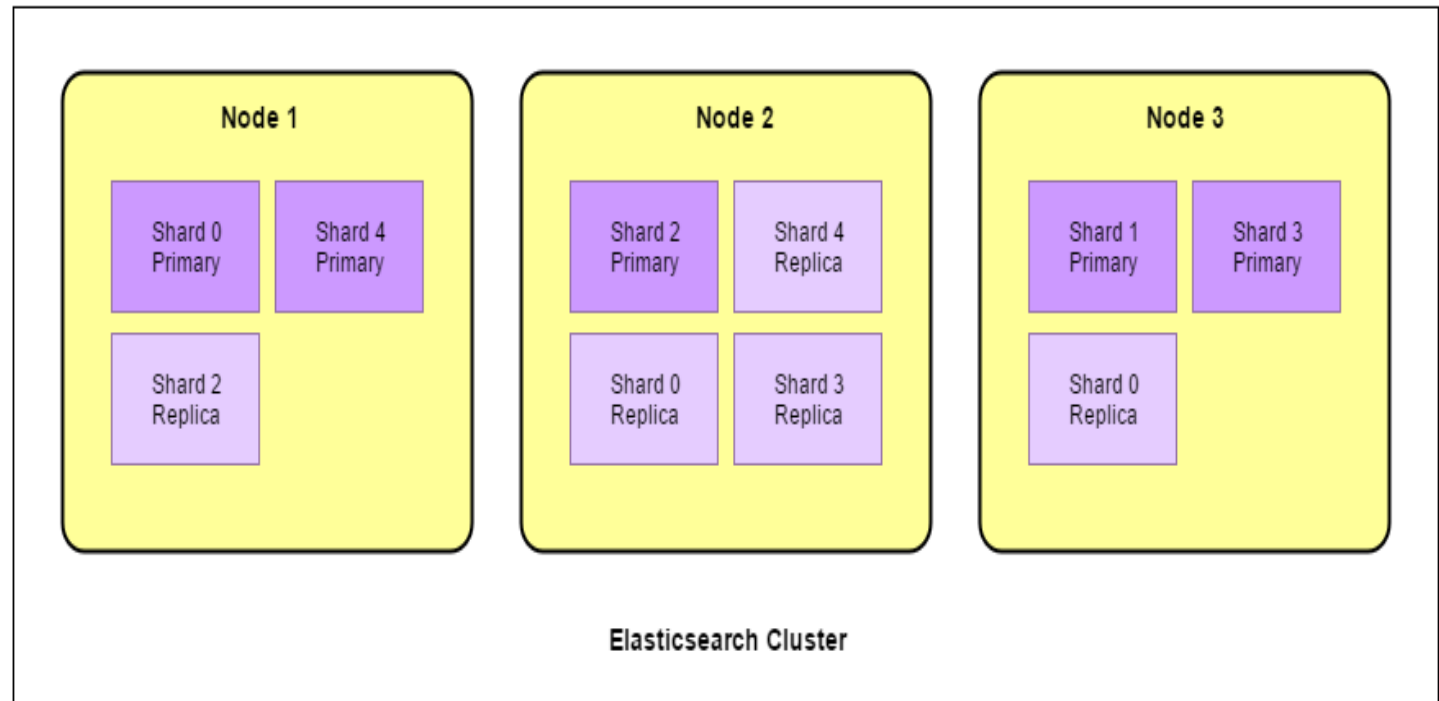
L. Selmi - BDE - Societal Challenge on
Transport



BIG DATA EUROPE

Empowering Communities
with Data Technologies

Elasticsearch Cluster



L. Selmi - BDE - Societal Challenge on
Transport



BIG DATA EUROPE

Empowering Communities
with Data Technologies

Rserve

The R Project



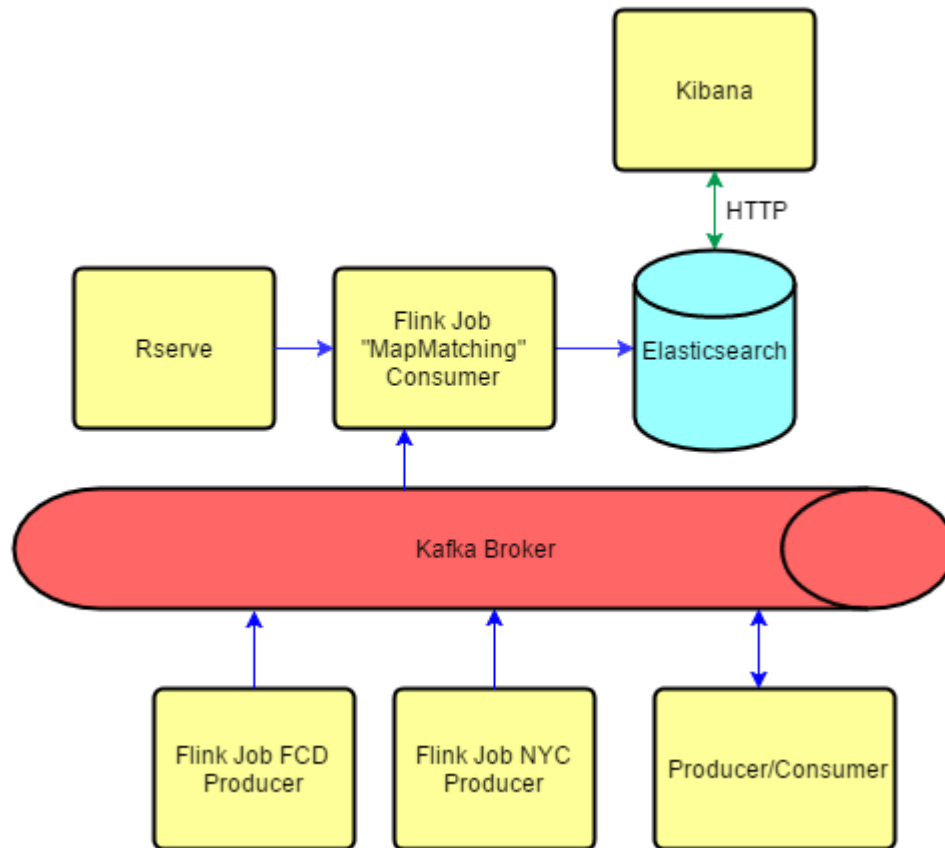
R is a free software environment for statistical computing. It is used in the pilot to run the map-matching and the prediction algorithms.



L. Selmi - BDE - Societal Challenge on
Transport



Pilot 1 Architecture



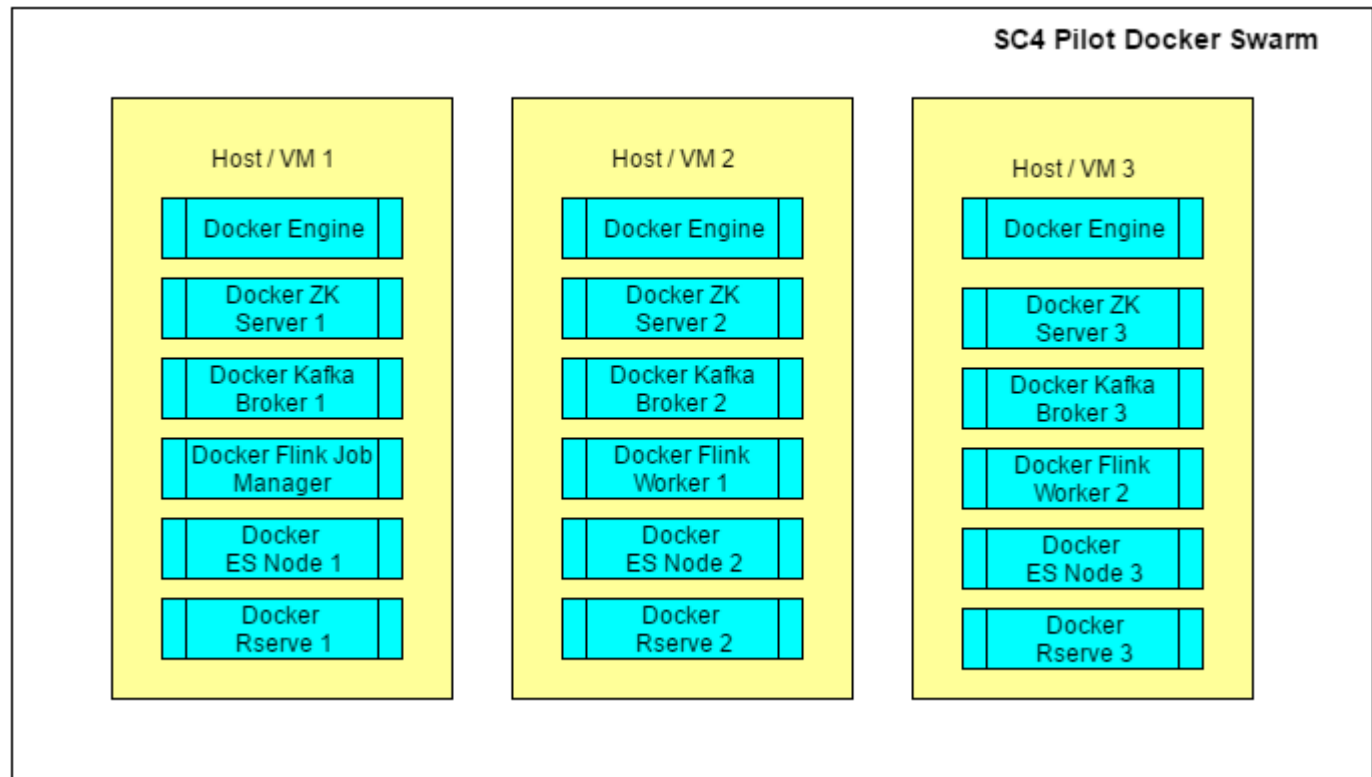


BIG DATA EUROPE

Empowering Communities
with Data Technologies



BDE Components



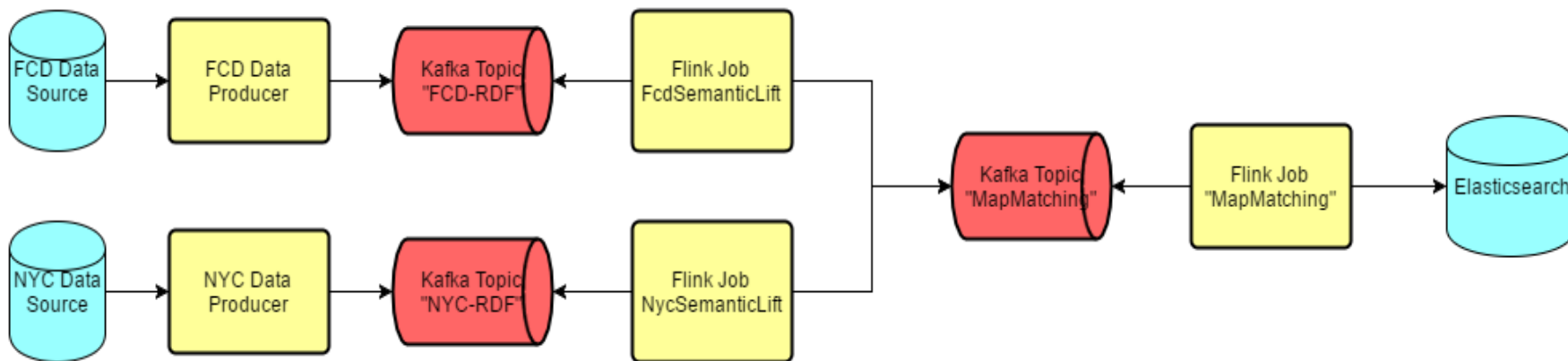
L. Selmi - BDE - Societal Challenge on
Transport



BIG DATA EUROPE

Empowering Communities
with Data Technologies

A Pipeline



L. Selmi - BDE - Societal Challenge on
Transport



SC4 - Pilot 1 Visualization

BIG DATA EUROPE

Empowering Communities
with Data Technologies



SC4 Pilot 1 can process real-time FCD data for map-matching and aggregations (e.g. average speed)



L. Selmi - BDE - Societal Challenge on
Transport



BIG DATA EUROPE

Empowering Communities
with Data Technologies

Future Work

- ⊙ Pilot 2: integration of the FCD historical data and prediction algorithm
- ⊙ Pilot 3: more data sets (BT sensors, NYC taxi data, OpenStreetMap)



L. Selmi - BDE - Societal Challenge on
Transport