CiberSecurity

Telefonica

STRATIO

Spark use case at Telefonica CBS

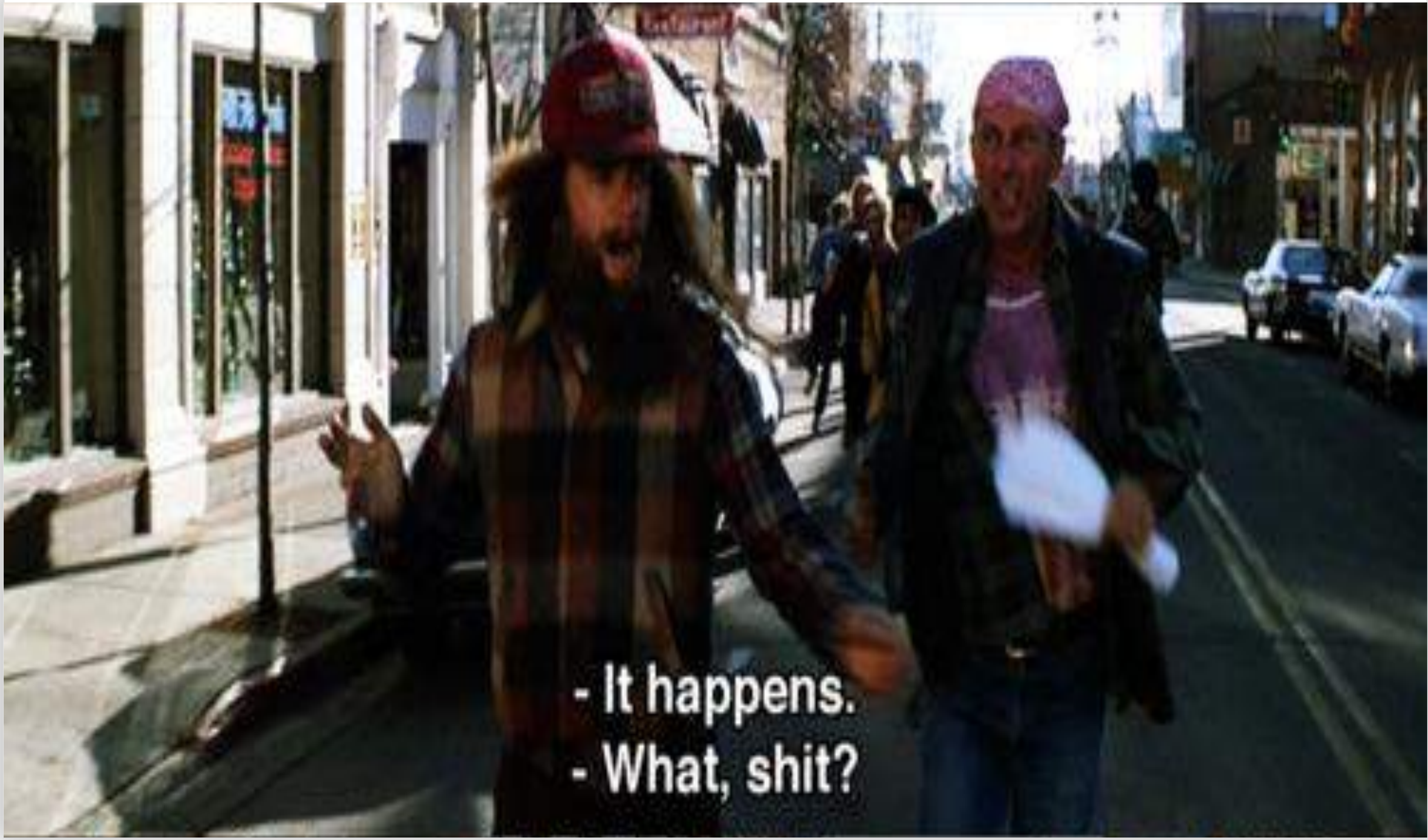
Telefónica Digital
Digital Services

WHOAMI

- Francisco J. Gomez
- Worker at Telefónica (Spain)
- Securityholic
- @ffranz

WHY

WHY



WHY

CiberSecurity

Spark use case at Telefonica CBS

STRATIO Telefonica

Contents

01 Telefonica

What Does Cyber-Security Mean?

What does it mean for us?

Cyber-Security in numbers

Show me the reality

Traditional Approach

New Approach

02 Our Skills

Our skills

What do we need ?

03 Stratio

Stratio distribution Spark

Architecture

Data adquisition

Data fusion

Batch

03 Upcoming Challenges

Stratio Streaming

SQL Like Interface

Sinfonier

CiberSecurity

Telefonica

What Does Cyber-Security Mean?

“Cybersecurity is the **collection of tools, policies**, security concepts, security safeguards, guidelines, risk management approaches, actions, training, best practices, assurance and technologies that can be used **to protect the cyber environment and organization and user’s assets**. Organization and user’s assets include connected computing devices, personnel, infrastructure, applications, services, telecommunications systems, and the totality of transmitted and/or stored information in the cyber environment. **Cybersecurity strives to ensure** the attainment and maintenance of the security properties of the organization and user’s assets against relevant security risks in the cyber environment. The general security objectives comprise the following: Availability; Integrity, which may include authenticity and non-repudiation; Confidentiality” ITU-T X.1205, Overview of cybersecurity

“(8) CYBERSECURITY THREAT.— The term “cybersecurity threat” means any action that may result in **unauthorized access to, manipulation of**, or impairment to **the integrity, confidentiality, or availability of an information** system or information stored on or transiting an information system, or **unauthorized exfiltration** of information stored on or transiting an information system.” DEPARTMENT OF HOMELAND SECURITY CYBERSECURITY AUTHORITY

What does it mean for us?

“Cybersecurity is the *collection of* ***tools, policies… capabilities*** *to protect the cyber environment and organization and user’s assets.* ***Cybersecurity strives to ensure unauthorized access to, manipulation of the integrity, confidentiality, or availability of an information, or unauthorized exfiltration of information.”***

Cyber-Security in numbers

- Hacktivism
- Cyber Crime
- Cyber Warfare
- Cyber Espionage

- DDoS (23%)
- SQLi (19%)
- Defacement (14%)
- Account Hijacking (9%)
- Unknown (18%)

Show me the reality

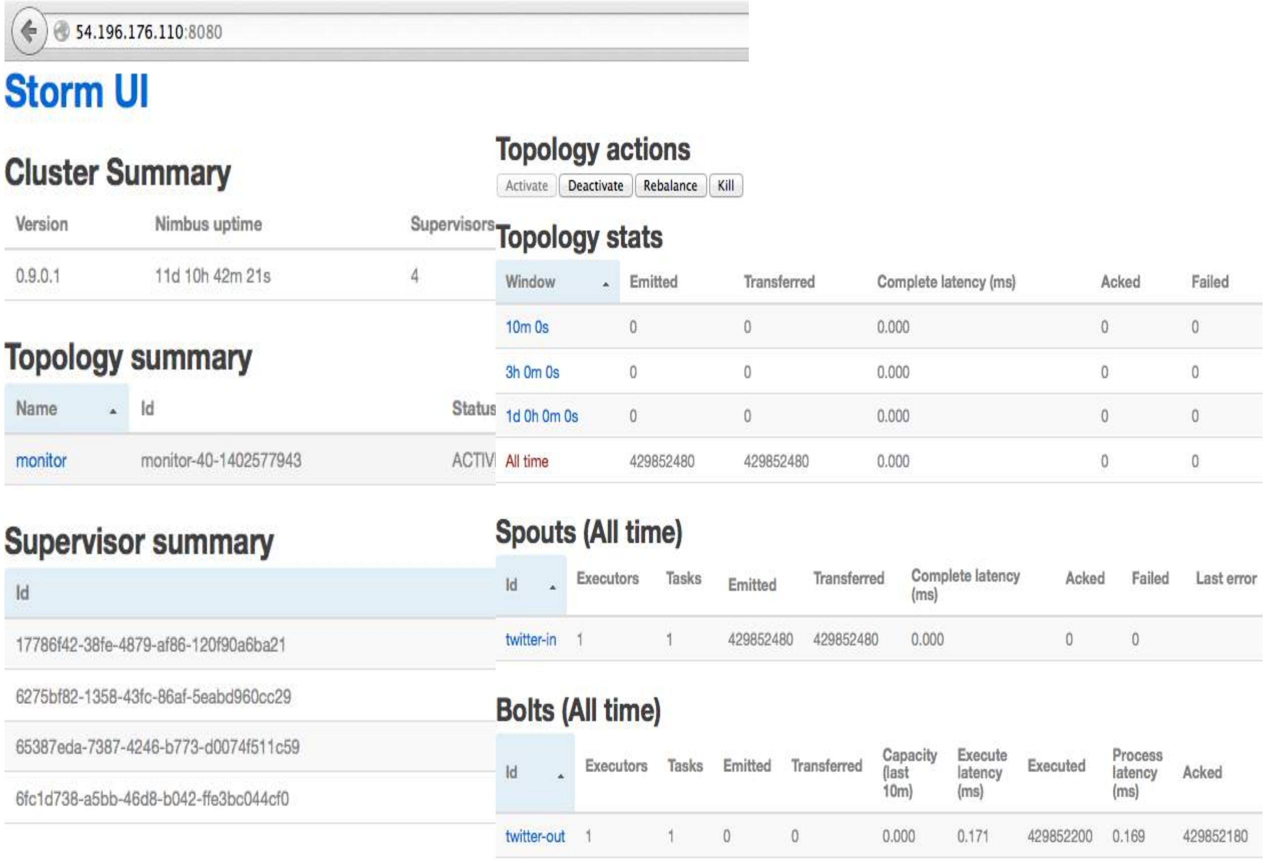
Storm UI World map

Canada
54.221.8.199 - Ashburn
United States
Mexico
Cuba
Venezuela
Colombia

Wordpress

Telefonica

Storm UI



54.196.176.110:8080

Storm UI

Cluster Summary

Version	Nimbus uptime	Supervisors
0.9.0.1	11d 10h 42m 21s	4

Topology summary

Name	Id	Status
monitor	monitor-40-1402577943	ACTIV

Supervisor summary

Id
17786f42-38fe-4879-af86-120f90a6ba21
6275bf82-1358-43fc-86af-5eabd960cc29
65387eda-7387-4246-b773-d0074f511c59
6fc1d738-a5bb-46d8-b042-ffe3bc044cf0

Topology actions

Topology stats

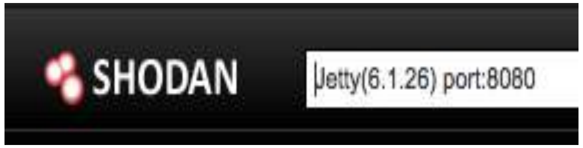
Window	Emitted	Transferred	Complete latency (ms)	Acked	Failed
10m 0s	0	0	0.000	0	0
3h 0m 0s	0	0	0.000	0	0
1d 0h 0m 0s	0	0	0.000	0	0
All time	429852480	429852480	0.000	0	0

Spouts (All time)

Id	Executors	Tasks	Emitted	Transferred	Complete latency (ms)	Acked	Failed	Last error
twitter-in	1	1	429852480	429852480	0.000	0	0	

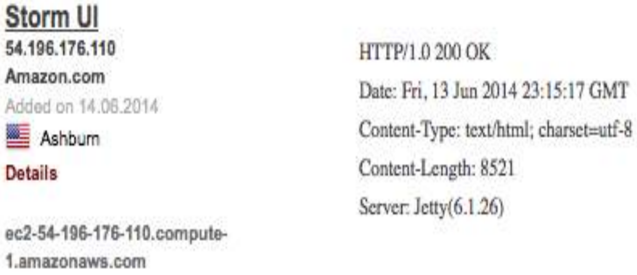
Bolts (All time)

Id	Executors	Tasks	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked
twitter-out	1	1	0	0	0.000	0.171	429852200	0.169	429852180



SHODAN

Jetty(6.1.26) port:8080



Storm UI
54.196.176.110
Amazon.com
Added on 14.06.2014
Ashburn
Details

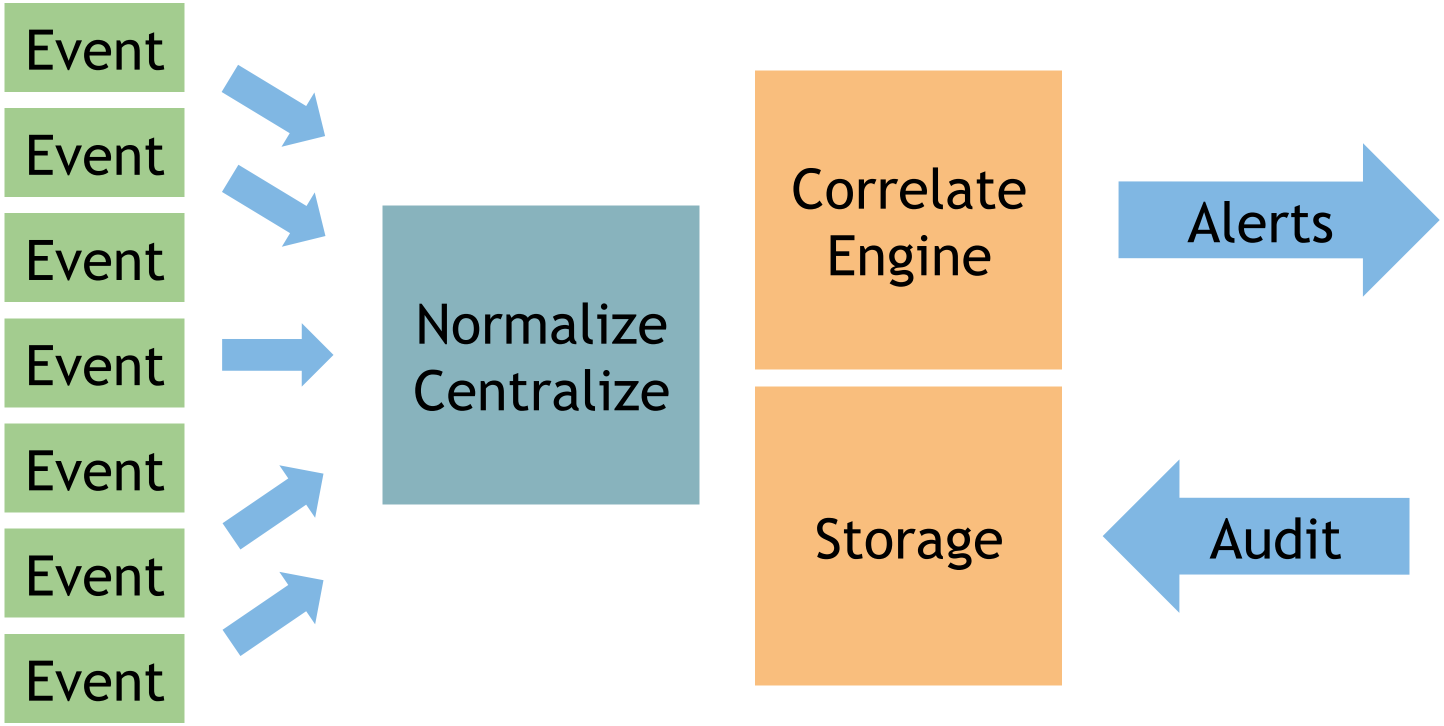
ec2-54-196-176-110.compute-1.amazonaws.com

```
HTTP/1.0 200 OK
Date: Fri, 13 Jun 2014 23:15:17 GMT
Content-Type: text/html; charset=utf-8
Content-Length: 8521
Server: Jetty(6.1.26)
```

Check point

Traditional Approach

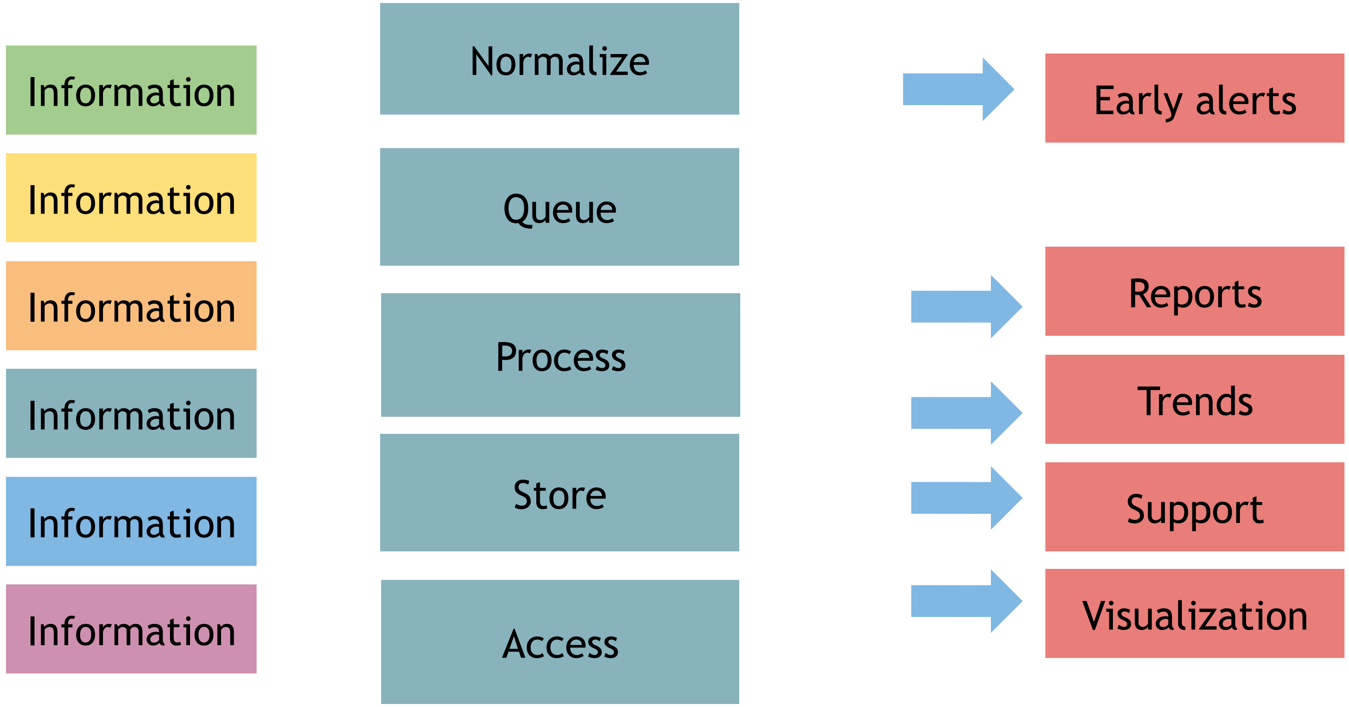
Events, Logs and Alerts: Correlation



New Approach

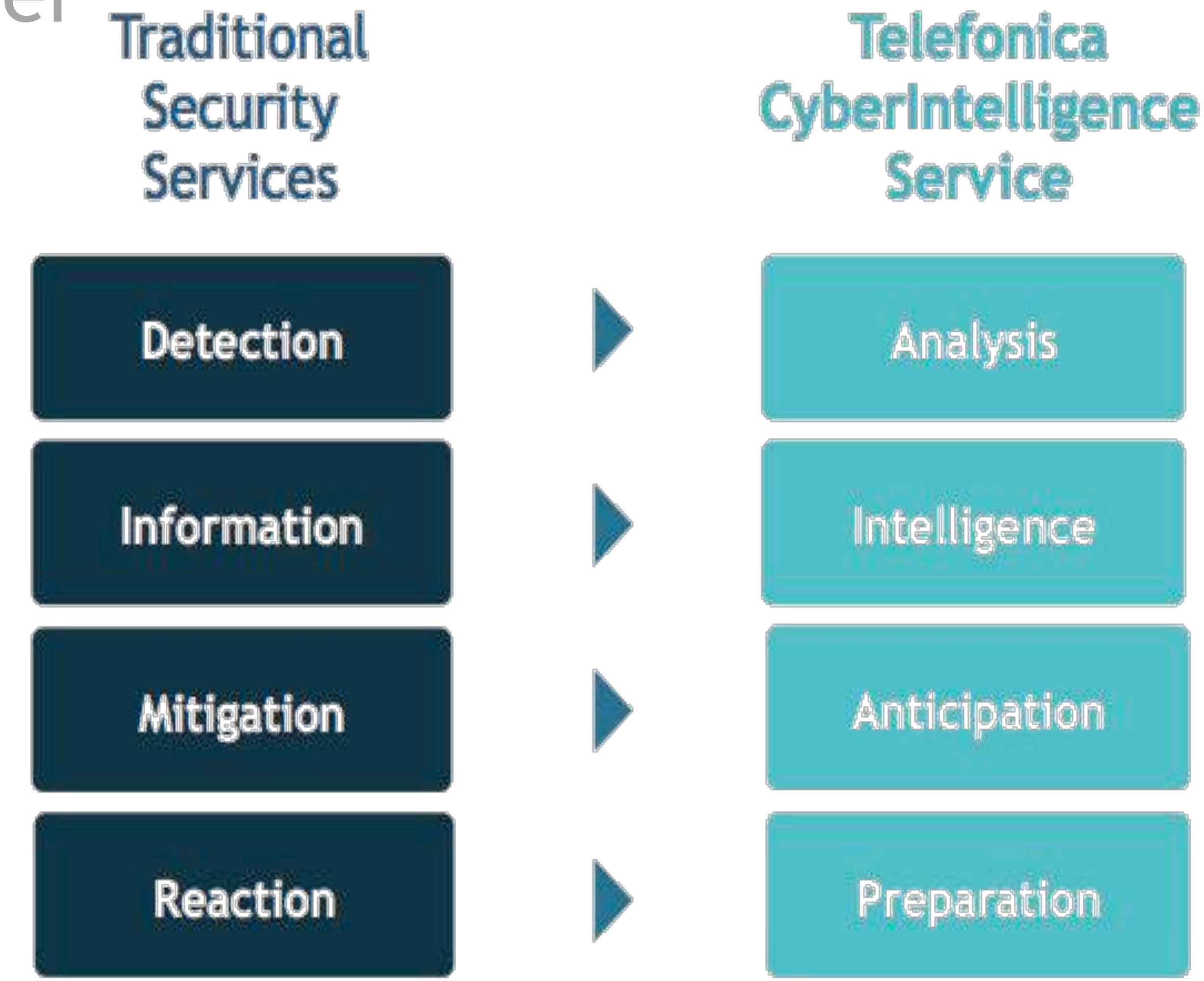
SKYNET
NEURAL NET-BASED ARTIFICIAL INTELLIGENCE
CYBERDYNE SYSTEMS CORPORATION

Context, Behavior, Anomalies: Processing and Storage



Telefonica

Before / After



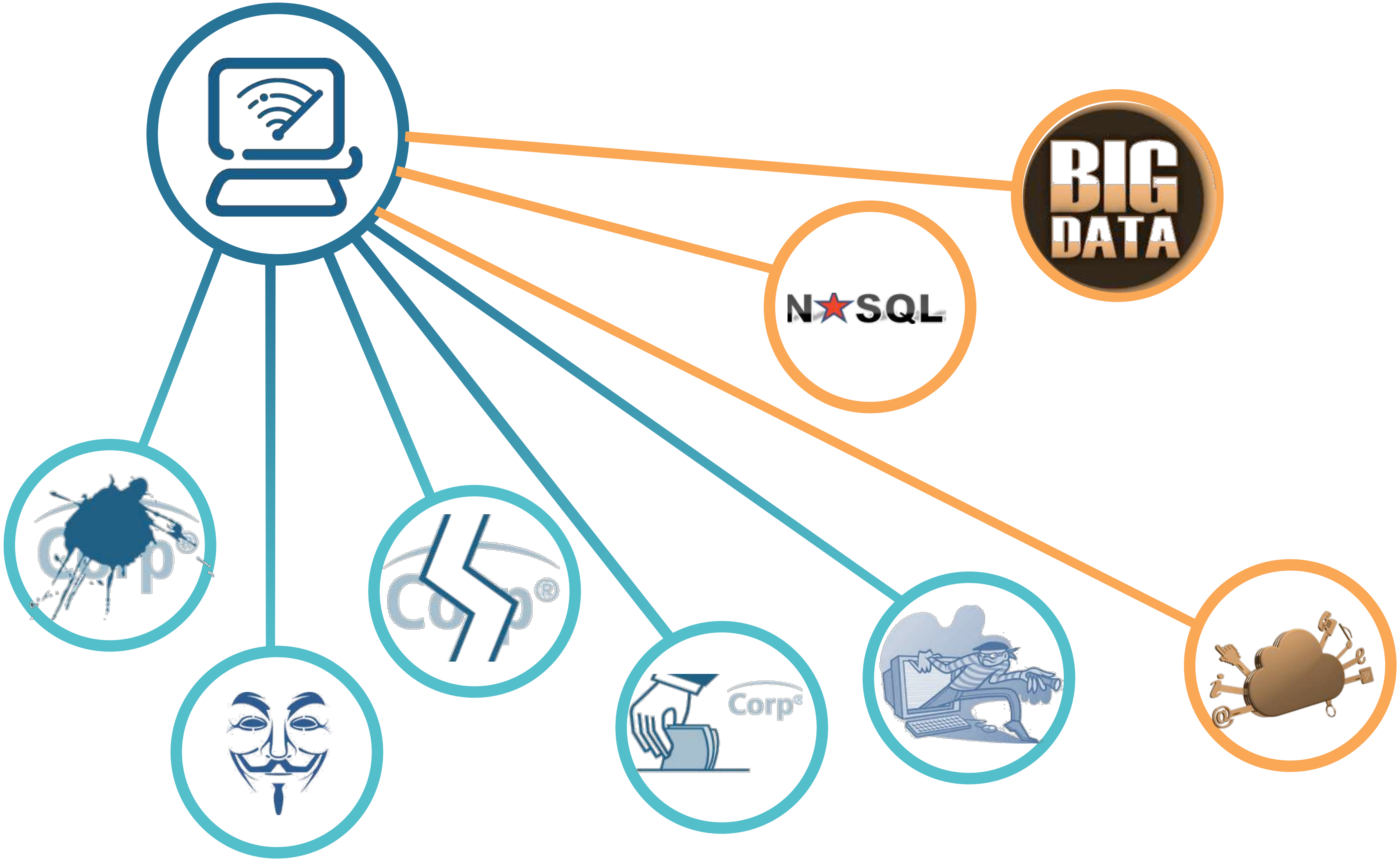
Traditional Security Services	Telefonica CyberIntelligence Service
Detection	Analysis
Information	Intelligence
Mitigation	Anticipation
Reaction	Preparation

CiberSecurity

Our Skills

OUR SKILLS

But...



OUR SKILLS

We need **skills** in:

- Big Data
- Cloud
- NoSql
- …

Detección de Amenazas

Stratio

Whoami

- Oscar Mendez Soto
- CEO Stratio
- CEO Paradigma tecnológico

- Pon aquí lo que quieras!!!

Stratio distribution Spark is:

A unifier data hub

Combine historical data and real time data streaming in a single query

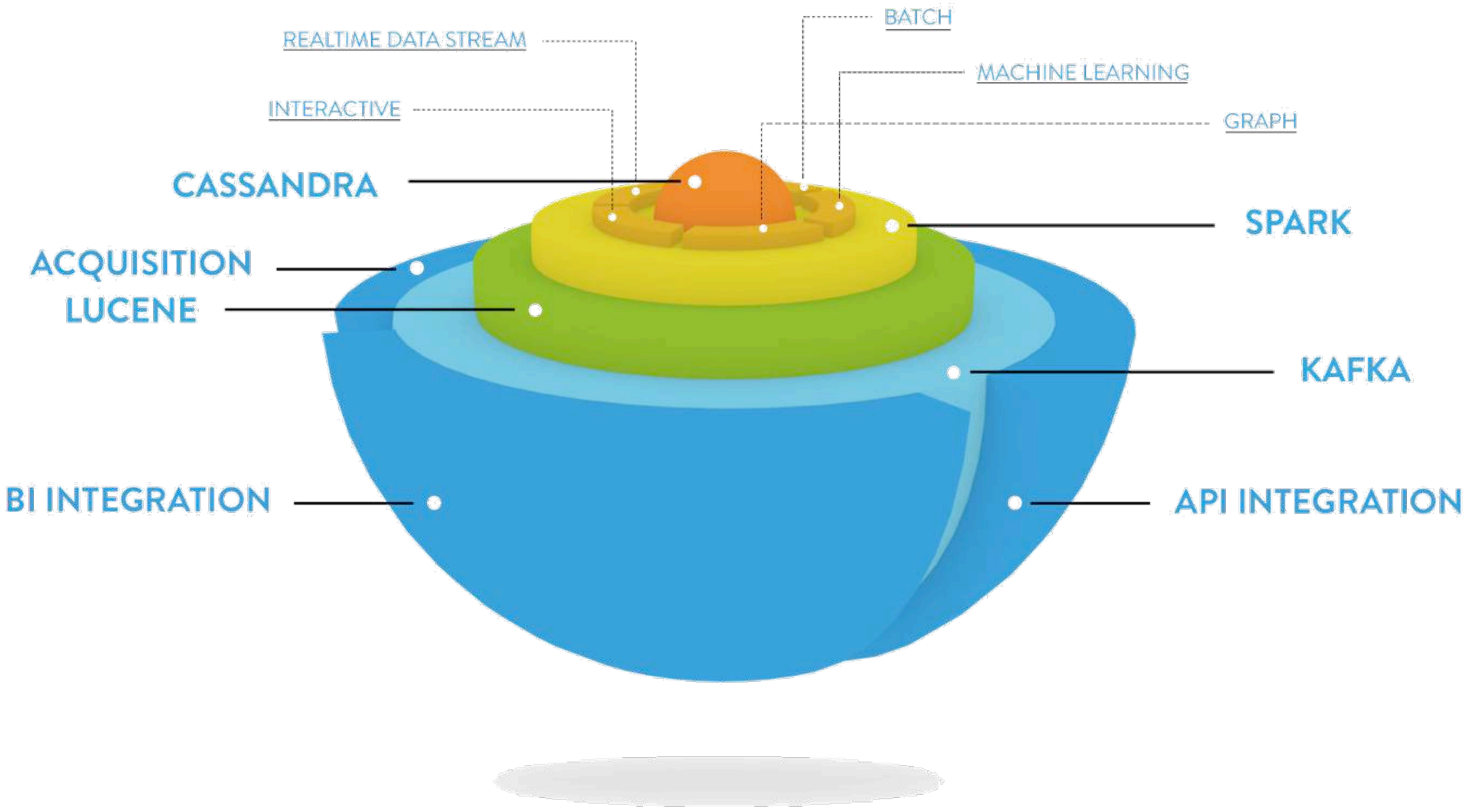
Multi-application Multi-Data

Concentrate all and any type of data into a single Data Hub that allows the implementation of any use case or application

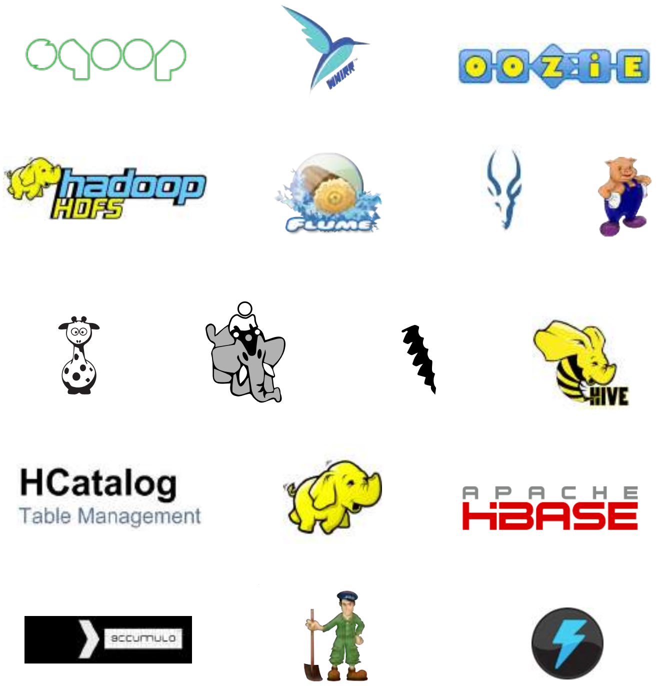
Big Data a child’s play

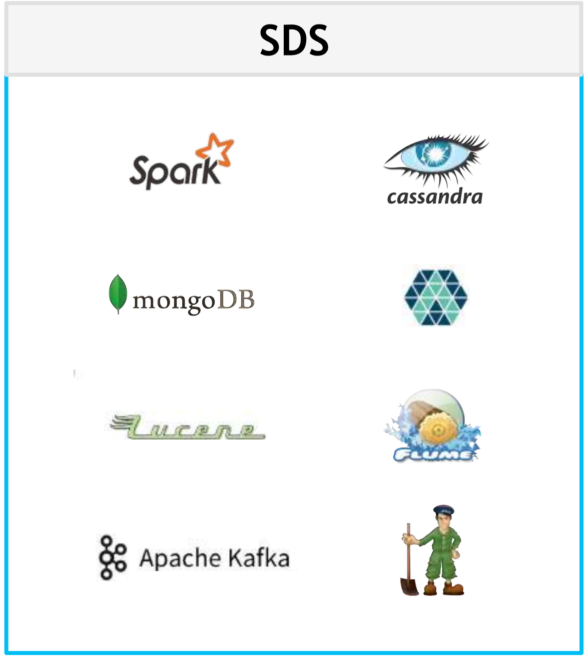
An easy SQL interface to access all the power and capabilities of the platform

Stratio distribution Spark



Fewer components:




SDS

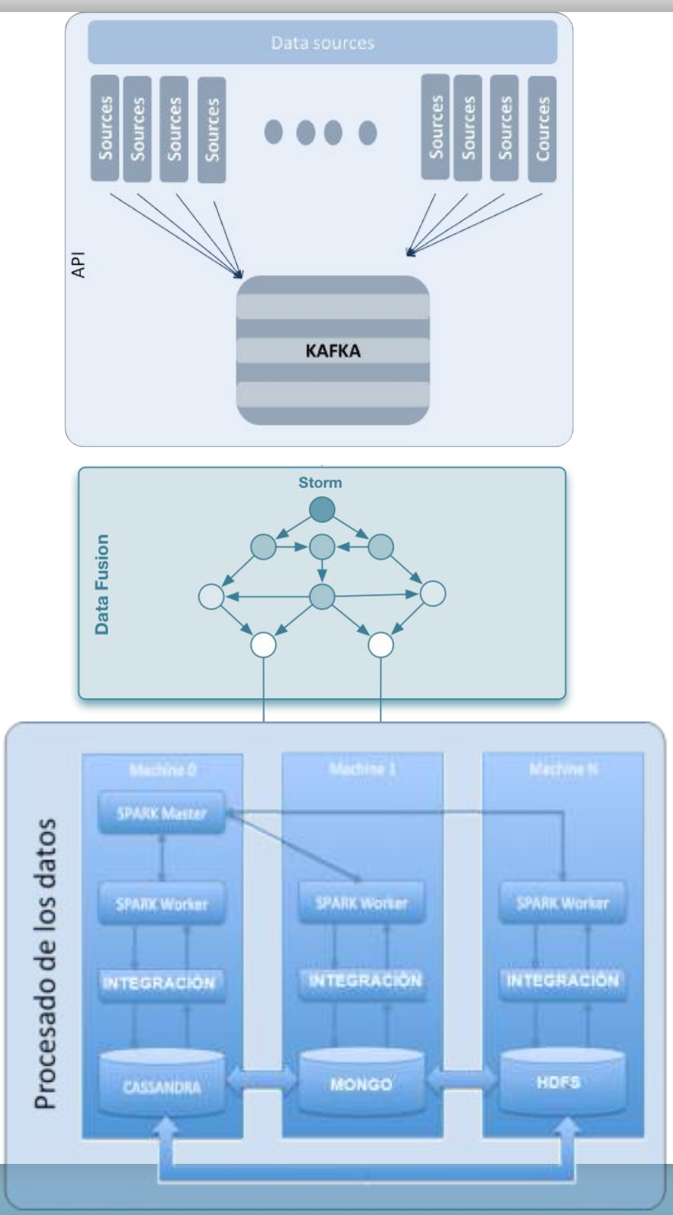
1/2 Components = 1/2 Odds of failure

Our architecture

We adapt our platform to this Telefonica project.

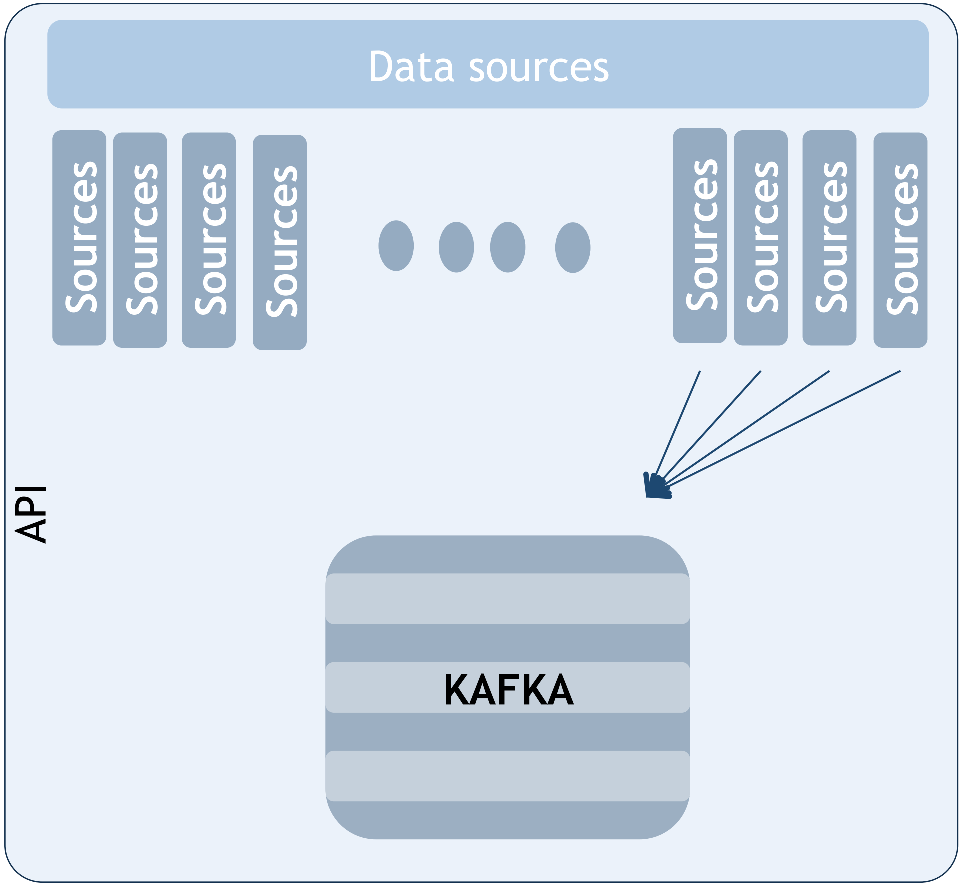
We have three step:

- Ingestion: We use Kafka
- Data fusion: We use Storm.
- Batch: We use Cassandra+Spark



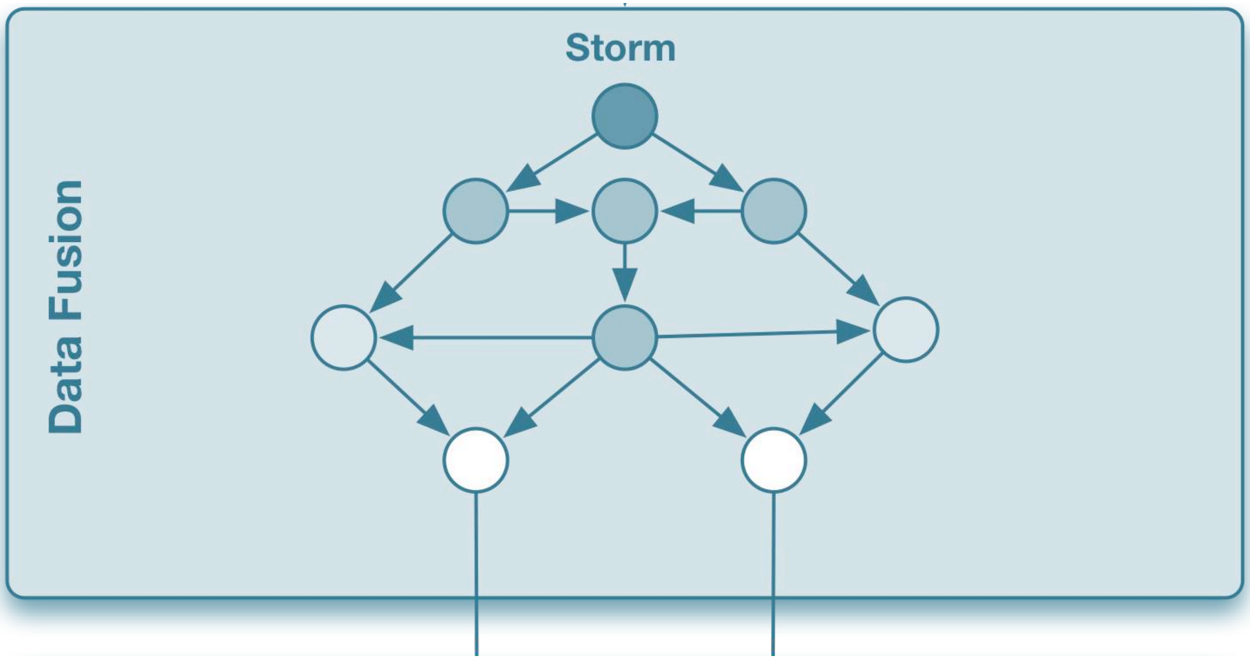
Data Adquisition

- Data are in several sources:
  - DNS traffic
  - IP
  - Social media
  - Underground sources
  - Goverment sources
  - …
- PULL Sources with the information.
- The data are going to Kafka.
- The volume is totally variable.



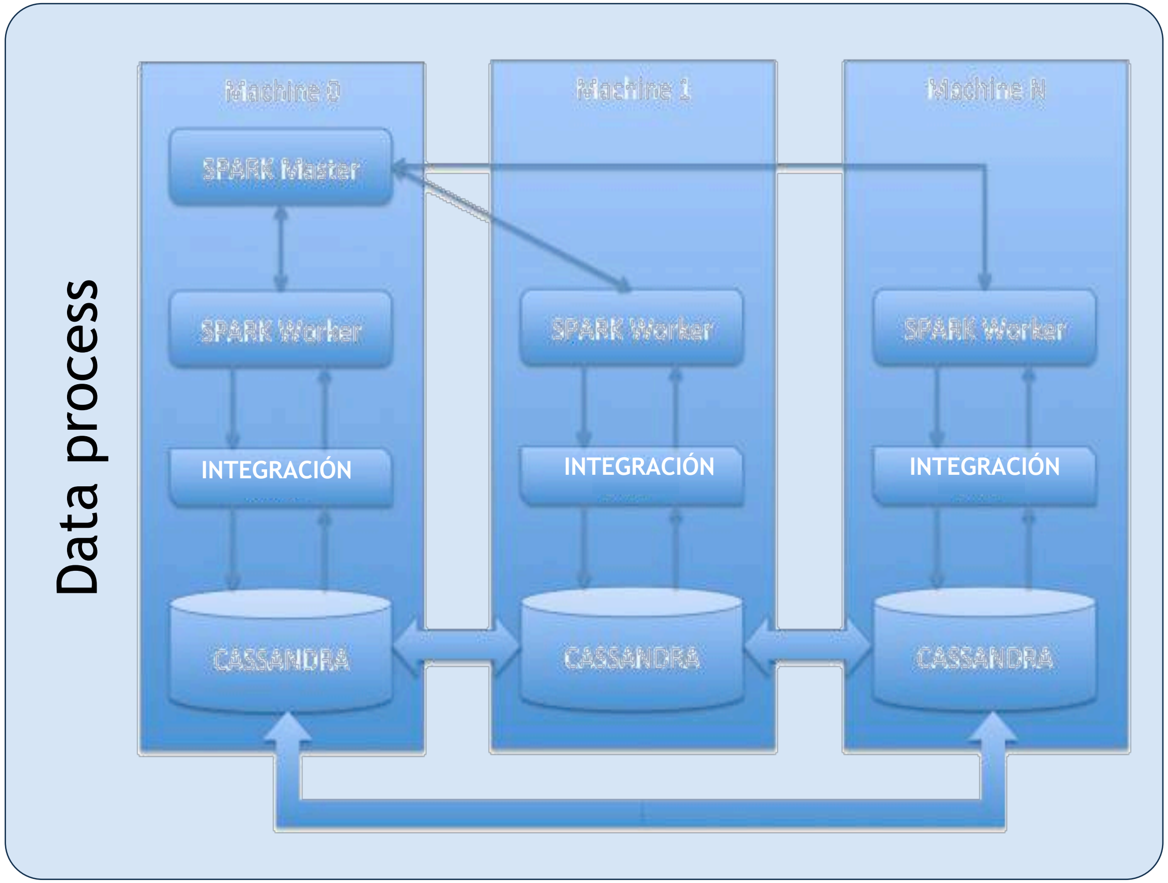
Data fusion

- We use Storm to process and normalized the information.
- The system must to generate alert to the customer.
- This use case required a Big Data component capable of processing the data and extract its information in real-time
  - Warnings and alerts are time-sensitive in order to deal efficiently with security attacks.

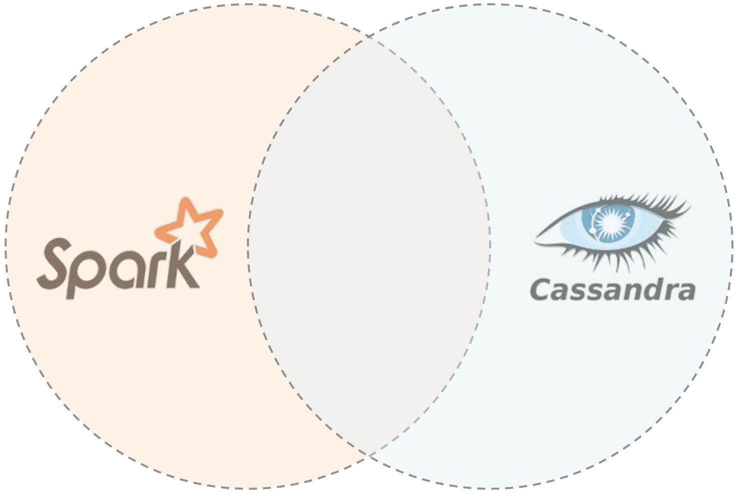


Batch

- The data are saved in Cassandra.
- We use Cassandra direcly for the easy queries.
- And we used Spark to extract the information not accesible to cassandra directly.



Spark+Cassandra

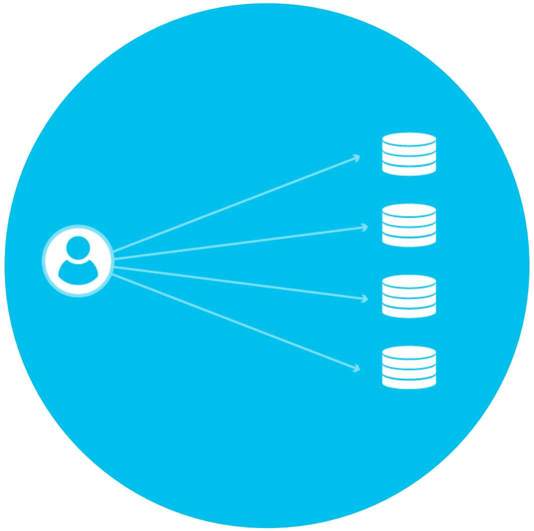


Spark working over Cassandra

"Two plus two is four? Sometimes… Sometimes it is five."

G. Orwell

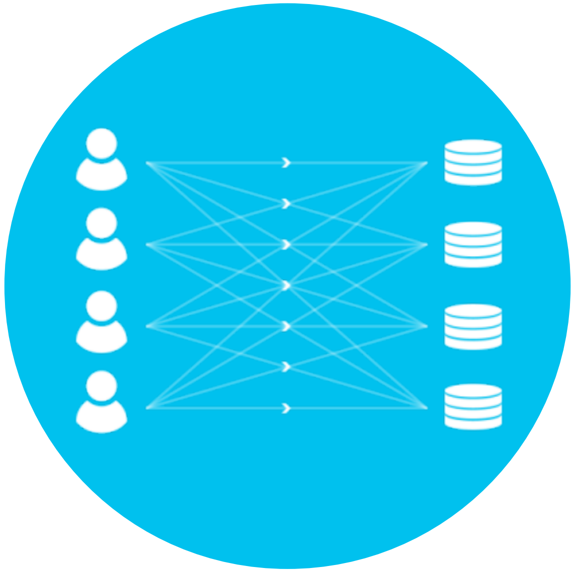
Spark+Cassandra



FEW USERS / HIGH VOLUME OF DATA

Usually when an analyst or a single user wants to access a big data repository, you need to distribute the information because it's too big. The main solution is to use tools based on MapReduce distributed processing like Spark.
But the process is very hard and the cluster can't support many users/many concurrent operations.

Spark+Cassandra

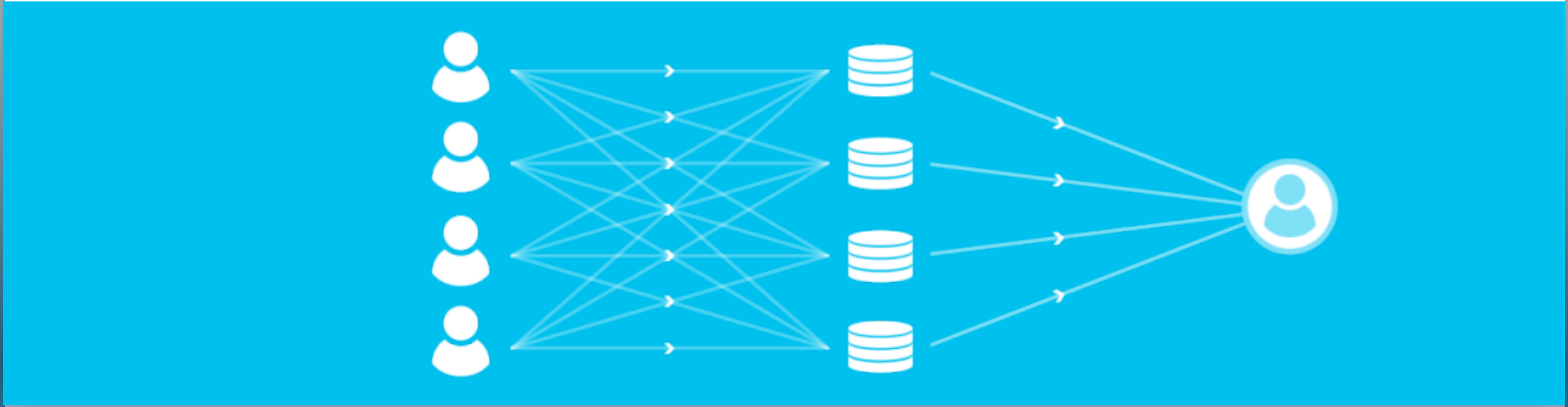


MANY USERS WITH HIGH VOLUME OF DATA

You have many users and a high volume of data. In this case, you must design the queries correctly, because you need a database that is prepared for specific queries. The system will work well for the predesigned queries. This is the perfect case for Cassandra. This case is common in a lot of samples… Machine to machine communication, financial transfers, mobile apps, log monitoring, network sensors, surveillance systems, ad hoc applications…

With Spark-Cassandra we are covering a more complete use case

A lot of users accessing a lot of data from applications or predefined reports. The needs of data analysts that can transform ,analyze, and **query openly** high volumes of data with a more powerful data manipulation tool in their hands.



Spark-Cassandra enables the implementation of any use case or application with any number of users

Cassandra

Applications or dashboards with many users and much data using a predefined set of queries, perfectly solved with Cassandra, using very few cluster resources.

Spark

BI applications or tools with few users (BI analysts or similar) executing open queries, perfectly solved with Spark over Cassandra using the remaining power of the cluster.

With Spark-Cassandra, Spark integrated with Cassandra, you combine the best of both solutions

Spark-Cassandra allows selecting just the initial data Spark needs from the Cassandra data store

With the integration Spark-Cassandra we can leverage the power of Cassandra's main indexes, and especially secondary indexes in order to only and efficiently fetch the data we need

... And moreover

In Spark-Cassandra we have **improved the use of Cassandra's secondary indexes** in order to speed up any interactive query, therefore we have maximum efficiency for the initial recovery of data.

In Spark-Cassandra we have **extended the filters and queries of Cassandra's secondary indexes** to any logical operation or almost any sql sentence

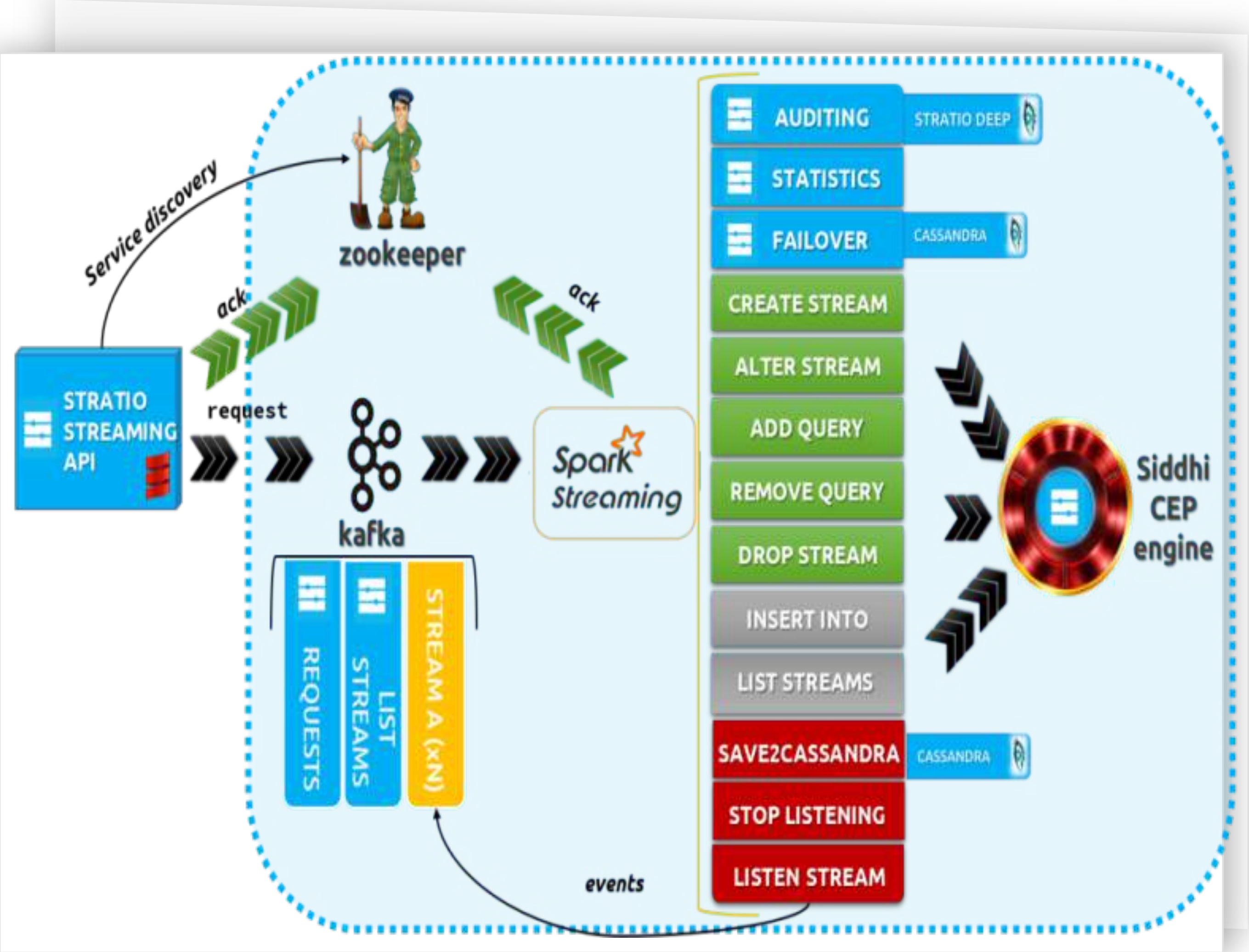
04

Upcoming challenges

Upcoming challenges

Data fusion

- Change Storm to Spark Streaming in data fusion layer.



Upcoming challenges

Use **SQL-Like interface** for analyst queries

An SQL-Like language to simplify the use and combination of historical data and data streaming

SQL-Like language that allows making interactive queries that combine queries over batch/historic data, with queries over real time data streaming in the easiest way.

- SQL + DSL (Linq) abstractions.
- Domain specific language: Business oriented, easy and scalable.
- Evolves and adapts to the business needs of any customer with UDFs (user defined functions).
- Script extension to allow imperative programming for smart scripts.
- Easy interface to all Stratio Dsitribution Spark modules.
- Meta is an SQL abstraction for distributed programming that combines stream and batch process (abstraction of tables and streams).

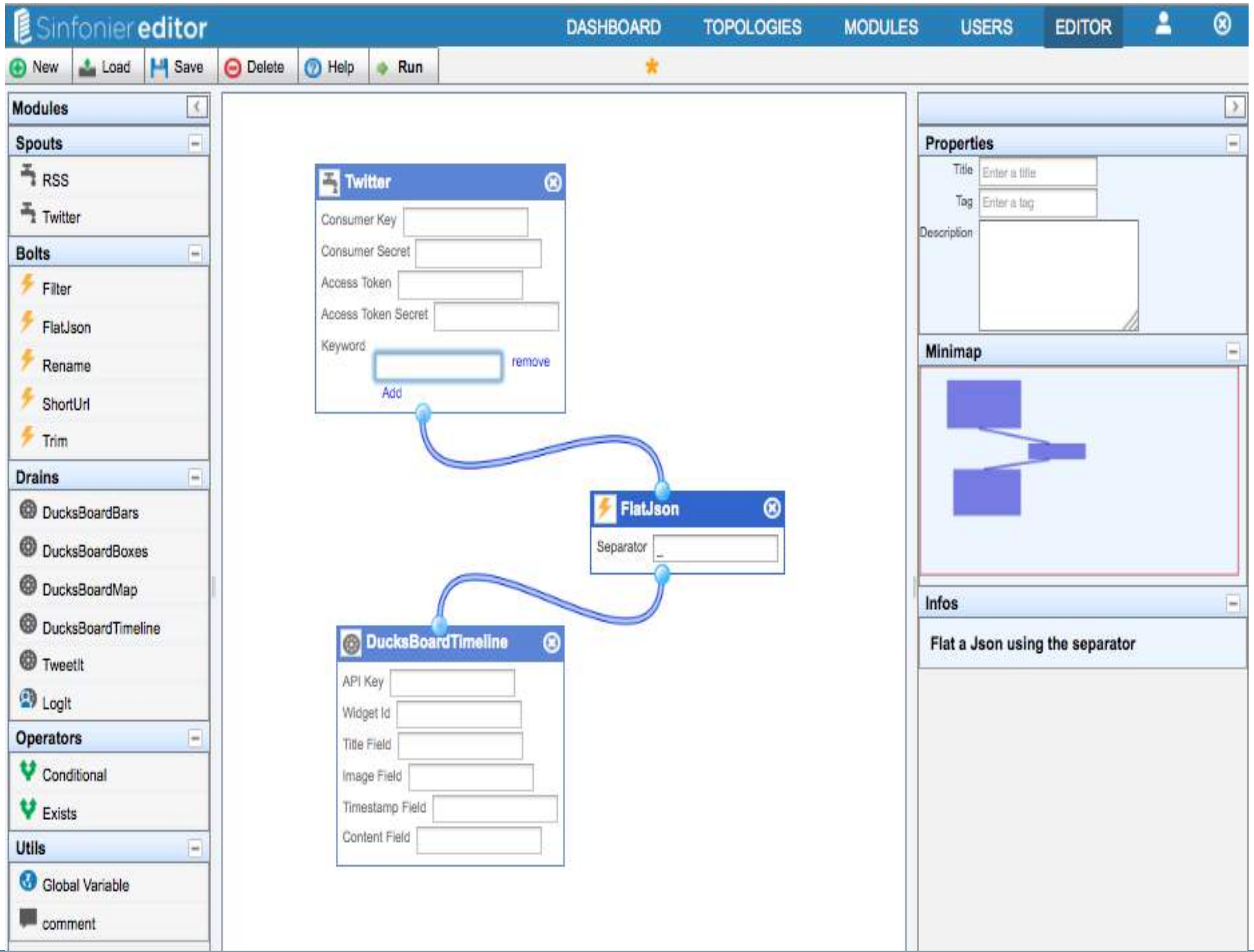
Upcoming challenges

Sinfonier

Upcoming challenges

Sinfonier:

Simplify Building Process



Flat a Json using the separator

Cierre

Telefonica