**Kafka Usage in SES**

For our project we will be using Kafka to receive data, Schema registry to understand those data, Kerberos to authenticate ourselves and have access.

SES use Kafka for a lot of different things, for our purpose we use it to retrieve SES-17 satellites telemetry data, but first let’s take a quick look of what is KAFKA.

1. **Kafka**

* Distributed events store
* Stream processing
* Written in JAVA & SCALA
* **Unified, Hight throughput, low latency, for real time data feeds**
* TCP Based

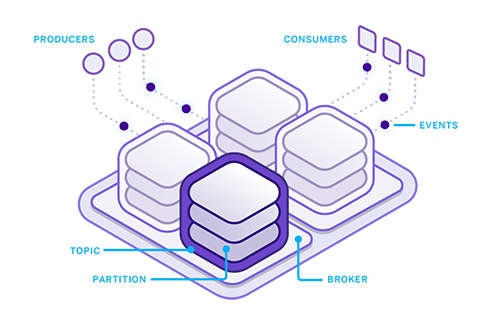
Vocabulary

* *Broker*: One or more server available in Kafka
* *Topics*: Message are stored and published in a category/feed named “topic”
* *Partitions*: Leader or replicas of a topic responsible for all writes and reads in a topic
* *Cluster*: Group of computers, each having one instance of Kafka broker
* *Message*: Contains the data and identified by a partition and offset number

Major API

* *Producer* (publish stream)
* *Consumer* (subscribe to topics)
* *Connector* (link topic to application)
* *Streams* (convert input into output)
* *Admin* (manage kafka)

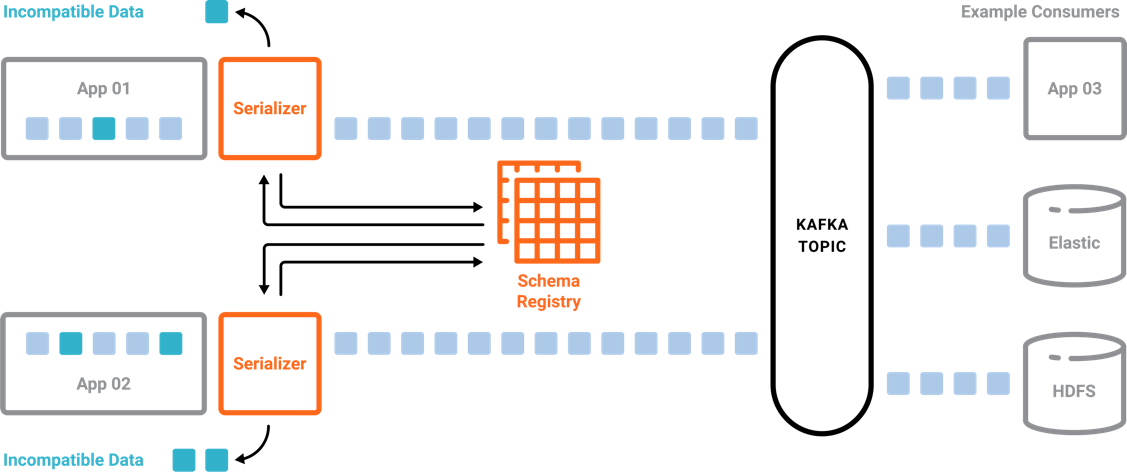
*It is possible to write your own consumer client in any programming language.*



1. **Schema Registry**

The Confluent Schema Registry component is an open-source component initially developed by Confluent. The Schema Registry allows the centralized management of all the schemas of a data platform and thus a better data governance. These schemas are used in the message serialisation and deserialization processes by the platform's clients.

* Lives outside and separately from a Kafka Broker
* Consumer and Producer of Kafka still talk to Kafka to read and publish data, next to that they can talk to schema registry to send and retrieve schemas that describe the data models for the message
* Each schema is registered by a unique ID (monotonically increasing and unique, but not necessarily consecutive)
* Serialize Data, most of the time Avro schemas are used in JSON format



Further Reading: (French tutorial)

<https://blog.ippon.fr/2019/11/18/confluent-schema-registry-un-premier-pas-vers-la-gouvernance-des-donnees>

1. **Kerberos**

According to Greek mythology Kerberos (Cerberus) was the gigantic, three-headed dog that guards the gates of the underworld to prevent the dead from leaving.

So when it comes to Computer Science, Kerberos is a network authentication protocol, and is currently the default authentication technology used by Microsoft to authenticate users to services within a local area network.

* Client: a user, a service of any software.
* Server: where the kerberos protected resource/service is hosted
* Key Distribution Center (KDC): The trusted third-party authentication service
* Server: where the kerberos protected resource/service is hosted
* Key Distribution Center (KDC): The trusted third-party authentication service

