# Kafka streams

Everything you do by a kafka consumer can be done kafka stream

Stream is nothing but a continuous flow of data (sequence of small data packets)

Its unbounded ever growing data, like lorry sending the gps positioning data continuously

When to use kafka streams

1. To process the data at run time and to identify fraudulent transactions in those cases we should use streams

Examples:-

1. Sensors sending transportation data means lorrys sending continuous latitude and longitude positions to kafka systems
2. Video streaming , online gaming
3. Transactions coming from Stock market like stock price updates will be sent to each and every second
4. Ecommerce orders
5. Data feeds such as social media activities, security and threat systems

Kafka streams is a separate processing module of kafka

* Kafka streams is a separate java library specifically designed for sole purpose of stream processing
* Input data must be in a topic
* Out of the box parallel processing capability, scalability, fault tolerance
* Working with streams tables, and interoperating with

them.

* I mean, you can mix and match your solutions with streams and tables and you can even convert a streamto a table and vice versa.
* It allows you to group your streams and compute continuously updating aggregates.
* You can join streams, tables, and a combination of both.
* You can create and manage fault tolerant, efficient local state stores. Creating windows of different types and dealing with all the time domain complexities such as even time, processing time, latecomers, high watermark, exactly ones processing etc.
* It also allows you to serve other microservices using request/response interface over and above your streams application.
* This feature is also known as Kafka streams interactive query.
* It gives you a set of tools for unit testing your application.
* The library gives you an easy to use DLS and also offers flexibility to extend and create your custom processors beyond what is provided.
* You can get an inherent fault tolerance and dynamic scalability.
* You can deploy your stream processing applications in containers and manage them in the Kubernetes Cluster.

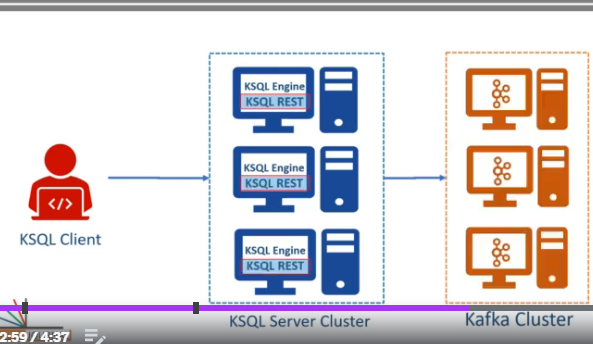
KSQL

KSQL is an SQL interface for kafka streams,

Ksql runs in 2 modes ,

1) interactive mode – having CLI run the query and get immediate response same like other sql interface

2) Head less mode – its non interactive – mostly used in production which allows to submit files and gets response later



Its same like plsql engine =KQSL Engine

## **KSQL allows you to**

In this scenario topic is considered as a table and you can do all operations what u can do on a table.

* You can do group by and aggregates on your Kafka topics.
* You can group an aggregate over time window.
* You can apply filters.
* You can join two topics
* You can sink the result of your query into another topic.

The possibilities are endless, and all you need to know and learn is the same SQL and some Kafka flavor to it.

So the KSQL for Kafka is one big step forward for Kafka to become a real time data warehouse.

Those days are not too far when you might see JDBC/ODBC connectors being available for KSQL and visualization tools like Tableau and QlikView to start connecting with the KSQL.

I'm not predicting anything, but that's the next logical step for Kafka and KSQL.

Great.