# System Architecture

Kafka Stack is the System Architecture considered for designing the data infrastructure on the cloud for a company whose main business is in processing images. Within the architecture, the Azure CosmoDB is considered as it provides On-the-fly scaling, Auto Indexing, and it supports Stored Procedures and Transactions which is a relevant tools for processing frequent Structured Query Language (SQL), and would be helpful for future plan to automate the extraction of key statistic process. The Architecture consists of components such as the Broker, Topic, Consumer, and Producer. A short description of what each of these components performs is as provided below.

* Broker
* A broker is a server that stores messages sent to the topics and serves consumer requests.
* Topic
* A topic is a queue of messages written by one or more producers and read by one or more consumers.

Producer

* A producer is an external process that sends records to a Kafka topic.
* Consumer
* A consumer is an external process that receives topic streams from a Kafka cluster.
* Client
* Client is a term used to refer to either producers and consumers.

# Process Flow

## Uploading of Images

In the daily operation, whenever customer/client uploads stream of images, the process first goes through the Load Balancer, followed by the Front-End Upload. After which, the Image Processing (Producers) sends these images to the Kafka Cluster through Node.js gateway, which is then picked up by one of the available Kafka Broker (Node Worker) at the point of time. Lastly, the Kafka Broker interacts with the CosmoDB to inserts these images over to the database.

## Retrieval of Images

In the daily operation, when Analyst would like to retrieve the current stream of images available within the CosmoDB. The retrieval of Images is handled by the Kafka Broker, where the stream of images is sent across using a Kafka stream to the Front-End Image Stream (Consumers). Lastly, these images would be made available to the Analyst for download.

## Business Intelligence Data Analytic

The extraction of key statistics which includes the number and type of images processed, and by which customers is powered by the Power BI, which interacts with the CosmoDB for the available dataset by establishing connection between it and perform the analytic at the background process. The Analyst is just required to interact with the Power BI to perform the extraction.

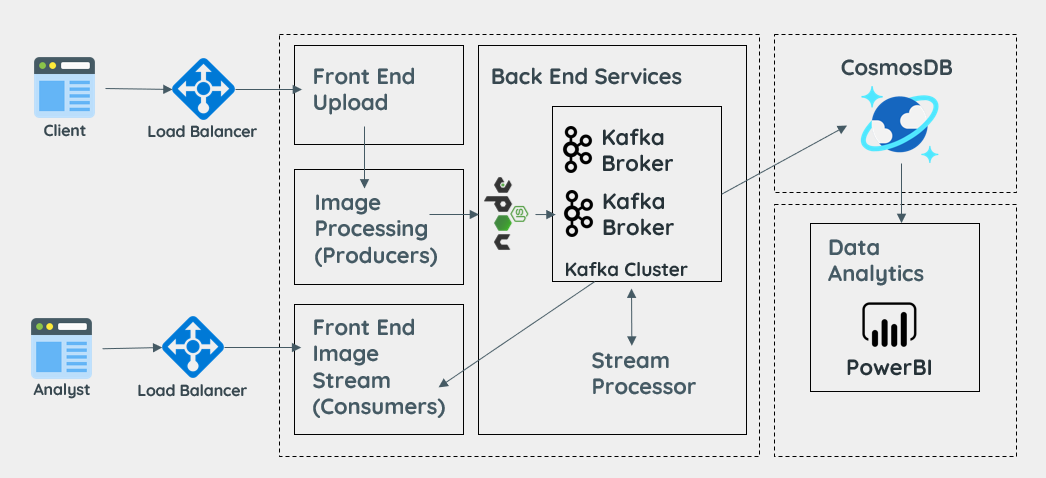


Figure 1: Kafka Stack Architecture

# Assumptions

1. In the future, if the company would like to perform other types of Data Visualisations.
2. The sizes of stream of images varies and the storage requirements to stores these images would be expected to be increase exponential overtime.
3. New Analyst would come join the company and has little to minimal knowledge on Data Analytics.