1. What is Apache Kafka

1. 
2.   
   **It has following two meanings**:   
   A picture containing text

   Description automatically generatedRealtime Stream of Data Pipeline.
3. **Real-time Example**: Smart Meters generating data related data every minute and sending to **Kafka Server**.  
   Text

   Description automatically generated with low confidence
4. Graphical user interface, text, application

   Description automatically generated
5. 
6. Apache Kafka adopted **Pub-Sub Messaging System Architecture** and works as an **Enterprise Messaging System**.
7. **Three components of a messaging system** 🡺 Producer(Publisher), Broker, & Consumer (Subscriber).

1. Graphical user interface, application

   Description automatically generated  
   A screenshot of a computer

   Description automatically generated

Kafka works as **Pub-Sub Messaging System** where we create a **Producer App** to create and send data **as a stream**.   
We install and configure **Kafka Server** to act as **a message Broker**.  
Finally, we create **consumer app** to process the **data stream** in **realtime**.

1. 
2. Kafka was initially conceptualized and developed at **Linkedin** and later on **open sourced** in 2011 to solve Data Integration
3. A diagram of a network

   Description automatically generated
4. A diagram of a computer system

   Description automatically generated
5. A close-up of a computer screen

   Description automatically generated
6. 
7. Timeline

   Description automatically generated

2. Apache Kafka Core Concepts

1. Text

   Description automatically generated with medium confidence
2. Producer is app sending data (msg, msg record) from small to medium sized.
   1. Msg may have any record structure schema but for kafka it is an **array of bytes.**
   2. **Example**:  
      A picture containing text

      Description automatically generated  
      Other examples may be like sending a table or result set from a query.   
      So, basically create appropriate Producer App as per your requirement.
3. **Consumer**:
   1. Recipient. Producer doesn’t send directly to consumer but via Broker. A close-up of a diagram

      Description automatically generated
4. **Broker**:
   1. Kafka Server as an agent or a broker to exchange messages.
5. **Cluster**:
   1. Kafka Cluster is a group of computers each running an instance of Kafka Server/Broker for a single purpose.
6. **Kafka Topic**:
   1. **Examples**: Current Load, Consumed Units, Input Fluctuations which one you want.
   2. Arbitrary but unique Name given to a data set. Think like a DB Table Name which is design time decision.
   3. Diagram

      Description automatically generated with low confidence
7. **Partitions**: A Smallest and independent portion of a topic
   1. A topic Size may be more than the size of a Kafka Server/Broker.
   2. **Solution**: P1 + P2 + … + Pn = Topic.  
      Where each partition will be assigned to one Kafka Server on a Kafka Cluster. Definitely we need Kafka Cluster.
   3. No of partition is design time decision by an architect.
8. **Partition Offset**:
   1. A screenshot of a computer

      Description automatically generated
9. **Consumer Group**:
   1. A group of consumers to share a workload.