Appache Kafka:

- An open-source distributed event streaming platform designed to handle real-time data feeds with high throughput, fault tolerance, and scalability.
- Widely used for building real-time data pipelines and streaming applications.

How Kafka Assists in Making Streaming Jobs:

- Real Time Data Ingestion
- Scalability
- Fault Tolerance
- Low Latency Rate

What Cases We Prefer Streaming Jobs?

- For applications like fraud detection, real-time recommendation engines, and monitoring systems, where decisions need to be made immediately based on the

latest data.

- For IoT data processing, real-time user activity tracking, or microservices that need to respond to events in real time.

- When building ETL (Extract, Transform, Load) pipelines where data needs to be processed and loaded into a data warehouse.

Key Kafka Components:

- **Brokers**: Servers that store and manage data, distributed across a Kafka cluster.

- **Zookeeper**: Manages broker coordination and metadata (being replaced by KRaft).

- **Kafka Server**: The component running

Kafka services, part of a distributed system.

- **Kafka Clients**: Applications that produce and consume data.

- **KRaft**: Kafka's new consensus protocol replacing Zookeeper.

- **Kafka Cluster**: A group of brokers working together to provide scalability and fault tolerance.