**Tools Used:**

Kinesis Data stream and Firehose

**Why Kinesis for real time Streaming:**

Kinesis used for processing high volume of data.

Kinisis can continuously capture gigabytes of data per second from hundreds of thousands of sources such as website clickstreams, database event streams, financial transactions, social media feeds, IT logs, and location-tracking events.

Amazon Kinesis enables to process and analyze data as it arrives and respond instantly instead of having to wait until all data is collected before the processing can begin.

**Technical overview:**

Kinesis acts as a highly available conduit to stream messages between data producers and data consumers. Data producers can be almost any source of data: system or web log data, social network data, financial trading information, geospatial data, mobile app data, or telemetry from connected IoT devices. Data consumers will typically fall into the category of data processing and storage applications such as Apache Hadoop, Apache Storm, and Amazon Simple Storage Service (S3), and ElasticSearch.

It has 2 service offerings, data Stream and Firehose stream.

The main difference between these 2 is :

Firehose can scale to gigabytes of streaming data per second, and allows for batching, encrypting and compressing of data. It should be noted that Firehose will automatically scale to meet demand, which is in contrast to Kinesis data Streams, for which we must manually provision enough capacity to meet anticipated needs.

**Use Case/Project requirement:**

SAP (Trace) ECC metadata has been received through TIBCO on the edge node in real time. Data has to be processed through real time stream and pass it S3 or store to RedShift in order to send the reminder (Alert) to the Admin and Lead users.

**Process Flow:**

Input Data -> Kinesis data Stream -> Kinesis Data Analytics -> Output to BI Tools

Input has been captured and send to Amazon Kinesis stream . Kinesis Data Streams used to continuously collect data and build analytics on top that and produce output to BI tools.

Input Data -> Kinesis Firehose -> Data Storage(Redshift/Amazon S3/Splunk/Amazon search Engine) -> Output to BI Tools

Input has been captured and send to Kinesis Firehose . This Firehose prepares and loads the data continuously to the destination(S3/Redshift). From this it produce output to BI tools.

**Lessons Learnt:**

Learned how to use Amazon Kinesis to get real-time data insights and integrate them with Amazon S3. Learned current architecture that enabled the move from a batch processing system to a real-time system overcoming the challenges of migrating existing batch data to streaming data and how it benefits in real-time analytics.