**Introduction:**

In today’s ‘Do it Now’ world, high velocity, veracity and huge volume of data from enormous sources having impact on real time decision making. Data from multiple structured/unstructured sources and varieties of system language deter the pace of information processing which eventually blocking organisational prompt decision making and action taking process.

The scope of the project is to design a system incorporating universal query language and real time event streaming architecture, which aims to speed up Big Data management and data processing system to make an organization from reactive to proactive decision maker.

**Timeline:**

Need to decide and finalize

**Purpose of interim report:**

Purpose of interim report is to show the process of the project work. In this interim report, we would like to represent our progress on proposed architecture or design of Big Data Management and processing system segregating on two part - 1. Universal query language and 2. Real time event streaming system.

**Progress:**

1. Universal Query Language:

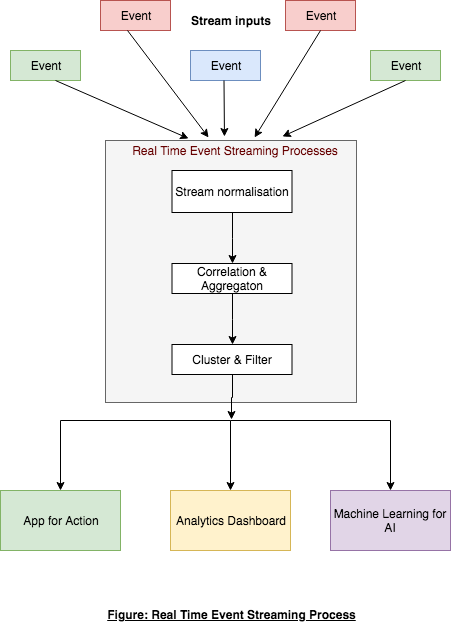
Need to include this part

1. Real time event streaming architecture:

Traditional system collect data from multiple sources and store it in RDBMS and hadoop cluster. From this storage system users needs to process/query to utilise these data for further action. However, such traditional method makes the value of the stored data low as it takes huge processing time. Such slow data processing system affect proactive analysis for event streaming data like preventing consumer churn, identify system failure or fraud detection.

Within the system, all the streaming data from different sensors, mobile phones or third party system are stored in messaging infrastructure. We just do not know which part of data are useful within our storage or how potential our dark data is.

Therefore, to overcome this and to work with only with the potential data rather spending time and processes on unnecessary data we can introduce a system which will normalize, aggregate, cluster and filter to the necessary data and feed it to applications. Following figure is a draft proposal for such real time event processing system -



In the first process, incoming stream input from multiple sources will be normalized to be feed to the next step. In next step, streaming data will be aggregated and correlated. In the last step, all the aggregated data will be clustered according to necessity of the data. In this step, all the dark data can be clustered or correlated in the same category. From this step users will get filtered data regardless unnecessary data to be fed into his system like Apps, Dashboard or ML system.

Key feature of the architecture:

1. Fast access to live data.
2. Scalable.
3. Fault tolerance.

**Task remaining**

**Conclusion**