**Review of Real time data processing**

# **Summary**

Real time data processing powers real time reporting of the aggregated, anonymized voice of Facebook users, analytics for mobile applications, and insights for Facebook page administrations. This ecosystem can handle hundreds of Gigabytes per second across hundreds of data pipelines.

In the design of a real time data system, there are five important points: Ease of use, Performance, Fault-tolerance, Scalability, and Correctness. They use Facebook’s Puma, Swift, and Stylus stream processing systems here.

**Contributions**

①Targeting seconds of latency. Seconds latency allows us to use a persistent message bus for data transport and it paved the way for fault tolerance, scalability, and multiple options for correctness in the stream processing systems Puma, Swift, and Stylus.

② Ease of use is as important as the other qualities.

③ By providing choices along the spectrum of correctness, the application builders can decide what they need.

④ Puma provides pre-computed query results for simple aggregation queries.Puma provides filtering and processing of Scribe streams. Swift provides checkpointing functionalities for Scribe. Stylus provides a function to estimate the event time low watermark with a given confidence interval.

**Comments**

**Flash Points:** ①First, targeting seconds of latency, not milliseconds, was an important design decision. Seconds is fast enough for all of the use cases we support.②ese of use.③ there is a spectrum of correctness.

**Limitations:** ① Improving the dynamic load balancing for the stream processing jobs.

② Considering alternate runtime environments for running stream processing backﬁll jobs.