



zynga

Realtime Analytics Via Stream Processing

Michael(Analytics Platform)

Rushan(Analytics Platform)

Agenda

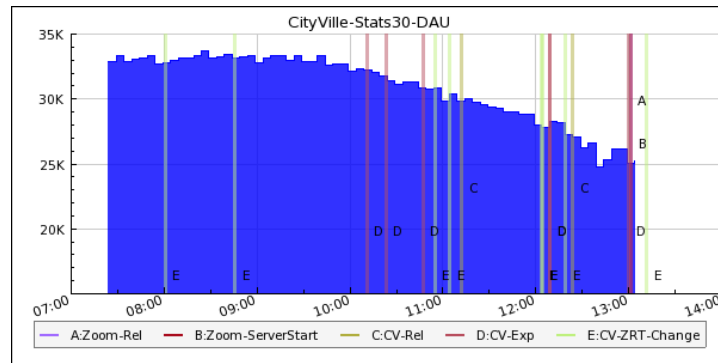
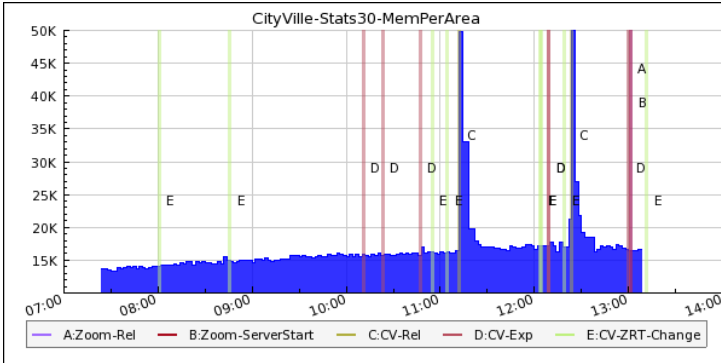
- **What is realtime analytics?**
- **How does streaming realtime analytics work?**
- **Zynga Streaming 3.0**
- **Challenges and Future Work**

What is Realtime Analytics?

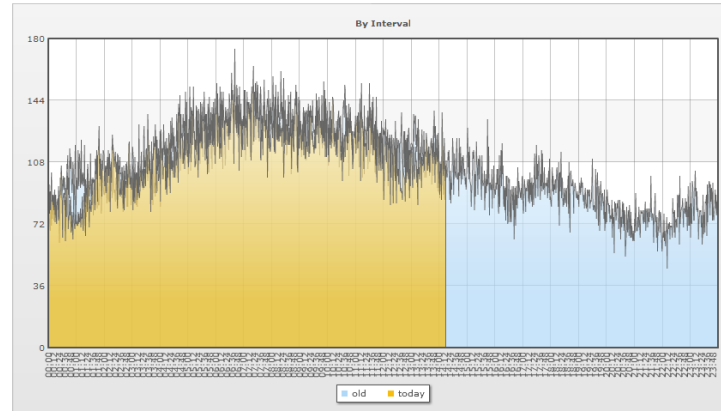
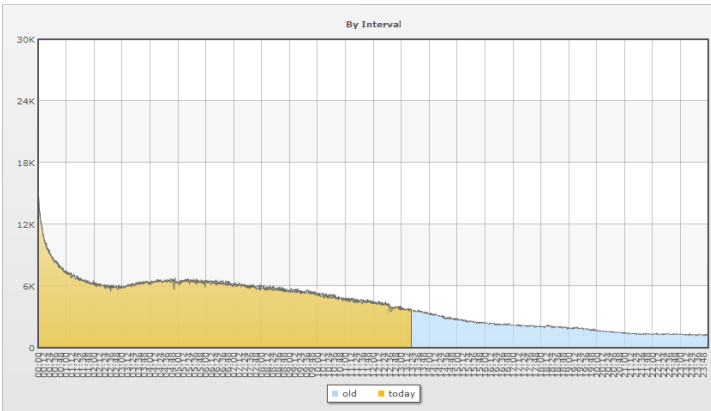
- **Real-time**
- **Aggregation based time window**
 - Continuous time window
 - Batched time window
- **Scale with data volume and throughput**
- **Publish-Subscribe Event Model**
- **Some examples:**
 - how many users installed by different source in the last minute?
 - What is the average number of messages sent in the last minute?
 - Aggregation is computed by all hierarchical levels – e.g. Kingdom, Phylum, Class, Family, and Genus

What do we use it for?

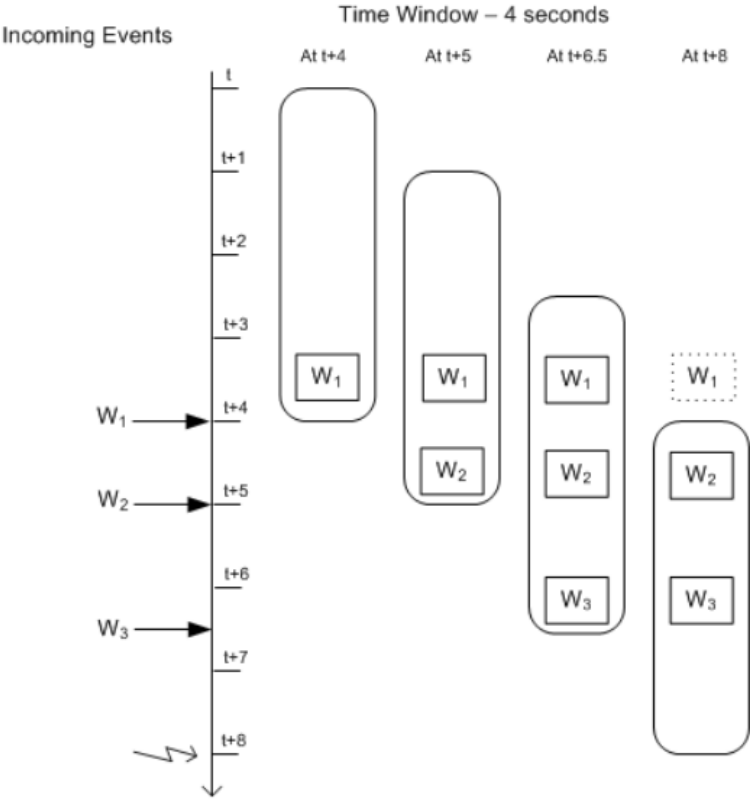
➤ Monitoring and Alerting



➤ Streaming reports



How does streaming time window work



Streaming 2.0(S2)

➤ Built on Esper – Open Source Project

- Time based Event processing
- High throughput
- Low latency
- Complex computation

➤ Limitation

- Esper does not scale – limited by local server memory.
- JVM memory tuning can be hard

Streaming 3.0(S3)

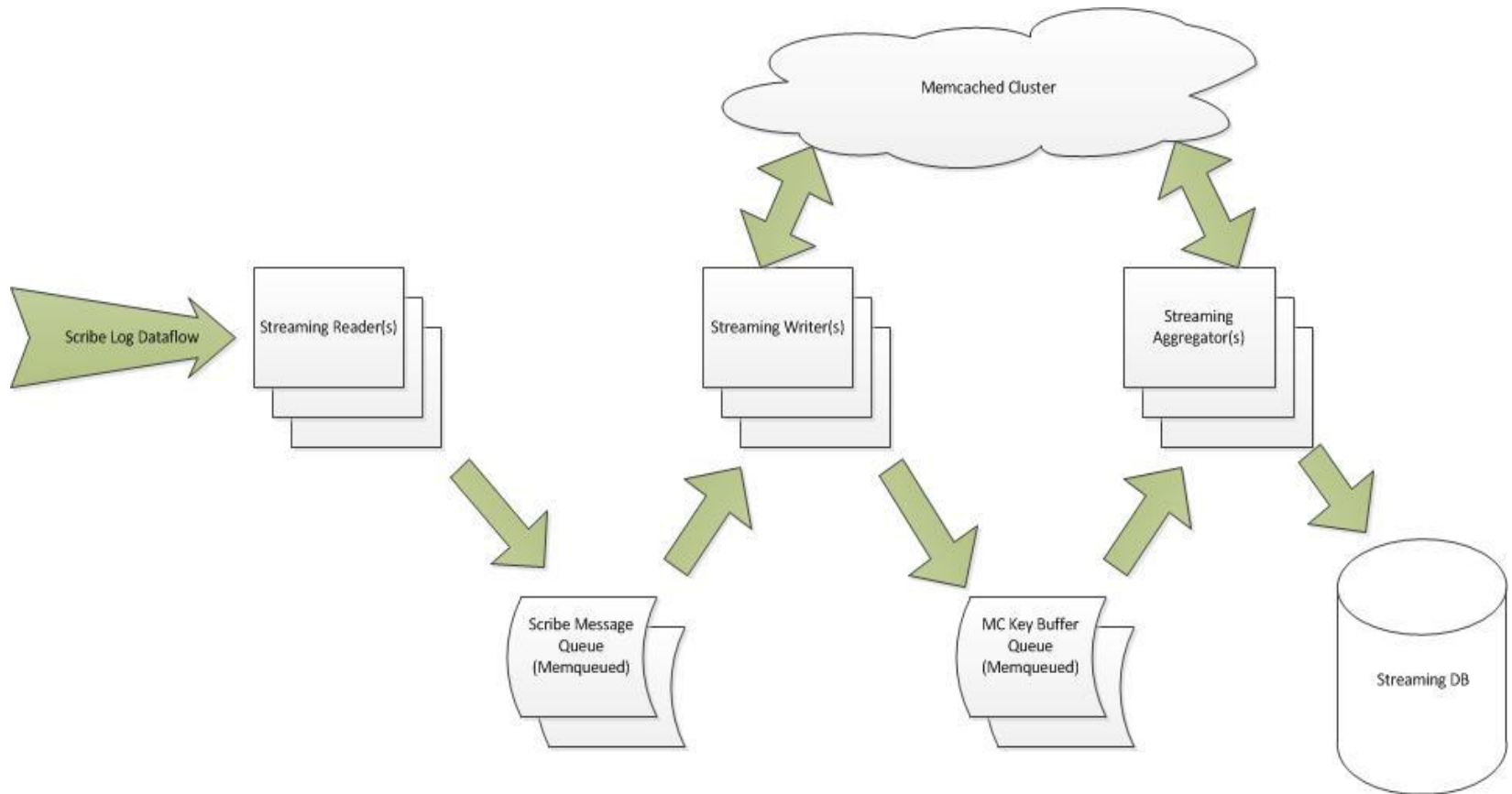
➤ Built in House

- Scales using exist technologies, and build whatever is needed
- Easy to debug, upgrade, and enhance
- Features built on demand

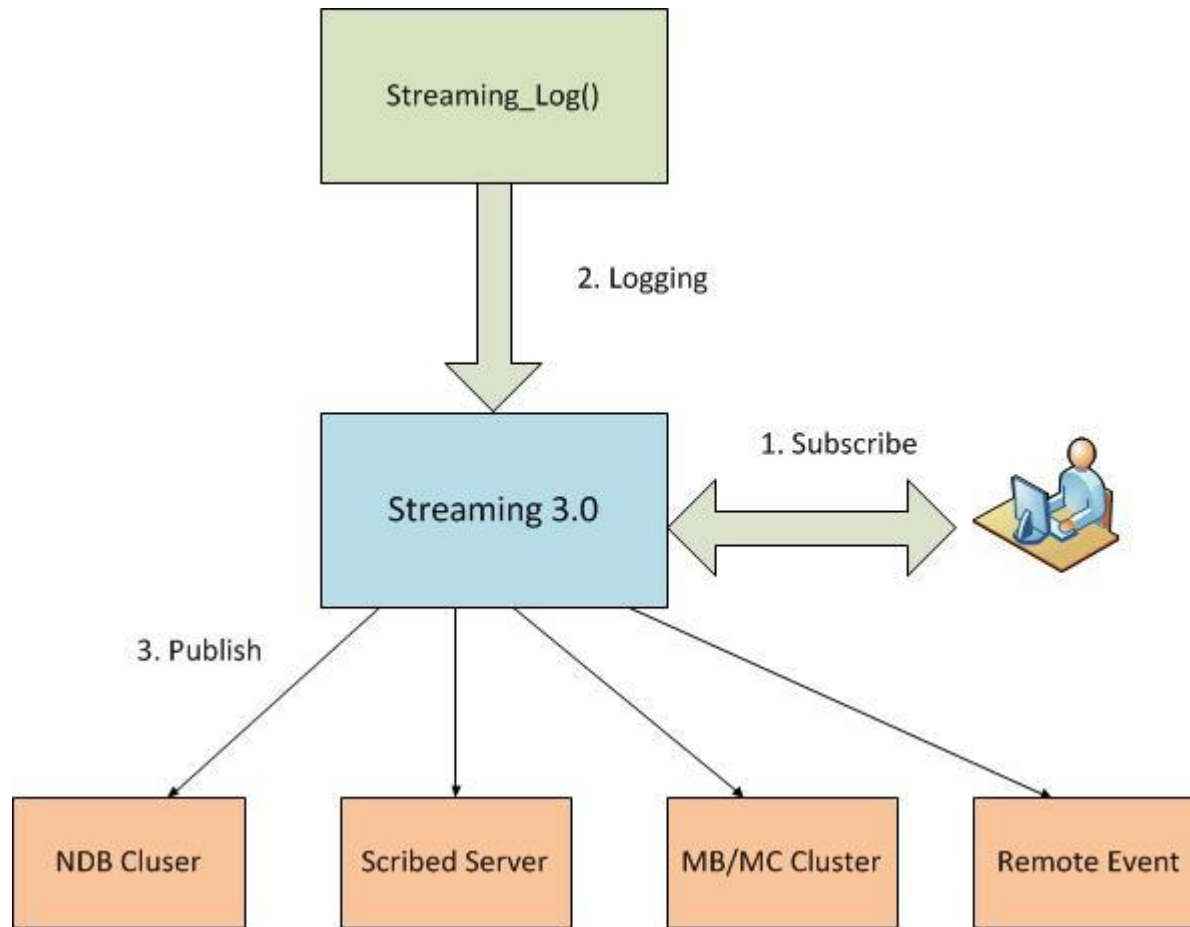
➤ Scaling Ability

- Input Data Volume/Rate
 - **65B** messages/day , or **45M** messages/min.
- Stored Data Volume/Rate
 - After aggregation, data volume drops down dramatically.
 - **1.8B** records per day dumped into MySQL DB, Or **1.3M** per minute.

Streaming 3.0 Architecture



Streaming 3.0 Publish-Subscribe Model



Challenges and Future Work

➤ Explore more publish channels

- Currently most event data goes to MySQL NDB cluster. We are exploring pushing data to messaging server, as well as some subscription API.

➤ Explore other ways to configure and deliver data

- Support EPL (Event Processing Language).