

Data Streaming Service

IT-DB-SAS 2020

2020 Achievements

- Limit Admin Access
 - CERT clusters ready (current version 2.0)
 - IT-Monitoring update to version 2.0
- ETL in Kafka – Easy to users
 - Based on Kafka Streams
 - Evaluate alternatives to lambda functions
 - Evaluate the integration with Web interface
 - Create a first pilot
 - JDBC Connector (NEW)

2020 Achievements

- Operations and configuration management
 - Remove automatic propagation of configuration changes by puppet (avoid automatic systems restarts)
 - Deprecation of cluster metadata database in favor of hiera (hadoop approach)
 - Define an alternative to propagate configuration changes
 - OC11 to logs (using Hadoop approach)
 - Test and move to Java11
 - Benchmark tests suite (NEW)

2020 Achievements

- Web Interface
 - Deploy the newer version (Improvement on ACLs and topics management)
 - Real-time preview of messages (WIP)
 - Improve connectors Interface and management.
 - Validate current functionalities with users to remove admin access to clusters (CERT)
 - Move to Openshift and integration with MALT-SSO (NEW)
- General Purpose
 - Improve monitoring and alerts on Connects (MISSING SNOW ALERTS)
 - Setup QA cluster for GP/GP-TN (NEW)

2020 Achievements

- Schema registry (open source) to enforce data format on GP and dedicated clusters (Summer student program)
- Open source Web interface and rest-api? (most a long term plan but requires to plan it now)
- Improve awareness of services
 - Interactive Training
 - Based on use cases
 - Consultancy

2020 Achievements

- User documentation
 - Connects, Streams and web interface
 - Kafka quickstart (NEW)
- Covid Tracing project (NEW)
 - Successful deployment of JDBC Connector for Oracle with Oracle Advanced Security enabled (NEW)
 - Patch of JDBC Connector to support OAS encryption (NEW)

Future Plans (2021)

- Limit Admin Access
 - IT-Monitoring still not updated to version 2.0
 - Validate with CERT if we can move to same Principal/Authorizer so the cluster is supported via webportal, and, subsequently, remove the Admin access to the cluster.
- ETL in Kafka – Easy to users
 - Create a first pilot
 - Make available to the users connectors' plugins that are already deployed (InfluxSink in PROD, JDBC both PROD/QA)
- Operations and configuration management
 - OC11 to logs (using Hadoop approach)
 - Test and move to Java11 (deploy java 11 in GP-GP/TN)

Future Plans (2021)

- General Purpose Clusters
 - Upgrade to Kafka 2.7.0 in order to:
 - Support delegation tokens
- Kafka Service that generates delegation tokens (similar to hadoopfetchdt)
- Web Interface
 - Real-time preview of messages
 - Validate current functionalities with users to remove admin access to clusters (CERT)

Future Plans (2021)

- Schema registry (open source) to enforce data format on GP and dedicated clusters
- Open source Web interface and rest-api? (most a long term plan but requires to plan it now)
- Improve awareness of services
 - Interactive Training
 - Based on use cases
 - Consultancy
- User documentation
 - Connects, Streams and web interface

Long Term Plan (5 years)

- Integrated data streaming platform user driven
 - Sourcing -> ETL, ANALYSIS, EVENT PROCESSING, ML -> Sinking
- Open source Web interface and rest-api?
- Improve awareness of services
 - Interactive Training
 - Based on use cases
- Standby Passive-Active clusters
 - Federation replication
- Co-hosting Kafka and Hadoop for specific use cases
- Providing only abstract interfaces (gateways) to the users and seamless integration between our internal services