

Apache Kyuubi 1.7.0 Release Note

DRAFT

Highlight

- Fully tested w/ Spark 3.1, 3.2 and 3.3, initial support oncoming 3.4
- Fully tested w/ Flink 1.14, 1.15 and 1.16
- Introduce Trino frontend (experimental)
- Introduce Spark Lineage plugin (experimental)
- Introduce Spark Hive connector (experimental)
- Spark engine supports arrow-based results serialization
- Spark engine supports executing Python/PySpark scripts(experimental)
- Several Kyuubi Helm Chart & Spark on K8s improvements
- REST and Batch API functions, clients and stability enhancements
- Major Spark Authz plugin improvements of including Apache Iceberg support
- Security enhancements including LDAP authentication, internal encrypted transports, etc.

重点特性

- 基于 Spark 3.1, 3.2 和 3.3 做了充分验证, 初步支持即将发布的 3.4
- 基于 Flink 1.14, 1.15 和 1.16 做了充分验证
- 新增 Trino 接入协议 (实验特性)
- 新增 Spark 血缘插件 (实验特性)
- 新增 Spark Hive 连接器 (实验特性)
- Spark 引擎支持基于 Arrow 序列化的高性能结果集传输
- Spark 引擎支持执行 Python/PySpark 脚本 (实验特性)
- 诸多 Kyuubi Helm Chart 和 Spark on K8s 改进
- REST 和 Batch 接口在功能、客户端和稳定性上进一步增强
- Spark 权限插件重大改进, 并包含了对 Apache Iceberg 的支持
- 包括LDAP认证、内部通信加密在内的诸多安全改进

Server

[KYUUBI #2887] Add a POLLING balance policy for engine pool

[KYUUBI #3089] Support custom event handler

[KYUUBI #3449] Change default server info provider to ENGINE

[KYUUBI #3545] [KYUUBI #3563] Support restrict spark configurations

[KYUUBI #3554] REST API functions/command-line tool enhancements

[KYUUBI #3577] Transfer connection url when opening session

[KYUUBI #3615] Retry opening the engine when encountering a special error

[KYUUBI #3658] Support SSL for Kyuubi thrift binary connection

[KYUUBI #3659] Support alternative keys in ConfigBuilder

[KYUUBI #3663] Support auto build Kubernetes client from env when Kyuubi running in Pod

[KYUUBI #3742] Add FileSessionConfAdvisor to manage session level configuration
[KYUUBI #3766] Support real user for KyuubiSession
[KYUUBI #3835] Allow to use spark-internal as resource for batch job
[KYUUBI #3839] Introduce signature mechanism to protect session variable on engine side
[KYUUBI #3844] Forward the server ip in openSession
[KYUUBI #3847] Add jdbc-shaded profile to support IDE debug
[KYUUBI #3863] Arrow-based results serialization
[KYUUBI #3867] Init SQL scripts should create table iff table does not exist
[KYUUBI #3887] Provide kyuubiServerPrincipal as alias for principal in JDBC parameter
[KYUUBI #3897] Supplying pluggable GroupProvider
[KYUUBI #3901] Introduce Trino frontend (experimental)
[KYUUBI #3922] Only the ApplicationInfo with non-empty id is valid for BatchJobSubmission
[KYUUBI #3923] Add dedicated batch session idle timeout
[KYUUBI #3926] Introduce antlr4 to parse query statement
[KYUUBI #3928] Application state mapping considers FinalApplicationStatus
[KYUUBI #3950] Fix the batch metadata in-consistent issue on open batch session failure
[KYUUBI #3975] Support to post batch session/operation event
[KYUUBI #3983] [KYUUBI #3982] Introduce refreshing user defaults configs
[KYUUBI #4021] Fix async start engine failure resulting in inaccurate operation metrics
[KYUUBI #4106] Introduce resource file uploading in batch creation via REST API
[KYUUBI #4119] Return app submission time for batch
[KYUUBI #4145] Change lock and polling seq_num path on service discovery
[KYUUBI #4151] Support to ignore subdomain when engine pool conditions are met
[KYUUBI #4152] Enhance LDAP authentication
[KYUUBI #4241] Only force close engine ref when open session failed
[KYUUBI #4322] MySQL URL configuration joiner should use `&` instead of `;`
[KYUUBI #4352] Support System.gc() with periodic GC interval
[KYUUBI #4360] Support to refresh the unlimited users for session limiter
[KYUUBI #4372] Support to return null value for OperationsResource RowSet
[KYUUBI #4390] Allow user to provide batch id on submitting batch job
[KYUUBI #4418] Allow disable async retry and fail fast on unrecoverable DB error
[KYUUBI #4419] Implement simple EngineSecuritySecretProvider

Client

[KYUUBI #3973] Fix conf fallback for `kyuubi-ctl list server`
[KYUUBI #4006] Fix the incorrect execution of the source command
[KYUUBI #4067] Reset the operation log before fetching new one
[KYUUBI #4216] Support to transfer client version for JDBC and REST client
[KYUUBI #4221] REST client for creating batch with uploading resource file
[KYUUBI #4311] Fix the wrong parsing of JVM parameters in JDBC url
[KYUUBI #4334] REST client should catch `NoHttpResponse Exception` and retry

Spark Engine

[KYUUBI #3128] Support CostMode for PlanOnlyStatement
[KYUUBI #3385] Set executor Pod name prefix if missing in Spark on K8s case

[KYUUBI #3441] Change default Spark version to 3.3.1
[KYUUBI #3487] Support Kyuubi as Spark JDBC dialect
[KYUUBI #3776] Revise Kyuubi Spark engine shaded
[KYUUBI #3780] Add support for executing Python/PySpark scripts (experimental)
[KYUUBI #3792] Engine UI support grace stop
[KYUUBI #3801] Correctly calculate active stages in SQLOperationListener
[KYUUBI #3851] Support auto set up spark.master when Kyuubi running inside K8s Pod
[KYUUBI #3885] Fix memory leak when using incremental collect mode
[KYUUBI #3892] Remove sensitive kyuubi.engine.credentials from spark conf
[KYUUBI #4035] Post SparkOperationEvent and show sessionId for statements
[KYUUBI #4150] Support to execute Scala statement synchronized to prevent conflicts
[KYUUBI #4316] Fix returned Timestamp values may lose precision
[KYUUBI #4336] Avoid listing all schemas for Spark session catalog on schema pruning
[KYUUBI #4407] Adapt SLF4J 2.x
[KYUUBI #4412] Align the session handle between server and engine for Spark engine

Spark Authz Plugin

[KYUUBI #3424] Access privilege checks for namespaces and tables of DatasourceV2
[KYUUBI #3515] Support Iceberg commands and skip apply row-filter to output tables
[KYUUBI #3325] Privilege checks for permanent views and skipping shadowed tables
[KYUUBI #3300] Support user group based policies from Ranger
[KYUUBI #3371] Ranger client support sending requests in batch
[KYUUBI #3581] Support row filter and data masking on permanent views
[KYUUBI #3607] Support {OWNER} variable defined in Ranger policy
[KYUUBI #3904] New Authz Plan Serde Layer
[KYUUBI #4076] Modified query plan should correctly report stats
[KYUUBI #4270] Register shutdown hook for plugin cleanup
[KYUUBI #4255] Handle describe relation for V2 relations
[KYUUBI #4262] Rename table requires ALTER privilege of the source table
[KYUUBI #4202] Fix reference resolution when data masking enabled for V2 relations
[KYUUBI #4437] Fix dependencies conflict by replacing `jersey-bundle` with `jersey-client`

Other Spark Extensions

[KYUUBI #2282] Introduce Spark Lineage plugin (experimental)
[KYUUBI #3365] Introduce Spark Hive connector (experimental)
[KYUUBI #3601] Support infer columns for rebalance and sort
[KYUUBI #3962] Add two conditions to decide if add shuffle before writing
[KYUUBI #3988] Final stage config isolation support write only
[KYUUBI #3893] UnspecifiedDistribution does not have default partitioning

Flink Engine

[KYUUBI #3514] Support Flink 1.16
[KYUUBI #3604] Propagate HADOOP_PROXY_USER on launching Flink engine
[KYUUBI #3717] Support Flink engine see primary keys

Other Notable Changes

- [KYUUBI #3473] Add Docker Compose based Kyuubi playground
- [KYUUBI #3903] Support generate `kyuubi-version-info.properties` on Windows
- [KYUUBI #4000] JDBC engine supports Apache Phoenix
- [KYUUBI #4288] Use eclipse-temurin:8-jdk-focal as default base image
- [KYUUBI #4320] Trino engine supports GetPrimaryKeys
- [KYUUBI #4473] Helm Chart improvements