



Ampool 2.3 (B17) Release Notes

Introduction

This document provides a brief overview of the various features and enhancements that are part of the Ampool 2.3 B17 release.

Ampool Hub

1. Azure Multi-resource group support

Ability to create a cloud credential with different resource groups specified for VM, Storage account, Vnet and Network security group. This cloud credential can be used to create AE clusters using resources from these different resource groups.

2. Hub user management enhancements

- a. Ability to perform various hub user operations like, add user, reset password etc. without SMTP configurations and sending emails.
- b. Group Admin user permission fixes

3. Private IP support

- a. Hub installation on Azure and AWS with Private IP (no Public IP's)
- b. AE cluster creation on Azure and AWS with Private IP (no Public IP's)

4. Log collection and filtering sensitive information from logs

- a. Generic REST API for log collection of various hub containers

5. RHEL 7 support on AWS

- a. Ability to install Hub on RHEL 7 on AWS
- b. Ability to specify CentOS/RHEL AMI names in AWS cloud credential

6. IAM Role support on AWS

- a. Installing Hub on vm with IAM role attached to it.
- b. Add support for installing AE using IAM role attached to instance in Hub



7. Security and Performance
 - a. Disable external API calls to Ampool Admin Hub
 - b. Disable external calls to HERE maps
 - c. Added view (eye icon) for cloud credentials form and create clusters page.
 - d. Encrypting database container password
 - e. Oracle driver installation is made optional on hub and AE
 - f. Updated memory config for Medium and Large Azure clusters

8. UI Re-branding and uniformity in Hub, Ampool Engine, Ampool Proxy UI

Ampool Engine

1. Query Federation
 - a. Ability to query tables across multiple data systems and those materialized in Ampool in a single SQL query
 - b. Transparent Caching capability for federated query execution based on the presto catalog or session configuration
 - c. Query rewriting capability to utilize Transparent Caching
 - d. Transparent Cache invalidation
 - i. The cached data is invalidated if the source table data is changed
 - ii. The cached data is invalidated using LFU (least frequently used), LRU, FIFO if there is no resource available
 - e. Query federation support was extended for Snowflake, ADLS data sources apart from MSSQL and other data sources that were already supported by Ampool Engine.
 - f. REST API support was enhanced to support query using the API's
 - g. Dynamic catalog support
 - i. Upon add/delete/update action on a data system in Ampool Engine, Presto catalog is created/deleted/updated accordingly for the data system without any manual steps

2. ADML (Ampool Data Markup Language)
 - a. Add support for data model creation and query rewriting for complex data models
 - b. Cost based decision making for query rewriting whether to rewrite query to be executed on denormalized table or not
 - c. Invalidation of denormalized table if change is detected in the constituent tables
 - d. REST API's for data models operation
 - e. Statistic for denormalized tables using distinct queries
 - f. Presto ingestion performance improvements
 - g. Add support for partition column for materialized view creation



3. Spark Job Server with Queuing support
 - a. SJS used on AE, ACP for Explicit caching, MQT, Export, Backup/Restore
 - b. SJS context Management
 - c. Support for Kerberised data sources
 - d. SJS HA support
 - e. Job Queuing and cancellation support

4. Tableau Connector
 - a. Tableau connector
 - b. Presto JDBC driver enhancements
 - c. TDVT validation of the connector
 - d. NULL value support for fixed Length column types
 - e. Ampool connector available as a beta connector from Tableau Online

5. Datasystems and Presto catalog support
 - a. Snowflake
 - i. Private key with and without encryption apart from user/password authentication
 - b. Teradata
 - c. Redshift
 - d. ADLS + HDInsight
 - e. S3 + Glue support enhanced with KMS on AWS
 - f. Enhanced data type support for various data systems
 - g. Adding support for the CSV file format for data loading from AWS Glue data source.

6. Presto Based Data Loading
 - a. Metadata browsing, Explicit caching and export support for few data systems
 - b. Change detection

7. Log collection and filtering sensitive information from logs
 - a. Added generic Ampool Agent component on all nodes for log collection
 - b. Generic REST API for log collection of various ae components
 - c. Filtering sensitive information like passwords, various credentials etc from logs



8. Observability and Monitoring
 - a. Netdata system monitoring of each AE node
 - b. Pulse UI for ADS monitoring
 - c. Presto, Spark UI for Query monitoring

9. AE deployment on RHEL 7 on AWS

10. AWS IAM Role support
 - a. Add support for accessing S3 + Glue with IAM role attached to instance
 - b. Add support for installing AE cluster with installer using IAM role attached to instance
 - c. Add support of backup and restore to/from AWS s3 using IAM role attached to instance
 - d. Add support to access Glue + S3 using IAM role attached to instance through presto glue catalog

11. AE UI Notifications window

12. Performance
 - a. Accelerated Network and generation 2 VM changes for Azure
 - b. Ingestion performance improvement by using Partition table layout for FTable in Ampool Connector

13. Security
 - a. Enable Firewall on AE nodes and whitelist only necessary ports



Ampool Engine Known Issues

1. Following Exception logged in ampool log

```
[severe 2020/09/30 18:45:47.348 UTC server-ip-172-31-35-246.us-west-2.compute.internal
<unicast receiver,ip-172-31-35-246-2336> tid=0x2c] Membership service failure: this member is
no longer in the view but is initiating connections
org.apache.geode.ForcedDisconnectException: this member is no longer in the view but is
initiating connections at
org.apache.geode.distributed.internal.membership.gms.mgr.GMSMembershipManager.forceDi
sconnect(GMSMembershipManager.java:2524) at
org.apache.geode.distributed.internal.membership.gms.membership.GMSJoinLeave.forceDisco
nnect(GMSJoinLeave.java:1005) at
org.apache.geode.distributed.internal.membership.gms.membership.GMSJoinLeave.processRe
moveRequest(GMSJoinLeave.java:642) at
```

Impact

The query execution/load will fail, if the disk gets full for a node and it will lose cluster membership.

Resolution

For this build the cluster will have to be sized considering this limitation.

The disk size needs to be minimum of 3 times the data ingested in Ampool.

2. Ampool Engine limitation for multi-user operations.

Data system registered by a user are not visible to other AE users.

Impact

Users are not able to perform data system operations like 'Create and Load' etc for data systems not registered by him.

Workaround

The 'Create and Load' operation has to be done by user who registers the Data system for explicit caching.

OR

Other users can use the CTAS (Create table in Ampool catalog from the datasystem catalog) presto query to do the explicit caching.