Mike Rumore

AbboTT Labs  Lake Forest, IL

R/SQL Architecture

Table of Contents

[1 Apollo Requirements 1](#_Toc22576000)

[1.1 General Requirements 1](#_Toc22576001)

[1.2 input.csv 2](#_Toc22576002)

[1.3 config.csv 2](#_Toc22576003)

[1.3.1 DX via SAML Connection 3](#_Toc22576004)

[1.3.2 DX via Custom Connection 3](#_Toc22576005)

[1.3.3 Reliability via DSN Connection 4](#_Toc22576006)

[1.3.4 Reliability via NON-DSN Connection 4](#_Toc22576007)

[1.3.5 DX via Spark using YARN 4](#_Toc22576008)

[1.3.6 DX via SPARK using Apache Livy 4](#_Toc22576009)

[1.3.7 IDA 5](#_Toc22576010)

[1.4 results.csv 5](#_Toc22576011)

[1.5 chart\_data.csv 7](#_Toc22576012)

[1.6 errors.csv 7](#_Toc22576013)

[2 R/SQL Script Architecture 7](#_Toc22576014)

[2.1 R/SQL Flat Script Layout 7](#_Toc22576015)

[2.2 R/SQL Ad Hoc Script Layout 7](#_Toc22576016)

[2.3 RLIB 7](#_Toc22576017)

[2.4 Testing 7](#_Toc22576018)

[3 Algorithms 7](#_Toc22576019)

[3.1 Current Algorithms 7](#_Toc22576020)

[3.2 Running Algorithms 7](#_Toc22576021)

[3.3 Adding a New Algorithm 7](#_Toc22576022)

[3.4 Testing an Algorithm 7](#_Toc22576023)

# Apollo Requirements

## General Requirements

Following are the requirements for R algorithms to be executed against BDAA/Dx data source using Athena

* Algorithm client shall create the input parameter (input.csv) file and make it available to R algorithm
* Athena and on-premise Reliability database connections would be written in a separate R script and the algorithms could source the connection
* script. R algorithms should not be hard-coding the connection strings/database credentials as part of algorithm logic/code.
* Algorithms shall loop thru all the algorithms defined using the group routine, process the data and output results for each algorithm.
* Algorithm shall output both flagged and healthy instrument data for each algorithm.
* Get available instruments in the last 24 hours, from all the logs that the algorithm is querying and identify healthy/flagged conditions for that data set.
* Algorithm execution results shall be written to results.csv file
* Algorithm execution errors shall be written to errors.csv file

Additional requirements/assumptions:

* Execution timeouts need to be defined for the algorithms? (not implemented)
* Additional libraries required for the Algorithm(s): Checkpoint package install option shall be used to install the required packages, as needed

## input.csv

The input.csv file format is shown next. This file is created by Apollo.

| **Column Name** | **Column Type** | **Comments** |
| --- | --- | --- |
| ALGORITHM\_NAME | VARCHAR(200) | Algorithm Name, as entered in the algorithm definition |
| PHM\_PATTERNS\_SK | NUMBER | PHM\_PATTERNS\_SK from PHM\_PATTERNS table |
| PARAMETER\_NAME | VARCHAR(500) | Algorithm Parameter names, as configured in the algorithm definition |
| PARAMETER\_VALUE | VARCHAR(4000) | Algorithm Parameters values, as configured in the algorithm definition |

## config.csv

The config.csv file format is shown next. This file is created by Apollo. There are several different formats depending on the methods used to connect to the databases. Examples are given for each database connection type. A config.csv file can have more than one set of parameters since the simultaneous connections to DX via Athena or via Spark, and to the Reliability database is common.

| **Column Name** | **Column Type** | **Comments** |
| --- | --- | --- |
| NAME | VARCHAR | Name of parameter used in R script |
| VALUE | VARCHAR | Value of parameter used in R |

For development test, the config.csv file must be created by hand. This process will be described in a later section on testing.

### DX via SAML Connection

| **Name** | **Value** |
| --- | --- |
| ATHENA\_JDBC\_DRIVER\_CLASS | com.simba.athena.jdbc.Driver |
| ATHENA\_JDBC\_CLASSPATH | C:/Users/rumormx/Documents/sandbox/phm/jlib/athena/AthenaJDBC41\_2.0.7.jar |
| ATHENA\_LOGLEVEL | 6 |
| ATHENA\_LOGPATH | logs |
| ATHENA\_WORKGROUP | add\_service\_dx\_readonly |
| ATHENA\_USERESULTSETSTREAMING | 0 |
| ATHENA\_S3OUTPUTLOCATION | s3://abt-bdaa-test-us-east-1-sandbox/athena/RUMORMX |
| ATHENA\_AWSCREDENTIALSPROVIDERCLASS | com.simba.athena.amazonaws.auth.profile.ProfileCredentialsProvider |
| ATHENA\_AWSCREDENTIALSPROVIDERARGUMENTS | saml |
| ATHENA\_S3\_STAGING\_DIR | s3://abt-bdaa-test-us-east-1-sandbox/athenaquerylog |
| ATHENA\_DB\_CONN\_STRING | jdbc:awsathena://awsregion=us-east-1 |
| ATHENA\_CREDENTIAL\_PROVIDER | saml |
| ATHENA\_DB\_USER | XXXXXXXX |
| ATHENA\_DB\_PASSWORD | YYYYYYYY |

### DX via Custom Connection

| **Name** | **Value** |
| --- | --- |
| ATHENA\_JDBC\_DRIVER\_CLASS | com.simba.athena.jdbc.Driver |
| ATHENA\_JDBC\_CLASSPATH | C:/Users/rumormx/Documents/sandbox/phm/jlib/athena/AthenaJDBC41\_2.0.7.jar |
| ATHENA\_LOGLEVEL | 6 |
| ATHENA\_LOGPATH | logs |
| ATHENA\_WORKGROUP | add\_service\_dx\_readonly |
| ATHENA\_USERESULTSETSTREAMING | 0 |
| ATHENA\_S3OUTPUTLOCATION | s3://abt-bdaa-test-us-east-1-sandbox/athena/RUMORMX |
| ATHENA\_AWSCREDENTIALSPROVIDERCLASS | com.simba.athena.amazonaws.auth.profile.ProfileCredentialsProvider |
| ATHENA\_AWSCREDENTIALSPROVIDERARGUMENTS | saml |
| ATHENA\_S3\_STAGING\_DIR | s3://abt-bdaa-test-us-east-1-sandbox/athenaquerylog |
| ATHENA\_DB\_CONN\_STRING | jdbc:awsathena://awsregion=us-east-1 |
| ATHENA\_CREDENTIAL\_PROVIDER | custom |
| ATHENA\_DB\_USER | XXXXXXXX |
| ATHENA\_DB\_PASSWORD | YYYYYYYY |

### Reliability via DSN Connection

| **Name** | **Value** |
| --- | --- |
| RELIABILITY\_DB\_SERVER | WQ00418P\SQLP01 |
| RELIABILITY\_DB\_NAME | Reliability |
| RELIABILITY\_DB\_USER | gsr\_reader |
| RELIABILITY\_DB\_PASSWORD | Rel11APP |
| RELIABILITY\_DB\_CONN\_TYPE | DSN |

### Reliability via NON-DSN Connection

| **Name** | **Value** |
| --- | --- |
| SQL\_SERVER\_ODBC\_DRIVER\_NAME | SQL Server |
| RELIABILITY\_DB\_SERVER\_NAME | WQ00418P\SQLP01 |
| RELIABILITY\_DB\_USER | Reliability |
| RELIABILITY\_DB\_PASSWORD | Rel11APP |
| RELIABILITY\_DB\_INSTANCE\_PORT | 63430 |
| RELIABILITY\_DB\_PORT | 1433 |
| RELIABILITY\_DB\_CONN\_TYPE | NON-DSN |

### DX via Spark using YARN

| **Name** | **Value** |
| --- | --- |
| SPARK\_HOME | /usr/lib/spark |
| HADOOP\_CONF\_DIR | /etc/hadoop/conf |
| YARN\_CONF\_DIR | /etc/hadoop/conf |
| SPARK\_EXECUTOR\_INSTANCES | 4 |
| SPARK\_EXECUTOR\_CORES | 4 |
| SPARK\_EXECUTOR\_MEMORY | 20GB |
| SPARK\_DYANMICALLOCATION\_ENABLED | false |
| SPARK\_MASTER | yarn-client |
| SPARK\_VERSION | 2.3.1 |

### DX via SPARK using Apache Livy

| **Name** | **Value** |
| --- | --- |
| SPARK\_HOME | /usr/lib/spark |
| HADOOP\_CONF\_DIR | /etc/hadoop/conf |
| YARN\_CONF\_DIR | /etc/hadoop/conf |
| SPARK\_EXECUTOR\_INSTANCES | 4 |
| SPARK\_EXECUTOR\_CORES | 4 |
| SPARK\_EXECUTOR\_MEMORY | 20GB |
| SPARK\_DYANMICALLOCATION\_ENABLED | false |
| LIVY\_MASTER | TBD |

### IDA

| **Name** | **Value** |
| --- | --- |
| IDA\_DB\_NAME | pabbto |
| IDA\_DB\_HOST | ux00147p-scan.oneabbott.com |
| IDA\_DB\_PORT | 1521 |
| IDA\_DB\_USER | SVC\_PHM\_CONNECT |
| IDA\_DB\_PASSWORD | svc\_phmc\_0771 |
| IDA\_JDBC\_CLASSPATH | ,C:/Users/rumormx/Documents/sandbox/phm/jlib/OJDBC-Full/ojdbc6.jar |
| IDA\_JDBC\_DSN\_FORMAT | jdbc:oracle:thin:@//%s:%s/%s |
| IDA\_JDBC\_DRIVER\_CLASS | oracle.jdbc.OracleDriver |

## results.csv

The results.csv file is created by the algorithm. The file format is shown next.

| **Name** | **Required?** | **Type** | **Comments** |
| --- | --- | --- | --- |
| PHM\_PATTERNS\_SK | Yes | NUMBER | Algorithm unique ID |
| PL | Yes | VARCHAR(5) | Product Line Code (e.g., 115, 116, 117, 205, 210, 214, etc.) |
| SN | Yes | VARCHAR(20) | Instrument Serial Number |
| FLAG\_DATE | Yes | VARCHAR(20) | Instrument local Date/Time of when this instrument has been flagged. Value should be in YYYYMMDDHH24MISS format  If instrument is marked as flagged (FLAG\_YN =1)  -> Flag Date - would be the max of instrument local date/time over a period of time that the algorithm ran against ->  format: YYYYMMDDHH24MISS  If instrument is marked as healthy (FLAG\_YN=0)  -> Flag Date -> would be the current day in YYYYMMDD format |
| CHART\_DATA\_VALUE | No | NUMBER(15,5) | Counter value. Required, If FLAG\_YN is 1. Make sure to set DEVICE\_VALUE to >= 1 if FLAG\_YN = 1. Maps to DEVICE\_VALUE in  PHM\_ALG\_OUTPUT table |
| FLAG\_YN | Yes | NUMBER(1) | Flag to identify if a instrument is healthy or unhealthy. Flagged -> 1; Healthy -> 0 |
| IHN\_LEVEL3\_DESC | No | VARCHAR(150) | 1) Algorithm has single experience code assigned for a PL.  -> If the results.csv file populate IHN\_LEVEL3\_DESC value then R processor takes it and persists as IHN\_LEVEL3\_DECSC in  PHM\_ALG\_OUTPUT table  -> If the results.csv file does not populate IHN\_LEVEL3\_DESC then R processor picks up what is assigned to IHN\_LEVEL3\_DESC  parameter in the algorithm definition and populates it in PHM\_ALG\_OUTPUT table  2) Algorithm has multiple experience codes assigned for a single PL  -> In this case, results.csv should always populate IHN\_LEVEL3\_DESC and it should also match with LEVEL 3 description assigned in  TCT. System will compare the TCT LEVEL 3 with IHN\_LEVEL3\_DESC column and pick up the right experience code.  -> If the R output (results.csv) does not populate IHN\_LEVEL3\_DESC value then leave this field as blank in PHM\_ALG\_OUTPUT table. |

## chart\_data.csv

| **Name** | **Required?** | **Type** | **Comments** |
| --- | --- | --- | --- |
| PHM\_PATTERNS\_SK | Yes | NUMBER | PHM\_PATTERNS\_SK (algorithm unique key), represents the algorithm |
| PL | Yes | VARCHAR2(5) | Product Line Code (Ex: 115, 116, 117, 205, 210, 214 etc.,) |
| SN | Yes | VARCHAR2(30) | Instrument Serial Number |
| FLAG\_DATE | Yes | VARCHAR2(20) | Instrument local Date/Time of when this instrument has been flagged. Value should be in YYYYMMDDHH24MISS format  If instrument is marked as flagged (FLAG\_YN =1)  -> Flag Date - would be the max of instrument local date/time over a period of time that the algorithm ran against -> format: YYYYMMDDHH24MISS If instrument is marked as healthy (FLAG\_YN=0) -> Flag Date -> would be the current day in YYYYMMDD format |
| CHART\_DATA\_VALUE | YES | NUMBER(15,5) | Counter value. Required, If FLAG\_YN is 1. Make sure to set DEVICE\_VALUE to >= 1 if FLAG\_YN = 1. Maps to DEVICE\_VALUE in PHM\_ALG\_OUTPUT table  Note: May not required to be populated by the algorithms. Deborah to confirm. |
| DATA\_SERIES | YES | VARCHAR2(5) | Data series value  For v5.0 scope, only WAM algorithms shall output chart\_data.csv file. DATA\_SERIES field will have **hourly** data points for WZ\_PROBE information (Ex: 1.2, 1.3, 2.1, 2.2) |

## errors.csv

| **Name** | **Required?** | **Type** | **Comments** |
| --- | --- | --- | --- |
| PHM\_PATTERNS\_SK | Yes | NUMBER | PHM\_PATTERNS\_SK (algorithm unique key), represents the algorithm |
| ERROR\_MESSAGE | Yes | VARCHAR2(4000) | Error message |

# R/SQL Script Architecture

## R/SQL Flat Script Layout

## R/SQL Ad Hoc Script Layout

## RLIB

## Testing

# Algorithms

## Current Algorithms

## Running Algorithms

## Adding a New Algorithm

## Testing an Algorithm