KPI Metrics Metadata Configuration Guide

An Open Source Asset for use with TIBCO® Data Virtualization

TIBCO Software empowers executives, developers, and business users with Fast Data solutions that make the right data available in real time for faster answers, better decisions, and smarter action. Over the past 15 years, thousands of businesses across the globe have relied on TIBCO technology to integrate their applications and ecosystems, analyze their data, and create real-time solutions. Learn how TIBCO turns data—big or small—into differentiation at www.tibco.com.

Project Name	AS Assets KPI Metrics
Document Location	This document is only valid on the day it was printed. The source of the document will be found in the ASAssets_KPI folder (https://github.com/TIBCOSoftware)
Purpose	Self-paced instructional

www.tibco.com

Global Headquarters
3303 Hillview Avenue
Palo Alto, CA 94304

Tel: +1 650-846-1000
+1 800-420-8450
Fax: +1 650-846-1005

Revision History

Version	Date	Author	Comments
1.0	Aug 30 2019	Mike Tinius	Initial revision
1.1	Nov 6 2019	Mike Tinius	Added reportResourceDatasourceLineage.
1.2	Dec 12 2019	Mike Tinius	Modified location and name of constant configuration file.
1.3	Jan 20 2020	Mike Tinius	Moved Published Resource info to “KPImetrics Data Dictionary v1.1.pdf”
1.4	Feb 25 2020	Mike Tinius	Update Cache_METADATA_TABLES to perform more efficiently.
1.5	Mar 12 2020	Mike Tinius	Removed METADATA_ALL_PRIVILEGE_STG.
1.6	Apr 6 2020	Mike Tinius	Added two new reports. reportMetadataAllCount and reportMetadataAllCountArch
1.7 – 2020.202	May 1 2020	Mike Tinius	Removed Archive views by adding partitioning to hold current and history in the same table.

Related Documents

Name	Version
How To Use Utilities.pdf	2020Q200
KPImetrics Configuration Guide vX.Y.pdf	2020Q202
KPImetrics Overview.pdf	2020Q200
KPImetrics Data Dictionary vX.Y.pdf	2020Q202
KPImetrics_Table_Relationship_Diagram.pptx	2020Q202
KPI Metrics Overview.pptx	2020Q200

Supported Versions

Name	Version
TIBCO® Data Virtualization	7.0.8 or later
AS Assets Utilities open source	2020Q200 or later

Table of Contents

1 Introduction

Purpose

The purpose of this document is to provide guidance on how configure and use the AS Assets KPI Metadata.

Audience

This document is intended to provide guidance for the following users:

- Data Virtualization Administrators – provides a guide for installation.
- Architects – provides the KPImetrics architecture.
- Data professionals – provides background on the published views and usage.
- Operations users – provides insight into triggers and procedures that are executed.
- Project Managers – provides general information on KPImetrics.

References

Product references are shown below. Any references to CIS or DV refer to the current TIBCO® Data Virtualization.

- TIBCO® Data Virtualization was formerly known as
  - Cisco Data Virtualization (DV)
  - Composite Information Server (CIS)

Overview

Please review the document "**KPImetrics Overview.pdf**".

2 Requirements

The following requirements and pre-requisites must be met:

- See requirements section in KPImetrics Configuration Guide vx.yy.pdf.

3 Use Cases

Metadata Metrics – The following use cases are examples of design-time metrics. Design-time is different than KPI metrics run-time metrics.

1. How many rows exist in each table? – data count.
    a. Count various types including the following:
        i. Project – Count the rows in each table for each project found in METADATA_CONST_NAME and nodehost and nodeport
            1. GROUP BY loaddate, projectnameid, projectname, nodehost, nodeport
        ii. Subtotal – Count the subtotal of rows for each nodehost and nodeport.
            1. GROUP BY loaddate, nodehost, nodeport
        iii. Total – Count the total rows in each table.
    b. When this view is invoked externally, the invoking report should sort by the following:
        i. ORDER BY viewname, loaddate DESC, counttype, nodehost, nodeport, projectnameid

    reportMetadataAllCount

2. How many views do not properly adhere to the layer rules? – compliance with architecture.
    a. Each layer should invoke the appropriate layer below it. Should never invoke source views.

    reportMetadataNonCompliantLayers

3. Which connector/adapter is used by which views

    reportMetadataDatasource

4. Source View is compliant with additional columns: source code, fetchtimestamp etc.

    reportMetadataNonCompliantColumns

5. # views by layers

    reportNumResourcesByLayer

6. Owner of views. Who has modified.

    vMetadataResource

7. # policy, roles, policy name, attributes, description

    vMetadataPolicy

    vMetadataPolicyAssignmnt

8. Metadata regarding access and authorization for a give resource associated with access groups.

vMetadataPrivilege

9. Report on what data sources are associated with a particular resource. For example, a user can view all of the published resources and their corresponding data source(s). This report will show actualprivileges.
    a. The report will only show combinedprivileges and inherited privileges for those projects where it was configured in pqInsert_METADATA_Constants "METADATA_CONST_LAYERS" section. Each layer requires a configuration of COMBINED_NO_USERS or COMBINED_WITH_USERS for that data to be present in the report.
    reportResourceDatasourceLineage

10. Report on what columns are associated with a particular resource and layer. For example, a user can view all of the published resources and their corresponding column(s).
    reportResourceColumn

11. Report on what resources are assigned privileges and what users are assigned to the privilege. When the privilege type is GROUP then the resource may have 0 or more users assigned to that group. When the privilege type is USER then there would be a single user assigned.
    a. The report will only show users for those projects where it was configured in pqInsert_METADATA_Constants "METADATA_CONST_LAYERS" section. Each layer requires a configuration of ACTUAL_WITH_USERS or COMBINED_WITH_USERS for that data to be present in the report.
    b. Note: for a report on just resource privileges use vMetadataPrivilege.
    reportMetadataPrivilegeUsers

4 Configuration

Configure Metadata Constants

Background Information:

The procedure “10_pqInsert_Metadata_Tables_METADATA_Constants” is used to configure various constants for a given “project”. A project has a base path which encompasses all of the layer folders and resources.

This procedure “DOES NOT” need to be executed manually. It will be executed each time the trigger “kpimetricsTrig_40_Cache_METADATA_TABLES” executes. The trigger executes Cache_METADATA_TABLES which in turn executes “10_pqInsert_Metadata_Tables_METADATA_Constants”. It does this so that all metadata is kept in synch with the same LOAD_DATE across all of the tables.

Instructions:

- Configure the following /shared/ASAssets/KPImetrics/Customize/pqInsert_METADATA_Constants.
- Configure the section “INSERT METADATA_CONST_NAME ROWS”
  - Modify the concatenated string below as needed. Add a row for each "project" name to capture metadata for.
  - PROJECT_NAME: A unique name that will be assigned a unique ID.
  - EXECUTE_FLAG: Y=execute this row. N=do not execute when triggered.
  - PROJECT_DESC: A description of the project path.
  - Maintain the existing structure with double pipe separating the line and single pipe separating a column.

```
SET projectName = 'TestSpoke';
SET METADATA_CONST_NAME_str = METADATA_CONST_NAME_str ||
PROJECT_NAME                    EXECUTE_FLAG            PROJECT_DESC
'|'||projectName        ||'|'||         'Y'             ||'|'||         'TestSpoke project desc'    ||
''; -- This is always the last line
```

- Configure the section “INSERT METADATA_CONST_PATH ROWS”
  - Modify the concatenated string below as needed. Add a row for each base path within the "project" to capture metadata for.
    - Modify projectName, pathSH, pathDS.
    - The variable "pathWS" is not currently supported for web services.
    - Modify the PROJECT_PATH and RESOURCE_TYPES as per your requirements.
  - Maintain the existing structure with double pipe separating the line and single pipe separating a column.

- PROJECT_NAME: A foreign key reference to METADATA_CONST_NAME which provides a unique name that will be assigned a PROJECT_NAME_ID that is unique.
- PROJECT_PATH: A unique key for this table which drives all of the processing for Cache_METADATA_TABLES procedure to load data.
- RESOURCE_TYPES: - A comma-separated list of resource types to process.
    - When using pathSH for shared area then [TABLE,PROCEDURE,TREE]
    - When using pathDS for /services/databases then [LINK]
  - NOTE: Web Services are not currently supported.

```
SET projectName = 'TestSpoke';
SET pathSH = '/shared/00_DataFederation/TestSpoke';
SET pathDS = '/services/databases/PWC/TestSpoke';
SET METADATA_CONST_PATHS_str = METADATA_CONST_PATHS_str ||
--PROJECT_NAME                PROJECT_PATH                          RESOURCE_TYPES
'||'||projectName        ||'''||    pathSH                ||'''||    'TABLE,PROCEDURE,TREE'        ||
'||'||projectName        ||'''||    pathDS                ||'''||    'LINK'                        ||
''; -- This is always the last line
```

- Configure the section “INSERT METADATA_CONST_LAYERS ROWS”
  - Modify the concatenated string below as needed. Only modify the layer type and parent path after the standard project path.
    - Modify projectName, pathSH, pathDS.
    - The variable "pathWS" is not currently supported for web services.
    - Modify the PROJECT_PATH, LAYER_TYPE, PARENT_PATH and GENERATE_LINEAGE as per your requirements.
  - Maintain the existing structure with double pipe separating the line and single pipe separating a column.
  - PROJECT_NAME: A foreign key reference to METADATA_CONST_NAME which provides a unique name that will be assigned a PROJECT_NAME_ID that is unique.
  - PROJECT_PATH: Provides a foreign key back to META_DRIVER table.
  - LAYER_TYPE: A unique string describing the layer to acquire metadata for.
  - PARENT_PATH: The actual path in DV which is associated with the LAYER_TYPE.
  - GENERATE_LINEAGE: Y=Generate lineage for this layer path. N=Do not generate lineage for this layer path.
  - EXCLUSION_LIST: A comma-separated list of paths or partial paths ending in a / that are to be excluded from the lineage generation. If a comma exists within a path then escape the comma with "_002C". e.g. /shared/my,path1/path2/ --> /shared/my_002Cpath1/path2/
  - ASSIGN_PRVILEGES: Provides the rules for assigning privileges on a per layer basis.
    - NO_PRIVILEGES - Do not assign any privileges for this layer
    - ACTUAL_NO_USERS - Assign actual privileges but do not invoke the getResourcePrivileges() api to get COMBINED or INHERITED. Do not retrieve users associated with groups.
    - ACTUAL_WITH_USERS - [DEFAULT] Assign actual privileges but do not invoke the getResourcePrivileges() api to get COMBINED or

INHERITED. Retrieve all users associated with a GROUP privilege.
    - COMBINED_NO_USERS - Invoke the getResourcePrivileges() api to get COMBINED and INHERITED privileges. Do not retrieve users associated with groups. Invoking the api will slow down the processing considerably.
    - COMBINED_WITH_USERS - Invoke the getResourcePrivileges() api to get COMBINED and INHERITED privileges. Retrieve all users associated with a GROUP privilege. Invoking the api will slow down the processing considerably.
- Rules:
    - A LAYER_TYPE that is a parent to a sub-folder is allowed and it will not cause duplication of resources. This concept will work in any layer including /shared and published /services/databases.
    - The table METADATA_CONST_LAYERS is validated for duplicates. If a duplicate layer and PARENT_PATH is found an exception is thrown.
    - Each LAYER_TYPE should have a unique name within a given PROJECT_NAME_ID.

For example,
1) Given the following layer type designations, there is a grandparent-parent-child folder relationship represented here:
Note: The number of levels/layers is NOT restricted.

LAYER_TYPE:	PARENT_PATH:
Note: 01_SourceViewLayer is a parent to 01_SourceViewLayer_svThirdParty	
01_SourceViewLayer	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer
Note: 01_SourceViewLayer_svThirdParty is a parent to 01_SourceViewLayer_svThirdParty_A and 01_SourceViewLayer_svThirdParty_B	
01_SourceViewLayer_svThirdParty	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer/012_svThirdParty
01_SourceViewLayer_svThirdParty_A	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer/012_svThirdParty/012_svThirdParty_A
01_SourceViewLayer_svThirdParty_B	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer/012_svThirdParty/012_svThirdParty_B

2) Given the following resources, the layer type will be assigned from the child (lowest folder) up to the grandparent (highest) folder.

LAYER_TYPE	RESOURCE_PATH
01_SourceViewLayer_svThirdParty_A	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer/012_svThirdParty/012_svThirdParty_A/012_svThirdParty_A1/customers
01_SourceViewLayer_svThirdParty_A	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer/012_svThirdParty/012_svThirdParty_A/customers
01_SourceViewLayer_svThirdParty_B	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer/012_svThirdParty/012_svThirdParty_B/012_svThirdParty_B1/customers
01_SourceViewLayer_svThirdParty_B	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer/012_svThirdParty/012_svThirdParty_B/012_svThirdParty_B2/customers
01_SourceViewLayer_svThirdParty_B	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer/012_svThirdParty/012_svThirdParty_B/customers
01_SourceViewLayer_svThirdParty	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer/012_svThirdParty/customers
01_SourceViewLayer	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer/011_svInternal/tutorial/customers
01_SourceViewLayer	/shared/00_DataFederation/TestSpoke/01_SourceViewLayer/DS_ORDERS/tutorial/customers

The following demonstrates how to setup the constants.

```
SET projectName = 'TestSpoke';
SET pathSH = '/shared/00_DataFederation/TestSpoke';
SET pathDS = '/services/databases/PWC/TestSpoke';
SET METADATA_CONST_LAYERS_str = METADATA_CONST_LAYERS_str ||
--PROJECT_NAME          PROJECT_PATH    LAYER_TYPE              PARENT_PATH             GENERATE_LINEAGE    EXCLUSION_LIST
        ASSIGN_PRVILEGES
'|'||projectName        ||'~'||         pathSH||'~'||   '00_DataSource'||'~'||      pathSH||'/00_DataSource'||'~'||                         'N'||  '~'||       '~'||  '~'||
        'ACTUAL_WITH_USERS'||
'|'||projectName        ||'~'||         pathSH||'~'||   '01_SourceViewLayer'    ||'~'||pathSH||'/01_SourceViewLayer'    ||'~'||          'N'||  '~'||       '~'||  '~'||
        'ACTUAL_WITH_USERS'||
'|'||projectName        ||'~'||         pathSH||'~'||   '02_ConformingViewLayer'||'~'||pathSH||'/02_ConformingViewLayer'||'~'||          'N'||  '~'||       '~'||  '~'||
        'ACTUAL_WITH_USERS'||
'|'||projectName        ||'~'||         pathSH||'~'||   '031_CommonEntityModel'||'~'||pathSH||'/03_CommonModelLayer/031_CommonEntityModel'||'~'||'N'||  '~'||       '~'||
'~'||       'ACTUAL_WITH_USERS'||
'|'||projectName        ||'~'||         pathSH||'~'||   '032_CommonDimensionalModel'||'~'||pathSH||'/03_CommonModelLayer/032_CommonDimensionalModel'
||'~'||'N'||  '~'||           '~'||  '~'||       'ACTUAL_WITH_USERS'||
```

```
'|'||projectName      ||'''||          pathSH||'''||   '033_CommonAnalyticalModel'||'''||pathSH||'/03_CommonModelLayer/033_CommonAnalyticalModel'||'''||'N'||  '|'||
        ''||  '''||       'ACTUAL_WITH_USERS'||
'|'||projectName      ||'''||          pathSH||'''||   '034_CommonIntegrationModel'||'''||pathSH||'/03_CommonModelLayer/034_CommonIntegrationModel'||'''||'N'||
'''||      ''||  '''||       'ACTUAL_WITH_USERS'||
'|'||projectName      ||'''||          pathSH||'''||   '041_BusinessDemandModel'||'''||pathSH||'/04_BusinessDeliveryLayer/041_BusinessDemandModel'||'''||'N'||  '|'||
        ''||  '''||       'ACTUAL_WITH_USERS'||
'|'||projectName      ||'''||          pathSH||'''||   '042_BusinessDemandView'||'''||pathSH||'/04_BusinessDeliveryLayer/042_BusinessDemandView'||'''||'Y'||  '|'||
        ''||  '''||       'ACTUAL_WITH_USERS'||
'|'||projectName      ||'''||          pathDS||'''||   'PublishedDS_tutorial'      ||'''||          pathDS                      ||'''||'Y'||  '|'||   ''||  '|'||
        'ACTUAL_WITH_USERS'||
''; -- This is always the last line
```

- Configure the section “INSERT METADATA_CONST_VALIDATE ROWS”
  - Modify the concatenated string below as needed.
    - Modify projectName, pathSH, pathDS.
    - The variable "pathWS" is not currently supported for web services.
    - Modify the PROJECT_PATH, LAYER_TYPE, RULE_TYPE and RULE_DESC as per your requirements.
  - Maintain the existing structure with double pipe separating the line and single pipe separating a column.
  - PROJECT_NAME: A foreign key reference to METADATA_CONST_NAME which provides a unique name that will be assigned a PROJECT_NAME_ID that is unique.
  - PROJECT_PATH: Provides a foreign key back to META_DRIVER table.
  - LAYER_TYPE: A valid layer name found in the table METADATA_CONST_LAYERS.
  - RULE_TYPE: Valid values=[ENFORCE_LAYER|ENFORCE_COLUMN]
  - RULE_DESC: Enforce the rule type.
    - When RULE_TYPE=ENFORCE_COLUMN
      a. Enforces which columns must be present in all of the views for a given layer type. Comma-separated list of case-sensative column names.
    - When RULE_TYPE=ENFORCE_LAYER
      a. Enforces which source layer resource can invoke which target layer resource. Comma-separated list of valid LAYER_TYPES.
      b. If a resource can invoke another resource in the same layer then add its own layer to the list.

```
SET projectName = 'TestSpoke';
SET pathSH = '/shared/00_DataFederation/TestSpoke';
SET pathDS = '/services/databases/PWC/TestSpoke';
SET METADATA_CONST_VALIDATE_str = METADATA_CONST_VALIDATE_str ||
--PROJECT_NAME     PROJECT_PATH     LAYER_TYPE            RULE_TYPE                       RULE_DESC
'|'||projectName     ||'''||pathSH  ||'''||          '01_SourceViewLayer'    ||'''||'ENFORCE_LAYER' ||'''||          '00_DataSource'||
'|'||projectName     ||'''||pathSH  ||'''||          '01_SourceViewLayer'    ||'''||'ENFORCE_COLUMN'||'''||          'fetchTimeStamp,systemSourceCode'||
'|'||projectName     ||'''||pathSH  ||'''||          '02_ConformingViewLayer'||'''||'ENFORCE_LAYER'||'''||          '01_SourceViewLayer'||
'|'||projectName     ||'''||pathSH  ||'''||          '031_CommonEntityModel'||'''||'ENFORCE_LAYER'||'''||          '02_ConformingViewLayer'||
'|'||projectName     ||'''||pathSH  ||'''||          '032_CommonDimensionalModel'||'''||'ENFORCE_LAYER'||'''||  '02_ConformingViewLayer'||
'|'||projectName     ||'''||pathSH  ||'''||          '033_CommonAnalyticalModel'||'''||'ENFORCE_LAYER'||'''||    '02_ConformingViewLayer'||
'|'||projectName     ||'''||pathSH  ||'''||          '034_CommonIntegrationModel'||'''||   'ENFORCE_LAYER'||'''||  '02_ConformingViewLayer'||
'|'||projectName     ||'''||pathSH  ||'''||          '041_BusinessDemandModel'||'''||'ENFORCE_LAYER'||'''||
'031_CommonEntityModel,032_CommonDimensionalModel,034_CommonIntegrationModel,041_BusinessDemandModel'||
'|'||projectName     ||'''||pathSH  ||'''||          '042_BusinessDemandView'||'''||'ENFORCE_LAYER'||'''||          '041_BusinessDemandModel'||
'|'||projectName     ||'''||pathDS  ||'''||          'PublishedDS_tutorial'    ||'''||'ENFORCE_LAYER' ||'''||          '042_BusinessDemandView'||
''; -- This is always the last line
```

Configure Trigger

Enabling triggers starts the processing of KPI metadata data. The trigger “kpimetricsTrig_40_Cache_METADATA_TABLES” is turned off by default. It must be turned on in order to begin the processing of

1. Modify /shared/ASAssets/KPImetrics/Customize/**defaultTriggersToEnable** and change the trigger kpimetricsTrig_40_Cache_METADATA_TABLES from OFF to ON if you want to capture metadata.
2. When updateTriggers is executed, it will turn on and off the trigger automatically according to how the trigger is set in defaultTriggersToEnable.

5 KPImetrics Metadata Resources

Configuration Resources

This section outlines the resources that are used for configuration of KPImetrics Metadata.

Published Resources

This section outlines the resources that are published under the ASAssets virtual database to expose metrics data. Resources are organized under catalogs and schemas based upon their functionality.

Please review the document "***KPImetrics Data Dictionary vX.Y.pdf***" for details about published tables, procedures and columns.

Configuration Parameters

The following configuration parameters found in "**commonValues**" are important for Metadata collection.

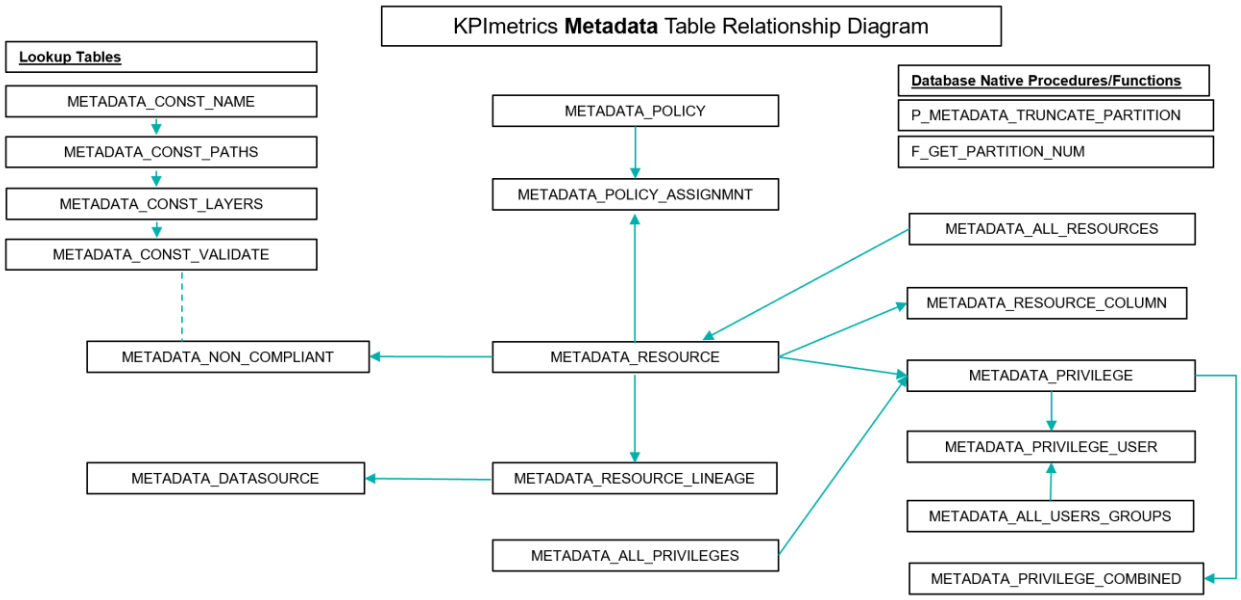
1. **Cluster Awareness Configuration:**
    a. The following variables are used to designate a dedicated time keeper host and port. None of the ClusterSafeCache procedures should be executed on a dedicated time keeper node because general processing and work is never routed to this node. It's only job in the cluster is to service triggers and perform deployments. However, rule #4 is the exception to this recommendation. Execute the following procedure on the "dedicated" time keeper node to get the hostname and port: ../Physical/Metadata/System/Helpers/pGetSystemInformation
    b. Rules:
        i. SINGLE NODE: Set the hostname and port to null when this environment is a single node and not a cluster.
        ii. CLUSTER NO TIMEKEEPER: Set the hostname and port to null when this environment is a cluster and any node in the cluster may be the timekeeper and all nodes do work.
        iii. CLUSTER WITH TIMEKEEPER: Set the hostname and port to the the dedicated timekeeper node and port when the environment is a cluster and the node has been dedicated as a timekeeper and does not do work.
        iv. CLUSTER WITH TIMEKEEPR CAPTURE: Set the hostname and port to null when the environment is a cluster and metadata should be captured for the timekeeper even though it is dedicated and does not perform work. Note: Using this option will require running the KPImetrics procedures on this node thus requiring it do work.
    c. **dedicatedTimeKeeperHostname** – dediated time keeper host
    d. **dedicatedTimeKeeperPort** – dedicated time keeper port.

e. **totalNumberWorkerNodes** – designates the total number of worker nodes used for processing.
f. Rules:
    i. SINGLE NODE: Single node (stand-alone) environment. No cluster. The total should be the default of 1. Example: totalNumberWorkerNodes=1
    ii. CLUSTER NO TIMEKEEPER: Cluster environment with NO "dedicated" timekeeper. The total is the total number of nodes in the cluster. Example: a 3-node cluster with NO dedicated timekeeper. totalNumberWorkerNodes=3
    iii. CLUSTER WITH TIMEKEEPER: Cluster environment with a "dedicated" timekeeper. The total is the total number of nodes minus 1. Example: a 3-node cluster with a dedicated timekeeper. totalNumberWorkerNodes=2
    iv. CLUSTER WITH TIMEKEEPR CAPTURE: Cluster environment with a "dedicated" timekeeper where KPImetrics runs on the timekeeper node to capture metrics and metadata. The total is the total number of nodes in the cluster. Example: a 3-node cluster with a dedicated timekeeper. totalNumberWorkerNodes=3

2. **Purge History:**
    a. **purgeMetadata** – RULE: purgeMetadata - Purge tables: All METADATA_% tables. A valid value must be between 1 and 366. Anything else will throw an exception. The Cache_METADATA_TABLES will always purge the current day prior to loading new data. The amount of data is always the current day + the history. To keep 3 days would be the current day - 2 days of history. Example: To keep 3 days of metadata the value will be 3. The other partitions will be truncated as per the formula: [*startDate*=CURRENT_TIMESTAMP *endDate*=DATEADD('day', (366-3), CURRENT_TIMESTAMP)]. The startDate and endDate are passed into the P_METADATA_TRUNCATE_PARTITION function. All partitions are truncated from the current day until -3 days from current day within the 366 interval partitioning scheme. Therefore, there are 3 total days of data available to query. Rule executed by .../Metadata/System/ClusterSafeCache/Cache_METADATA_TABLES().

Metadata Data Source Tables

The following provides a description for the database tables used by KPImetrics Metadata.



Metadata Data Source Tables and Procedures

Location: /shared/ASAssets/KPImetrics/Physical/Metadata/KPI_<database_type>_<version>

The KPImetrics module provides data source for all currently supported storage database platforms under /shared/ASAssets/KPImetrics/Physical/Metadata.

Currently the KPImetrics module includes the following KPImetrics data sources

- /shared/ASAssets/KPImetrics/Physical/Metadata/KPI_oracle_<version>
- /shared/ASAssets/KPImetrics/Physical/Metadata/KPI_sqlserver_<version>
    - NOTE: SQL Server 2016 or higher or SQL Serve Azure SQL Pass is required due to syntax requirements with truncating partitions.

The following tables have been created in DVPI schema to capture the required data. Each table has a corresponding archive table.

RULES:

- Only one load set of data is stored at any given point in time in the main metadata tables.
- Each node in a cluster will contain its own set of metadata rows therefore, NODE_HOST and NODE_PORT are a part of every key. Even though the resource name will be the same, the RESOURCE_ID may be different on any given node. Be sure to do reporting based on a particular NODE_HOST and NODE_PORT.

Table Name	Description

METADATA_ALL_PRIVILEGES	This table contains the resource and privilege pool of privileges from METADATA_ALL_PRVILEGES_STG and METADATA_ALL_RESOURCES. It is possible to have a resource that does not have privileges in which case the privilege is NONE for that resource. KEY: LOAD_DATE, RESOURCE_ID, NAME_TYPE, NAME_ID, DOMAIN_NAME, PRIVILEGE, NODE_HOST, NODE_PORT
METADATA_ALL_RESOURCES	This table contains the pool of system.ALL_RESOURCES, system.ALL_TABLES, system.ALL_PROCEDURES, system.ALL_WSDL_OPERATIONS, system.ALL_COLUMNS, and system.ALL_PARAMETERS. The RESOURCE_ORGIN columns defines which table the data came from so that it can be queried appropriately when processing data. KEY: LOAD_DATE, RESOURCE_ID, NAME_ID, NAME_TYPE, PRIVILEGE, NODE_HOST, NODE_PORT
METADATA_ALL_USERS_GROUPS	This table contains the list of a all domain groups and the users associated with those groups. Therefore, the username will be repeated within the table for each group it is a member of. This is not the same as system.ALL_RESOURCES. This table is created by getting a list of all domains and then getting the users for each domain. This table is used with assigning users to privileges. It is more efficient than an API call to achieve the same capability.
METADATA_CONST_NAME	This table contains a unique base project path that drives all of the metadata capture for all of the tables. Only metadata is captured the project paths present in this table. The trigger specified below along with the procedure it invokes is the only mechanism for capturing metadata for all of the metadata tables listed here. LOAD_DATE: The timestamp of the latest metadata load. PROJECT_NAME_ID: A unique sequence id for each project name. PROJECT_NAME: A unique name that will be assigned a PROJECT_NAME_ID that is unique. ENVIRONMENT_NAME: The environment nickname from commonValues.cisServerNickname. EXECUTE_FLAG: Y=execute this row. N=do not execute when triggered. PROJECT_DESC: A description of the project path.

	RESOURCE_TYPES: TABLE,PROCEDURE - A comma-separated list of resource types to process. Currently only TABLE and PROCEDURE are valid. EXECUTE_STATUS: The status of the latest load. SUCCESS or EXCEPTION which includes the exception message. NODE_HOST: Indicates which hostname/node the processing took place on. Multiple hosts/nodes in a cluster. NODE_PORT: Indicates the port of the DV server in which the processing took place on. TRIGGER: /KPImetrics/Physical/Metadata/System/ClusterSafeTriggers/kpimetricsTrig_40_Cache_METADATA_TABLES→ Cache_METADATA_TABLES KEY: LOAD_DATE, PROJECT_NAME_ID, PROJECT_NAME, NODE_HOST, NODE_PORT
METADATA_CONST_PATHS	This table contains a list of base project paths that drives all of the metadata capture for all of the tables. Only metadata is captured the project paths present in this table. PROJECT_PATH: A unique key for this table which drives all of the processing for Cache_METADATA_TABLES procedure to load data. RESOURCE_TYPES: TABLE,PROCEDURE,LINK - A comma-separated list of resource types to process. KEY: LOAD_DATE, PROJECT_NAME_ID, PROJECT_PATH, NODE_HOST, NODE_PORT
METADATA_CONST_LAYERS	This table contains the valid layer types for each project path. A layer type has a corresponding parent path within the project path that it correlates to. PROJECT_PATH: Provides a foreign key back to METADATA_CONST_NAME table. LAYER_TYPE: A unique string describing the layer to acquire metadata for. PARENT_PATH: The actual path in DV which is associated with the LAYER_TYPE. KEY: LOAD_DATE, PROJECT_NAME_ID, LAYER_TYPE, NODE_HOST, NODE_PORT
METADATA_CONST_VALIDATE	This table contains the layer validation rules. The rules provide for enforcing columns within views and which views can invoke views in specific layers. PROJECT_PATH: Provides a foreign key back to METADATA_CONST_NAME table. LAYER_TYPE: A valid layer name found in the table METADATA_CONST_LAYERS.

	RULE_TYPE: Valid values=[ENFORCE_LAYER|ENFORCE_COLUMN] RULE_DESC: Enforce the rule type. When RULE_TYPE=ENFORCE_COLUMN Enforces which columns must be present in all of the views for a given layer type. Comma-separated list of case-sensative column names. When RULE_TYPE=ENFORCE_LAYER Enforces which source layer resource can invoke which target layer resource. Comma-separated list of valid LAYER_TYPES. If a resource can invoke another resource in the same layer then add its own layer to the list. KEY: LOAD_DATE, PROJECT_NAME_ID, LAYER_TYPE, RULE_TYPE, NODE_HOST, NODE_PORT
METADATA_RESOURCE	This is the core table which all other tables reference. This table contains a row for each TABLE and PROCEDURE resource found within the specified PROJECT_PATH in the METADATA_CONST_NAME table. KEY: LOAD_DATE, PROJECT_NAME_ID, RESOURCE_ID, NODE_HOST, NODE_PORT
METADATA_RESOURCE_COLUMN	This table contains all of the COLUMNS referenced by the RESOURCE_ID in METADATA_RESOURCE. KEY: LOAD_DATE, PROJECT_NAME_ID, RESOURCE_ID, COLUMN_NAME, NODE_HOST, NODE_PORT
METADATA_RESOURCE_LINEAGE	This table contains the lineage for each resource in each layer. This will be a very large table. KEY: LOAD_DATE, PROJECT_NAME_ID, RESOURCE_ID, LINEAGE_ORDER, LAYER_TYPE, NODE_HOST, NODE_PORT
METADATA_DATASOURCE	This table contains the all of the datasource information for a given project path. KEY: LOAD_DATE, PROJECT_NAME_ID, DATASOURCE_ID, NODE_HOST, NODE_PORT
METADATA_NON_COMPLIANT	This table contains information on column and layer compliancy based on the METADATA_CONST_VALIDATE rules tables. KEY: LOAD_DATE, PROJECT_NAME_ID, RESOURCE_ID, LINEAGE_ORDER, NON_COMPLIANT_REASON, NODE_HOST, NODE_PORT
METADATA_POLICY	This table contains RBS [rule-based security] and CBS [column-based security] rows for a given project path. KEY: LOAD_DATE, PROJECT_NAME_ID, POLICY_ID, NODE_HOST, NODE_PORT
METADATA_POLICY_ASSIGNMNT	This table contains the assignments for a policy.

	KEY: LOAD_DATE, PROJECT_NAME_ID, RESOURCE_ID, POLICY_ID, NODE_HOST, NODE_PORT
METADATA_PRIVILEGE	This table contains the assigned privileges for all of the resources in a given project path. KEY: LOAD_DATE, PROJECT_NAME_ID, RESOURCE_ID, NAME, NAME_TYPE, DOMAIN_NAME, USER_NAME, NODE_HOST, NODE_PORT
METADATA_PRIVILEGE_USER	This table contains a many to many relationships between METADATA_PRIVILEGE and METADATA_ALL_USERS_GROUPS. KEY: LOAD_DATE, PROJECT_NAME_ID, PRIVILEGE_ID, USER_PK, NODE_HOST, NODE_PORT

Metadata System Triggers and Load Scripts

Location: /shared/ASAssets/KPImetrics/Physical/Metadata/System

/ClusterSafeCache

/ClusterSafeTriggers

/Helpers

This section provides a quick summary of all triggers, their schedules and how they execute in a cluster.

Note: "all nodes" and cluster dedicated timekeeper…

The reference to "**all nodes**" refers to all working nodes in a cluster except if there is a dedicated timekeeper. If there is no dedicated timekeeper then one of the nodes is nominated to be a timekeeper. KPImetrics will execute on that node.

When there is a dedicated timekeeper, then KPImetrics procedures will not execute on those nodes as configured in commonValues.dedicatedTimeKeeperHostname and commonValues.dedicatedTimeKeeperPort.

For "only once per cluster", whichever node is the timekeeper nominates a single node in the cluster to perform the work.

Trigger Name	Trigger Schedule	Trigger Period	Cluster execution
kpimetricsTrig_40_Cache_METADATA_TABLES	10:30 PM	1 day	all nodes

This section lists all triggers and load scripts that have been defined to execute various KPImetrics procedures at regular intervals. The default execution frequencies are listed for each trigger. The load scripts have been created to load and aggregate raw data into processed KPImetrics metadata.

Trigger [schedule] → Script Name → View name	Description
Schedule: [1 day, 10:30 pm] **kpimetricsTrig_40_Cache_METADATA_TABLES →** /shared/ASAssets/KPImetrics/Physical/Metadata/System/ClusterSafeCache/Cache_METADATA_TABLES → /shared/ASAssets/KPImetrics/Customize/pqInsert_METADATA_Constants	This trigger executes the Cache_METADATA_TABLES procedure. This procedure is used to capture all the metadata for all of the metadata tables. Exceptions: Emails will be sent if there are exceptions. Review the following view (table) for issues: /services/databases/ASAssets/KPImetrics/workflow/vCISWorkflowStatus Uses the same ALL_RESOURCE data from METRICS_ALL_RESOURCES_STG which gets cached every 2 hours. The data would be current as of 9 pm. This alleviates the need to recache data that was already cached. Therefore, there is a **dependency** on **Cache_ALL_RESOURCES** completing for a given node.

Load Script Procedure Architecture

The following provides a description for the load script architecture.

Location: /shared/ASAssets/KPImetrics/Physical/Metadata/System/Cache_METADATA_TABLES

Architecture

This section describes the architecture of the load script.

One word is used to simplify the entire process “PARTITIONING”. Database partitioning is used to ELIMINATE “ARCHIVING” to separate tables and then consequently “PURGING” with delete statements thos addition archive tables. There still is the concept of purge but it is much more efficient in that it uses a truncate on the partition which is effectively a metadata call and thus very fast.

The benefit of ths strategy is three-fold. The process reduces time in the purge process and archive process which can account for hours of time with very large tables. The third benefit is that only one set of views is required for both current and history which makes reporting easier.

With partitioning, it is very important to include the partition number in the query in order to realize partition elimination and thus increase performance.

Partitioning is advantageous because the same table contains the current and history broken down by partitions. The current “DAY” contains the current partition of rows. Partitioning allows for a sliding window of days of data. This strategy allows for a total of 366 days (accounting for leap year) to hold data. The number of commonValues.purgeMetadata will define how many days you want to keep and will purge the other days via the sliding window.

Sliding Window Example:

Assumption: commonValues.purgeMetadata=120 days (4 months)

Current Day=May 1 2020 = partition 122 [current day is the day the data is loaded]

Sliding window looks backwards for 120 days to preserve data. Therefore, data is preserved between Jan 3 2020 [partition=3] thru May 1 2020 [partition=122].

The load process will automatically truncate the current day partition just prior to loading it. In addition, it will truncate all the partitions looking forward until beginning date of the sliding window thus preserving the history.

Pre-processing section:

1. A gatekeeper control name “METADATA_TRUNCATE_ALL_PARTITIONS” is inserted when all nodes begin. The first node is takes control first and the other nodes wait.
    a. All other nodes will wait until the processing is complete. If the processing takes longer than 60 tries * 60 second pause [1 hour], then the node will throw an exception as follows: [gateKeeper] Time expired waiting for a chance to delete rows for controlName=[METADATA_TRUNCATE_ALL_PARTITIONS]
    b. Since truncate happens very fast, this should never happen.
2. **Truncate partition** – Due to the code path, each node will execute the truncate. The first node acutally does the work and performs the truncate according to the sliding window. The procedure “P_METADATA_TRUNCATE_PARTITION” is invoked to perform the truncate. The other nodes detect a zero count and bypass the operation.

3. **Truncate tables** – The first node through the gatekeeper will be responsible for truncating the following tables [METADATA_ALL_RESOURCES, METADATA_ALL_PRIVILEGES, METADATA_PRIVILEGE_COMBINED] in order to maximize efficiency. The first node who reaches this point in the code will take control. All other nodes will wait at the gatekeeper() until the first node has finished. The other nodes will detect a zero count and bypass the truncate.
4. A gatekeeperCheck() is invoked after truncate to wait for all nodes to synchronize the truncate process. All nodes wait to complete the control name "METADATA_TRUNCATE_ALL_PARTITIONS". Since there is no data in the tables at this point, the other nodes will check for a row count and simply take no action. They will complete that section of code very quickly and synchronize with the first node. Once all nodes are synchronized, they will move on to the next section of processing.
5. All nodes will participate in the remainder sections which begin the insert of records.
6. Insert all resources into METADATA_ALL_RESOURCES
    a. If METRICS_ALL_RESOURCES is current then use that table otherwise get it from /shared/ASAssets/KPImetrics/Physical/Metadata/System/ALL_RESOURCES
7. Insert all privileges into METADATA_ALL_PRIVILEGES
    a. Query /services/databases/system/ALL_PRIVILEGES, ALL_USERS and ALL_GROUPS
8. Insert configuration records from /shared/ASAssets/KPImetrics/Customize/pqInsert_METADATA_Constants
    a. METADATA_CONST_NAME
    b. METADATA_CONST_PATHS
    c. METADATA_CONST_LAYERS
    d. METADATA_CONST_VALIDATE
9. Insert all users and groups into METADATA_ALL_USERS_GROUPS
    a. Query /shared/ASAssets/Utilities/repository/"user"/getDomainUsers(null) du LEFT OUTER JOIN /services/databases/system/ALL_GROUPS
10. Insert project level metadata. Only capture what is configured by pqInsert_METADATA_Constants
    a. METADATA_RESOURCE
    b. METADATA_RESOURCE_COLUMN
    c. METADATA_DATASOURCE
    d. METADATA_RESOURCE_LINEAGE
    e. METADATA_POLICY
    f. METADATA_POLICY_ASSIGNMNT
    g. METADATA_PRIVILEGE
    h. METADATA_PRIVILEGE_USER
    i. METADATA_PRIVILEGE_COMBINED - used to updated combined and inherited privileges in METADATA_PRIVILEGE – must be configured in pqInsert_METADATA_Constants

j. METADATA_NON_COMPLIANT – logs rule types of “ENFORCE_LAYER” and “ENFORCE_COLUMN” to determine compliancy.

11. Table statistics are executed for all tables

Metadata Partitioning Strategy:

The following diagram shows what the metadata partitioning strategy is. Using a function “F_GET_PARTITION_NUM”, the Cache_METADATA_TALES procedure uses that to insert the data into the tables. The data is directed to the correct partition based on the day of the year and accounting for leap year. Specifically, partition 60 is reserved for 2/29 on the leap year. It otherwise would be empty on a non-leap year.