Welcome to our webinar!

data virtuality

- This webinar starts in a moment - please stay tuned
- This webinar will be recorded
- You will get the slides, recording and SQL snippets
- During the webinar, you may ask questions using the Q&A button - you may ask questions anonymously
- You can hop on and off on all the sessions

Day 1	14:00 - 15:00 Automate tasks using SYS and SYSADMIN schemas 15:15 - 16:00 The DV REST API and how to use it 16:15 - 17:00 Security
Day 2	15:00 - 16:00 Synchronization Methods 16:15 - 17:15 Working with Connectors
Day 3	15:00 - 16:00 Self-service Troubleshooting 16:15 - 17:15 Job Management Outlook

DATA VIRTUALITY

Management using SYS, SYSADMIN and SYSLOG schemas

Agenda

- Motivation
- Overview of system schemas
- SYSADMIN schema
- SYSLOG schema
- SYS schema
- Summary

Motivation - Why this course?

- Help users understand what information can be gathered
- How to use system tables and procedures to automate recurring tasks
- Operate Data Virtuality more efficiently
- Explore new topics or refresh memory
- Learn something you previously haven’t had the time for

Organization

Learning objectives

- How to get information about the state of Data Virtuality's components
- How to trace changes and find logged information
- How to actually manage Data Virtuality with tasks that might be helpful
- How to find metadata across all data sources and virtual schemas that are available

Please note that we cannot cover every possible table, column and procedure. Instead, we focus on what might actually help you operate Data Virtuality more efficiently and how to make use of all the tools available.

System schemas - Overview

Overview - Available schemas

System schemas

- SYS (queryable meta information across Data Virtuality)
- SYSADMIN (the current ‘state’ of Data Virtuality)
- SYSLOG (history of Data Virtuality objects and logging information)
- SYSADMIN_VDB (SSH tunnel-related objects)
- INFORMATION_SCHEMA (metadata schema for some front end connections)
- pg_catalog (metadata schema for some front end connections)

This track will only cover **SYS**, **SYSADMIN** and **SYSLOG** as the others are not of interest or not really meant for active user interaction.

SYSADMIN schema

System schemas - SYSADMIN

Overview

- This is the main schema that deals with operating Data Virtuality
- Many tables provide insights into the current state, for example
  - Available data sources
  - Scheduled Jobs
  - User permissions
- There are also a lot of procedures that help with task automation of recurring processes and with changing components like adding a batch update or refreshing a data source’s metadata.

We will look at the tables and procedures based on the component they belong.

SYSADMIN - Data source related objects

- Tables **Connections** and **Data Sources** provide information about available data sources.
  Remember that a data source consists of a connection (the main information about how to connect) and the data source (the information how to work with the connection)
- Procedures **removeDataSource()** and **removeConnection()** to effectively drop a data source.
  Best Practice: run the in above order and alway together
- Procedures **createConnection()** and **createDataSource()** to effectively add a new data source. If anything is wrong (credentials, host, database,..) then createConnection will still succeed but createDataSource will fail. You end up only with a new connection
- Procedures **importConnection()** and **importDataSource()** are similar to their create..() counterparts.
  But importDataSource() will <u>always succeed</u> and you might add a new data source in *failed* state.

SYSADMIN - Data sources related tasks

- Find failed data sources
- List all data sources of a certain type
- List all connections that have the same connection information
- List all data sources added within the last X days or by the a specific user
- Refresh a data source’s metadata for re-reading available tables and views from the source
  - use **refreshDataSource()**
- Drop and recreate the same data source to effectively edit it
- Advanced: Script data source creation for self-service BI scenarios

SYSADMIN - How to add a new data source

SYSADMIN.createConnection()

- name
- jbossCLITemplateName
- connectionOrResourceAdapterProperties

SYSADMIN.createDataSource()

- name
- translator
- modelProperties

Demo - Working with data sources

SYSADMIN - Job and schedule related objects

Tables **ScheduleJobs** and **Schedules** contain the information about set up jobs and their schedules. Dependent schedules can be gathered from here.

The **JobEmailNotifications** table allows to see all specified email notifications for jobs and the global setting (entry with jobId *NULL*).

Management is done by the sets of procedures:

- deleteSchedulerJob, enableSchedulerJob, createSQLJob
- createSchedule, updateSchedule, deleteSchedule, enableSchedule
- various procedures for pre-defined types of jobs (materialization, batch, ...)

SYSADMIN - Job related tasks

- Set a retry counter for all existing jobs (that don’t have one)
- Calculate and set a timeout based on previous running times of successful job runs
- Find jobs violating sequential execution policy, ie. jobs that are set to be able to run in parallel
- Add a schedule to an existing job (based on the job name), especially a run-once schedule
- Temporarily disable all jobs (e.g.: during a maintenance) and re-enable them again
- Classify jobs based on their successful run
    - Failed
    - Succeeded and succeeded but needed a retry
    - Interrupted
    - Not_Started

Demo - Working with jobs

SYSADMIN - RBAC related objects

- There are tables that store what is configured
  - Roles
  - Users, UserRoles
  - Permissions
- … and procedures to add these items
  - addRole(), addUser(), addUserRole()
  - deleteRole(), deleteUser, deleteUserRole()
  - setPermissions()
  - changeUserPwd()

SYSADMIN - RBAC related tasks

- Find suspicious role assignments
  - who should really have the admin-role or super-admin-role?
- Enforce permission settings

  Make sure a certain role only has the permissions for which clearance was given
- Validate that required column masking and row-level security is set
- Remove unwanted permissions
- Script the generation of new users, roles and their permissions

Demo - Managing RBAC

SYSLOG schema

System schemas - SYSLOG

Overview

- This is the main schema that deals with traceability and logging
- There is one table for each component that can be created in Data Virtuality
  - **ConnectionHistory**, **DataSourceHistory**, **ViewDefinitionHistory** and **ProcDefinitionHistory**
  - **UserHistory**, **RoleHistory** and **PermissionHistory**
  - **ScheduleJobHistory** and **ScheduleHistory**
- There are also tables that log running
  - queries (**QueryLogs**)
  - jobs (**JobLogs**)
  - and performance counters (**QueryPerformanceLog** and **SystemPerformanceLog**)

SYSLOG - typical tasks

- Find changes of a specific object
- List all data sources (or views, procedures, ...) that were deleted
- List all operations that happened today
- Get peaks for executed queries (heap, cpu, buffers)
- Get a per-minute sum of all used resources
- Find all operations performed by a specific account
- Get the max number of concurrent sessions on the system

Demo - Trace changes using SYSLOG

SYS schema

System Schemas - SYS

Overview

- This is the main schema that contains metadata across the board
- The available tables provide information regarding
  - Data sources and virtual schemas
  - Their tables, views and procedures
  - Their tables’ and views’ columns and their data types
  - Their procedures’ parameters (in and out) and their data types
  - Key columns from data sources (these will influence the generated query plans)

Demo - Browse metadata via SYS

Summary

- Overview of system schemas and what they are used for
- How to use the SYSADMIN schema for operating Data Virtuality
- How to use the SYSLOG schema to trace the status of the system configuration and its performance
- How to use the SYS schema to browse the metadata across all data sources

Questions?

Where to get help

Reference Guide section about system tables and procedures
https://datavirtuality.com/docs/#doc=Reference%20Guide&version=2.3&page=System_Schema.html

Reference Guide section about procedural language
https://datavirtuality.com/docs/#doc=Reference%20Guide&version=2.3&page=Procedural_Language.html

Help Center
https://support.datavirtuality.com/hc

Community
https://support.datavirtuality.com/hc/en-us/community/topics

Thank you!

Please feel free to contact us at:
info@datavirtuality.com

or

visit us at:
datavirtuality.com