Data Virtualization Quick Start

|  |  |
| --- | --- |
| Instructor Frederick Wright Phone 617-672-8550 Email [fwright@eatonvance.com](mailto:fwright@eatonvance.com) Office Location Tower I – 8th Manager Dave Reid | Course Overview This presentation gives an overview of what data virtualization is, a glossary of commonly used terms, the structure of the development environment and internal procedures surrounding change control processes associated with Data Virtualization at Eaton Vance, and a brief road map for the future.  All software components of the technology stack referenced in this presentation are available on [\\evnt30\EV01Shrdata\IT](file:///\\evnt30\EV01Shrdata\IT) - Apps\Software Self-Help Resources Tutorials, www.dvknowledge.com  Support, [www.tibco.com](http://www.tibco.com) (registration required) Technology Stack These are the software components referenced in this presentation.   * Kettle/Pentaho – ETL Tool, Unlicensed Version * CIS Server – ODBC/JDBC Compliant Virtualized Database Server * CIS Studio – Developer’s Studio for creating and publishing virtual resources * Business Directory – web-based portal for managing data governance metadata and making it browsable by end uses. * RazorSQL / FlySpeed – general purpose GUI query tools for querying the virtual database. Any ODBC/JDBC client can be used. * ODBC/JDBC client drivers  Environment Default port always used, Domain is always ***eatonvance.com***. Authentication is always NTLM  *CIS*   * + Dev – bed-996-257     - LDAP Group: CISDEV-SHERPA-USER     - LDAP Group: CIS Developer   + Uat – bed-960-257     - LDAP Group: CISUAT-SHERPA-USER   + PROD – cis     - LDAP Group: CISPROD-SHERPA-USER   Kettle/Pentaho   * + Dev – bed-996-259   + Uat – bed-960-259   + Prod – kettle   Ingestion / Persistent Storage   * + Dev – bed-994-53   + UAT – bed-940-48   + PROD – bed-400-48   + Hadoop/Hive - TBD   Data Ingestion Landing Zone: [***\\bed-700-01\sherpa$***](file:///\\bed-700-01\sherpa$)  Deployment Request System: drs/ |

# Overview of CIS

Introspection – the process by which the metadata of a given data asset is registered with CIS

Resource Types

* + Folders, Packaged Queries, Views, Data Sources, Scripts/Procedures, Custom Functions, Triggers, Parameterized Queries.

Layered Architecture

* + Common vs Branch
  + Data Architecture Team owns the creation of all data sources
  + CamelCase is the standard for all resource names
  + Reserved Keywords, embedded spaces, prohibited

Publishing

In order for a resource to be made available for consumption by an end user or process, it must be published, either as a database resource, a web service resource, or both. All resources meeting the criteria of the ODATA standard will automatically be made available as an ODATA service. Refer to www.odata.org.

Deploying

Change Control is governed entirely by the Release Engineering team, and consists of submitting a Composite Archive (.car file) which contains the artifacts to be migrated from the development environment to the UAT environment. *At this point, the Data Architecture team owns the process of creating a .car file.*

**Performance Tuning of CIS Queries**

Push Query

Query Plan

DataShip

Cardinality Statistics

Caching

**Future Plans**

TIBCO Acquisition

Integration with Github for Version Control

Enabling of DataShip

Capabilities Extensions

Access to Calvert and Paraport employees

Business Directory / Data Governance