

# Podium Reference Guide

Version 1.3

April 27, 2015

# **CONTENTS**

<u>1</u>	INTRODUCTION	4
<u>2</u>	PODIUM BASICS	4
2.1 2.2	LOGGING IN USER MANAGEMENT	4
<u>3</u>	NAVIGATING PODIUM	6
3.1 3.2	DATA FLOW IN PODIUM USE CASE (GUIDE BY EXAMPLE)	6
<u>4</u>	DISCOVERY MODULE	8
4.2 4.3 4.4 4.5	4.1.1 HOW TO EDIT DATA SOURCE METADATA 4.1.2 HOW TO CREATE METADATA TAGS 4.1.3 HOW TO DELETE METADATA TAGS 4.1.4 HOW TO CREATE METADATA PROPERTIES 4.1.5 HOW TO DELETE METADATA PROPERTIES ENTITIES FIELDS 4.3.1 REAL TIME RESPONSE: SAMPLE DATA SOURCE HIERARCHY GLOBAL SEARCH	11 11 12 13 14 15 20 21
4.6 <u>5</u>	HIVE QUERY  SHOP FOR DATA	24 25
<u>s</u> 6	DATA SOURCE MODULE	28
	LOAD DATA: NEW SOURCE FROM JDBC 6.1.1 LOADING DATA LOAD DATA: NEW FLAT FILE SOURCE USING FDL LOAD DATA: NEW MAINFRAME DATA	28 30 32 36
<u>7</u>	SECURITY MODULE	40
7.1	USERS 7.1.1 HOW TO ADD A USER 7.1.2 USER PREFERENCES	<b>4</b> 0 41 42

8.1	Podium	SOURCE TYPE PROPERTIES	47
<u>8</u>	APPEN	IDIX	47
	7.2.2	HOW TO DELETE A GROUP	46
	7.2.1	How to Add a Group	44
7.2	GROUP	S	44
	7.1.3	How to Disable or Delete a User	43

# 1 Introduction

Podium, a data management platform from Podium Data enables users to source, store and provision data in a secure and scalable environment. Podium was designed to provide features for managing, discovering and preparing data.

Through an intuitive web based user interface, users can source data into Podium, browse and tag metadata to find and associate the fields they are interested in, and then select these fields for creation of custom views of their own datasets.

# 2 PODIUM BASICS

#### 2.1 Logging In

Users will gain access to data once they have entered the User ID and Password provided.



# 2.2 USER MANAGEMENT

Podium interacts with two distinct users, an Administrator and a Data User. While this can also be the same individual, the functions are different. For example, an *Administrator* can import metadata, load source data, create users and groups while the *Data User* can "discover" the data and metadata that has been loaded into Podium. Users are defined by the roles they are assigned and the groups they are members of (more details later). The Data User can be assigned as a "Master Analyst" or an "Analyst" while the Administrator is assigned an "Admin" role.

The table below summarizes Functions and Roles:

Role Function	Analyst	Master Analyst	Administrator
Load Data	N	Υ	Υ
Load / Modify Metadata	N	Υ	Υ
Browse / Search	Υ	Y	Υ
Shop For Data	Υ	Υ	Y
Data Preparation/Transformation	N	Υ	Υ
Data Export	N	Υ	Y
Create Users	N	N	Υ
Data Groups / Roles	N	N	Y

# 3 Navigating Podium

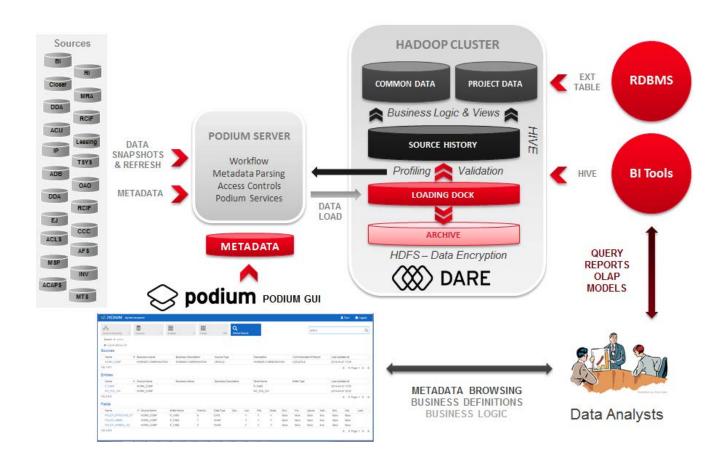
#### 3.1 DATA FLOW IN PODIUM

Before diving deep into the user interface of Podium, below is a description of what happens under the covers that will give a better understanding of what's at their fingertips. The default behavior when ingesting data in Podium, is to:

- Archive and compress a copy in its original source format (for auditing, etc.)
- Validate the data
- Profile the data
- Partition the data
- Expose the data via HIVE, Oracle External table or Teradata SQL-H table
- Encrypt the data (optional)

The users of Podium can feel confident in the data because of the exhaustive validation and profiling. The overhead of these steps is minimized by Podiums smart memory management and the leveraging of HADOOP and map reduce.

A visual flow of the data in Podium is shown in the following diagram:



As the users are "discovering data" in Podium, users can review business and technical metadata, data validation issues, data profiling statistics, and sample data to ensure they fully understand the data they are working with.

# 3.2 USE CASE (GUIDE BY EXAMPLE)

This reference guide will provide Use Cases with examples. Podium has been designed to be extremely intuitive which enables data self-service. Specifically, over time, your Data Lake Platform managed by Podium will grow to include hundreds of sources, thousands of entities and hundreds of thousands of fields. Managed by Podium, this data is secured and accessible via the Podium interface. Users of the system will want to search, browse and discover fields of data that are required for particular business initiatives. These initiatives include, but are not limited to:

- Analytic Data Sets
  - Find and Prepare data for R, SAS or other analytics tools
- Business Intelligence Reporting
  - Find and Prepare data for direct access or export to mart like systems for use with BI tools
- Data Quality and Validation
  - Create data quality reports for data governance and source system repair
- Data Transformation / ELT Offload
  - Quantify and Transform data for downstream use
    - Data Warehouse
    - Data Mart
    - Operational use

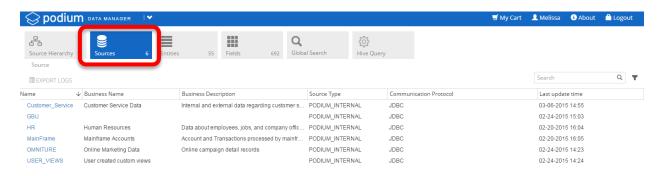
This guide will provide the reader and user with the ability to load, browse, collaborate and prepare data sets.

# 4 DISCOVERY MODULE

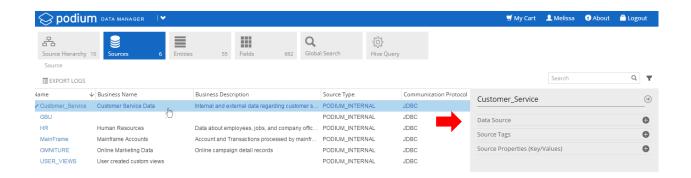


## 4.1 Sources

Once a user is logged in, the system automatically directs them to the Sources view. This provides a view of all the Sources of data and metadata the user has access to search, browse and/or edit.



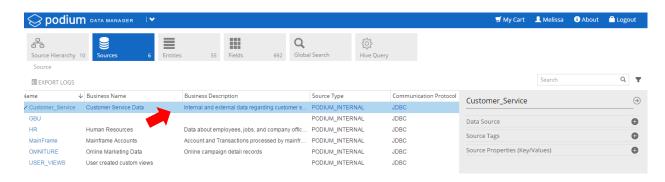
In order to view or update (depending on access level) the row for the data source should be highlighted. User must highlight the row for the Source by clicking anywhere within the row, but not on the hyperlink.



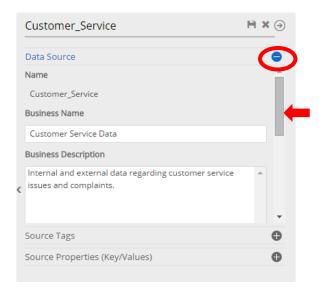
Once highlighted the metadata properties pane for the data source will show to the right. Under the data source view users can view or update the following:

Data Source Metadata		Searchable Field	Analyst Access	Admin Access
Name –	Podium name defined when Source is created in the Data Source module. Cannot contain spaces.	Y	View	View
Business Name –	Standard name for the source. Can contain spaces.	Y	View	Edit
Business Description –	Standard description of the source	Y	View	Edit
• Source Type —	Describes type of technical system data was originally sourced from. The various system types are:  o File o SQL Server o Oracle o Mainframe o Podium_Internal – the data stored in Podium's "receiving" area on the data lake. o Hive o Teradata o PostgreSQL o DB2 o MySQL	N	View	Edit
Communication     Protocol –	Technology used to bring data source into Podium. Data transfer types:  o FTP o S3 o HDFS o LocalFile o JDBC o SFTP	N	View	Edit
Last Updated –	The date and time metadata was last updated	N	View	View

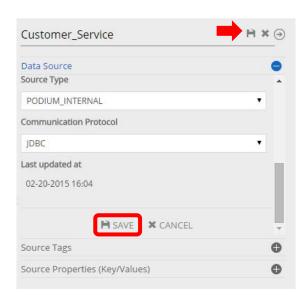
#### 4.1.1 How to Edit Data Source Metadata



Once highlighted the metadata pane for the data source will pop up to the right of the screen. Click on the "+" to see all Data Source metadata. Use the scrollbar to the right to navigate.



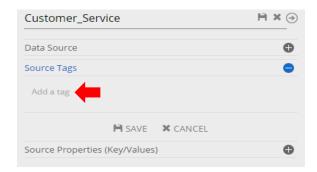
Type in edit. Hit "Save" at bottom of pane or icon on upper right corner of pane. *Note: must click "Save" for changes to be captured.* 



Source Tags Metadata	Searchable	Analyst	Admin
	Field	Access	Access
Under the data source view the user can view and/or add tags to the data source. Tags associated with the source will be shown.	Y	View	Edit

## 4.1.2 How to Create Metadata Tags

Once highlighted the metadata pane for the data source will pop up to the right of the screen. Click on the "+" to see all Source Tag metadata. Click on "Add a tag".

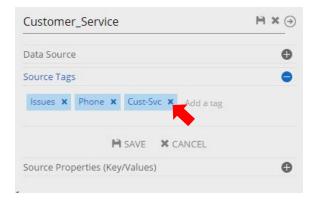


Type in new Tag. Hit Enter or Return to register the new tag. The newly entered Tag will be highlighted in blue. Hit Save or icon in order for newly created Tag to take effect.



#### 4.1.3 How to Delete Metadata Tags

Click on the "+" to view Source Tags. Click on the "x" next to the tag to be deleted. Hit Save or icon in order for changes to take effect.

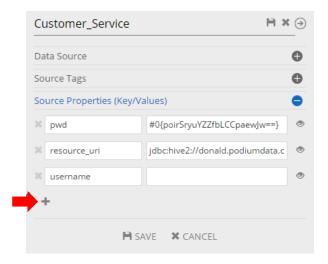


	Searchable	Analyst	Admin
Source Properties (Keys/Values) Metadata	Field	Access	Access
User can view and/or input a Key/Value pair that consists of	N	View	Edit
a name and value code that will allow access different types			
of properties within the data.			

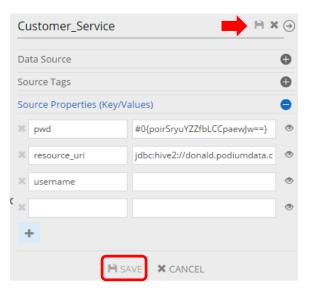
#### 4.1.4 How to Create Metadata Properties

Once highlighted the metadata pane for the data source will pop up to the right of the screen. Click on the "+" to see all Source Properties metadata.

Click on the "+" under the last listed Key/Value.



A new row with blank fields will pop up. Input Key and Value information. Once completed hit "Save" or icon in order for changes to take effect.

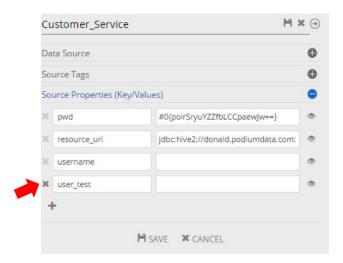


In the event that more space is needed to input Value information, click on the icon. A window will pop up. Once data is inputted, close window.



# 4.1.5 How to Delete Metadata Properties

Click on the "+" to view Source Properties. Click on the "x" next to the key/value to be deleted. Hit Save or icon in order for changes to take effect.

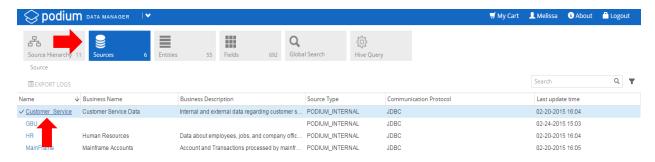


NOTE: The process for completing tags will be the same for editing the metadata for Entities and Fields.

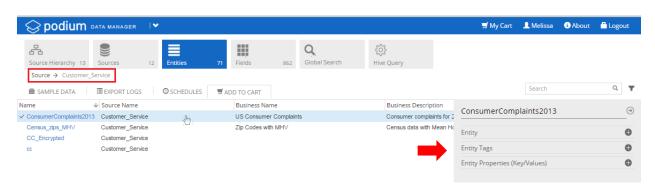
# 4.2 ENTITIES

Entities view provides a list of the collection of tables loaded from the data sources. As with the Sources view, each Entity has its own metadata properties that can be viewed and/or edited.

User can navigate to Entities section by clicking on the hyperlink on the source to see the entities associated with that source or by selecting Entities from the top navigation row.



The following are viewable and/or editable metadata properties within Entities:



Entity Metadata		Searchable Field	Analyst Access	Admin Access
Name –	Created when data is uploaded	N	View	View
Business Name –	Standard name in definition of the source	Y	View	Edit
Business Description –	Standard description of the source	Υ	View	Edit
Short Name –	Optional alternate name for the entity	N	View	Edit
<ul> <li>Entity Type –</li> </ul>			View	Edit
<ul> <li>Last Updated –</li> </ul>	Provides date and time metadata was last updated	N	View	View

Entity Tags Metadata	Searchable	Analyst	Admin
	Field	Access	Access
User can view and/or add tags to the data source, which will be searchable throughout system.	Y	Edit	Edit

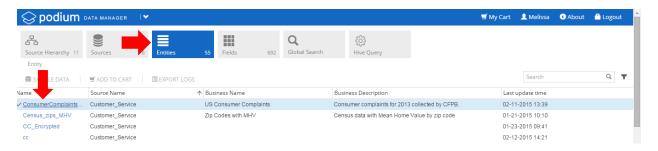
Entity Properties (Keys/Values) Metadata	Searchable Field	Analyst Access	Admin Access
User can view and/or input a Key/Value pair that consists of	N	View	Edit
a name and value code that will allow access different types			
of properties within the data.			

NOTE: Please refer to previous section on how to create and delete metadata properties.

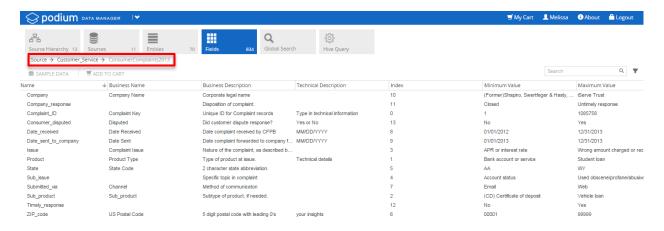
#### 4.3 FIELDS

Fields view provides a list of attributes that comprise the individual Entity. Each Field listed has its own metadata that can be viewed and/or edited. The fields also have statistical information about the content of the data loaded (data profile).

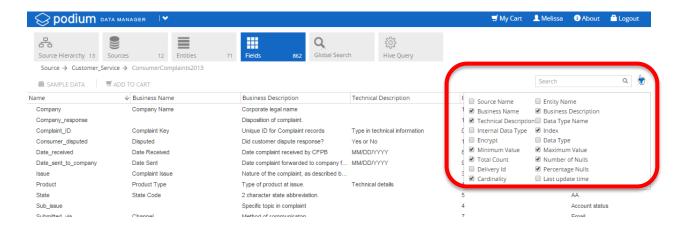
User can navigate to Fields section by clicking on the hyperlink on the entity to see the fields associated with that source or by selecting Fields from the top navigation row.



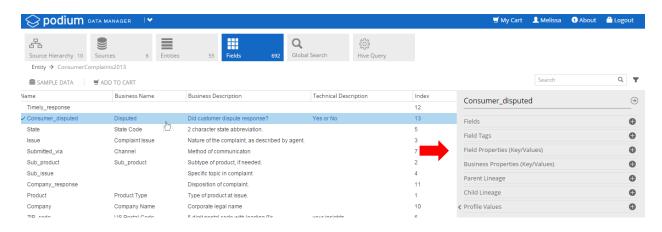
Fields have a wide variety of associated metadata, including business names and definitions, technical descriptions and format information, and profiling statistics that are calculated when the data is loaded, as show on the screen below:



As described earlier, the metadata attributes to display can be customized by selecting the "filter" icon on the upper right of the display grid.



As with other Podium objects, selecting a specific field opens the detailed metadata pane (shown below).



The following are viewable and/or editable metadata within Fields:

Fields Metadata		Searchable Field	Analyst Access	Admin Access
Name –	Created when data is uploaded	N	View	View
Business Name –	Standard name in definition of the source	Y	View	Edit
Business Description –	Standard description of the source	Y	View	Edit
Technical Description –	Freeform field to describe technical characteristics of the data	Y	View	Edit
Data Type –	The data type of the field in its source system. Types are described in the Appendix.	N	View	Edit

Internal Data Type -	<ul> <li>The data type of the field as stored in Podium receiving</li> <li>○ Integer</li> <li>○ Decimal</li> <li>○ Date</li> <li>○ String</li> <li>○ CLOB</li> <li>○ Boolean</li> </ul>	N	View	Edit
• Index –	Column position of field in table	N	View	Edit
Encrypt –	Indicates whether or not field was encrypted when loaded	N	View	Edit
Sensitive –	Indicates whether the field contains sensitive data. Used by data export to control obfuscation and masking.	N	View	
<ul> <li>Last Updated –</li> </ul>	Provides date and time metadata was last updated	N	View	View

Field Tags Metadata	Searchable Field	Analyst Access	Admin Access
User can view and/or add tags to the field, which will be searchable throughout system.	Y	Edit	Edit
Field Dreportion (Vov/Volume) Metadata	Searchable	Analyst	Admin

Field Properties (Key/Values) Metadata	Field	Access	Access
User can view and/or add tags to the field, which will be searchable throughout system.	Y	View	Edit

Business Properties (Key/Values) Metadata	Searchable	Analyst	Admin
	Field	Access	Access
User can view and/or add tags to the field, which will be searchable throughout system.	Y	View	Edit

Parent Lineage Metadata	Searchable	Analyst	Admin
	Field	Access	Access
Shows the source of the field data through a dropdown hierarchy.	N	View	View

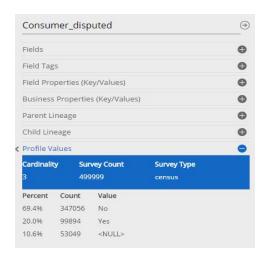


# Child Lineage Metadata Searchable Field Analyst Admin Access Shows the source of the field data through a dropdown hierarchy; identifies any other Podium data lake objects using this field. View View hierarchy:



Profile Values	Searchable	Analyst	Admin
	Field	Access	Access
Provides a snapshot view of how many distinct data values are found within the field (cardinality), survey count and survey type.	Y	View	View

User is able to view the breakdown of each cardinality by Percent, Count and Value. Values that are presented are based on the latest data loaded (by date last calendar load).



This distribution of values will provides the following information:

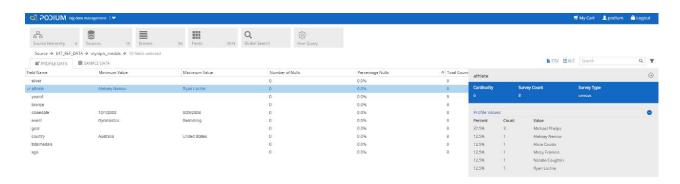
Cardinality – Uniqueness of values for that field. If the cardinality is an estimate based on a sample survey, there will be a wavy equal sign (≈) in front of the number.

#### Survey Type -

Describes the method for creating the distribution. "Census" means that every value in the field was counted for an exact distribution. "Sample" means that there were too many unique values in the field to be efficiently counted, and a sample of values were used to estimate the cardinality and distribution. "Log10Bucketing" is used for high-cardinality numeric values and counts the number of values in the range specified in the Survey Count.

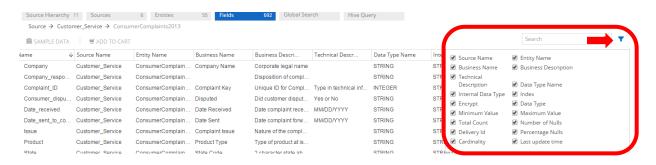
**Survey Count** – Provides the total number used to create the distribution.

**Profile Values** – This provides the actual values, count and percentage.

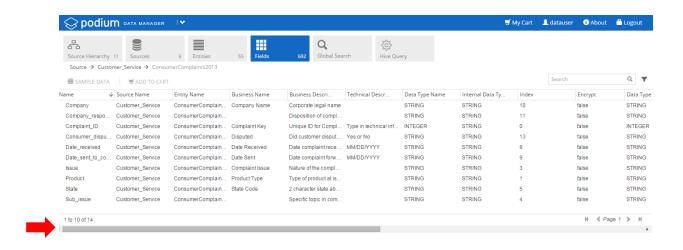


Users can limit the amount of information that is seen at a point in time by filtering the attributes being displayed. If the user wishes to permanently change the view, they must make changes to settings under User Preferences.

To make changes to current view, click on filter icon. A window will pop up displaying the attributes being viewed. User must unclick the attributes they do not wish to view. Once selections are made, click the filter icon again.



In the event the user has more attributes that than can fit on the screen, they can be seen by scrolling horizontally. Horizontal scrolling can be done by either using the horizontal scroll bar at the bottom of the screen, using the left or right arrows, or using gestures on a Mac.

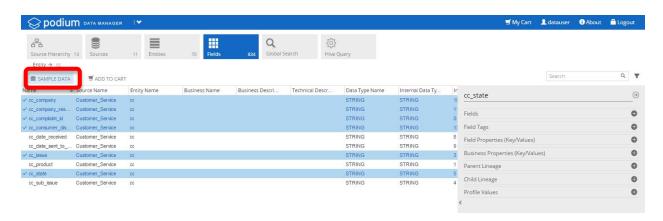


# 4.3.1 Real Time Response: Sample Data

It is often useful to see sample records when selecting data for analysis or preparation. The field view allows sample records to be reviewed interactively with the sample data function.

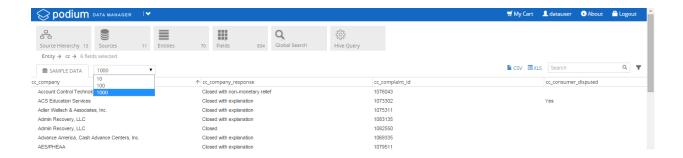
To view sample data, multi-select the fields (standard multi-select via shift-click or control-click) and click on the Sample Data link.

#### Multi-Selected Fields:



Clicking on the Sample Data button displays sample records for the highlighted fields, as illustrated below:

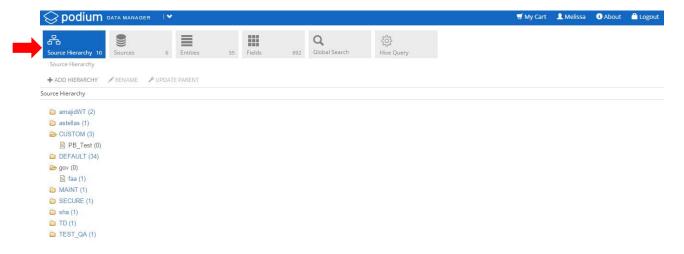
#### Sample Data:



By default 1000 records are displayed. The number of sample records can be varied through the drop-down menu above the sample data grid.

#### 4.4 Source Hierarchy

Source Hierarchy is set up as a folder system that allows for easy navigation of all data sources. Provides an organized view of Data Sources viewable and accessible to the user based on the role/function assigned.

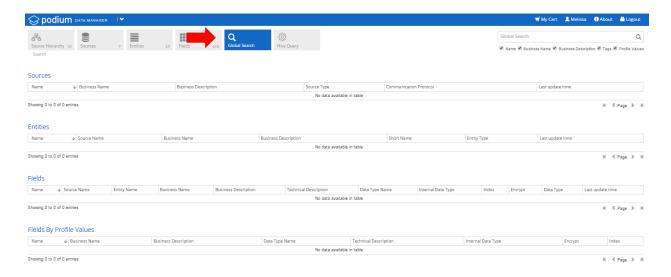


Within this view the user will be able to:

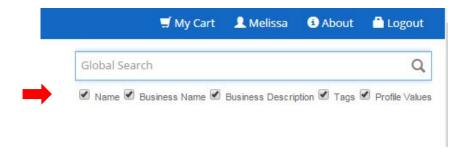
- Add Hierarchy User creates a new hierarchy, once created it a new parent level will be displayed with the name assigned
- Rename User renames the parent or child level data source
- Update Parent User can update the parent or child level hierarchy of data sources

#### 4.5 GLOBAL SEARCH

Users can conduct a "Global Search" for objects within Podium.



The search capability does a "fuzzy" search by default. Users have the ability to search the following: (selectable below the search bar)



Name: Searches the Object Name (Source Hierarchy, Source, Entity and Field)

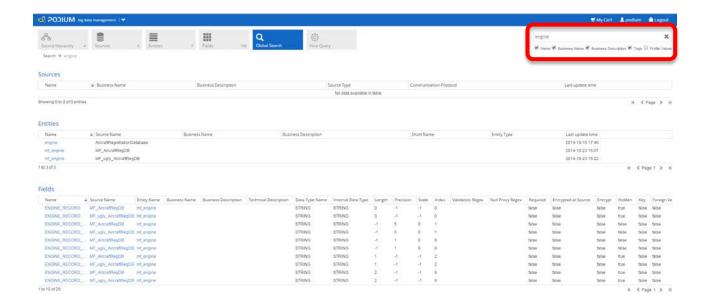
Business Name: Searches the Business Name field within each object.

**Business Description:** Searches the Business Description field within each object.

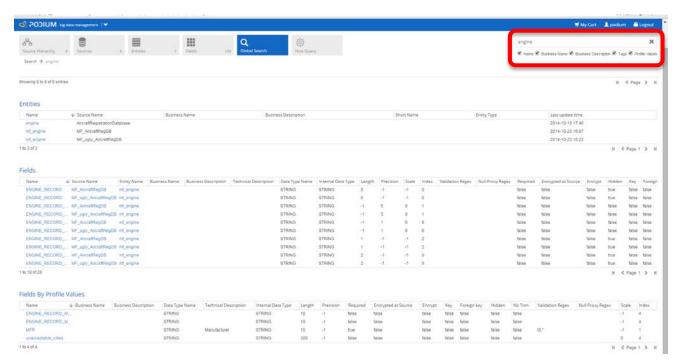
**Tags:** Searches all tags within each object.

**Profile Values:** Podium stores actual field data values as a part of profiling activities and these values are available for searching. NOTE: If an expected value does NOT return it does not mean it doesn't exist in the data, as a sample of values may be taken for a particular field dependent upon the cardinality.

A user simply types in a value for searching and all objects meeting that criteria will return to the results window. For example, the search below is for the string "ENGINE" in all searchable attributes except for Profile Values.



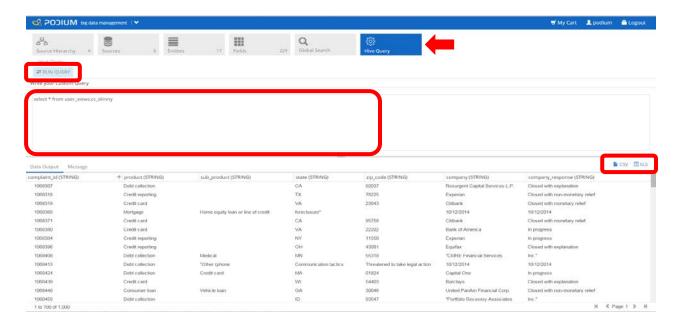
The same search with Profile Values selected:



At this point, users can select each object in the search results to get more detailed information such as profile data, sample data etc. This will help validate that fields are in fact the required fields the users are interested in.

#### 4.6 HIVE QUERY

The Podium Hive window allows users to run most HIVE commands from directly in Podium. The results are displayed in the Podium grid format. Here is an example of a simple select:



Podium uses the same object names for HIVE objects. Sources correspond to a HIVE schema, entities to HIVE tables, and fields to HIVE columns. This allows familiar HQL syntax (which is based on SQL) to be used to write HIVE queries.

Queries results are limited to 1000 rows. These results can be downloaded to your local computer as a CSV (comma separated values) or XLS (spreadsheet) file by clicking on the appropriate button above the result window on the right. The query will be re-executed before the data is downloaded.

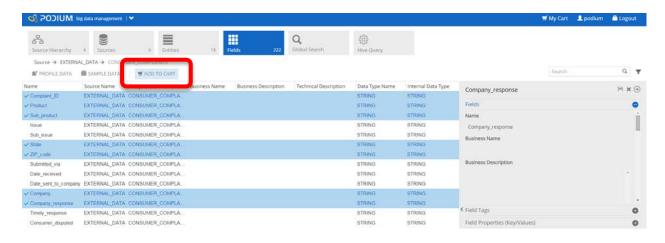
# **5** Shop for Data

Ultimately, the goal of Podium is to provide the user with a self-service interface to create the datasets they require to accomplish their jobs. This may include Data Governance, Business Intelligence, Data Provisioning (for data marts, warehouses etc.) and Analytics.

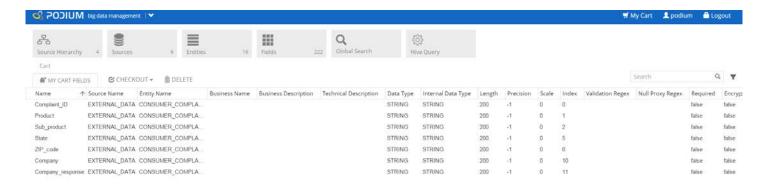
Podium's Shop for Data (SFD) feature enables to users to create custom views using a shopping cart metaphor. As users browse data fields, they can add them to their "shopping cart." When ready, users "check out" the data which creates a custom HIVE view of this data for further processing in Podium. The HIVE view is also available in HCATALOG for any tool that can connect to it (such as BI tools or RDBMS). Alternatively, SFD allows the data to be exported to an external file.

For example, suppose the user is interested in several fields in the "Consumer\_Complaints" entity shown below. They want to create a simple view of this entity containing the 7 columns they have "discovered."

The user selects (multi-selects) the fields:



The user clicks *ADD TO CART* and these fields are placed in the cart. The user can continue discovering data and adding fields, but for this Use Case, we will the go to the cart to review and checkout by clicking *MY CART FIELDS*.

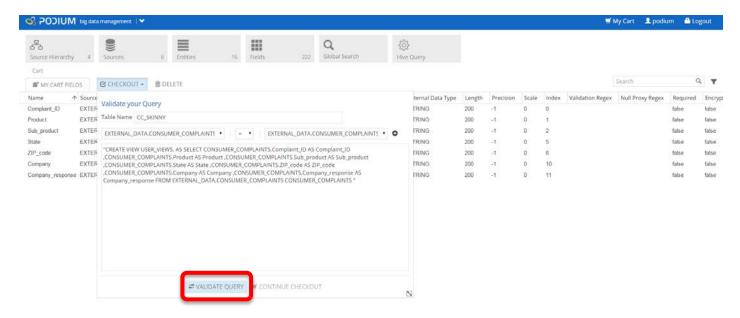


The user confirms that these are the fields they want, and click *CHECKOUT*. This opens a checkout window in which the user can do the following:

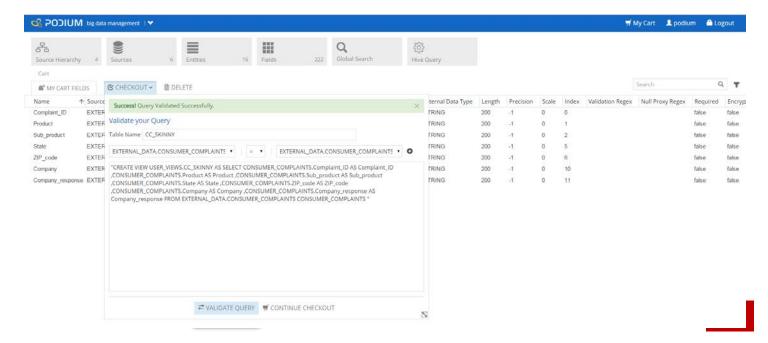
- Name the view (required)
- Add JOINS (optional)
- Add customized functions (optional)

The checkout process will validate the view definition by executing in HIVE. All HIVE syntax and functions are supported. For a more information, check out: https://cwiki.apache.org/confluence/display/Hive/LanguageManual

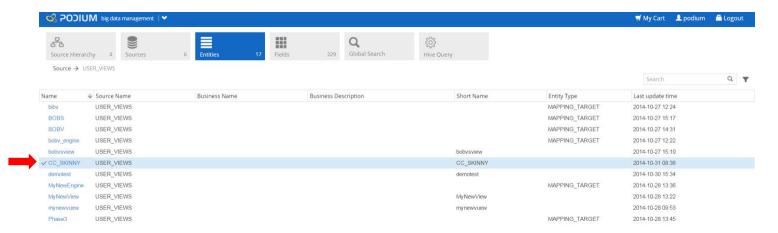
Below is how NAME of VIEW appears, just before VALIDATE QUERY is clicked:



Once *VALIDATE QUERY* is clicked, you will notice the name of the view is now in the definition in the cart, ready to be created:



Once the validation is successful, the user will click *CONTINUE CHECKOUT* and the view will be created in source: USER\_VIEWS.



This VIEW can now be queried with standard HIVE QL, exported and used to connect directly on any BI or analytics tool.

# 6 DATA SOURCE MODULE

Once logged in, users can navigate to the Data Source Module by clicking on the arrow on the top of the screen.



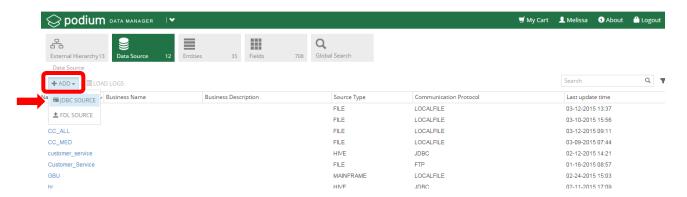
Note: Data Source Module is only accessible by Admin level users



The following will describe the various ways to load metadata from the PODIUM User Interface, including a variety of file formats and connecting to a relational database.

#### 6.1 Load Data: New Source from JDBC

From the Data Source Module (Green) section of the Podium UI select "ADD" and then select the option "JDBC SOURCE".



When adding a JDBC data source with the Podium UI, it will be done via a JDBC connection to any JDBC source database.

The wizard will guide the user through the following steps to establish connection and the selection of entity for import into Podium:

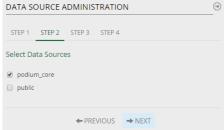
# NOTE: refer to the Appendix for a full listing of Podium Source Type properties, definitions and values

# STEP 1

- Enter the following connection information:
  - Source Type
    - Choose from: SQLServer, Oracle, Hive, Teradata, PostgreSQL
  - JDBC url input connection string
  - Username
  - Password
- Once information is entered select "TEST CONNECTION". System will provide a message verifying that connection was sucessful.
- "NEXT" should be highlighted and selected

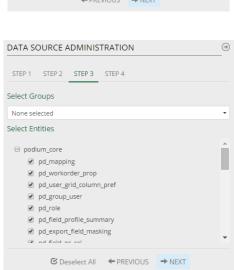
#### STEP<sub>2</sub>

- Select the Data Source (database/schema) to be used
- Click "NEXT"



#### STEP 3

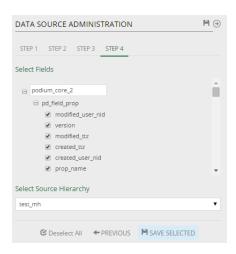
- Select Groups, if applicable, from the dropdown options. To see and use the data source, it must be added to a group the user belongs to.
- Select Entities
  - Ensure that all entities are visible expanding the view by clicking the "+"
  - Use scroll bar to the right in order to view full list of entities
- Once tables to load are selected, click on "NEXT"





#### STEP 4

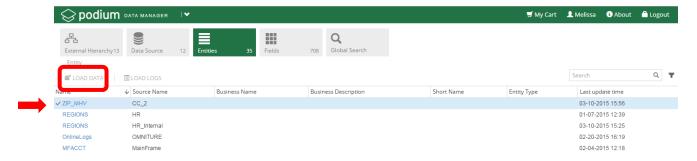
- Select Field(s)
  - Ensure that all entities are visible by expanding the view, by clicking the "+"
  - Use scroll bar to the right in order to view full list of fields
- Select Source Hierarchy for data
- Click "SAVE SELECTED" at the bottom or the save icon on upper right corner



The metadata of the new source will now be visible in the DATA SOURCE and DISCOVERY screens. However, no sample or profiling information is visible until the data is loaded for the entity.

#### 6.1.1 Loading Data

To do load data into an entity, highlight the entity and click "Load Data" on the upper left of the data source screen.



This will begin the data load. Depending on the amount of data and the speed of the connection to the source system, this may take minutes or hours. The status of the load is shown in the third column. To monitor the status, click the "Reload Data Logs" button above the Status column.



Data loading commences in two steps. The first step moves the data from the source to a staging area on HDFS called "Loading Dock." This is completed when the "Record Count" displays a number other than "0."

The second step reformats the data to the HDFS standard format, performs many data validation tests, calculates profile statistics for every field, and registers the data loaded in HCATALOG for querying. This step uses MapReduce to execute all these tasks in parallel, but takes at least 30 seconds to complete due to the overhead of launching MapReduce.

Upon completion, status will show as "FINISHED" or "FAILED" and the log will show the results of the action:



**Good** records have been validated as having correct record structures and no issues with data type violations or problems with the contents of the data.

**Bad** records do not match the specified record layout, such as having the correct number of columns, delimiters, headers and trailers. The most common reasons for bad records are:

- Incorrect layout information (wrong fixed length bytes or delimiter)
- Embedded delimiters within fields such as a common in the middle of a description or address field.
- Non-ascii or control characters causing record breaks

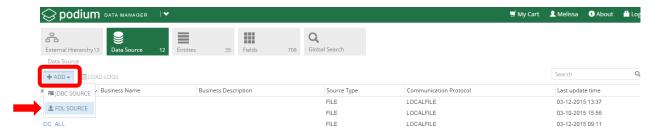
**Ugly** records match the record format but some of the field data is problematic. Most common reasons for ugly records are:

- Data type inconsistencies (e.g., non-numeric data in a field defined as numeric)
- Control characters in a field (these cause issues within the Hadoop code stack)

#### 6.2 LOAD DATA: NEW FLAT FILE SOURCE USING FDL

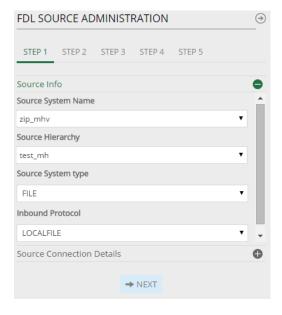
Select DATA SOURCE Tab in PODIUM to navigate to the Data Source Administration screens.

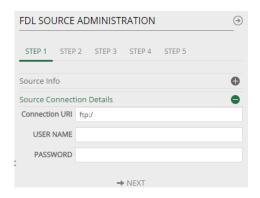
#### Select ADD> FDL SOURCE



#### STEP 1

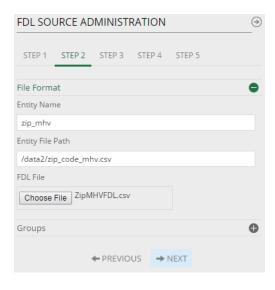
- Type in Source System Name
- Select Source Hierarchy
- Select "FILE" for Source System Type
  - Optional Source System types include: SQLServer, Oracle, MainFrame, Hive, Teradata, PostgreSQL, DB2, MySQL
- Select "LOCALFILE" for Inbound Protocol
  - Optional Inbound Protocol: FTP, S3, HDFS, SFTP
  - These Protocols require User Name and Password for authentication
- Click NEXT





## STEP<sub>2</sub>

- Type in Entity Name
- Enter Entity File Path where the test file is located
- Choose Field FDL File by clicking on "Choose File" button and selecting file to open
- Select Groups, if applicable
- Click NEXT



## STEP 3

Click on the "+" to expand the following sections to capture information:



- File Information
  - Character Encoding
    - select from: US\_ASCII, LATIN\_1,
       UTF\_8, UTF\_16LE, UTF\_16BE,
       UTF\_32LE, UTF\_32BE, US\_EBCDIC
  - Record Layout
    - select from: FIXED\_LENGTH, FIXED\_LENGTH\_TERMINATED, VARIABLE\_LENGTH\_TERMINATED, MAINFRAME\_VARIABLE\_BLOCKED



NOTE: refer to the Appendix for a full listing of Podium Source Type properties, definitions and values

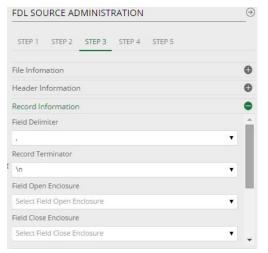
- Header Information
  - o Byte Count and Validation Regex
  - o Line Count

- Record Information
  - Field Delimiter
    - select from "," or "|" or ";"
  - Record Terminator
    - select from "\n" or "\r\n"
  - Field Open Enclosure
    - select from ' or " or <<</p>
  - Field Close Enclosure
    - select from ' or " or >>
  - Record Validation Regex
  - o Quote Scheme
  - Fixed Byte Count
- Trailer Information
  - Byte Count and Validation Regex
  - o Line Count
- Click NEXT

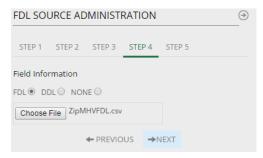
#### STEP 4

- Select "FDL" in Field Information
- Choose Field FDL File by clicking on "Choose File" button and selecting file to open
- Click NEXT







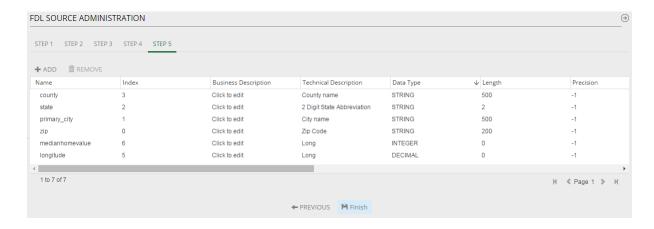


#### STEP 5

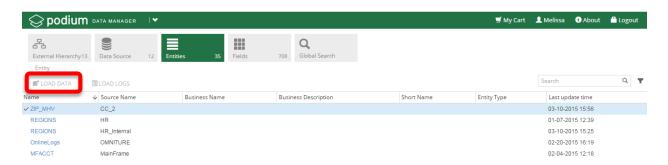
Provides the user a view of the file layout. Fields can be manually added or edited prior to loading data.

 Click "+ADD" to enter a new field or double click each field below to move into edit mode

- Enter field:
  - Name (required),
  - Business Description (optional)
  - o Technical Description
  - Data Type (required)
  - o Length
  - o Precision
- Click FINISH



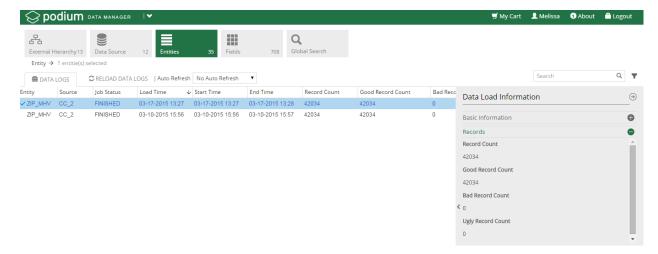
Once the previous step is successfully completed, navigate to ENTITIES screen to validate loaded data. Click LOAD DATA.



Once loading, click "RELOAD DATA LOGS" to refresh logs to see status.



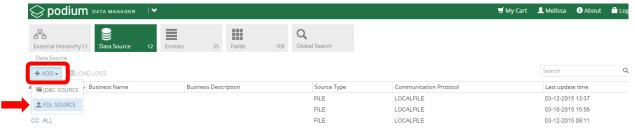
When Job Status indicates it is FINISHED, you can expand the "+" on the Records to view log record logs are bad and/or ugly.



#### 6.3 LOAD DATA: NEW MAINFRAME DATA

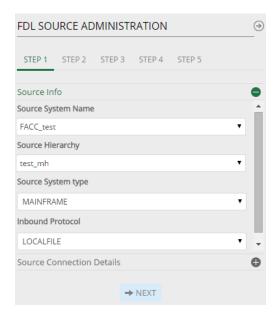
Select DATA SOURCE Tab in PODIUM to navigate to the Data Source Administration screens.



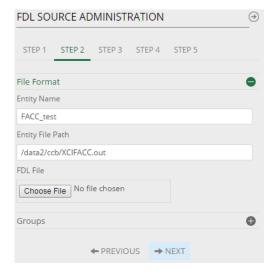


#### STEP 1

- Type in Source System Name
- Select Source Hierarchy
- Select "MAINFRAME" for Source System Type
  - Optional Source System types include: SQLServer, Oracle, File, Hive, Teradata, PostgreSQL, DB2, MySQL
- Select "LOCALFILE" for Inbound Protocol
  - Optional Inbound Protocol: FTP, S3, HDFS, SFTP
- Click NEXT



- Type in Entity Name
- Enter Entity File Path where the test file is located
- Select Groups, if applicable
- Click NEXT

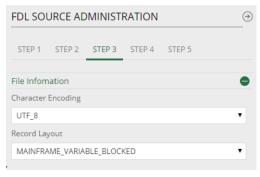


## STEP 3

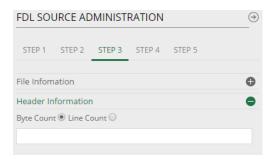
Click on the "+" to expand the following sections to capture information:



- File Information
  - Character Encoding Select US\_EBCDIC
  - Record Layout
    - select from: FIXED\_LENGTH,
       FIXED\_LENGTH\_TERMINATED,
       VARIABLE\_LENGTH\_TERMINATED,
       MAINFRAME\_VARIABLE\_BLOCKED

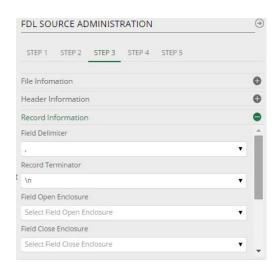


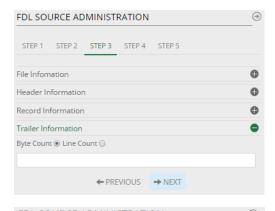
- Header Information
  - Byte Count and Validation Regex
  - Line Count

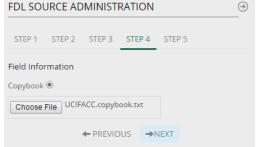


- Record Information
  - Field Delimiter
    - select from "," or "|" or ";"
  - Record Terminator
    - select from "\n" or "\r\n"
  - Field Open Enclosure
    - select from ' or " or <<</p>
  - Field Close Enclosure
    - select from ' or " or >>
  - Record Validation Regex
  - Quote Scheme
  - o Fixed Byte Count
- Trailer Information
  - Byte Count and Validation Regex
  - Line Count

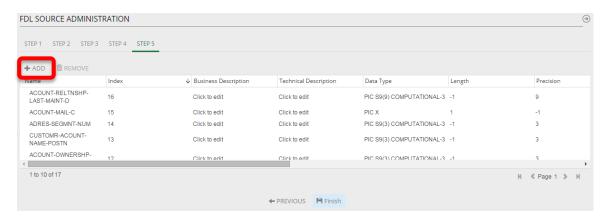
- Select "Copybook" in Field Information
- Choose file by clicking on "Choose File" button and selecting file to open
- Click NEXT







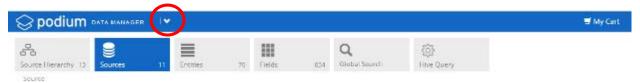
- Click "+ADD" to enter a new field or double click each field below to move into edit mode
- Enter field:
  - Name (required),
  - o Business Description (optional)
  - o Technical Description
  - Data Type (required)
  - o Length
  - o Precision
- Click FINISH.



Once the previous step is successfully completed, navigate to ENTITIES screen to validate loaded data. Follow instructions as detailed in previous section.

# 7 SECURITY MODULE

Once logged in users can navigate to the Security Module by clicking on the arrow on the top of the screen.



Note: Security Module is only accessible by Admin and Master\_Analyst level users



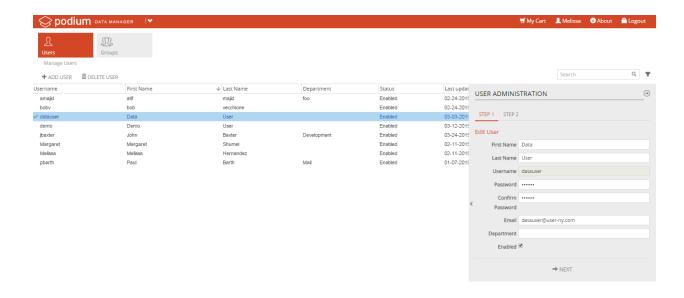
The Security Module allows for the management of Users and the creation and assignment of Users to Groups to user.

#### 7.1 USERS

Once in the Security Module, the use will be directed to the Users tab



In order to view the metadata on the individual user, the row must be hightlighted. Metadata fields for the individual are visible under the USER ADMINISTRATION pane. STEP 1 allows for edits to the User, STEP 2 allows for adding users to Groups.

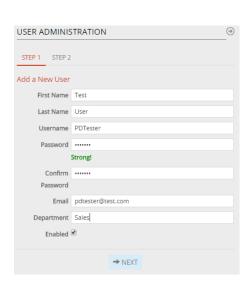


#### 7.1.1 How to Add a User

To add a User to the Podium UI, under User tab click on "+ ADD USER". The USER ADMINISTRATION pane will open up to the right hand side of the screen.



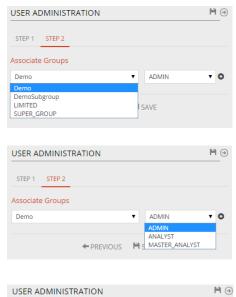
- Input the following profile information for New User:
  - o First Name
  - o Last Name
  - o Username
  - Password system will indicate if password selected is weak, medium or strong
  - o Confirm Password
  - o Email
  - Department
- Ensure "Enable" box is checked off
- Click "NEXT"



## STEP<sub>2</sub>

If New User is to be associated with any Group(s):

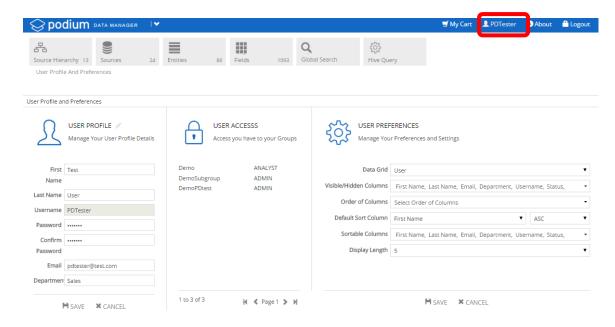
- Select Group from dropdown
- Select the access level the User will have as it pertatins to the Group
- Click on the "+" to add
- If the New User needs to be added to multiple Groups, repeat the step above
- Hit SAVE, once all Groups for the New User have been assigned





#### 7.1.2 User Preferences

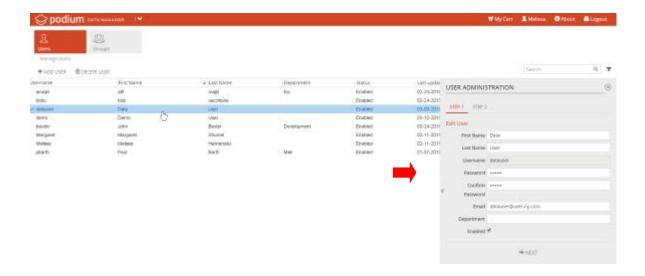
Once a New User has been setup in the Podium UI system, they are able to view their own Profile, Group Access levels and system preferences by clicking on their Username when logged in.



#### 7.1.3 How to Disable or Delete a User

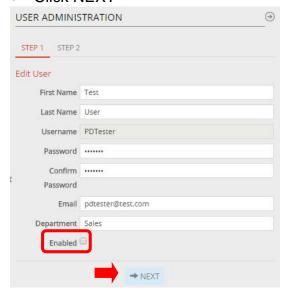
Access to the Podium UI can be either temporarily disabled for or User or they can be completely removed from the system.

In the event that access to Podium should be temporary disabled for a User, on the User tab, locate the User and highlight the row.



## STEP 1

- Disable access by unchecking the "Enable" box
- Click NEXT



## STEP<sub>2</sub>

Click SAVE



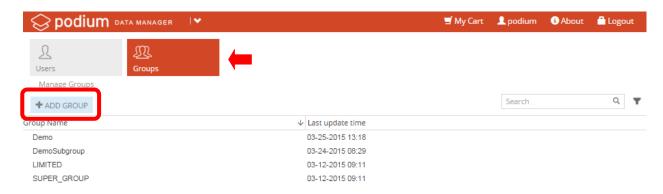
To remove access and delete User from Podium system, on the User tab, locate the User and highlight the row. Click on "DELETE USER".



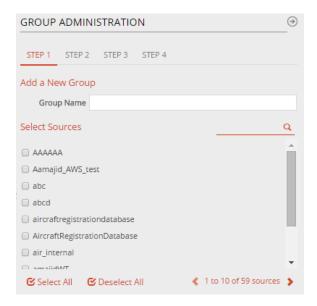
## 7.2 GROUPS

## 7.2.1 How to Add a Group

In order to add a Group(s) to Podium, on the Groups tab, click on "+ADD GROUP". The GROUP ADMINISTRATION pane will open on the right of the screen.



- Input the Group Name
- Select Sources Datasets that are available to New Group are listed
  - Selection can be made by:
    - Clicking each one individudally
    - Using the "Select All" option on the bottom left of the pane
    - Running a search using the Search bar
       Note: bottom right indicates total number of datasets available
- Click NEXT



- Select Entitites all available entities for the datasets selected are visible
- Selection can be made by:
  - o Clicking each one individudally
  - Using the "Select All" option on the bottom left of the pane
  - Running a search using the Search bar
- Click NEXT

- Add existing groups to this group by selecting a group from the drop down, which provides a list of Groups currently in Podium Select Entitites – all available entities for the datasets selected are visible
- Click on "+"
- Selection will be captured under Subgroups
- Click NEXT

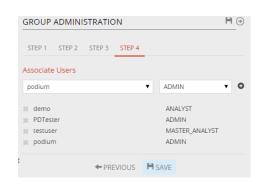






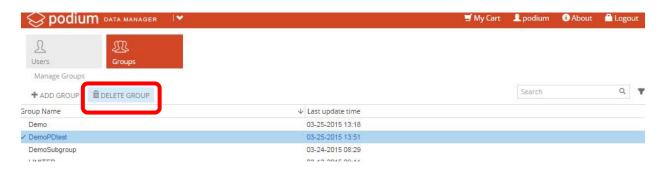
Select User(s) is to be associated with the New Group

- Select User from dropdown list
- Select the access level the User will have as it pertatins to the Group
- Click on the "+" to add
- If mulitple Users need to be added the New Group, repeat the step above
- Hit SAVE, once all Users have been assigned



## 7.2.2 How to Delete a Group

In order to delete a Group(s), on the Groups tab, highlight the row to be deleted. Click on "DELETE GROUP".



# 8 APPENDIX

# 8.1 PODIUM SOURCE TYPE PROPERTIES

Podium Property	Description	Values	
src.comm.protocol	communications protocol used to pull data from	S3, LOCALFILE, HDFS, FTP	
	source	FILE 001 0FD1/FD 02 101 F	
src.sys.type	source system type	FILE, SQLSERVER, ORACLE, MAINFRAME, PODIUM_INTERNAL, HIVE, TERADATA, POSTGRESQL, DB2, MYSQL	
src.entity.name	name of entity/table	user specified	
src.file.glob	wild-card glob of data files for this source (or select statement for jdbc source)	User defined (JDBC - system defined, user modifiable)	
unicode.byte.order.mark	Unicode byte order mark	UTF_32LE_BOM, UTF_32BE_BOM, UTF_8_BOM, UTF_16LE_BOM, UTF_16BE_BOM	
unicode.byte.order.mark.incidenc	defines expected BOM incidence	NEVER, SOMETIMES, ALWAYS	
header.byte.count	optional integer that specifies the fixed number of bytes in the file header	user specified	
header.line.count	optional integer that specifies a fixed number of lines in the file header	user specified	
header.defines.field.names	boolean which indicates whether or not the header specifies the field names	TRUE / FALSE	
header.record.count.regex	optional regular expression with a capturing group for extracting the record count from the header	Regular Expression	
header.validation.pattern.is.regex	boolean flag which controls whether or not header validation pattern is a regular expression	TRUE / FALSE	
header.validation.pattern	pattern for validating header either exact string match or regular expression	user specified	
trailer.byte.count	optional integer that specifies a fixed number of bytes in the file trailer	user specified	
trailer.line.count	optional integer that specifies a fixed number of lines in the file header where line is terminated by \n or \r\n	user specified	
trailer.record.count.regex	optional regular expression with a capturing group for extracting the record count from the trailer	Regular Expression	
trailer.validation.pattern.is.regex	boolean flag which controls whether or not trailer validation pattern is a regular expression	TRUE / FALSE	
trailer.validation.pattern	pattern for validating trailer either exact string match or regular expression	user specified	
trailer.control.z.at.eof.incidence	incidence of ControlZ/0x1A MS-DOS end of text file indicator	NEVER, SOMETIMES, ALWAYS	
record.characterset	characterset definition	US_ASCII, LATIN_1, UTF_8, UTF_16LE, UTF_16BE, UTF_32LE, UTF_32BE, US_EBCDIC	
record.charencoding.confidence	confidence in the char encoding	decimal in the range 0.0-1.0	
record.layout	defined how the record is laid out (fixed, delimitted etc)	FIXED_LENGTH, FIXED_LENGTH_TERMINATED, VARIABLE_LENGTH_TERMINAT ED, MAINFRAME_VARIABLE_BLOCK ED	

Podium Property	Description	Values
record.fixed.byte.count	fixed record byte length for FIXED_LENGTH and FIXED_LENGTH_TERMINATED records	user defined (CCB - utility will determine)
record.min.byte.count	minimum valid record byte length	user defined
record.max.byte.count	maximum valid record byte length	user defined
record.record.terminator	record terminator char sequence for FIXED_LENGTH_TERMINATED and VARIABLE_LENGTH_TERMINATED	user defined
record.field.delimiter	char sequence for delimiting fields for delimited record types	user defined
record.last.field.has.delimiter	boolean indicating whether there is a final field delimiter in the record	TRUE / FALSE
record.validation.regex	optional regular expression used to validate a record	Regular Expression
record.open.quote	char sequence used for opening an enclosed/quoted field	user defined
record.close.quote	char sequence used for closing an enclosed/quoted field	user defined
default.field.embedded.enclosure.s cheme	how to deal with an enclosure inside an enclosed/quoted field	NONE or DOUBLE_EMBEDDED_ENCLOS URE
default.field.enclosure.incidence	incidence of enclosed fields	NEVER, SOMETIMES, ALWAYS
default.field.nullif.pattern.is.regex	boolean flag which controls whether or not nullif pattern is a regular expression	TRUE / FALSE
default.field.nullif.pattern	char sequence which represents the NULL value exact match or regular expression	user defined
default.field.trim.left	boolean controlling whether or not whitespace should be trimmed from left of fields	TRUE / FALSE
default.field.trim.right	boolean controlling whether or not whitespace should be trimmed from right of fields	TRUE / FALSE
default.field.allow.whitespace	boolean controlling whether or not embedded whitespace is allowed in fields	TRUE / FALSE
default.field.allow.control.chars	boolean controlling whether or not ASCII/Unicode control chars 0x00-0x1F + 0x7F are allowed in fields	TRUE / FALSE
default.field.allow.non.ascii.chars	boolean controlling whether or not non 7-bit ASCII chars are allowed in fields	TRUE / FALSE
default.field.min.legal.char.length	minimum number of chars allowed in a field	user defined
default.field.max.legal.char.length	maximum number of chars allowed in a field	user defined
default.date.format	date format string in Java/JODA date format or vendor-specific date format	user defined
default.min.legal.date	minimum legal date as yyyy-MM-dd ISO 8601 date	user defined
default.max.legal.date	maximum legal date as yyyy-MM-dd ISO 8601 date	user defined
field.trim.left	boolean controlling whether or not whitespace should be trimmed from the left of this specific field	TRUE / FALSE
field.trim.right	boolean controlling whether or not whitespace should be trimmed from the right of this specific field	TRUE / FALSE
field.allow.whitespace	boolean controlling whether or not embedded whitespace is allowed in this specific field	TRUE / FALSE

Podium Property	Description	Values
field.allow.control.chars	boolean controlling whether or not ASCII/Unicode control chars 0x00-0x1F + 0x7F are allowed in this specific field	TRUE / FALSE
field.allow.non.ascii.chars	boolean controlling whether or not non 7-bit ASCII chars are allowed in this specific field	TRUE / FALSE
field.enclosure.incidence	incidence of enclosures for this specific field	NEVER, SOMETIMES, ALWAYS
field.min.legal.char.length	minimum number of chars allowed in this specific field	user defined
field.max.legal.char.length	maximum number of chars allowed in this specific field	user defined
field.min.legal.value	minimum legal value for this specific field interpretation depends upon field type	user defined
field.max.legal.value	minimum legal value for this specific field interpretation depends upon field type	user defined
field.encryption.type.tag	optional tag to identify the type of animal that is being encrypted	user defined
field.date.format	date format string in Java/JODA date format or vendor-specific date format for this specific field	user defined
enable.profiling	boolean controlling whether or not profiling should be performed on this source/entity/field	TRUE / FALSE
enable.validation	boolean controlling whether or not field validation should be performed	TRUE / FALSE
enable.archiving	whether or not inbound files should be archived	TRUE / FALSE
use.single.receiving.mapper	use a single receiving mapper instead of MapReduce in the cluster	TRUE / FALSE
cobol.copybook	the text of the copybook	System defined
cobol.branch.wiring.instructions	branch wiring instructions to provide guidance for decoding COBOL copybooks with REDEFINES	user defined
cobol.trunc.bin.enabled	COBOL compiler/environment option TRUNC(BIN) for truncation of BINARY/COMP-1 numeric items	TRUE / FALSE
cobol.supports.single.byte.binary	whether or not COBOL compiler supports single byte BINARY/COMP-1 numeric items non IBM mainframe	TRUE / FALSE
cobol.sync.alignment.enabled	COBOL compiler/environment option for SYNC word alignment	TRUE / FALSE
cobol.unsigned.packed.decimal.uses.sign.nybble	whether or not the sign nybble holds a digit for unsigned PACKED/COMP-3	TRUE / FALSE
cobol.little.endian	set to TRUE if the source COBOL machine architecture is little endian non IBM mainframe	TRUE / FALSE
cobol.allow.overflow.redefines	set to TRUE if you want to allow REDEFINES branches which are larger than the original storage allocation	TRUE / FALSE
cobol.allow.underflow.redefines	whether or not you want to allow REDEFINES branches which are smaller than the original storage allocation	TRUE / FALSE
cobol.numproc.nopfd.enabled	COBOL compiler/environment option NUMPROC(NOPFD) relating to valid/invalid sign nybbles in PACKED/COMP-3 numeric items	TRUE / FALSE
field validation regular expression	special class enabling regular expression checks at a field level	Regular Expression