

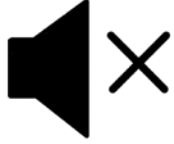
10-27-2020

Lineage of Databricks in EDC

Sachin Jain

sjain@informatica.com

Housekeeping Tips



- Today's Webinar is scheduled for **1 hour**
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the **Q&A option** and we will respond at the end of the presentation
- The webinar is **being recorded** and will be available to view on our **INFASupport YouTube channel** and **Success Portal**. The link will be emailed as well.
- Please take time to complete the **post-webinar survey** and provide your feedback and suggestions for upcoming topics.

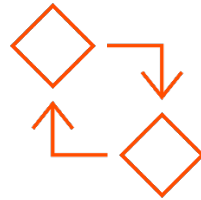
Feature Rich Success Portal



Bootstrap trial and
POC Customers



Enriched Customer
Onboarding
experience



Product Learning
Paths and Weekly
Expert Sessions



Informatica
Concierge with
Chatbot integrations



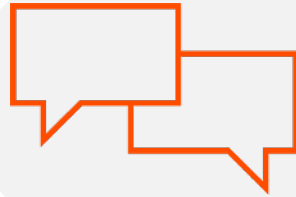
Tailored training and
content
recommendations

More Information



Success Portal

<https://success.informatica.com>



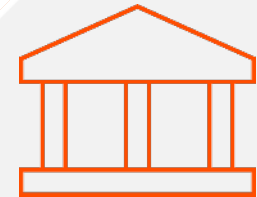
Communities & Support

<https://network.informatica.com>



Documentation

<https://docs.informatica.com>



University

<https://www.informatica.com/in/services-and-training/informatica-university.html>

Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informativa product or functionality described today remain at the sole discretion of Informativa and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

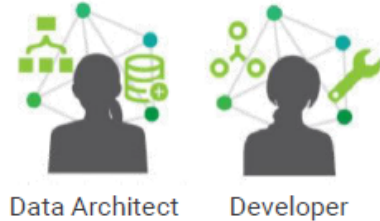


Agenda

- Business Use Case
- DEI configurations and Mapping Details
- Scanner Configurations
- Demo

Change Impact Analysis

Example scenario



John is a solution architect at a retail chain. They will be implementing a new ERP system and will need to assess the effort involved in retiring the data assets from their legacy system and identifying all the downstream impacts of this effort.

John will work with the project team to search the data bases associated with the legacy system and view all the upstream and downstream relationships and lineage.

This will provide them an inventory of data assets that must be retired, replaced or integrated with new systems. The data can be exported from EDC for further analysis. And the legacy assets can be annotated to reduce analysis time by other project teams

Search

Identify

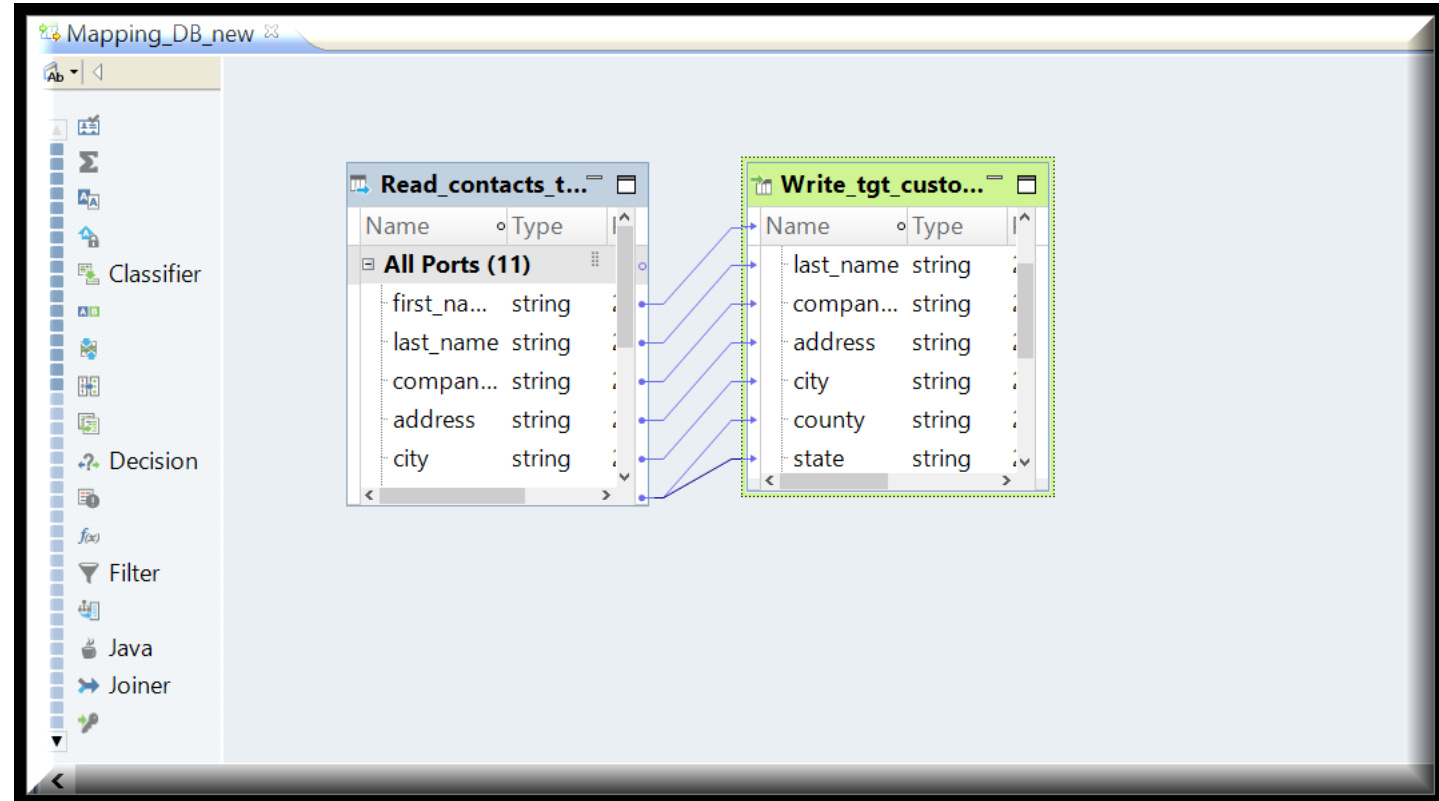
View
Relationships

Analyze
Transformation

Export Change
Impact

Simple DEI/BDM Mapping

- Databricks Cluster Should be Up and Running
- Cluster Configuration
- JDBC connection details
- Simple mapping (delta tables)
- Mapping should be deployed to Application



BDM/DEI Scanner Configurations in EDC

The screenshot displays the Informatica EDC interface for configuring a BDM resource named 'BDM_sachin'. The 'Resource' tab is active, and the 'Metadata Load Settings' sub-tab is selected. The 'Source Metadata (Enabled)' section is expanded, showing 'Additional properties' and 'Advanced Properties'. The 'Connection Assignment Properties' section is highlighted with a red box, containing the following settings:

Property	Value
Auto Assign Connections	true
Enable Reference Resources	Yes
Retain Unresolved Reference Assets	true

Other visible settings include 'Detailed Lineage (Applicable for Target version 10.2.1 onwards)' set to 'true', 'Memory' set to 'Medium', and 'JVM Options' set to '---'. The 'Edit' and 'Run' buttons are located in the top right corner of the configuration area.

Monitor The Status in EDC

The screenshot displays the Informatica EDC Monitoring interface. At the top, there are navigation tabs for Overview, Library, Assistance, and Monitoring. The current view is 'Monitoring', showing 'Last refreshed at 10/10/2020 07:55:26 AM'. The interface includes a 'Task Summary' section with a donut chart showing 136 tasks, a 'Task distribution' section with a donut chart showing 136 tasks, and a 'Tasks' table listing individual task runs.

Task Status Legend:

- All
- Cancelled
- Failed
- Paused
- Completed

Task distribution Legend:

- All
- Metadata Load
- Composite Domain Inference
- DataDomain Sync
- Delete
- SystemLocationScanner
- CustomMetadataScanner
- Profile Run
- Profile Result Fetch
- DDPScanner
- Purge
- Similarity Discovery

Tasks Table:

Type	Resource Name	Schedule	Triggered by	Status	Start time	End time	Run time (HH:MM:SS)	Next Schedule	Log URL
Metadata Load	BDM_Acctn	---	Manual	Completed	Fri, Oct 16, 2020, 7:42:13 AM India	Fri, Oct 16, 2020, 7:48:52 AM India	00:06:39	---	Log Location
Metadata Load	DemarcUser	---	Manual	Completed	Fri, Oct 16, 2020, 7:38:05 AM India	Fri, Oct 16, 2020, 7:45:28 AM India	00:06:22	---	Log Location
Metadata Load	ChangeNotification	---	Manual	Completed	Fri, Oct 16, 2020, 7:00:44 AM India	Fri, Oct 16, 2020, 7:04:26 AM India	00:03:42	Fri, Oct 16, 2020, 8:23:23 AM India	Log Location
Metadata Load	CustomResource_C...	---	Schedule	Failed	Fri, Oct 16, 2020, 1:00:54 AM India	Fri, Oct 16, 2020, 1:00:59 AM India	00:00:05	---	Log Location
File Script	CustomResource_C...	---	Manual	Completed	Fri, Oct 16, 2020, 1:00:54 AM India	Fri, Oct 16, 2020, 1:00:59 AM India	00:00:05	---	Log Location
File Script	TestResource_C...	---	Manual	Completed	Thu, Sep 24, 2020, 12:10:52 PM Ind.	Thu, Sep 24, 2020, 12:14:00 PM Ind.	00:03:08	Sat, Oct 17, 2020, 1:00:00 AM India	Log Location
File Script	TestResource_C...	---	Manual	Failed	Thu, Sep 24, 2020, 12:09:37 PM Ind.	Thu, Sep 24, 2020, 12:10:03 PM Ind.	00:00:26	---	Log Location
Metadata Load	PC_Test	---	Manual	Completed	Thu, Sep 17, 2020, 4:20:34 PM Indi.	Thu, Sep 17, 2020, 4:20:43 PM Indi.	00:00:07	---	Log Location
Metadata Load	PC_EBID	---	Manual	Completed	Thu, Sep 16, 2020, 3:42:29 PM Indi.	Thu, Sep 16, 2020, 3:48:24 PM Indi.	00:03:54	---	Log Location
Metadata Load	PC_EBID	---	Manual	Completed	Tue, Sep 8, 2020, 1:55:54 PM India	Tue, Sep 8, 2020, 2:05:24 PM India	00:09:29	---	Log Location

Reference Resources

The screenshot displays the Informatica Library interface. The left sidebar shows a navigation menu with 'Reference Resources' highlighted. The main content area shows a table of Reference Resources under the heading 'Reference Resources (6)'. The table has columns for 'Primary Resource', 'Description', and 'Resource Type'. The resource 'BDM_sachin_BDMScanner_Databricks_JDBC_conn_default' is highlighted with a red box. Below the table, there is a pagination control showing 'Page 1 of 2' and a 'View 1 - 4 of 5' indicator.

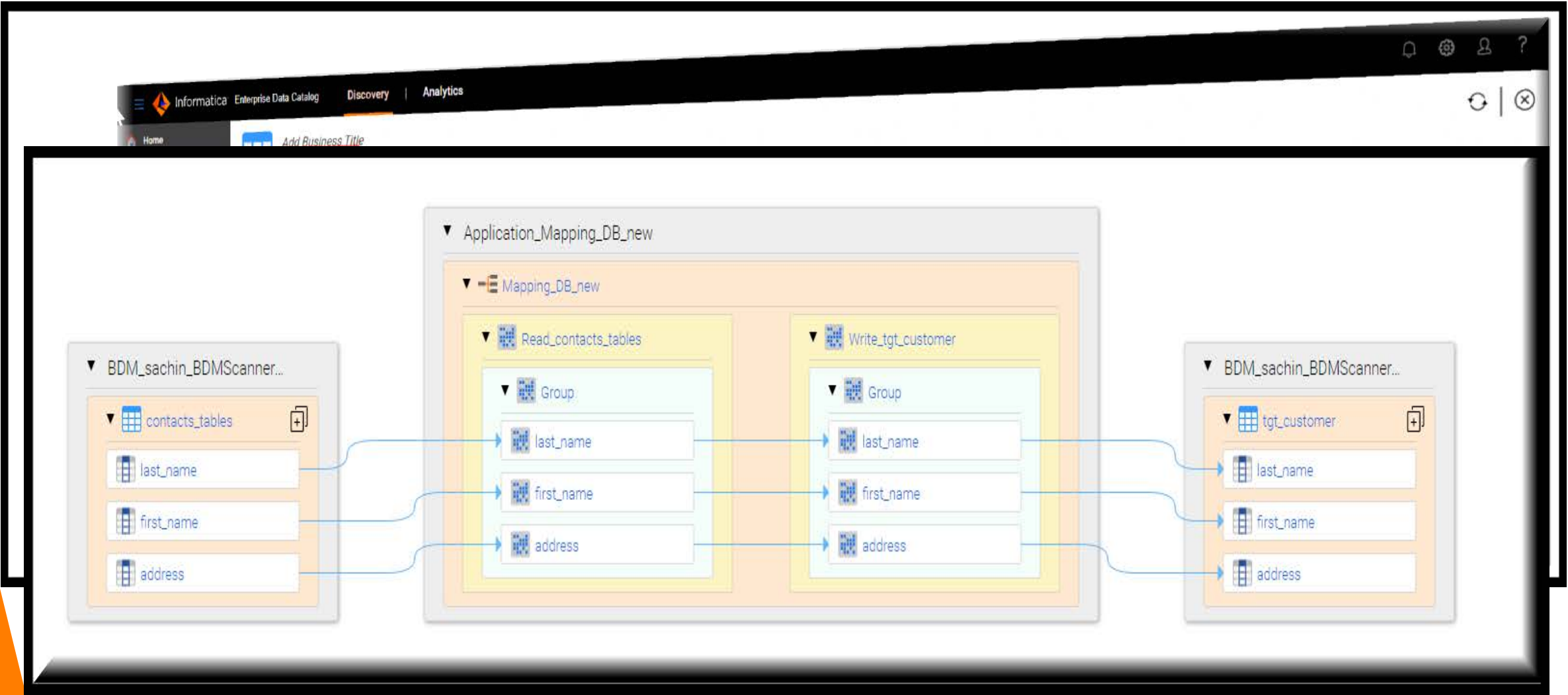
Primary Resource	Description	Resource Type
BDM_sachin		Informatica Platform
BDM_sachin_BDMScanner_sachin_connection		Reference resource
BDM_sachin_BDMScanner_Sales_sachin		Reference resource
BDM_sachin_BDMScanner_Databricks_JDBC_conn_default		Reference resource
BDM_sachin_BDMScanner_Redshift_Sachin_ebf17206		Reference resource
DEI_1041		Informatica Platform
DQScanner		Informatica Data Quality
PC_EBRD		PowerCenter
PC_Test		PowerCenter
pc_scanner		PowerCenter

Auto Assign Connections

The screenshot shows the Informatica Connection Assignment interface. The 'Connection Assignment' tab is active, and the 'Auto Assigned Connections' sub-tab is selected. A table lists various connections and their assignment status.

Resource Name	Connection	Resource Type	Schema/End-point	Resource Name	Resource Type	Assignment Type
<input type="checkbox"/> AWS_Glue <ul style="list-style-type: none"> <input type="checkbox"/> amazons3_infa-gcs_akhila → Unassigned <input type="checkbox"/> amazons3_infa-gcs-redshift-staging_/ → Unassigned <input type="checkbox"/> amazons3_infa-gcs_neelima → Unassigned <input type="checkbox"/> amazons3_infagcs_CDI-E → Unassigned <input type="checkbox"/> Athena_us-east-1_AwsDataCatalog → Unassigned <input type="checkbox"/> amazons3_athena-examples-us-east-1_elb → Unassigned <input type="checkbox"/> amazons3_infa-gcs_/ → Unassigned 	<input type="checkbox"/> BDM_sachin <ul style="list-style-type: none"> <input type="checkbox"/> Databricks_JDBC_conn.default → Unassigned <input type="checkbox"/> sachin_connection → Unassigned <input type="checkbox"/> Redshift_Sachin.ebf17206 → rsqa → Redshift_ankit → Amazon Redshift → Assigned <input type="checkbox"/> Sales_sachin → C02658711J → Oracle_Sales_DB → Oracle → Assigned <input type="checkbox"/> Databricks_JDBC_conn → default → BDM_sachin_BDMScanner_Databricks_JDBC... → Reference resource → Assigned 					

End-to-End Lineage



Demo



Informatica™