

# Information Governance Catalog - Lineage

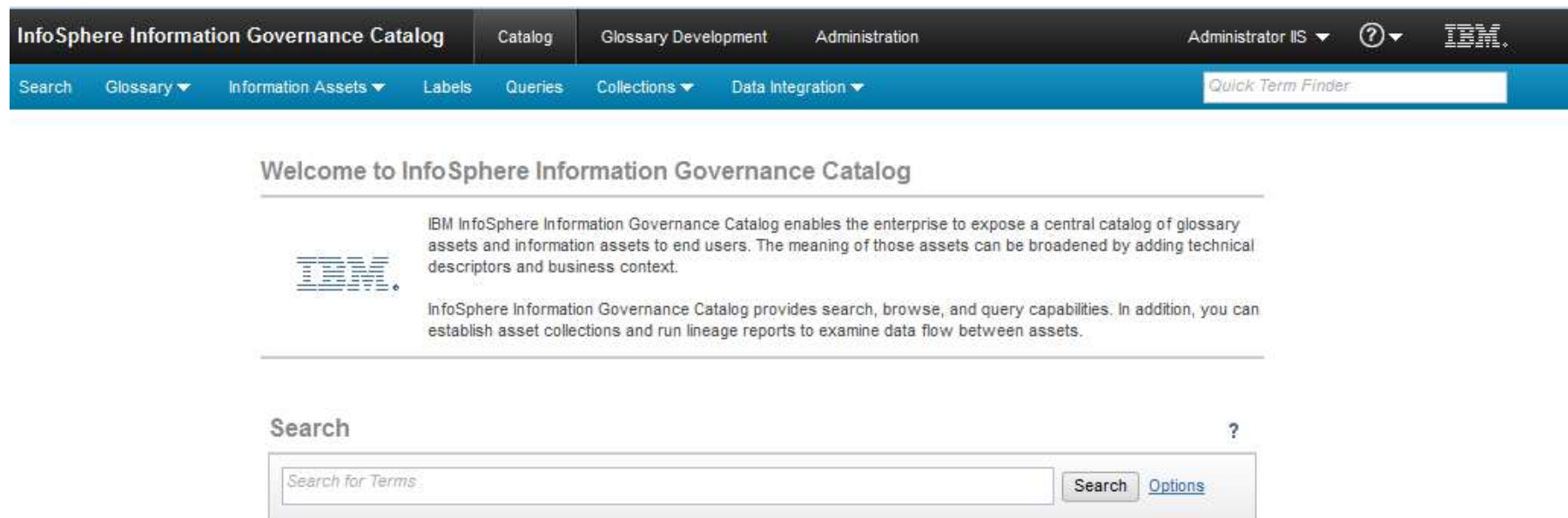


## Objectives

- Understand the Information Governance Catalog Assets
  - Queries
  - Collections and Labels
  - Lineage

## Information Governance Catalog

- Allows you to understand where information came from and where it is used
- A key enabler to regulatory compliance and the IBM Data Governance Maturity Model
- Cross-tool reporting on:
  - Data movement and lineage
  - Business meaning
  - Impact of changes
  - Dependencies
  - Data lineage for BI Reports



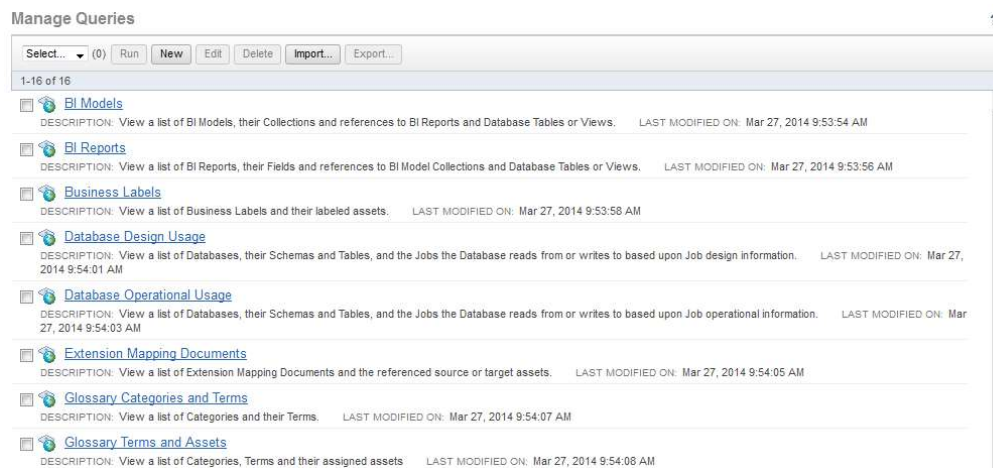
The screenshot displays the IBM InfoSphere Information Governance Catalog web application. The top navigation bar includes links for 'InfoSphere Information Governance Catalog', 'Catalog', 'Glossary Development', and 'Administration'. On the right, it shows 'Administrator IIS' and a help icon. Below this is a secondary navigation bar with 'Search', 'Glossary', 'Information Assets', 'Labels', 'Queries', 'Collections', and 'Data Integration'. A 'Quick Term Finder' search box is also present. The main content area features a 'Welcome to InfoSphere Information Governance Catalog' heading, followed by an IBM logo and a paragraph explaining the catalog's purpose: 'IBM InfoSphere Information Governance Catalog enables the enterprise to expose a central catalog of glossary assets and information assets to end users. The meaning of those assets can be broadened by adding technical descriptors and business context.' Below this, another paragraph states: 'InfoSphere Information Governance Catalog provides search, browse, and query capabilities. In addition, you can establish asset collections and run lineage reports to examine data flow between assets.' At the bottom, there is a 'Search' section with a search input field labeled 'Search for Terms', a 'Search' button, and a link to 'Options'.

## Information Governance Catalog Queries

*Gain visibility to a more complete view of information assets*

### ■ Explore

- Jobs, reports, databases, files, tables, columns, terms, stewards, servers
- View data rule descriptions, policies, stewards, and rule expressions
- View data relationships and insights discovered by InfoSphere Discovery
- Graphical or report view of asset relationships
- Simple and advanced search
  - Search by various asset classes
  - Save and share your searches
  - Graphical query builder
  - Many out-of-the-box queries



## Smart Hover

- Innovative technology for viewing metadata details without losing context of the object currently being viewed

### Query Results: Job Design Information

The screenshot shows the 'Query Results: Job Design Information' interface. At the top, there are buttons for 'Select...', 'Edit', 'Add to Collection', and 'Detect Lineage Relationships'. Below this is a pagination bar showing '1-50 of 55' and 'Page 1 of 2'. The main area is divided into 'Reads from Data Items' and 'Writes to Data Items' tabs. A list of jobs is displayed, with the job 'JK\_STAGING\_COMBINE\_CST\_INPUT\_01\_Extract\_Data' circled in red. A 'Smart Hover' popup is open over this job, showing its metadata:

- Context:** IS-SERVER.IBM.COM » dstage1
- Type:** Parallel
- Image:** (icon)
- Job Operational Information:** (icon)

The popup also includes icons for editing, deleting, and other actions. The background list of jobs includes:

- IDR2FILE\_ACCOUNTS\_BY\_STATE\_FileReaderJob
- JK\_BANK2\_CHECKING\_LOYALTY\_PROMO
- JK\_STAGING\_COMBINE\_CST\_INPUT\_01\_Extract\_Data
- JK\_STAGING\_COMBINE\_CST\_INPUT\_02\_Investigate
- JK\_STAGING\_COMBINE\_CST\_INPUT\_03\_Standardize\_Branches

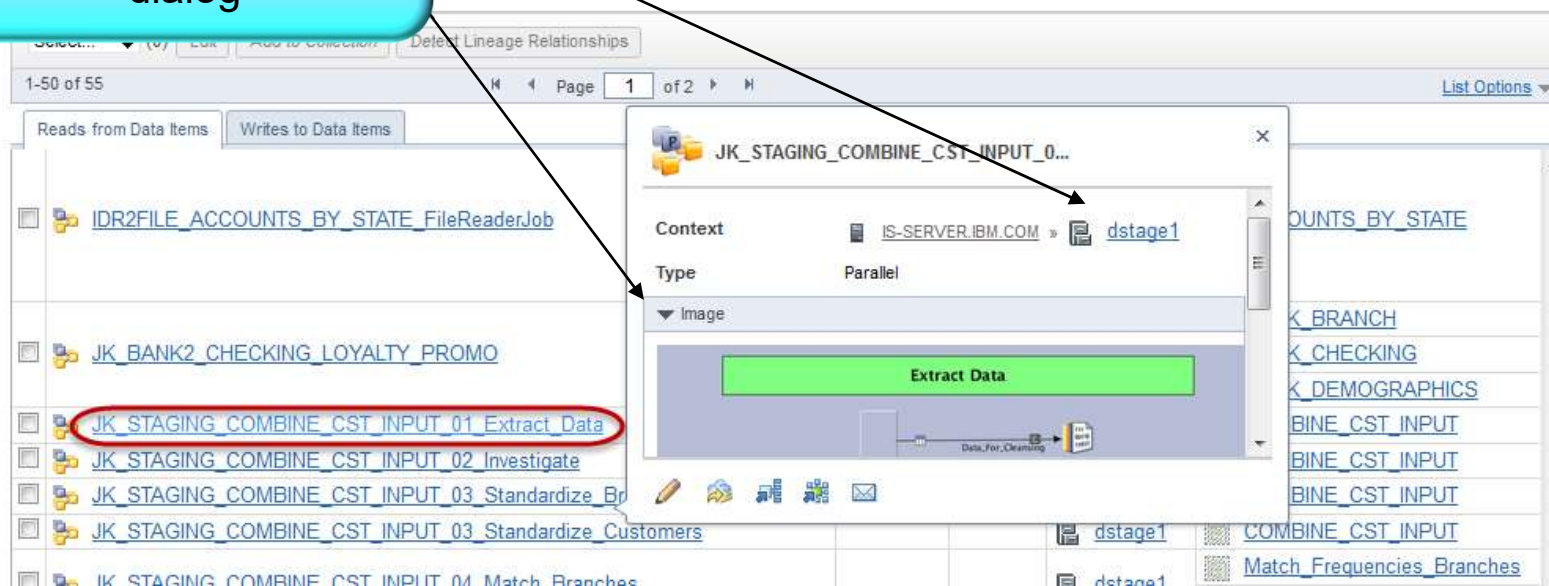
On the right side, a list of data items is visible, including:

- ACCOUNTS\_SCD
- ACCOUNTS\_BY\_STATE
- JK\_BRANCH
- JK\_CHECKING
- JK\_DEMOGRAPHICS
- COMBINE\_CST\_INPUT

## Smart Hover (continued)

- Open the twistie to see even further details (the types of detail available depends on the object that you are hovering on).

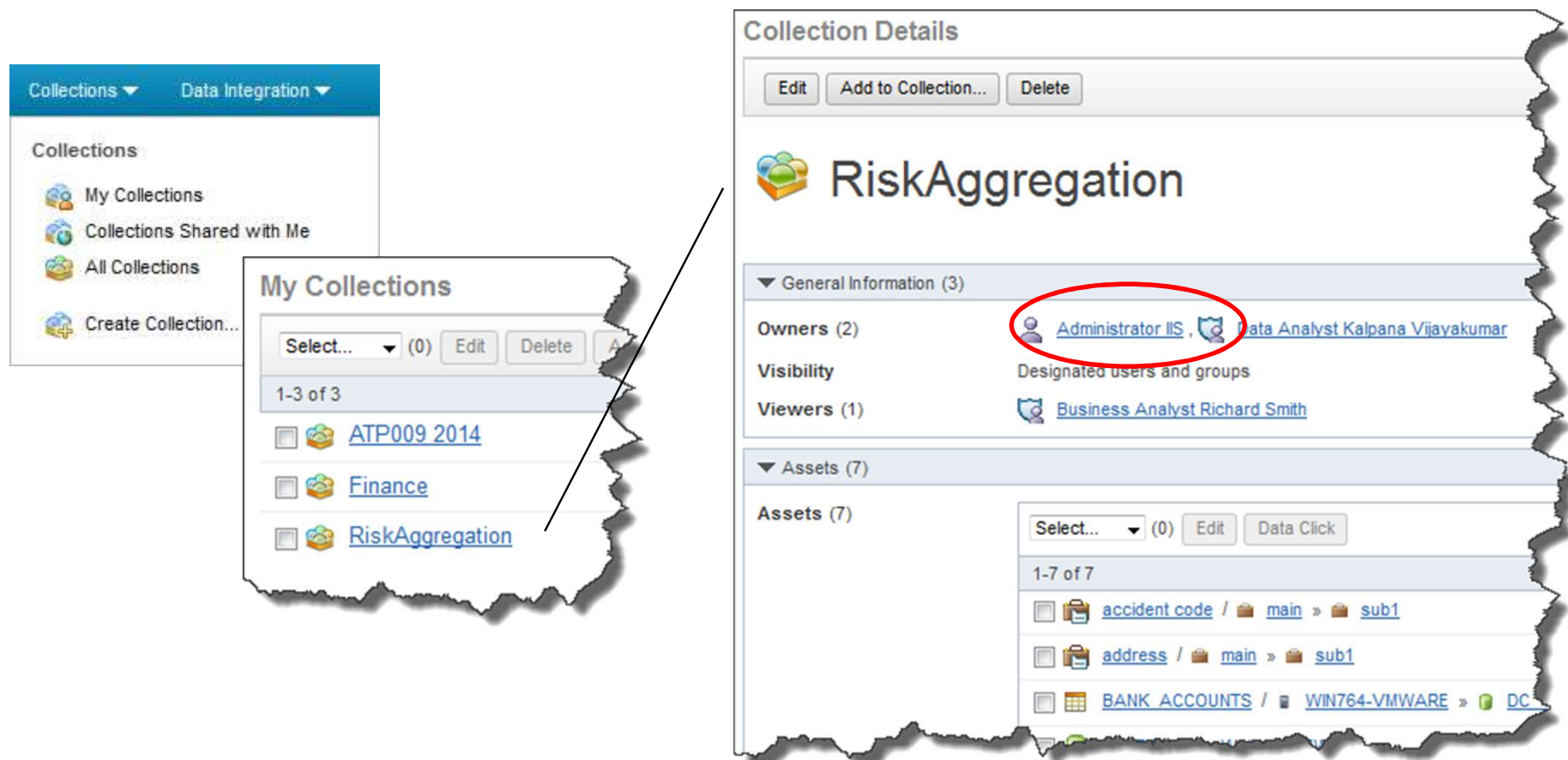
Expand or navigate to additional details from within the Smart Hover dialog





## Collections

- “Sets” of related assets
- Provides a direct “focus” for subject matter experts and teammates



The screenshot displays the IBM InfoSphere Information Server Collections interface. On the left, a sidebar shows navigation options: Collections, Data Integration, My Collections, Collections Shared with Me, All Collections, and Create Collection... The main area is titled 'My Collections' and lists three collections: ATP009 2014, Finance, and RiskAggregation. A red circle highlights the 'RiskAggregation' collection. On the right, the 'Collection Details' page for 'RiskAggregation' is shown. It includes buttons for Edit, Add to Collection..., and Delete. The collection is owned by 'Administrator IIS' and 'Data Analyst Kalpana Vijayakumar'. The visibility is set to 'Designated users and groups'. The viewers list includes 'Business Analyst Richard Smith'. The assets section shows 7 assets, including 'accident code', 'address', and 'BANK ACCOUNTS'.

## Collections (continued)

- Collections can be "shared" with other users as editors or viewers

Owner edits the Collection Properties...

General Information

**Owners (1)**

Select... (0) Remove from List Add User Find by U

1-2 of 2

- ☒ Business Analyst Richard Smith
- ☐ Administrator IIS

☐ All users and groups ☒ Designated users and groups

**Viewers (0)**

Select... (0) Remove from List Add User Find by U

1-1 of 1

- ☒ Data Analyst Kalpana Vijayakumar

...and then Business Analyst Richard or Data Analyst Kalpana logs in...

Collections Data Integration

**Collections**

- My Collections
- Collections Shared with Me
- All Collections
- Create Collection...

Collections Shared with Me

Select... (0) Edit Delete Add to Collection

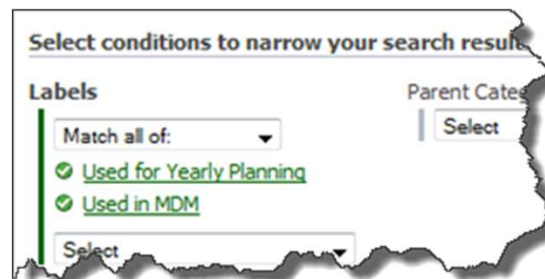
1-2 of 2

- ☐ ATP009 2014
- ☐ RiskAggregation



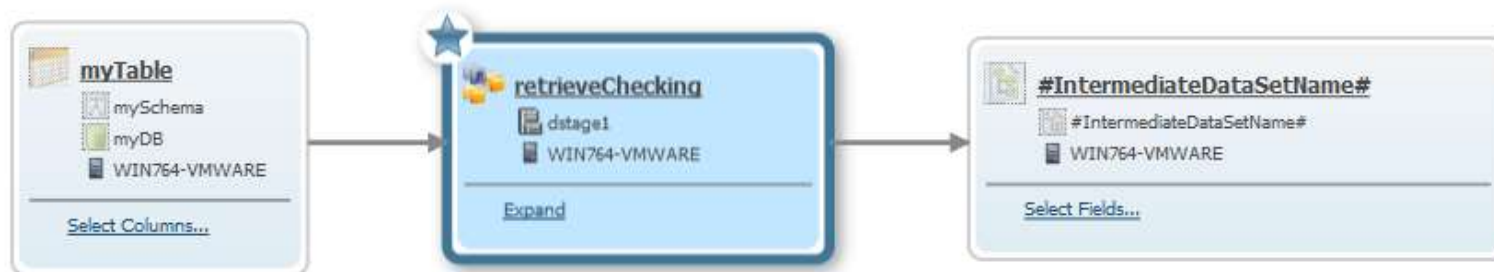
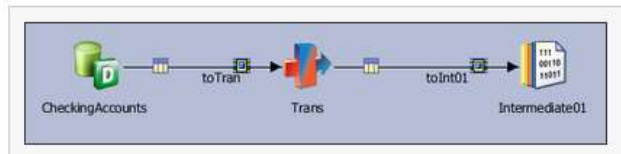
## Collections (continued)

- How are Collections different from Labels?
  - Collections have "owners"
  - Every user can create their own Collections!
    - Labels can only be created by Administrators and assigned by Authors
  - Collections can be shared with other users as co-owners or viewers
  - Collections themselves can be governed (and have custom attributes)
  - Collections can be "nested"
  
- How are Labels different from Collections?
  - Labels are like "sticky" notes. Attach to any object
  - Seen by everyone. Assigned only by Authors. Created only by Administrators
  
  - Labels have unique properties for complex multi-label filtered searches



## Enhanced Lineage

- Write a Job, get Lineage...



## Enhanced Lineage (continued)

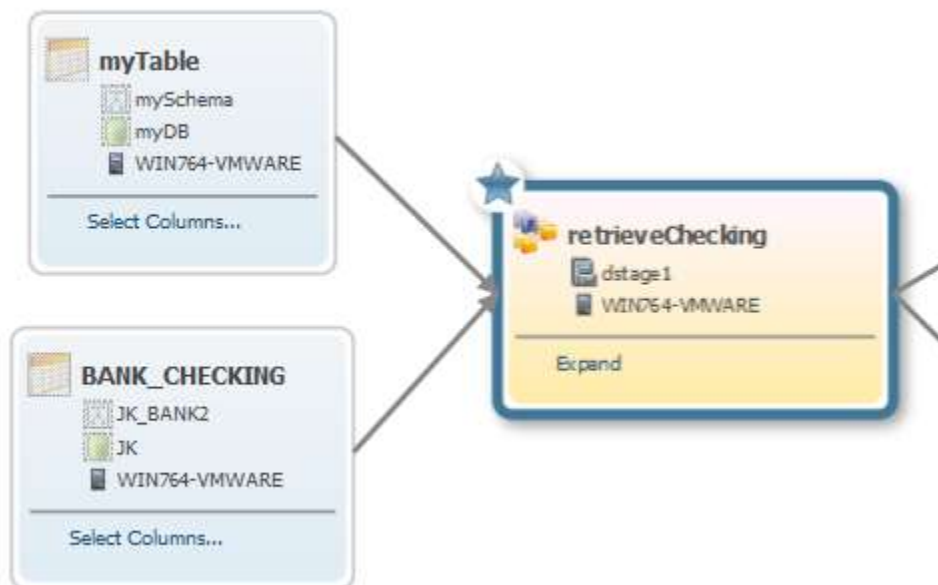
- Introducing Virtual Assets
  - Job detail immediately illustrates “reads from” and/or “writes to” Virtual Assets
  - Examples of Virtual Assets (sources):

### Design based Virtual Asset using Job Parameter Defaults...

	Parameter name	Type	Default Value
1	database	String	myDB
2	schema	String	mySchema
3	tablename	String	myTable

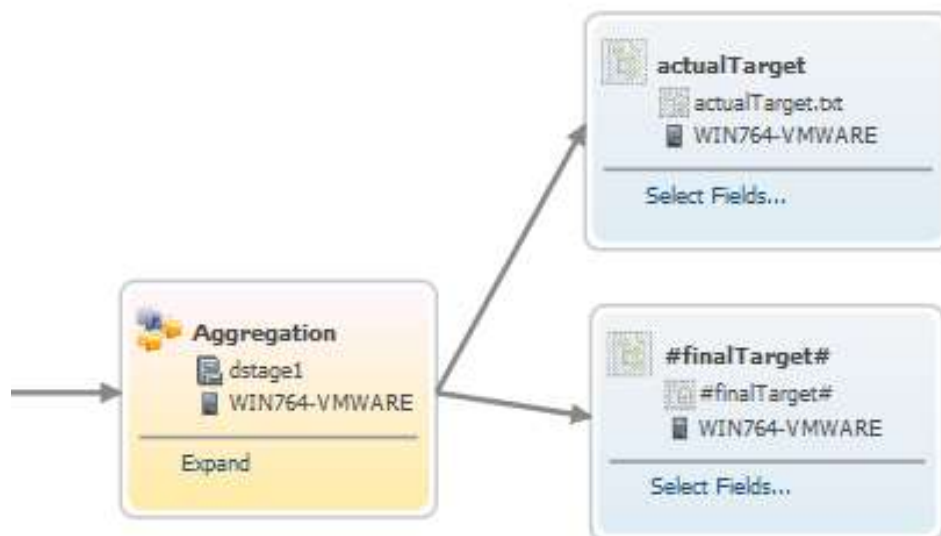
### Operational Metadata based Virtual Asset using actual run-time values

```
Starting Job retrieveChecking.
database = JK
schema = JK_BANK2
tablename = BANK_CHECKING
IntermediateDataSetName = dataset01.ds
DSJobController = runLineageLab
```



## Enhanced Lineage (continued)

- Introducing Virtual Assets (continued)
  - Examples of Virtual Assets (targets)



Operational Metadata  
based Virtual Asset using  
actual run-time values

```
Starting Job Aggregation.  
finalTarget = /tmp/actualTarget.txt  
DSJobController = runLineageLab
```

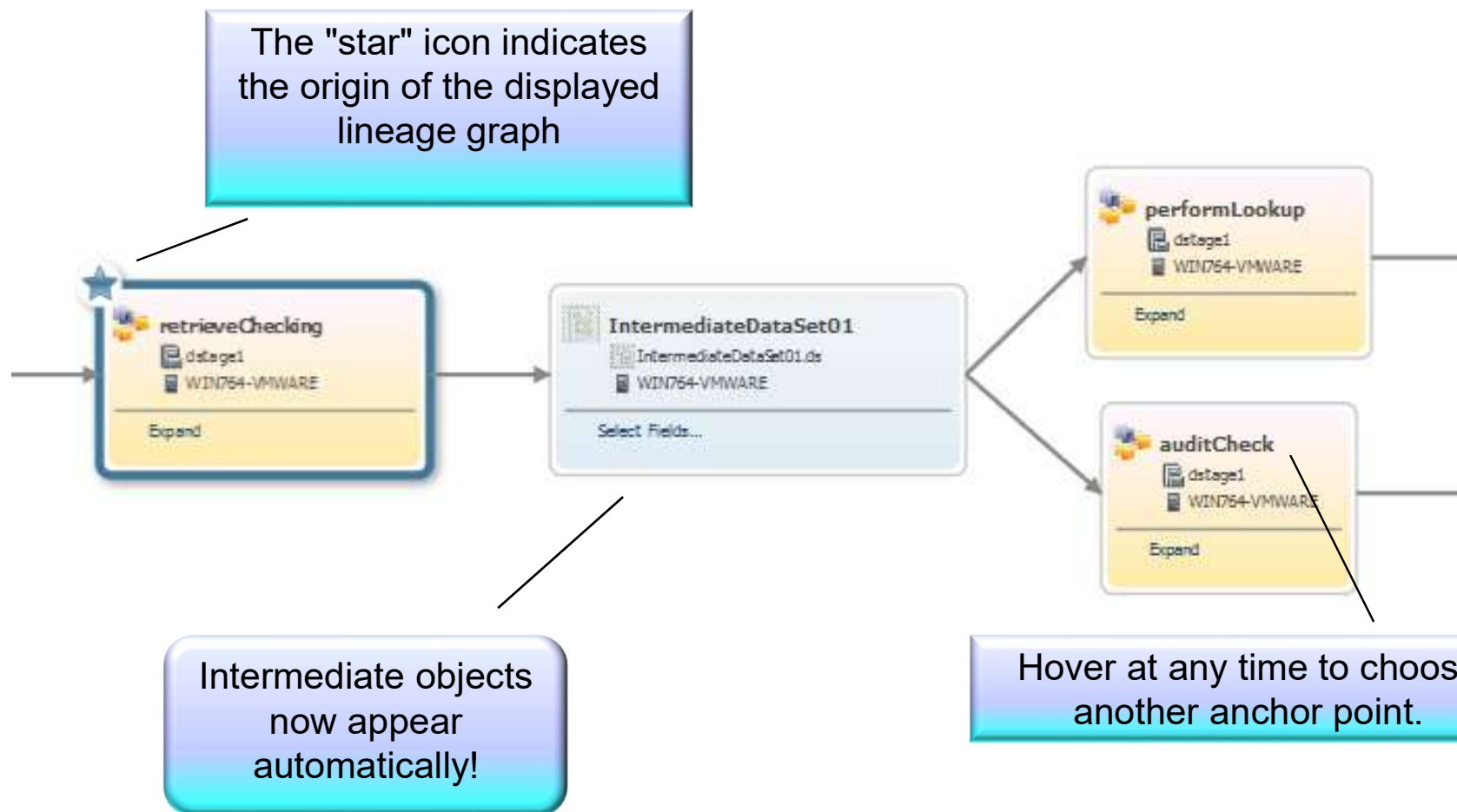
Design based Virtual Asset  
using Job Parameter that  
has no default value!

- Virtual Assets are only generated for Stage Types that otherwise support normal "stitching" technology

	Parameter name	Prompt	Type	Default Value
1	finalTarget	finalTarget	String	

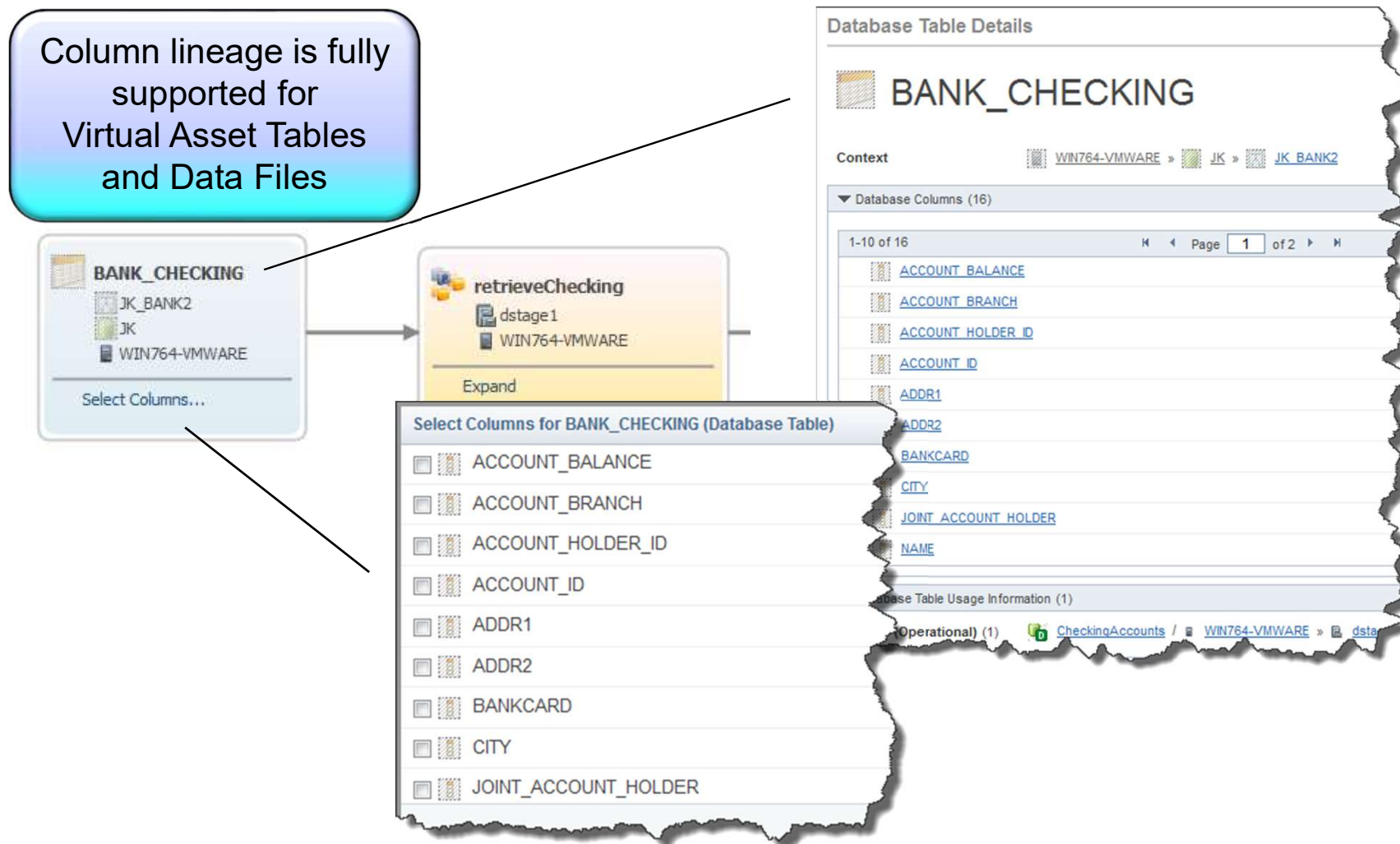
## Enhanced Lineage (continued)

- Virtual Assets illustrate intermediate objects between correctly linked Jobs



## Enhanced Lineage (continued)

- Virtual Assets can be viewed and have columns





## Enhanced Lineage (continued)

- Virtual Assets are referenced from Stage detail pages and also have their own detail page

### Sequential File Stage Details

Edit
Add to Collection...

**finalTarget**

Context
WIN764-VMWARE » DataClick »

General Information

Stage Usage Information (5)

Input Links (1)
DSLink6 / WIN764-VMWARE » DataClick »

Previous Stages (1)
Aggregation / WIN764-VMWARE »

Writes to (Design) (1)
#finalTarget#

Writes to (Operational) (2)
actualTarget
actualTarget2

### Data File Structure Details

**#finalTarget#**

Context
WIN764-VMWARE » #finalTarget#

Data File Fields (2)
ACCOUNT BALANCE
STATE

General Information

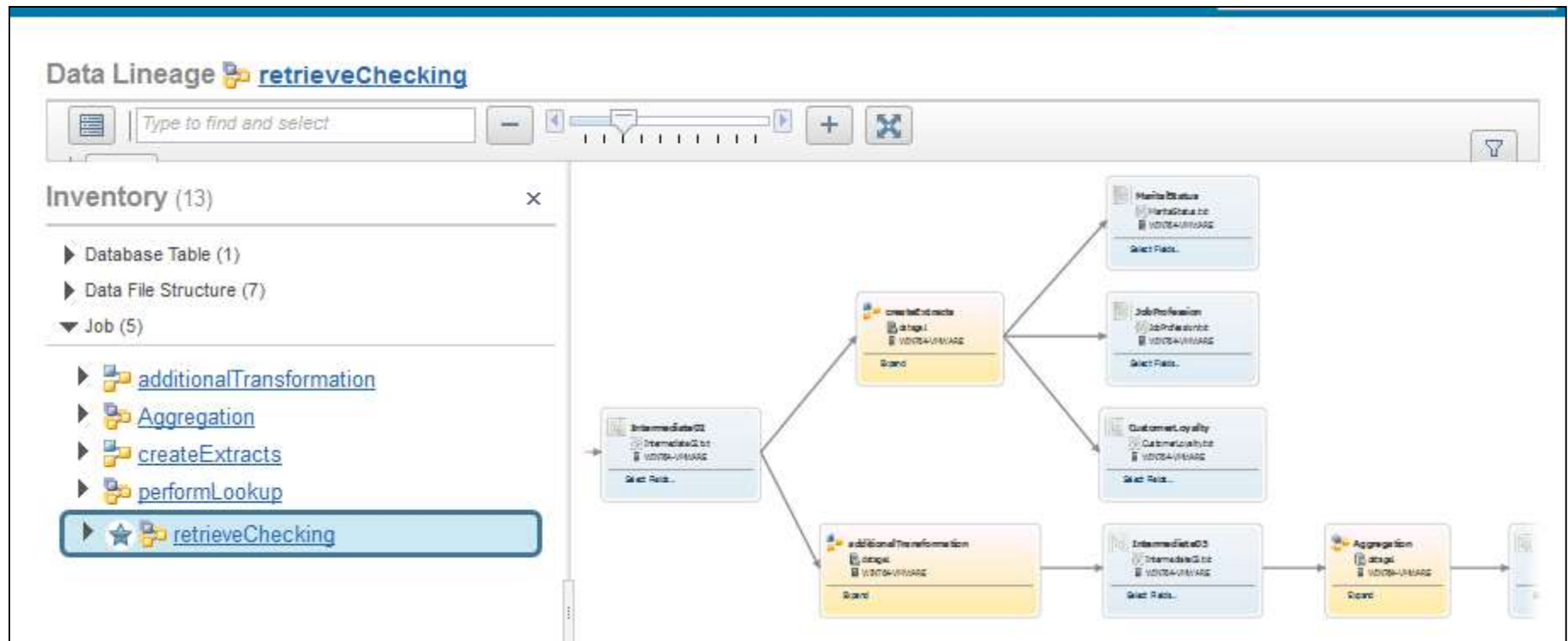
Include for Business Lineage
True

Data File Structure Usage Information (1)

Written by (Design) (1)
finalTarget / WIN764-VMWARE » DataClick » Aggregation

## Enhanced Lineage (continued)

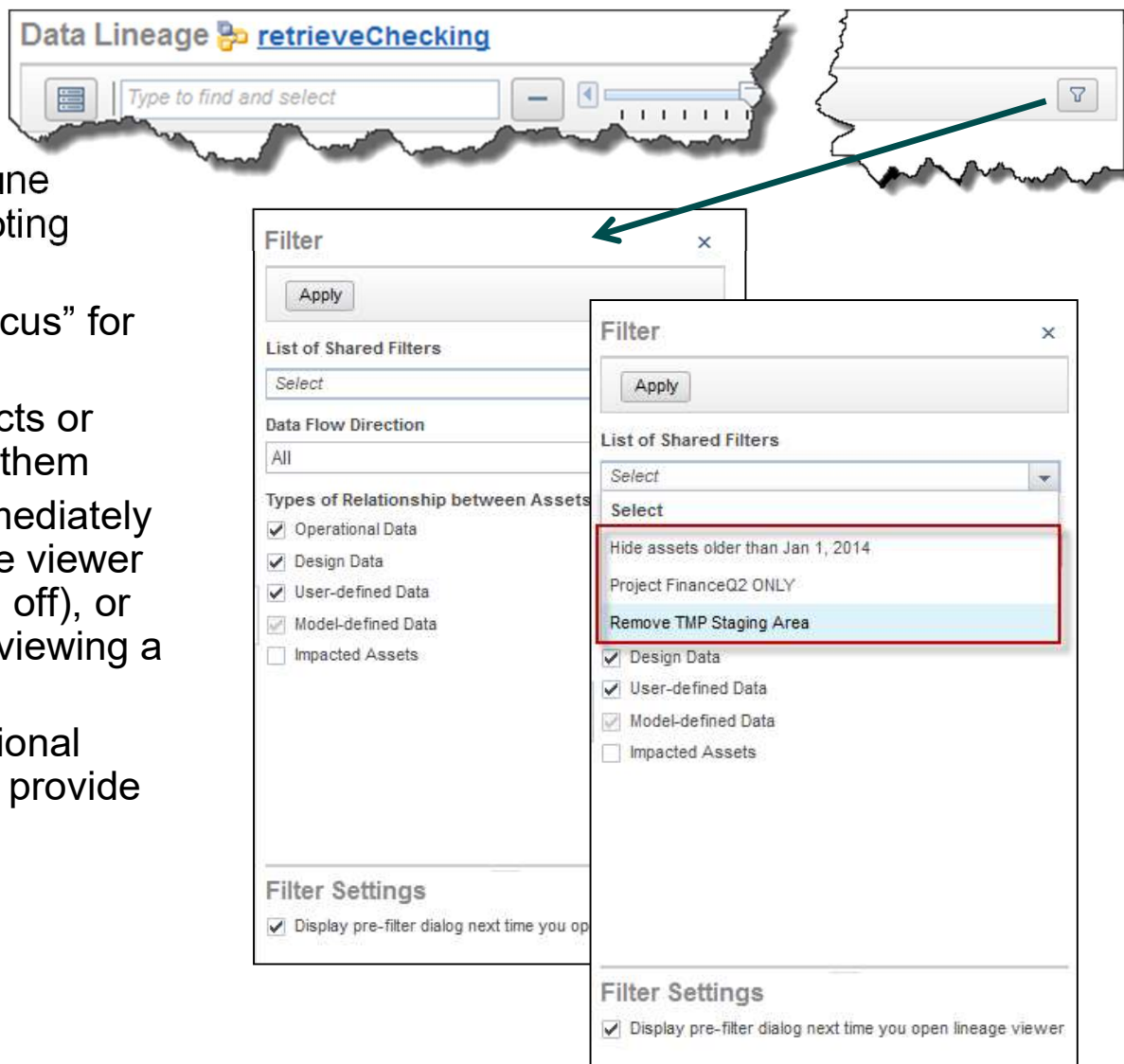
- Display for lineage reports



## Enhanced Lineage (continued)

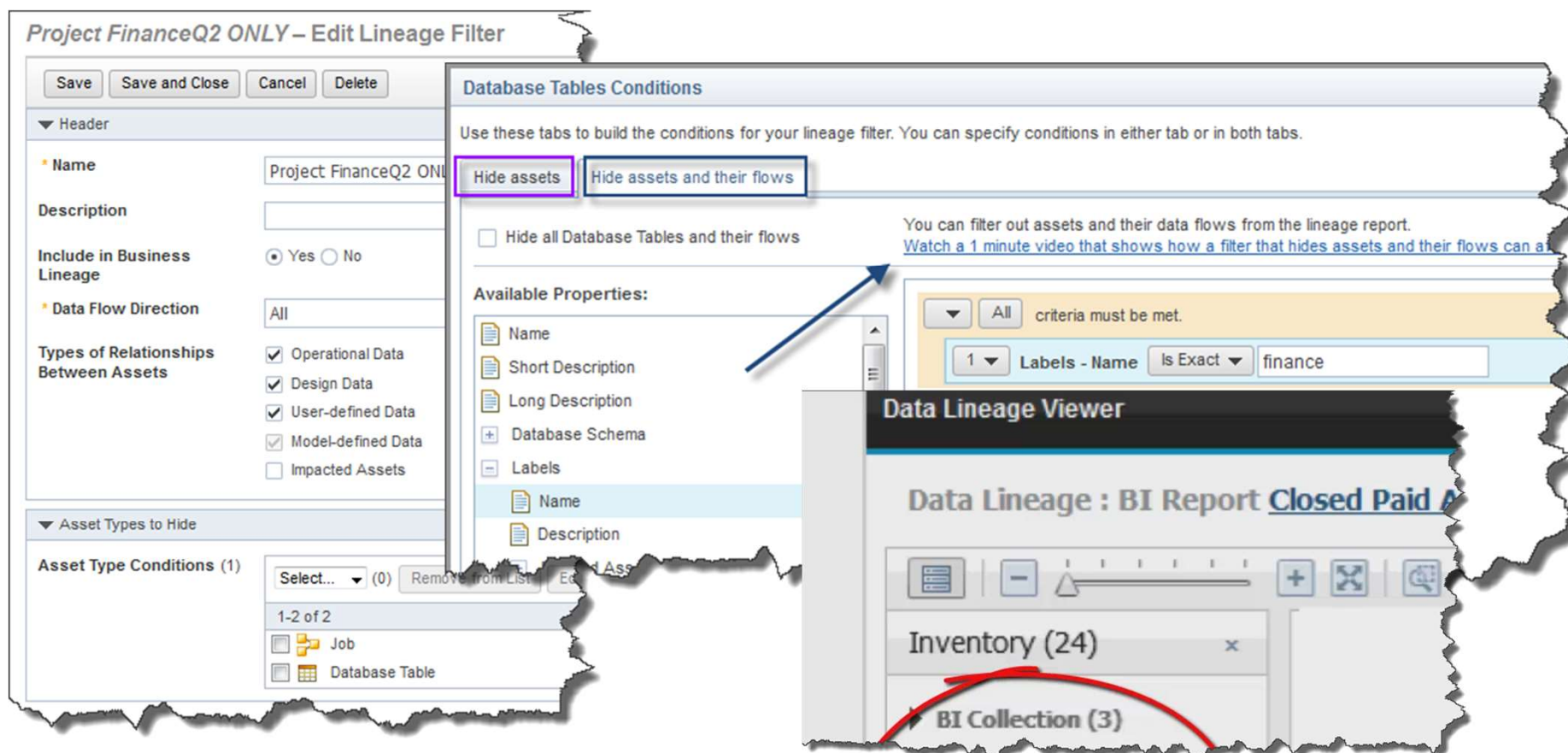
### ■ Lineage Filtering

- Quickly and easily fine tune lineage graphs - like pivoting a spreadsheet
- Enable more detailed “focus” for lineage graph review
- Hide only individual objects or everything connected to them
- Filter dialog appears immediately upon entering the lineage viewer (optionally can be turned off), or can be introduced while viewing a graph
- 11.3 re-introduces directional filtering and continues to provide relationship selection



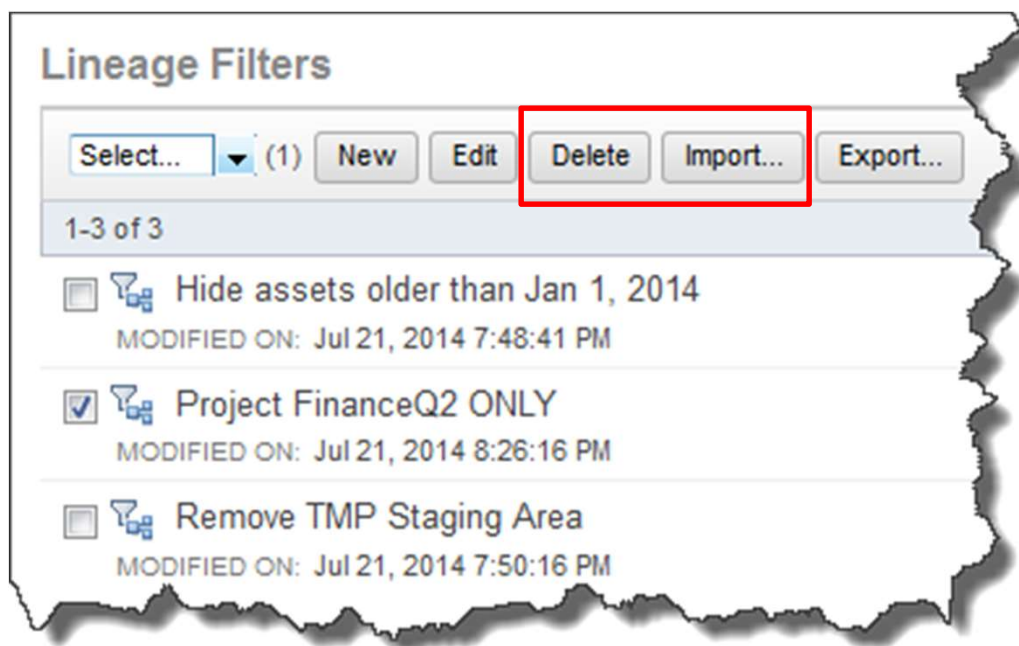
## Enhanced Lineage (continued)

- Create Lineage Filters
  - Filters are created by the Administrator and available for all users
  - Utilize the full power of the Query Tool for complex filtering



## Enhanced Lineage (continued)

- Organize your Filters
  - Import/Export similar to Information Governance Catalog Queries
  - You can choose names for your filters that clearly identify the objects and/or lineage paths being hidden from view



## Enhanced Lineage (continued)

- Using filters

