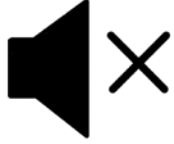October 1, 2020

Advanced Scanners Session for EDC Customers

Gaurav Pathak, Vice President, Product Management
Louis-Noel, Trapadoux, Principal Product Manager

Informatica™

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.
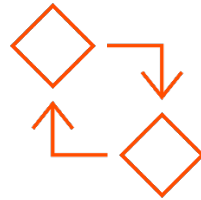
Informatica

Feature Rich Success Portal

Bootstrap trial and POC Customers

Enriched Customer Onboarding experience

Product Learning Paths and Weekly Expert Sessions

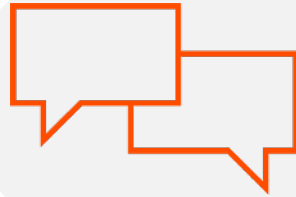Informatica Concierge with Chatbot integrations

Tailored training and content recommendations

Informatica®

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
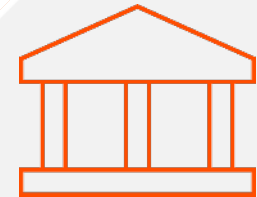
Informatica

Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica 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.

Informatica

Speakers

Gaurav Pathak

Vice President
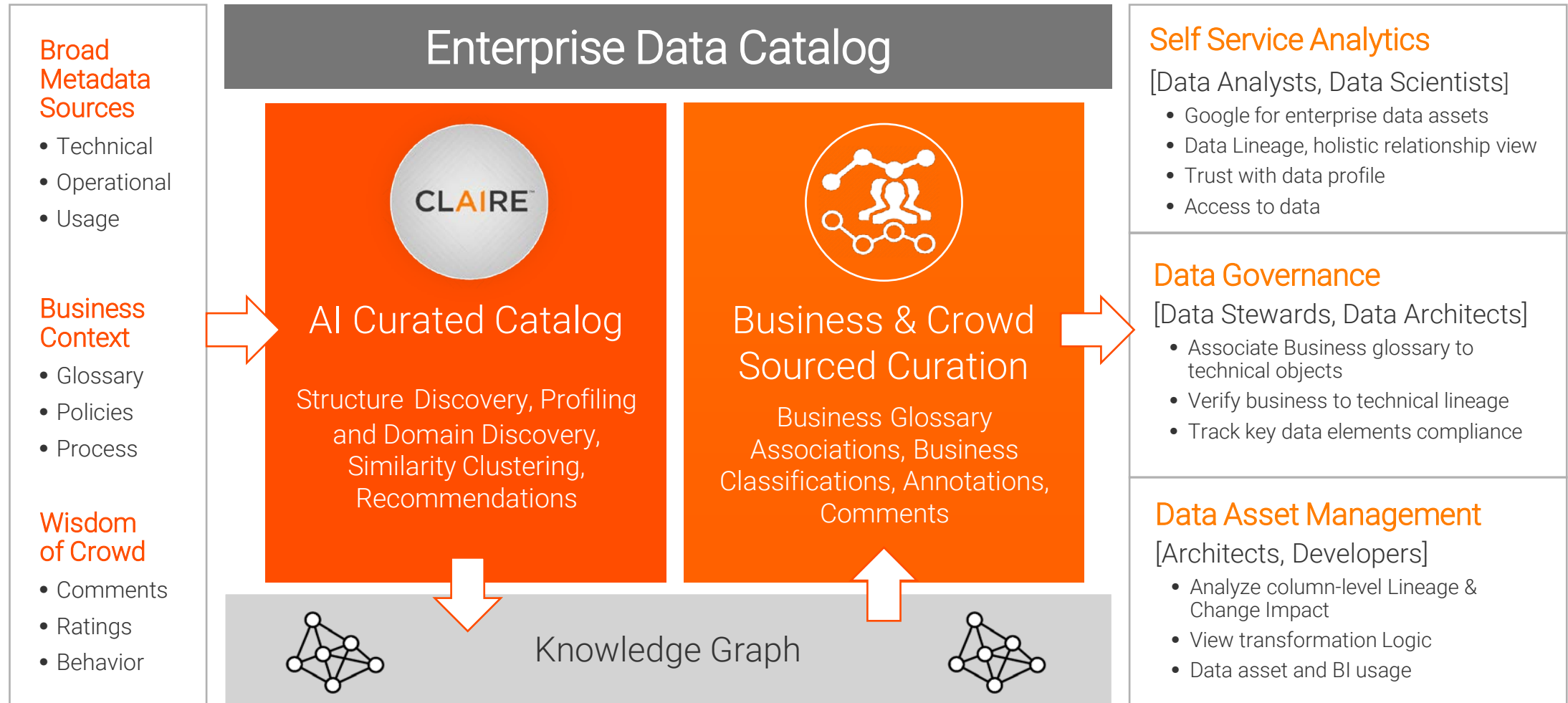Product Management
Metadata and CLAIRE
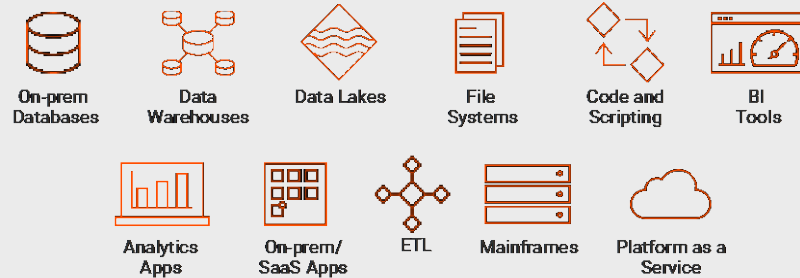
**Louis-Noel Trapadoux**

Principal Product Manager
EDC

Enterprise Data Catalog Powered by CLAIRE™

Broad Metadata Sources
- Technical
- Operational
- Usage

Business Context
- Glossary
- Policies
- Process

Wisdom of Crowd
- Comments
- Ratings
- Behavior

Enterprise Data Catalog

AI Curated Catalog

Structure Discovery, Profiling and Domain Discovery, Similarity Clustering, Recommendations

Business & Crowd Sourced Curation

Business Glossary Associations, Business Classifications, Annotations, Comments

Knowledge Graph

Self Service Analytics
[Data Analysts, Data Scientists]
- Google for enterprise data assets
- Data Lineage, holistic relationship view
- Trust with data profile
- Access to data

Data Governance
[Data Stewards, Data Architects]
- Associate Business glossary to technical objects
- Verify business to technical lineage
- Track key data elements compliance

Data Asset Management
[Architects, Developers]
- Analyze column-level Lineage & Change Impact
- View transformation Logic
- Data asset and BI usage

Informatica™

Technical Challenges

It's difficult and nearly **impossible to catalog all of our enterprise data** including legacy on-premises systems and newer Cloud enterprise and analytic applications
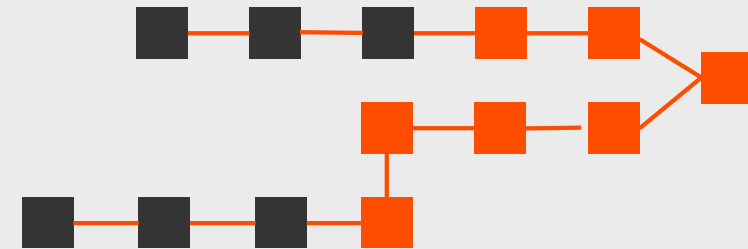
The ability to understand data through **data lineage is typically incomplete** (there exist black boxes) especially for scripts, code, legacy mainframe systems, multi-vendor ETL tools, and BI applications
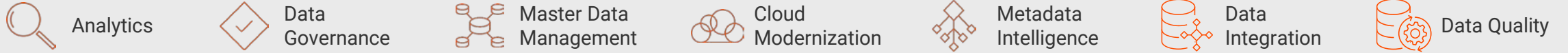
Informatica

Informatica + Compact Solutions

Industry's first and only catalog of catalogs across all enterprise data with the broadest metadata connectivity

- **Extends the industry's most comprehensive and detailed data lineage** (i.e. no "black boxes") to understand the provenance of all enterprise data and impact of changes as companies digitally transform and modernize.

- **Expands the industry's broadest metadata connectivity** to catalog all types of data in support of analytics, data governance and privacy, customer experience, and data warehouse modernization initiatives.

- **Establishes the industry's only single vendor metadata management solution** to simplify procurement, deployment, maintenance and support.

Informatica

The Catalog of Catalogs

| Analytics | Data Governance | Master Data Management | Cloud Modernization | Metadata Intelligence | Data Integration | Data Quality |

Open APIs, Full Access

Enterprise Data Catalog

- Discovery
- Profiling
- Lineage
- Impact Analysis

- Semantic Search
- Domain Discovery
- Similarity Clustering
- Business Term Association

- Relationships
- Business Context
- Glossary Integration
- Custom Annotations

- Reviews/Ratings
- Questions/Answers
- Data Certifications
- Change Notifications

Knowledge Graph + AI/ML

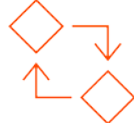## CLAIRE™

Breadth of Active Metadata

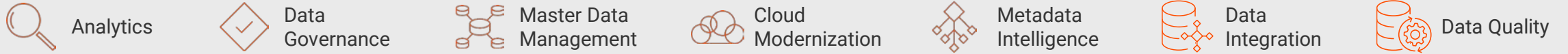| On-prem Databases | Data Warehouses | Data Lakes | File Systems | Code and Scripting | Statistical & BI Tools | Analytics Apps | On-prem/ SaaS Apps | ETL | Mainframes | Platform as a Service |

The Catalog of Catalogs: One Vendor, One Solution

Analytics	Data Governance	Master Data Management	Cloud Modernization	Metadata Intelligence	Data Integration	Data Quality

Open APIs, Full Access

- Discovery
- Profiling
- **Lineage**
- **Impact Analysis**

- Semantic Search
- Domain Discovery
- Similarity Clustering
- Business Term Association

Enterprise Data Catalog
+ Compact Solutions

- Relationships
- Business Context
- Glossary Integration
- Custom Annotations

- Reviews/Ratings
- Questions/Answers
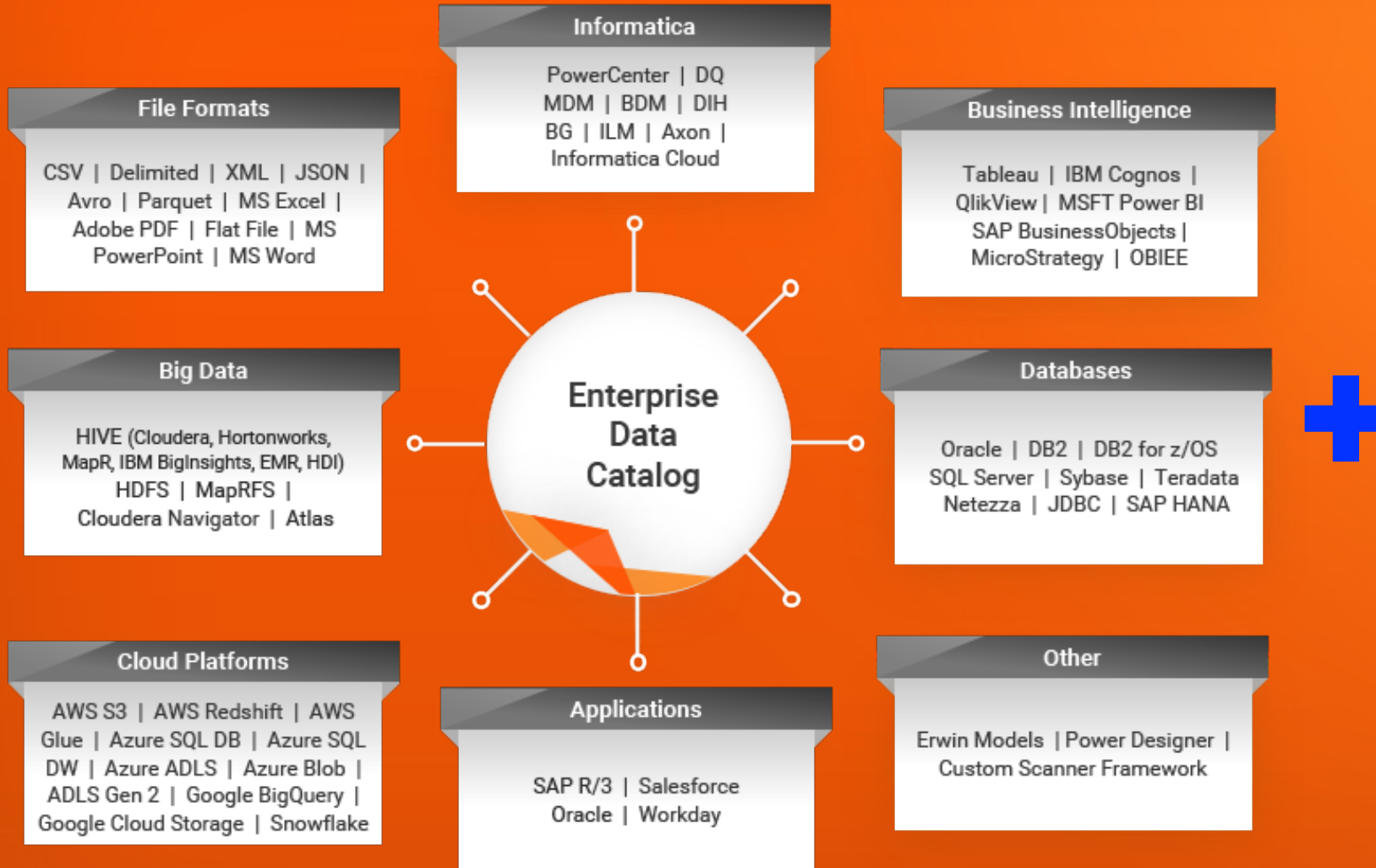- Data Certifications
- Change Notifications

Knowledge Graph + AI/ML

CLAIRE™

Breadth of Active Metadata

On-prem Databases	Data Warehouses	Data Lakes	File Systems	Code and Scripting	Statistical & BI Tools	Analytics Apps	On-prem/ SaaS Apps	ETL	Mainframes	Platform as a Service

Broadest and Most Complete Metadata Connectivity

File Formats

CSV | Delimited | XML | JSON | Avro | Parquet | MS Excel | Adobe PDF | Flat File | MS PowerPoint | MS Word

Informatica

PowerCenter | DQ MDM | BDM | DIH BG | ILM | Axon | Informatica Cloud

Business Intelligence

Tableau | IBM Cognos | QlikView | MSFT Power BI SAP BusinessObjects | MicroStrategy | OBIEE

Big Data

HIVE (Cloudera, Hortonworks, MapR, IBM BigInsights, EMR, HDI) HDFS | MapRFS | Cloudera Navigator | Atlas

Enterprise Data Catalog

Databases

Oracle | DB2 | DB2 for z/OS SQL Server | Sybase | Teradata Netezza | JDBC | SAP HANA

Cloud Platforms

AWS S3 | AWS Redshift | AWS Glue | Azure SQL DB | Azure SQL DW | Azure ADLS | Azure Blob | ADLS Gen 2 | Google BigQuery | Google Cloud Storage | Snowflake

Applications

SAP R/3 | Salesforce Oracle | Workday

Other

Erwin Models | Power Designer | Custom Scanner Framework

+

EDC Advanced Scanners

Code and Scripting

- Oracle
- SQL Server
- Teradata
- Netezza
- IBM DB2
- Sybase ASE

ETL Tools

- IBM Datastage
- Microsoft SSIS

Mainframes

- COBOL
- JCL

Statistical and BI Tools

- SAS
- Microsoft SSAS
- Microsoft SSRS
- SAP BW
- SAP BW4HANA

Usecases

Data Lineage: A Business Imperative

Data Lineage traces data from source to destination, covering the entire lifecycle of data. It includes information about changes to data during its journey.

Requirements:

○ Regulatory Compliance

○ Data Quality

○ Data Governance

○ Data Analytics

○ Data Privacy and Security

Data Lineage: The Foundational Use Case

Increasingly "IT" use cases are coming to the forefront…

- **Dev Operations:** Change Management & Impact Analysis - what-if analyses for changes

- **Operational Efficiency:** Eliminate proliferation, duplication, data silos, reduce costs

- **DW/Apps Modernization:** Complete understanding of the data landscape to enable app modernization & cloud migration

…and AI use cases

- **Explainable AI & AI Governance**: Track and assess data used to train models, govern AI projects. Support Explainable AI. Ensure training data variety.
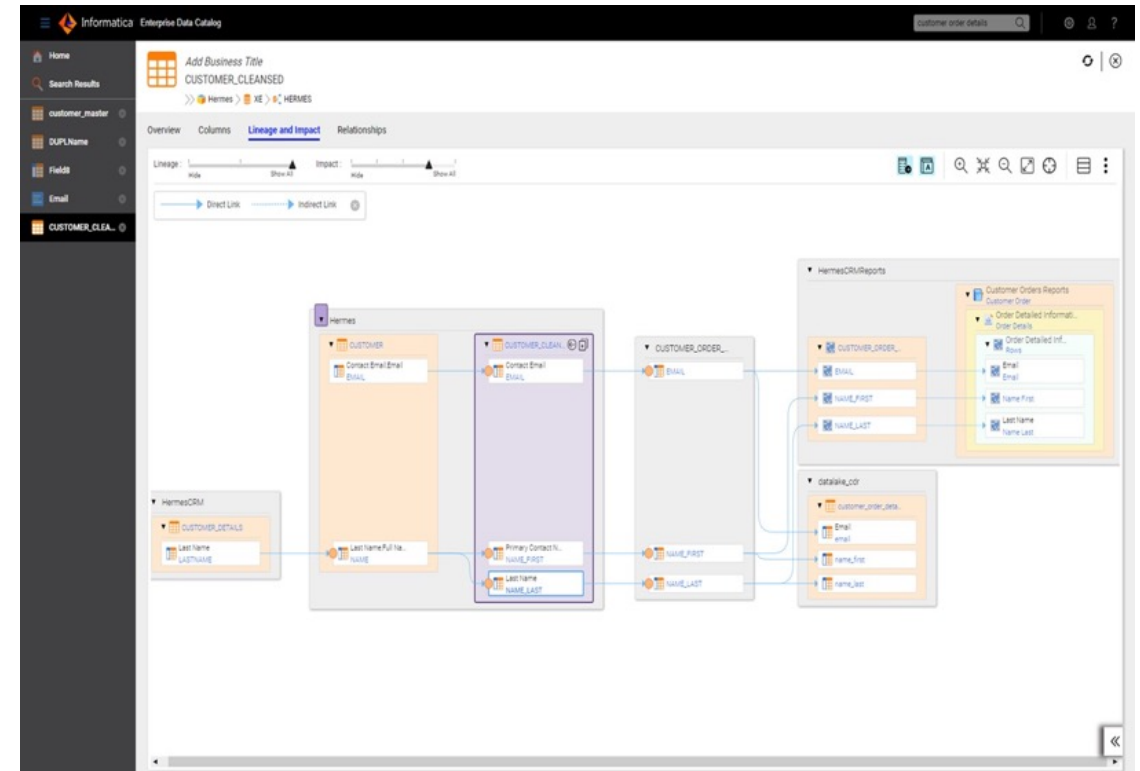
Enterprise Data Catalog Advanced Scanners

Extract metadata and data lineage with in-depth details

- Parse code from various stored procedures and multi-vendor ETL tools

- Obtain automatic lineage and data relationships at scale

- Extract deep metadata from both static and dynamic code

- Obtain complete visibility into the procedure calls with parameter tracking, dynamic SQL generation from values based on parameters, database queries and more

Informatica

Advanced scanner availability

Category	Source	Standard scanner	Advanced scanner
Database	Oracle DB	🟩	🟩
	MS SQL Server	🟩	🟩
	Teradata	🟩	🟩
	Netezza	🟩	🟩
	Sybase ASE	🟩	🟩
	IBM DB2 (LUW)	🟩	🟩
ETL	SSIS	⬜	🟩
	DataStage	⬜	🟩
Reporting/Statistics	SSAS	⬜	🟩
	SSRS	⬜	🟩
	SAS	⬜	🟩
	SAP BW	⬜	🟩
	SAP BW4HANA	⬜	🟩
Mainframe	Cobol	⬜	🟩
	JCL	⬜	🟩

- Standard and Advanced scanner only available for select databases
  - Standard scanners fetch simple object metadata and are mandatory.
  - Advanced Scanners are required for extracting lineage metadata.
- Advanced scanners cover:
  - Third Party ETL scanners (other than INFA)
  - Selected list of reporting scanners
  - Mainframe scanners
- We will be releasing more advanced scanners overtime

Informatica

Standard vs Advanced DB scanners

Standard DB scanners	Advanced DB scanners

Object metadata
- Tables
- Views
- Materialized views
- Synonym
- Trigger definitions
- Procedure definitions (no lineage)
- Function definitions (no lineage)
Lineage
- Views and Synonyms to Tables
• **Limited** Lineage from database scripts (available only for Oracle, Teradata, Hive, DB2)
    • Summary level only
    • Table level only
Profiling
- Available
	Code Lineage
	- Generated from Procedures/Functions SQL parsing
	- Detailed lineage for SQL statement at table/view/synonym and field level
	- Advanced visualization available for complex SQL
	- Support parsing SQL dynamically generated
	
	• Lineage from database scripts (available for Oracle, SQL Server, Teradata, Netezza, DB2)
	    • Detailed lineage for SQL statements at table/view/synonym and field level
	    • Support dynamic SQL
	    • Overcome "select *" limitations
	    • Support Loader/Export scripts

💡 **Advanced Scanners are required for any customer interested in data lineage**. Standard Scanners provide metadata extraction for simple objects – tables, columns, views. But Data Lineage requires metadata from parsing SQL Code, Stored Procedures, SQL Scripts that move data – this is where Advanced Scanners come in.

Informatica
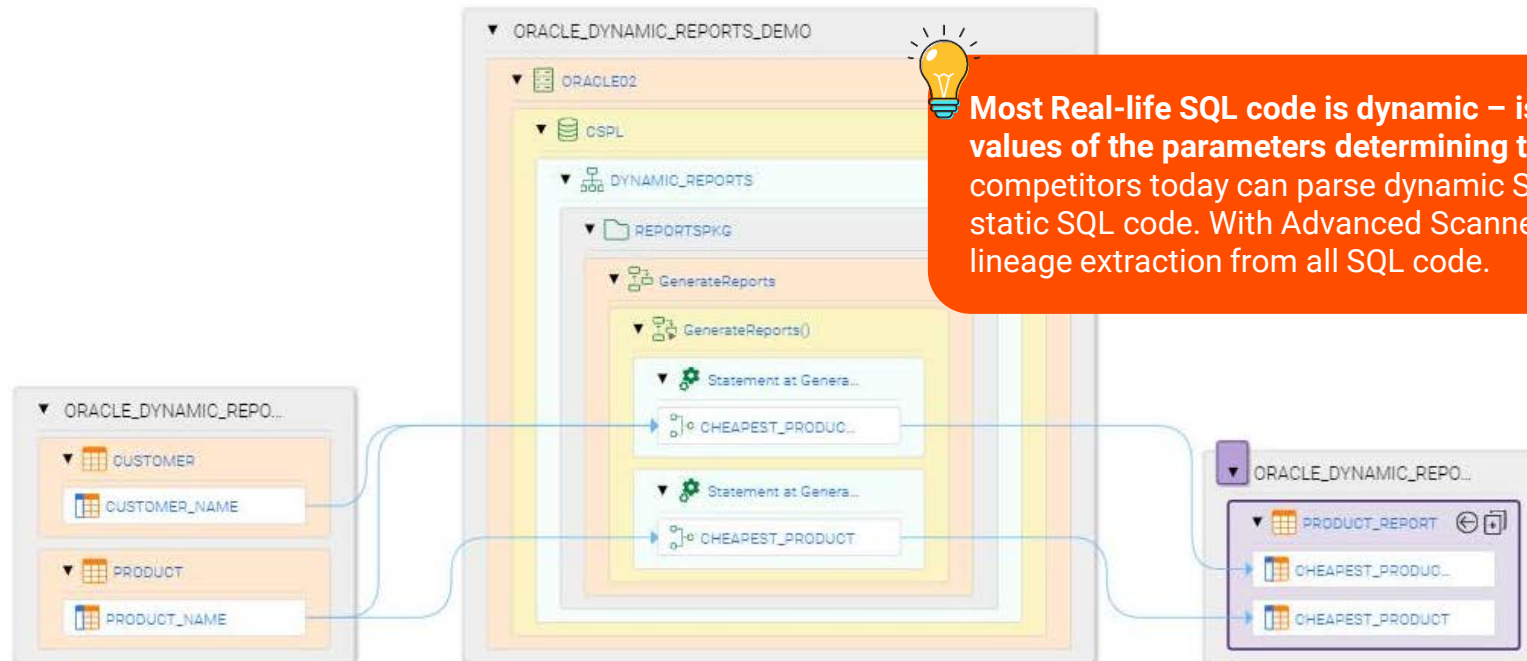
Dynamic SQL support with Advanced Scanners

```
create table ReportDefinition ( ReportTable varchar(100), ReportColum
insert into ReportDefinition values ('CustomerReport', 'TotalNumberOf
    'select sum(NumberOfDepartments) from Customer');
insert into ReportDefinition values ('CustomerReport', 'MaxNumberOfDe
    'select max(NumberOfDepartments) from Customer');
insert into ReportDefinition values ('ProductReport', 'CheapestProduc
    'select ProductName from (select ProductName from Product order b
insert into ReportDefinition values ('ProductReport', 'CheapestProduc
    'SELECT Code FROM (SELECT REPORTSPKG.GetProductCode(ProductName,CustomerName) as Code
    FROM Product P JOIN Customer C ON P.CustID=C.ID ORDER BY P.Price DESC) WHERE rownum <= 1

// Expected Results
// Table "CustomerReport" with Columns: "TotalNumberOfDepartments", "MaxNumberOfDepartments"
// Table "ProductReport" with Columns: "CheapestProduct", "CheapestProductCode"
```

```
create or replace package ReportsPkg is
    procedure GenerateReports;
    procedure GenerateReport(vReportTable varchar);
    procedure CreateReportTable(vReportTable varchar);
    procedure GenerateReportColumn(vReportColumn ReportDefinition%ROWTYPE);
    function GetProductCode(vProductName varchar, vCustomerName varchar) return varchar;
end;
```

```
procedure GenerateReports is
begin
    for row in (select distinct ReportTable from ReportDefinition) loop
        CreateReportTable(row.ReportTable);
        GenerateReport(row.ReportTable);
    end loop;
end;
```

Most Real-life SQL code is dynamic – is heavily parameterized with values of the parameters determining the code path. None of the INFA competitors today can parse dynamic SQL, most cannot even parse static SQL code. With Advanced Scanners we support accurate data lineage extraction from all SQL code.
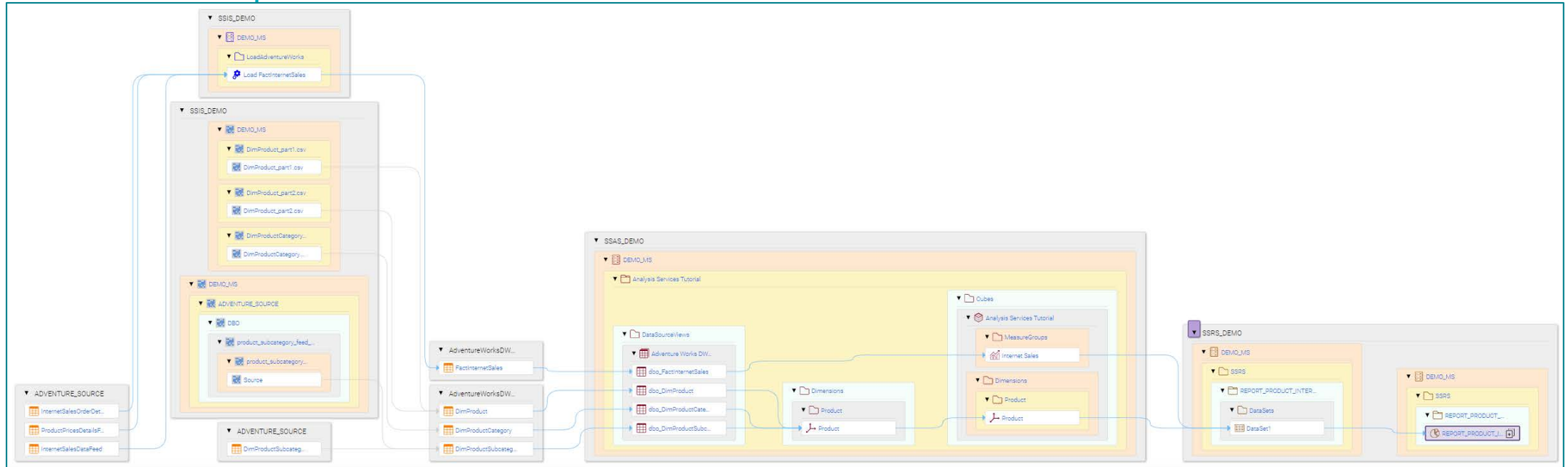
© Informatica. Prop

No Black Boxes – COBOL, JCL and SAS

Manage and govern all your enterprise data, improve change management and minimize risk of changes with end-to-end and detailed lineage and impact analysis (no black boxes).
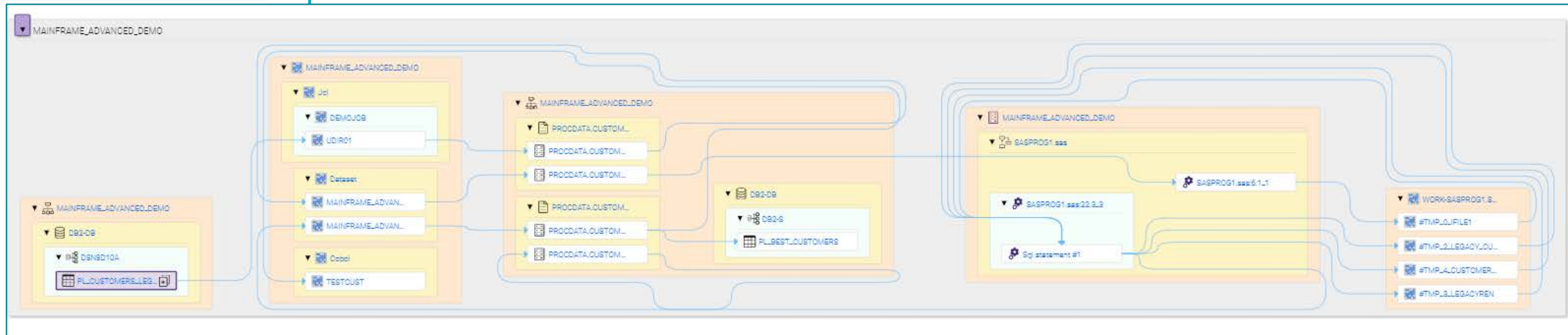
No Black Boxes – Microsoft SSIS, SSAS and SSRS

Manage and govern all your enterprise data, improve change management and minimize risk of changes with end-to-end and detailed lineage and impact analysis (no black boxes).

© Informatica. Proprietary and Confidential.

Informatica™

No Black Boxes – IBM DataStage

Manage and govern all your enterprise data, improve change management and minimize risk of changes with end-to-end and detailed lineage and impact analysis (no black boxes).
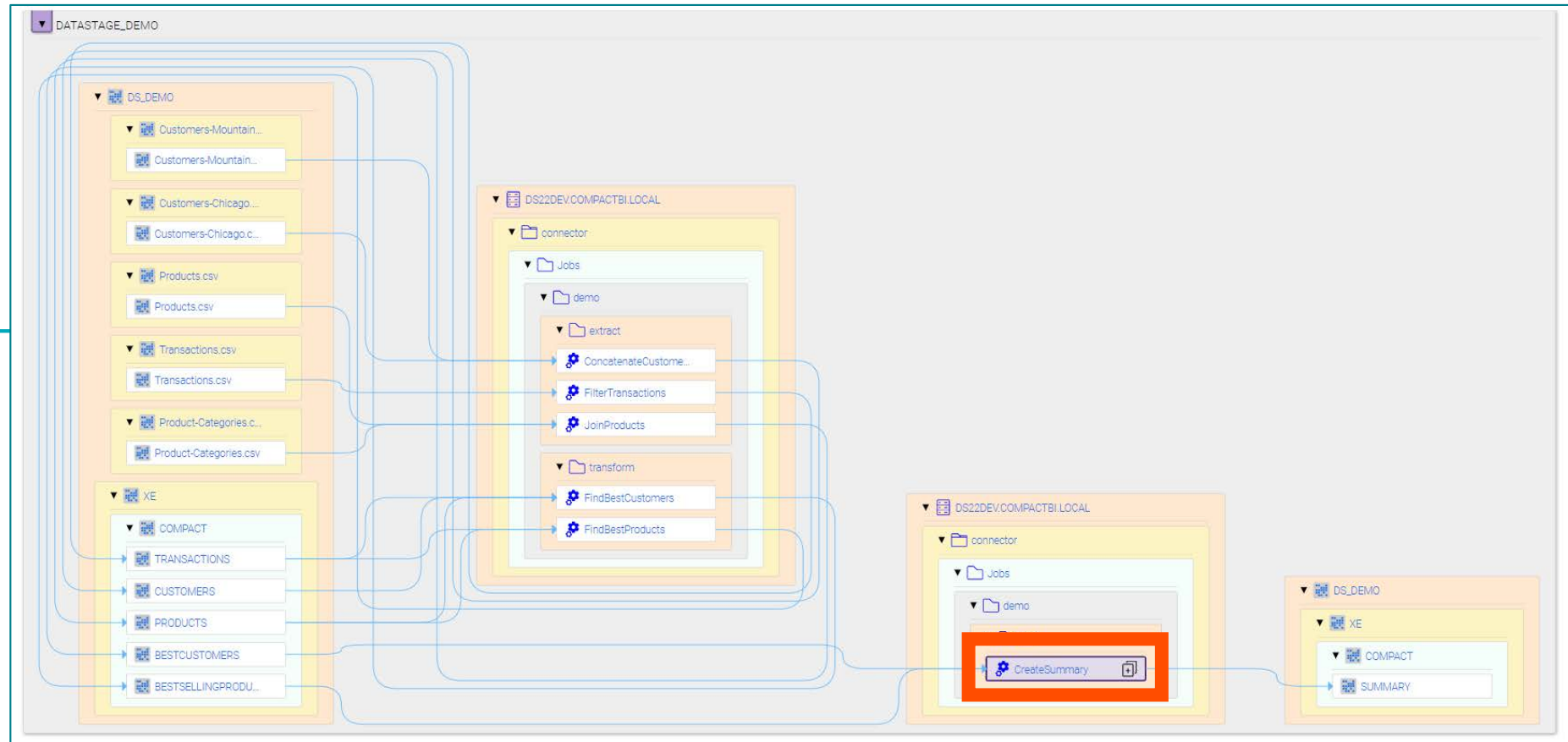
Deep Lineage Visualizations with Advanced Scanners

- Advanced lineage visualization with decomposition of SQL statements into individual transformations

- allow users to analyze Stored Procedures

- Availability of mapping report including list of transformation applied

Data Lineage is essential for regulatory compliance, root cause analysis, impact analysis, data migrations to cloud and establishing trust in data. The first three use cases require decomposition of code into understandable chunks. Advanced Scanners break down large blobs of SQL code into a data lineage subgraph for deeper analysis.
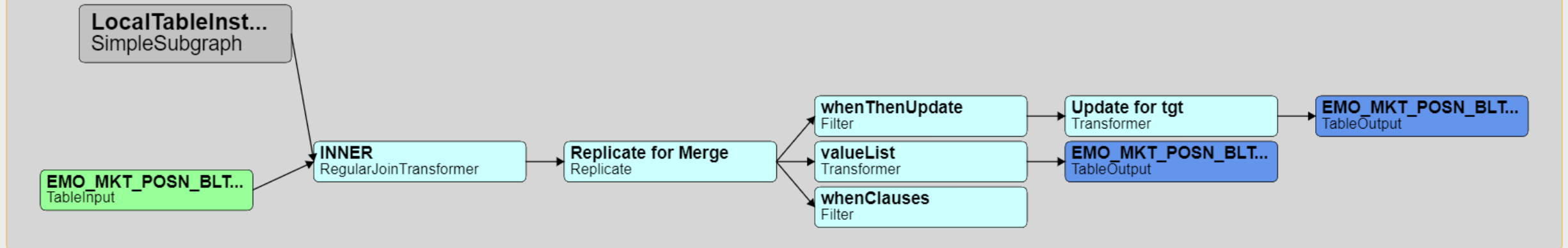
EDC Advanced Custom Metadata Loader

- Load custom object and lineage metadata into the catalog through a business-friendly process

- Allow ingesting metadata without creating manual models - directly ingest metadata

  - Relational databases

  - Microsoft Excel spreadsheets

  - File formats such as XML, JSON and CSV

- No development required – repeatable after configuration and setup

- Obtain complete auditing and governance control over the entire metadata extraction and loading process
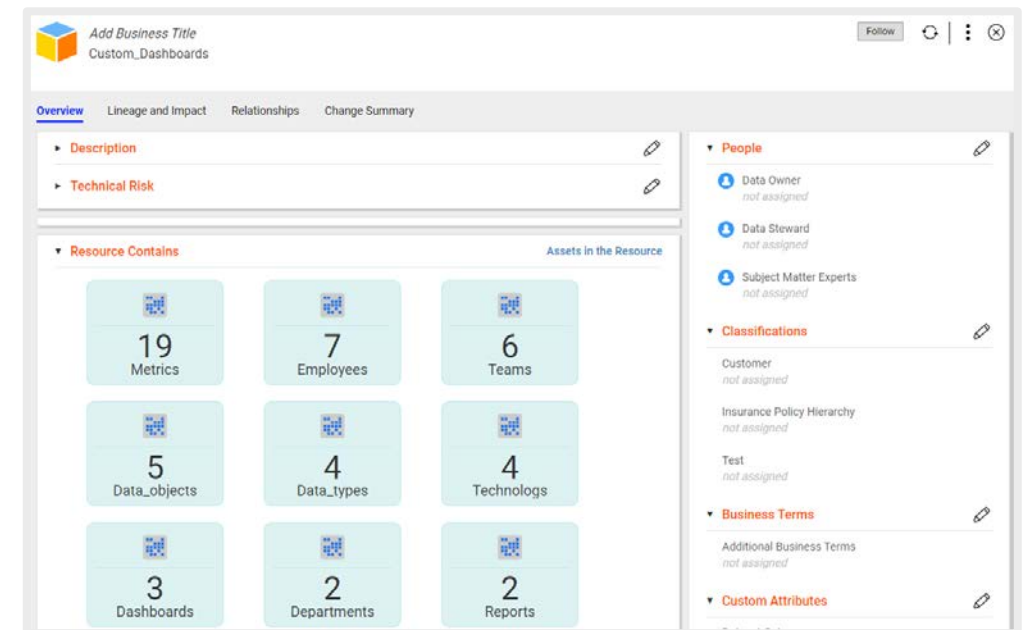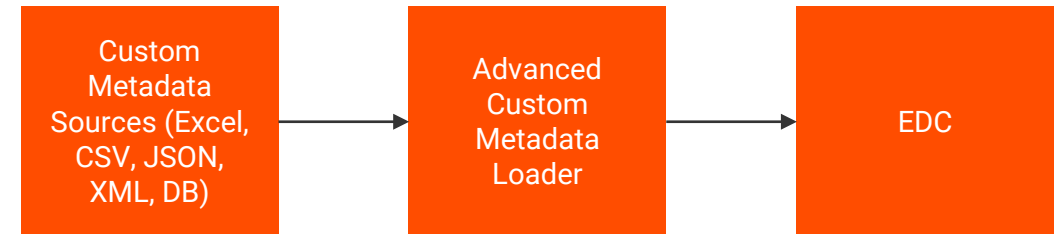
Leave no metadata behind

- **Breadth** – *scan everything you need* including stored procedures, mainframe, ETL, BI, analytical applications, embedded SQL buried everywhere and more

- **Depth** – *scan every single transformation* and every piece of logic, including dynamic SQL, hand-written scripts, database specific load/unload utilities and more

- **Trust** – *be sure that you get all the lineage and no surprises* with clear information about every situation when for some reason complete lineage could not be extracted

- **Integration** – *lineage is a critical part of the data governance story, but there is more* that EDC provides to create a complete data governance platform like profiling or glossary

DEMO

Learn More

- Read the [Informatica Enterprise Data Catalog Advanced Scanner datasheet](#)

- Download a free copy of "[Drive Your Business Forward With a Catalog of Catalogs](#)"

- Watch on-demand customer, partner and Informatica expert presentations on [CLAIREview](#) (the Informatica virtual summit in 2020)

- Visit us at [Informatica Enterprise Data Catalog Advanced Scanners](#)

Informatica

Thank You