



How to Become a Truly Data Driven Enterprise – Trends, Technologies and Scenarios

Erik van de Hoef, Manager Customer Success Unit, Data & AI, Microsoft

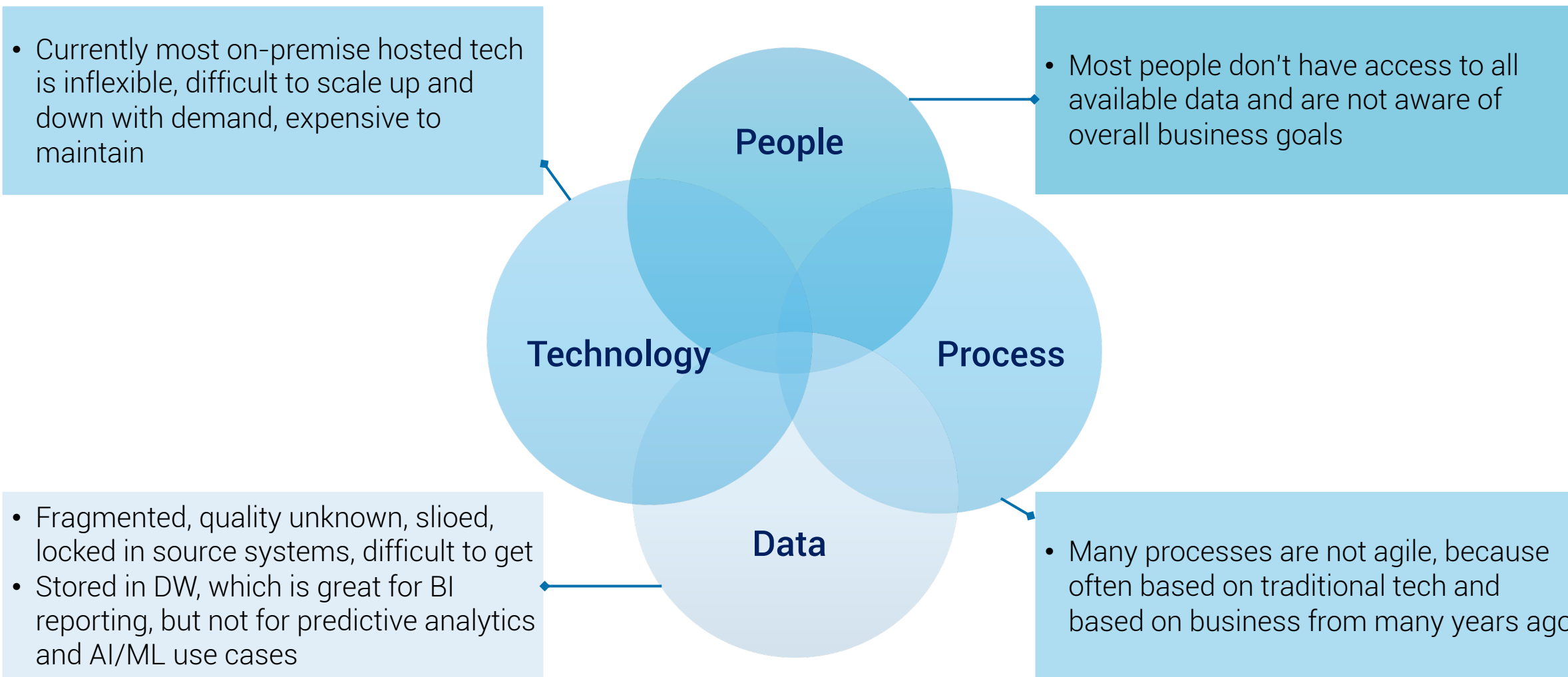
Rick Mutsaers, Principal Architect, Informatica

Shuvadeep Dutta, Principal – Architecture & Framework, AI & Data Engineering, Insights & Data, Capgemini Nederland B.V.

May 18th, 2020



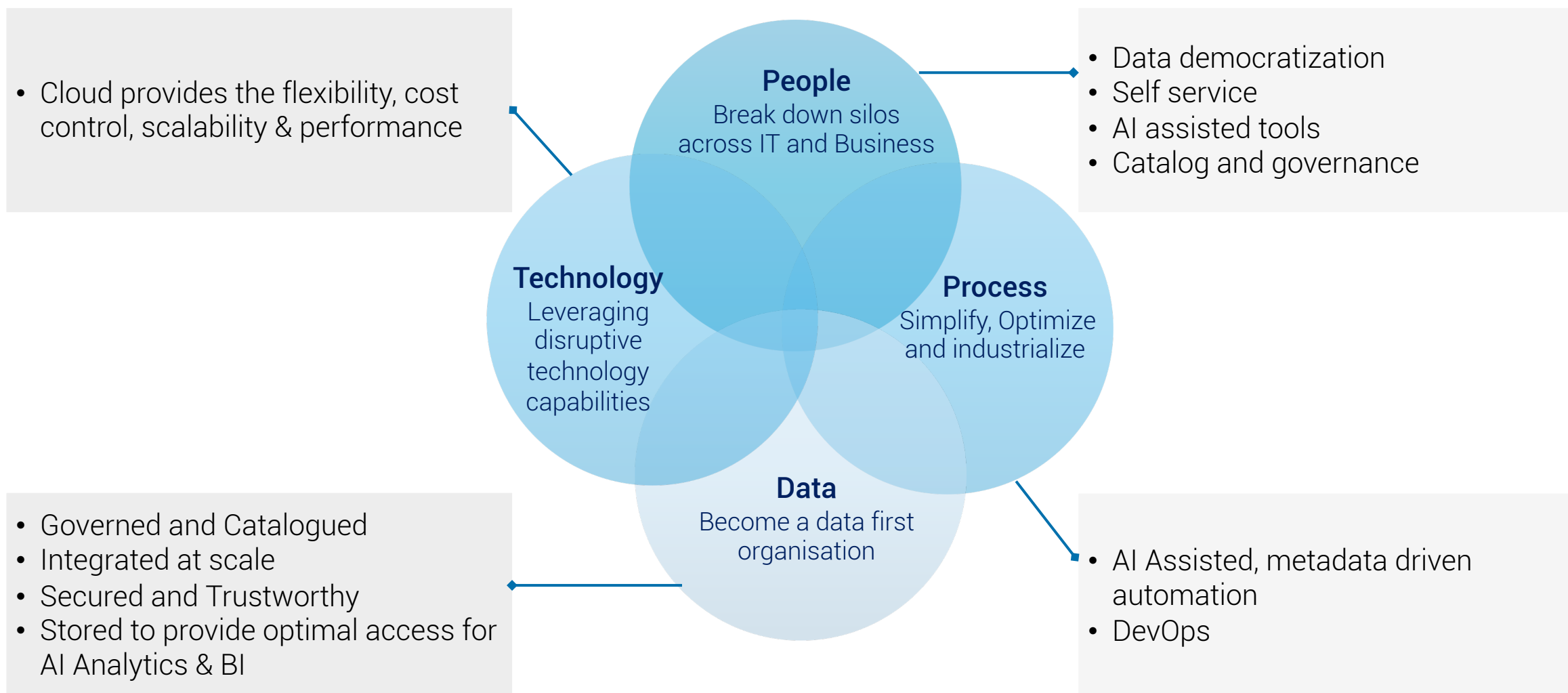
The Challenges



The TECHNOLOGY LANDSCAPE OF TODAY is unable to meet the agility and innovation required by the business and customer demands of today & tomorrow

Drivers to Modernize Enterprise Data Landscapes

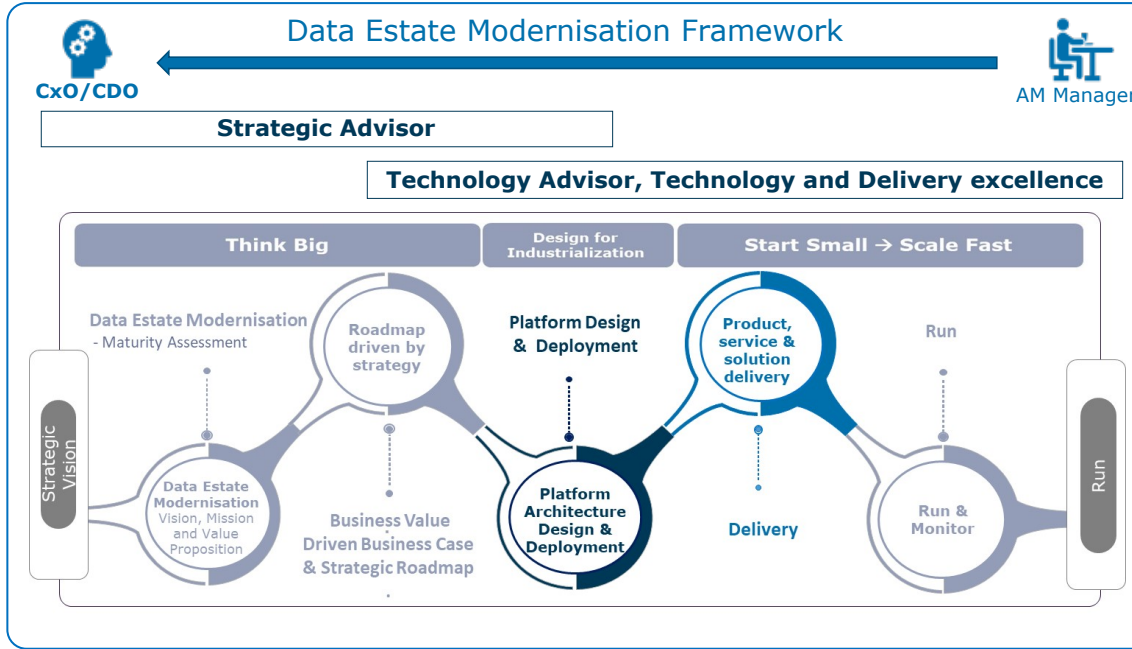
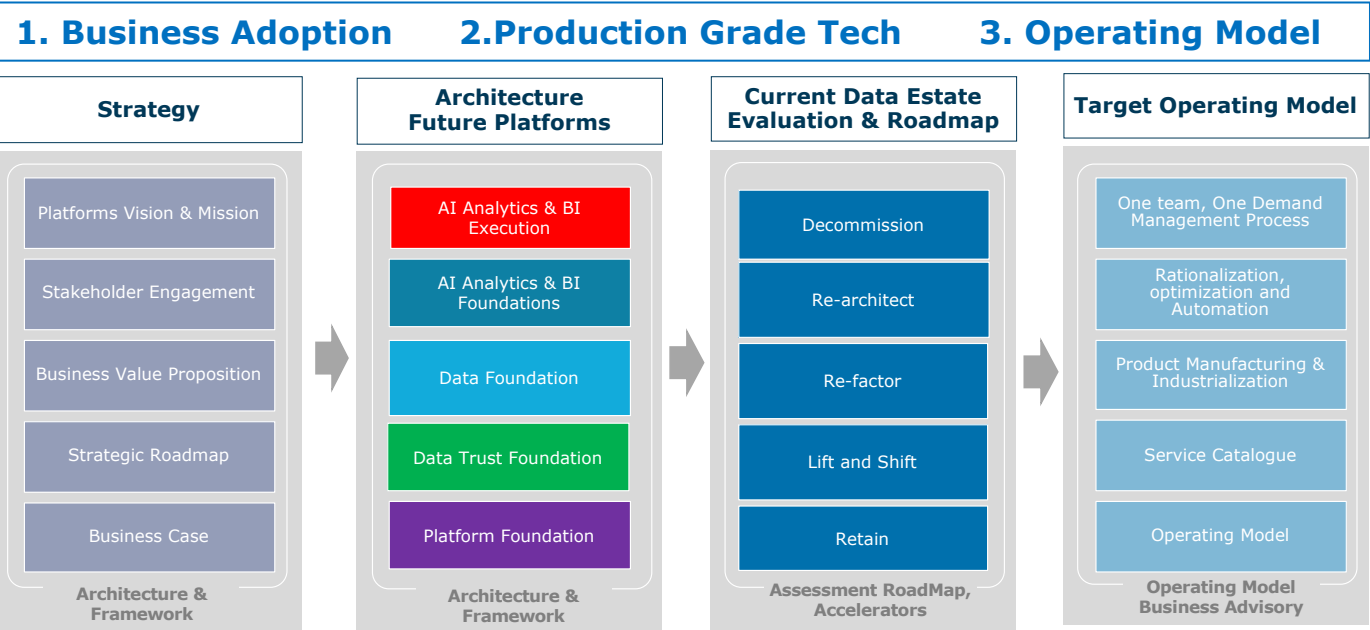
Powered by data, enabled by technology, and enacted by people





How Capgemini, Microsoft and Informatika can help ...

Define strategy, right architecture foundation and target operating model linked to the business outcomes. Capgemini provides a framework to help organize people, technology, processes and data to accelerate the data driven transformation journey





"Design for Industrialization"

Platform Design to Deployment

Architecture Future Data & AI Platform

AI Analytics & BI Execution

AI Analytics & BI Foundation

Data Foundation

Data Trust Foundation

Platform Foundation

Architecture & Framework

Reference Architectures

Architecture Playbooks

Vendor Partnerships



Self-service and Enterprise BI using Microsoft **Power BI**

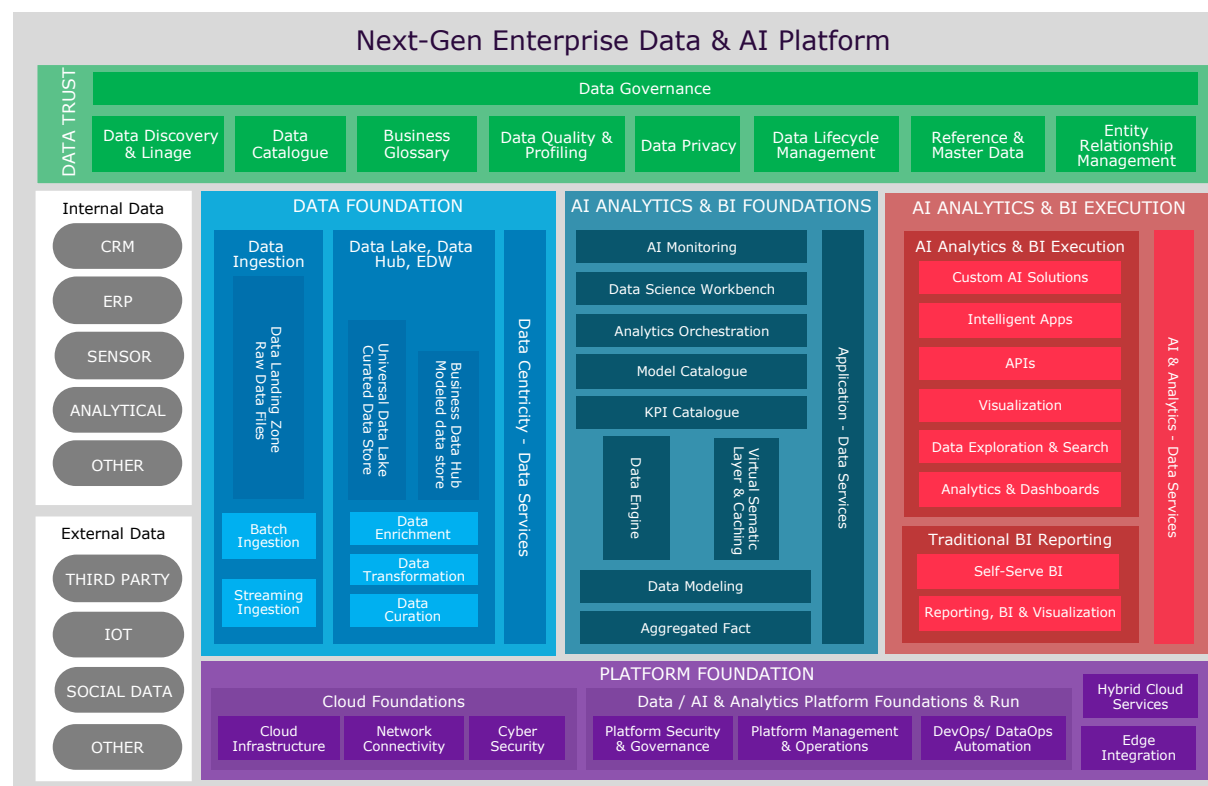
Intelligent app ecosystem and workflows using Microsoft **Power Platform**

Modernization and industrialization of AI Analytics and BI capabilities using **Azure Synapse Analytics** and **Azure Machine Learning**

Foundational Data Engineering capabilities using **Informatica** and **Azure data services**

Data trust strategy, process and tooling using **Informatica**

Platform security, operations management and governance using **Azure DevOps**, **Azure Security Centre**, **Azure Monitor** etc.



A conceptual illustration of Azure Synapse Analytics. The scene is set on a dark, isometric platform. In the center, a large blue cloud contains a hexagonal icon with a stylized 'S' and a circular arrow, representing the Synapse engine. To the left, there are icons for a document, a network of nodes, and two speech bubbles. Below these is a blue cylinder representing a data lake or storage. To the right of the cylinder is a grey server rack. Further right, a network of nodes connects to a donut chart and a 3D area chart with a green line. The text 'Azure Synapse Analytics' is written in white across the center.

Azure Synapse Analytics

Analytics & AI is the #1 investment for business leaders, however they struggle to maximize ROI

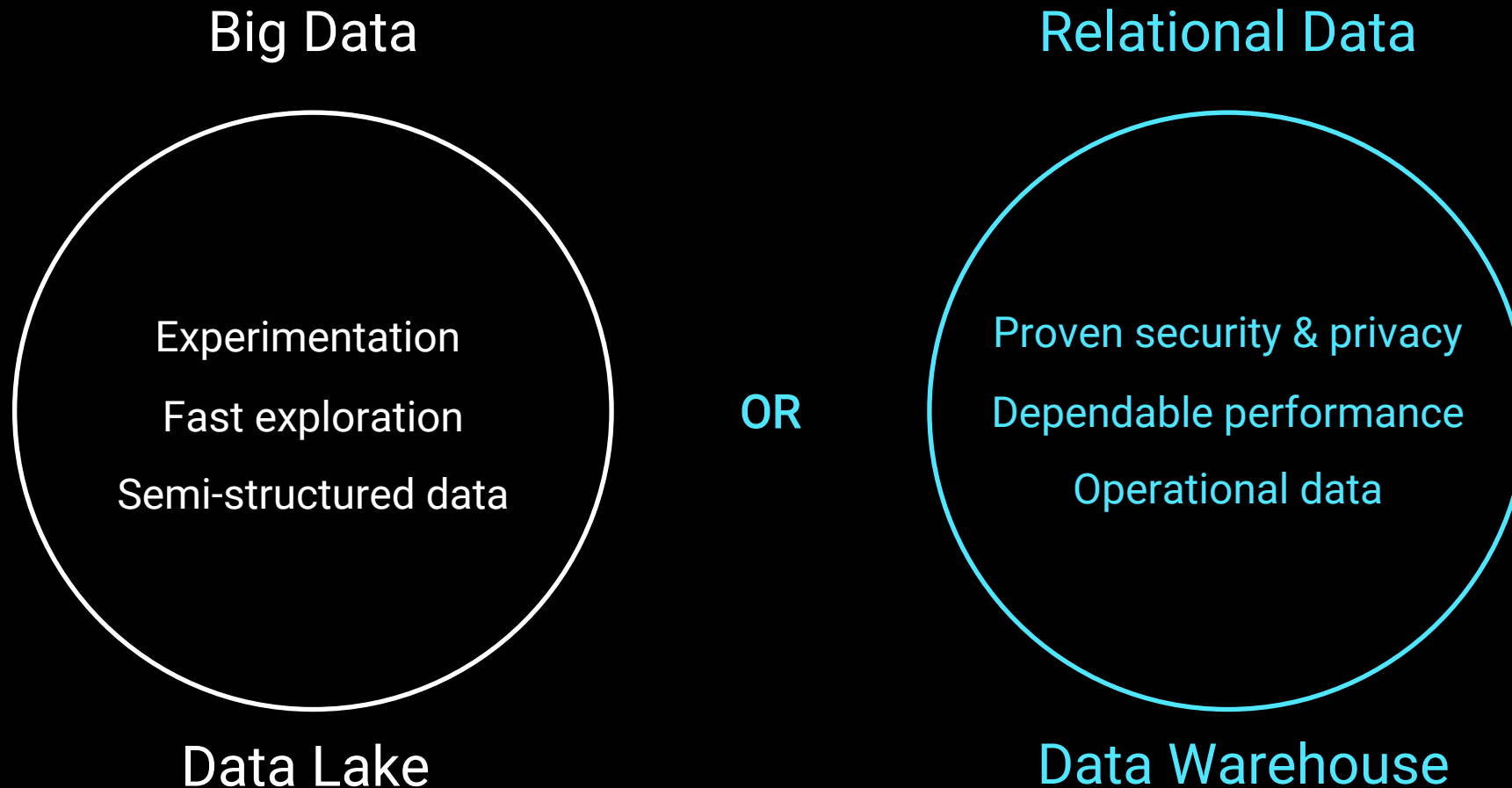
80%

report struggling to become mature users of data*

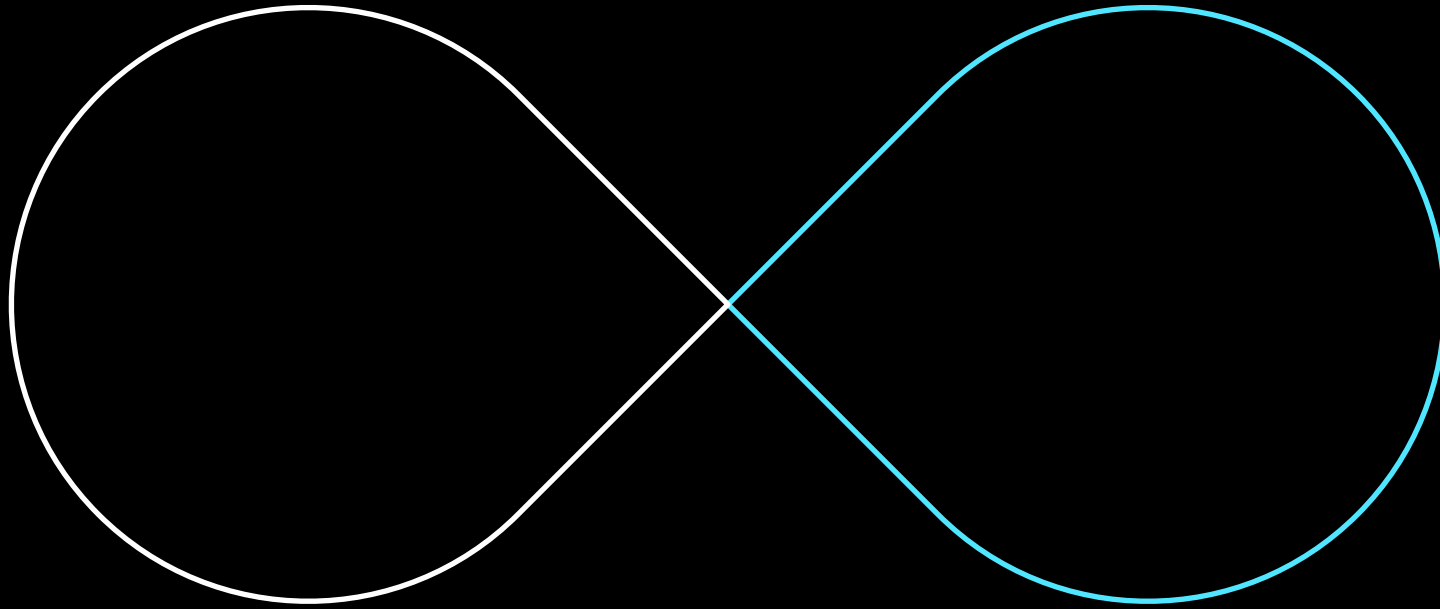
55%

report data silos and data management difficulties as roadblocks*

Businesses are forced to maintain two critical, yet independent analytics systems

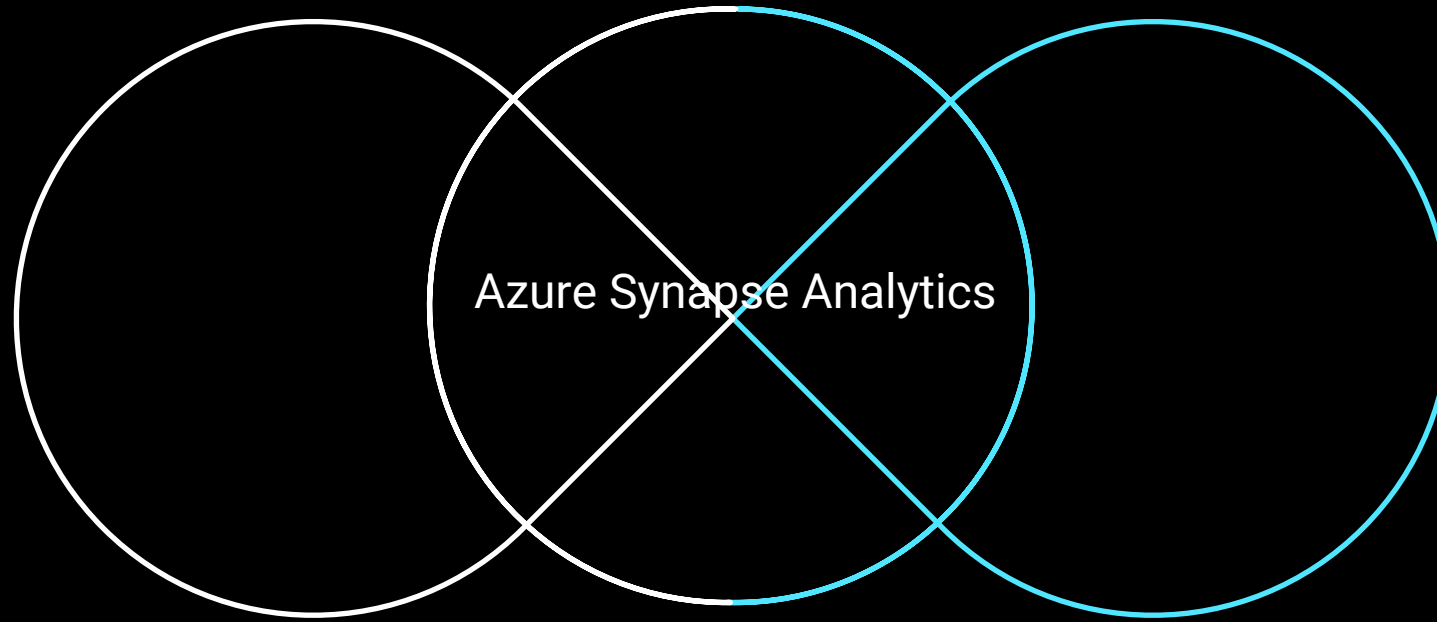


Azure brings these two worlds together, in a single service,
to provide limitless analytics

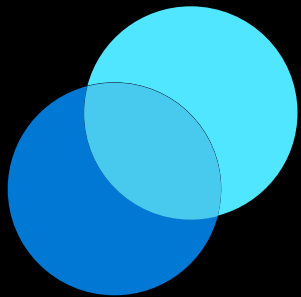


Data warehousing & big data analytics—all in one service

Azure brings these two worlds together, in a single service,
to provide limitless analytics



Data warehousing & big data analytics—all in one service



Introducing Azure Synapse Analytics

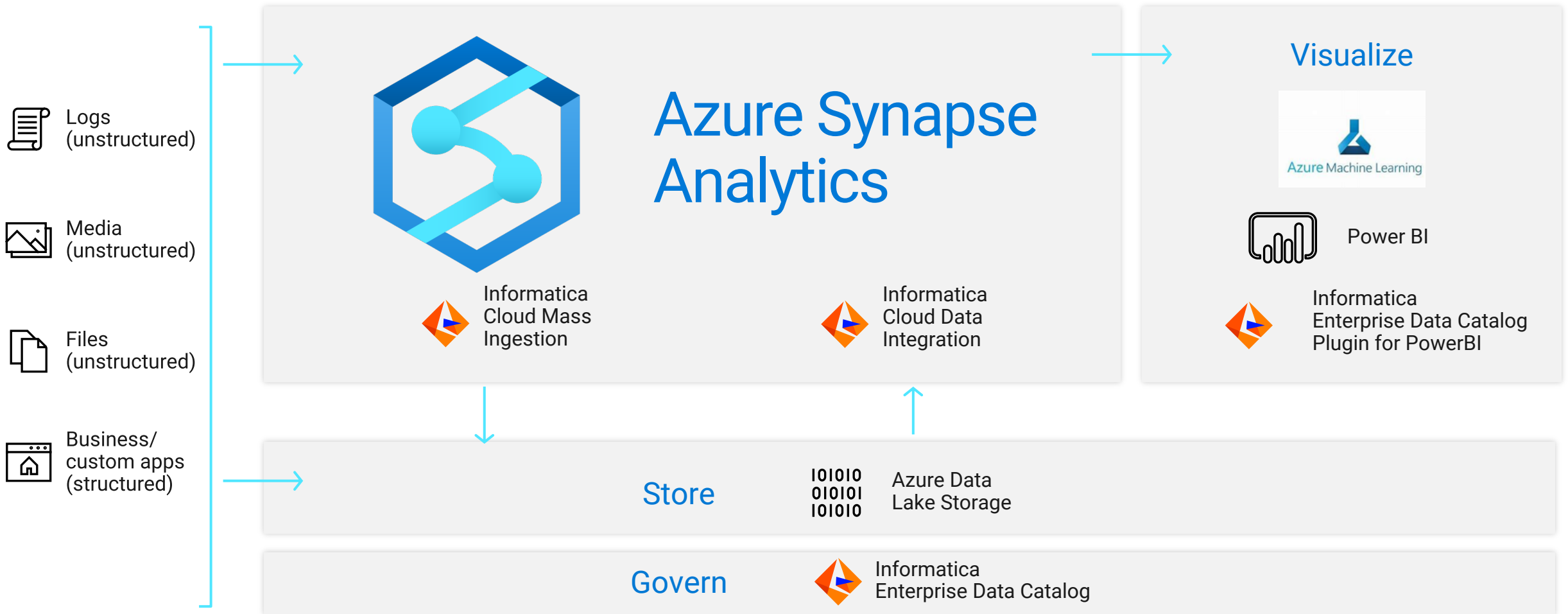
A **limitless** analytics service with **unmatched time to insight**, that delivers insights from all your data, **across data warehouses and big data** analytics systems, **with blazing speed**

Simply put, **Azure Synapse is Azure SQL Data Warehouse evolved**

We have taken the same industry leading data warehouse and elevated it to a whole new level of performance and capabilities

Azure Modern Data Warehouse

Best end-to-end ecosystem based on Azure and Informatica technology



Ecosystem



**Azure Synapse
Analytics**



Power BI



Azure Machine Learning

System Thinking 3.0



Informatica Intelligent Data Platform



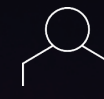
ANY TECHNOLOGY



ANY LATENCY



ANY DATA



ANY USER



MULTI-CLOUD/HYBRID



DATA INTEGRATION



DATA QUALITY



MASTER DATA
MANAGEMENT



DATA CATALOG &
GOVERNANCE



DATA PRIVACY &
SECURITY

CLAIRE™

MONITOR AND MANAGE

DATA ENGINE

CONNECTIVITY

INFORMATICA INTELLIGENT CLOUD SERVICES (CLOUD NATIVE)



ORACLE



teradata.



kubernetes



Provide insight in data assets

Enterprise data catalog as google for data

The screenshots illustrate the Informatica Enterprise Data Catalog interface, showing search results, asset details, and data lineage.

Search Results (1 - 20 of 183)

- Filter by:
 - Asset Type: All, Column (30), Data Domain (25), Dimension (20), Table (17)
 - Resource Type: All, SAP Business Objects (93), Oracle (28), Microsoft SQL Server (21), Salesforce (12), Hive (4)
 - Last Updated: From the beginning
 - Size: N/A (4), Under 500M (16), 1500M to 2000M (0)
 - Resource Name
 - Data Center
 - Remarks

Search Results:

- Customer_Order (Asset Type: Web Intelligence Model)
- CUSTOMER_STATUS (Asset Type: Column)
- Customer_Order (Asset Type: Web Intelligence Document)
- CUSTOMER_STATUS (Asset Type: Column)
- CUSTOMER_TIER (Asset Type: Column)
- CUSTOMER_STATUS (Asset Type: Column)
- Customer (Asset Type: Class)
- CUSTOMER_TIER (Asset Type: Column)
- CUSTOMER_TIER (Asset Type: Column)
- CUSTOMER_STATUS (Asset Type: Column)

Asset Details: cst_rtio_ops

- Description:** The Cost Ratio is an expression of a company's variable production costs as a percentage of sales, calculated as variable costs divided by total revenues. It compares costs that change with levels of production to the amount of revenues generated by production. This contrasts with fixed costs that remain constant regardless of production levels.
- Source Description:** BREAKING DOWN 'Cost Ratio'. Consideration of the Cost Ratio, which can alternately be calculated as 1 - contribution margin ratio, is one factor in determining profitability, since it indicates if a company is achieving, or maintaining, the desirable balance.
- Usage Context:** "Cost" is a word with a very strong association with monetary outlay. When the word "cost" is used, the listener/reader takes it as referring to money, unless the other resource being spent is specified. If we wanted to use "cost-effective" to say that something was very efficient in its use of time.
- Columns (first 5):**

Name	Display Name	Source Data Type
Material Cost	Cost_material	Integer
Material Cost	Cost_material	Integer
mainf_cost	Cost	Integer
Cost	admin_cost	Integer
G&A Cost	account	Integer

People:

- Data Owner: Bobby Shima
- Data Steward: Raul Black
- Stakeholders (54): Michael John, Jayson

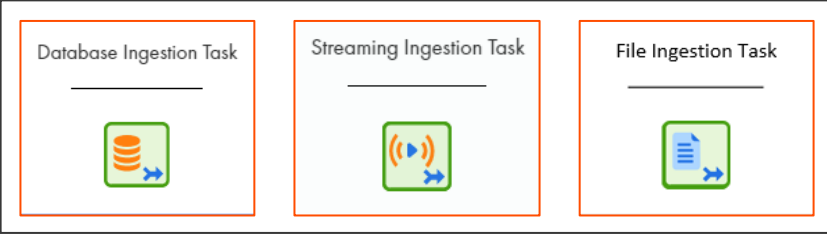
ATM Fraud Analysis

- Composite:** finalfraud_delta (5 stars, 2 reviews)
- Business Classification:** Department: Security Classification
- Business Type:** Actual Cost
- Properties:** Geography: Department: Approval Status: Tags:
- System Properties:**

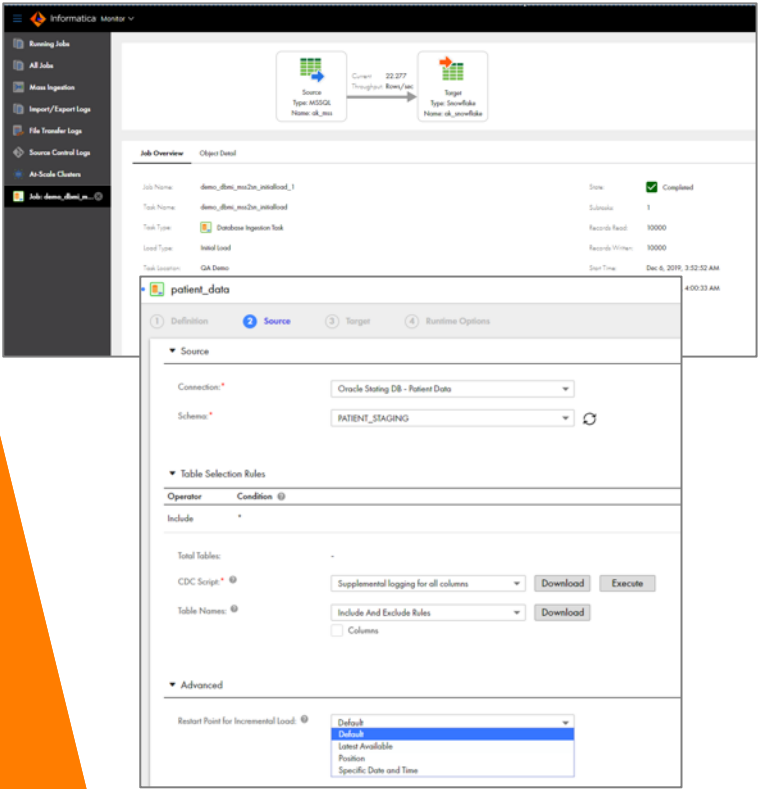
Lineage and Impact:

- Diagram showing data flow from source tables (bom_scanner_BOMScann..., databricks_azure_server) through intermediate tables (DataLakeRaw) to the final output (databricks_delta_table, finalfraud_delta).

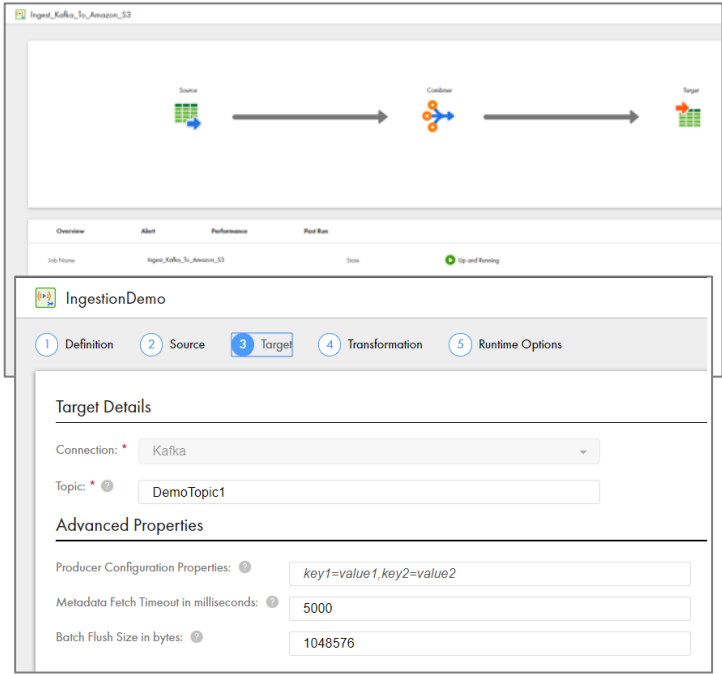
Ingest Data from Database, Streaming, Files



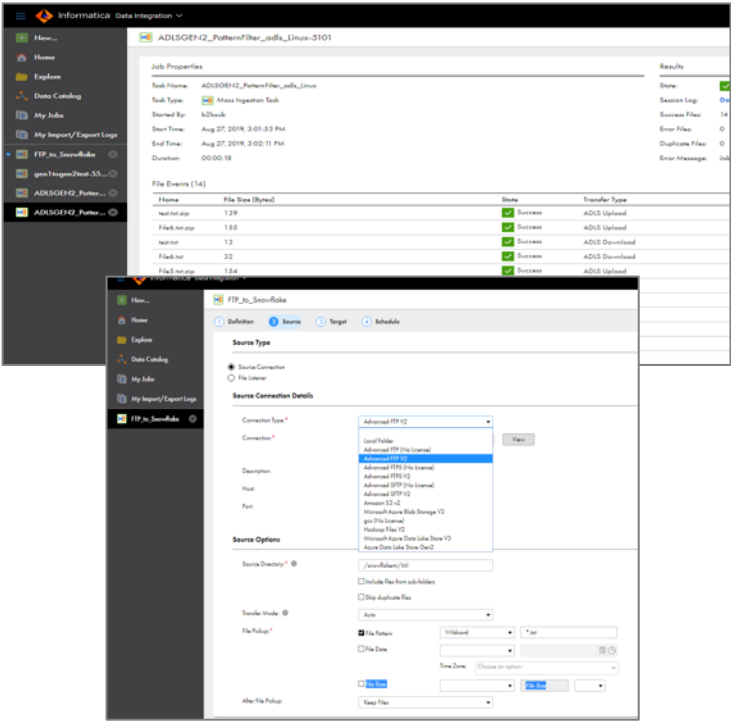
DATABASE INGESTION



STREAMING INGESTION



FILE INGESTION



Cloud Data Integration

Codeless Development Environment

The screenshot displays the Informatica Data Integration Codeless Development Environment. The main workspace shows a workflow design for 'm_load_MDM_Party'. The workflow consists of the following components: 'Source' (highlighted in the left sidebar), 'Target', 'Aggregator', 'Cleanser', 'Data Masking', 'Expression' (highlighted in the left sidebar), and 'Filter'. The workflow is currently in a 'Valid' state. A 'New Asset' dialog is open on the right, prompting the user to select the type of asset to create. The dialog lists several options under the 'Warehousing' category:

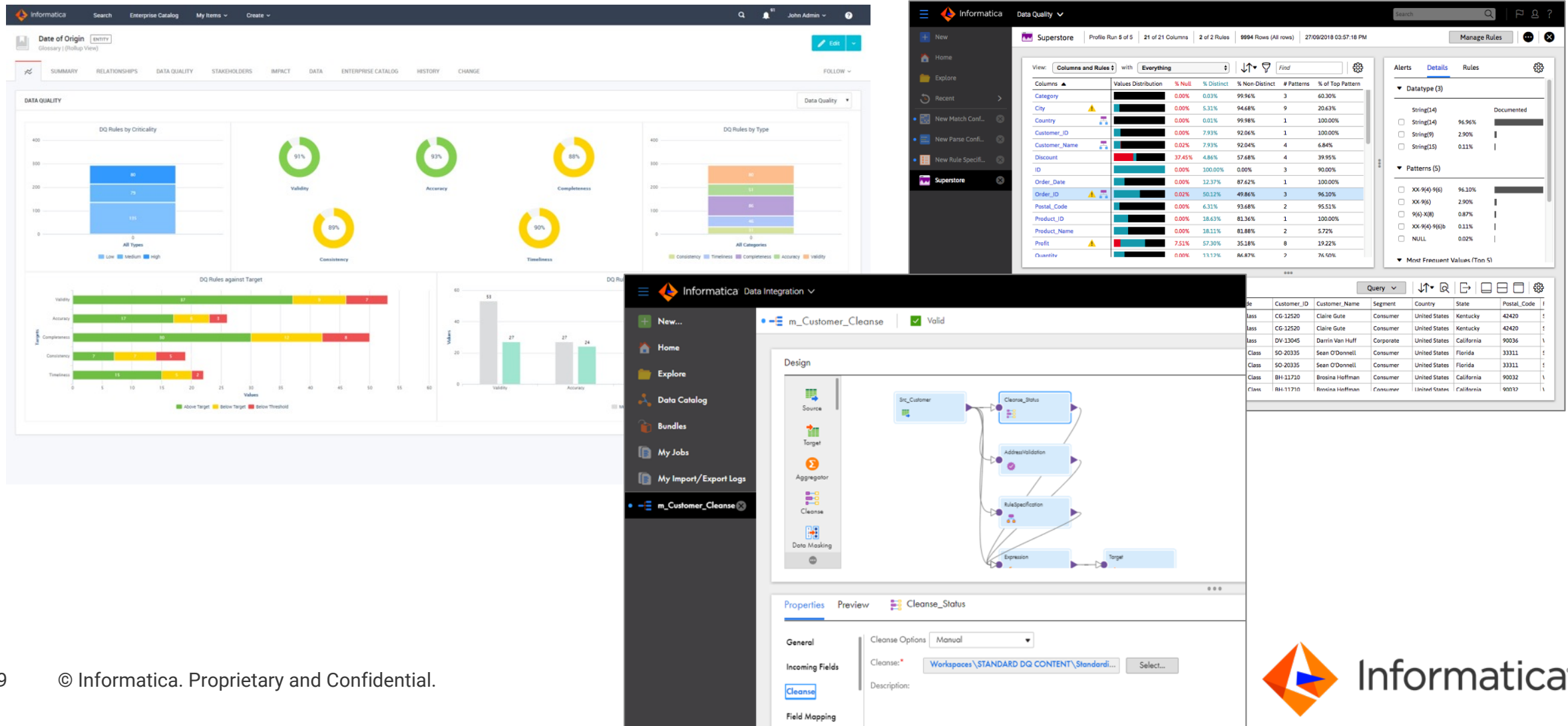
- Slowly Changing Dimension Type 1**: Source records are sent to the slowly changing dimension as inserts for new records or as updates for existing records.
- Slowly changing dimension Type 2**: Preserve changes for selected slow changing dimension fields. If one of selected fields changes in the source, a new record will be inserted into slow changing dimension with new validity dates and the old record will be updated with reset of its validity date and its status flag.
- Slowly changing dimension Type 3**: Manage both current and historical value of a certain attribute in a single record. History is limited to the attribute's previous value.
- Slowly changing dimension Type 6**: Track changes by combining techniques of Type 1, Type 2, and Type 3 (1 + 2 + 3 = 6).

The dialog also includes a 'Create' button and a 'Cancel' button.

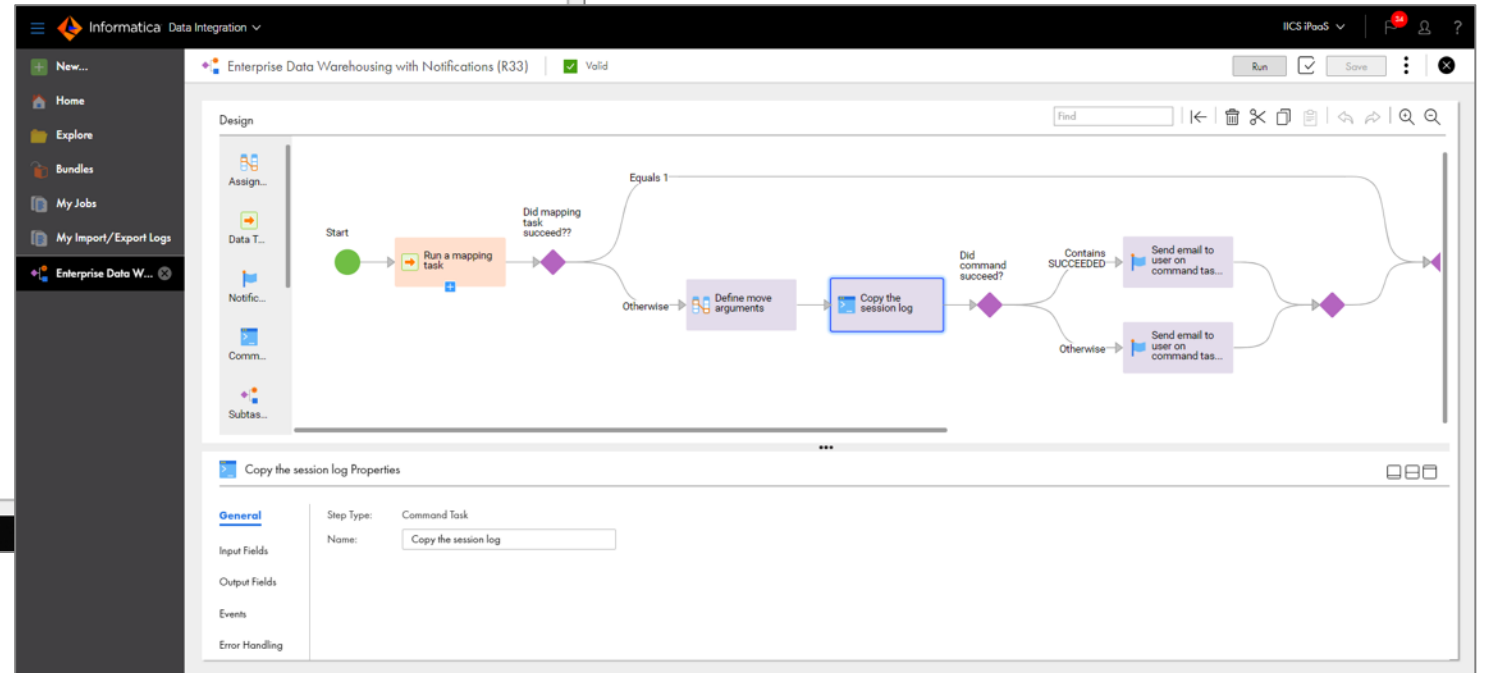
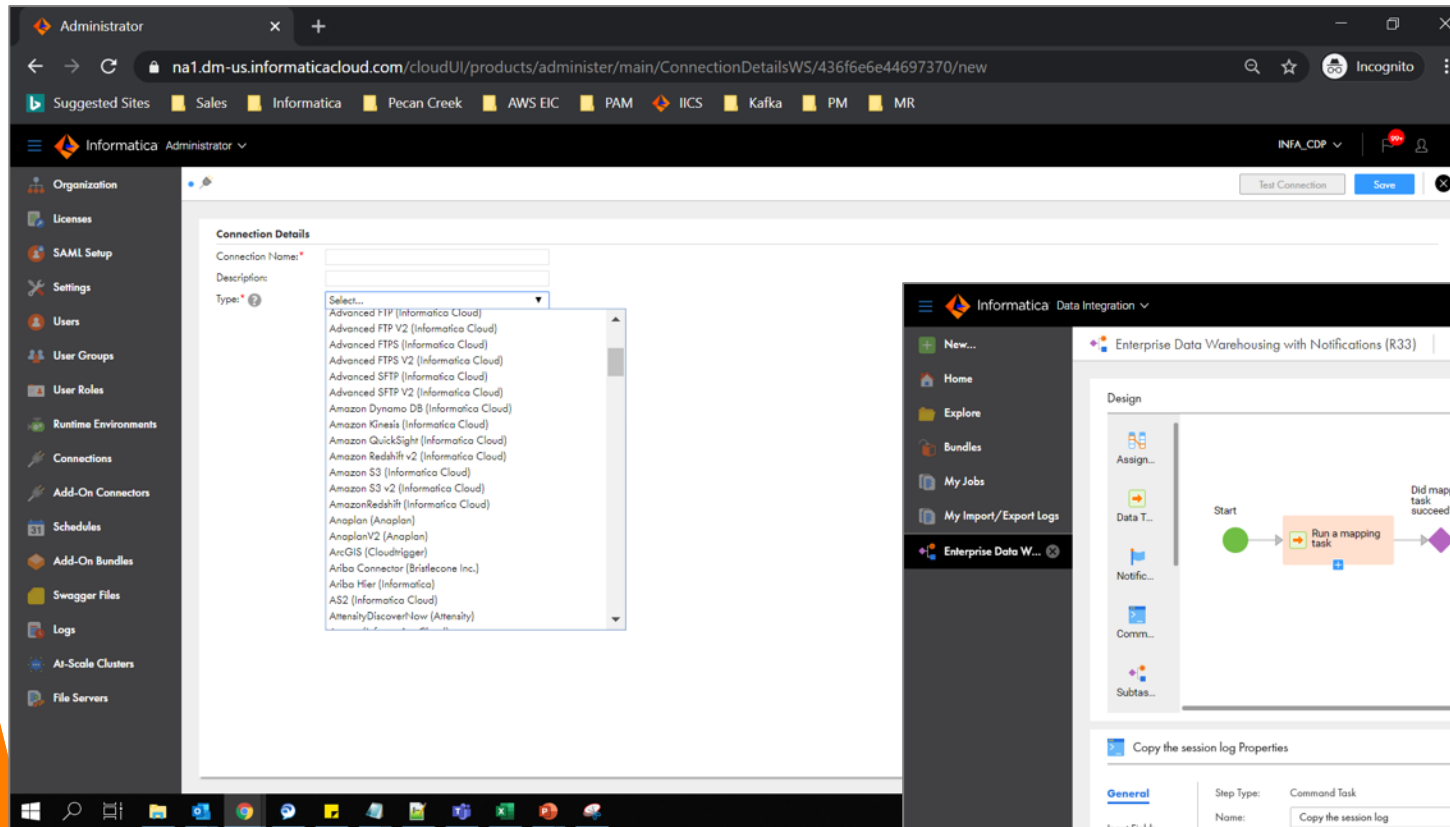
Cloud Data Integration

Enterprise business focused data quality

Define, measure and enhance data quality based on enterprise wide quality rules



Broad Connectivity and Robust Orchestration



Cloud Data Integration

Support for ETL and ELT modes

Advanced Session Properties

Session Property Name	Session Property Value
Allow Temporary View for Pushdown	Yes
Pushdown Optimization	Full
<input checked="" type="checkbox"/> Enable cross-schema pushdown optimization	

Mapping Image

SnowflakeSrc → Expression → SnowflakeTRG

Input Parameters

Name	Type	Value
\$\$SnowflakeSrc\$	Multi-Object Source	Source Connection: SnowflakeV2_Pushdown_Azure
		Source Object: CUSTOMER
\$\$SnowflakeTRG\$	Target	Target Connection: SnowflakeV2_Pushdown_Azure
		Selected Object: CUSTOMER2
		Operation: Update

Task Schedule

Schedules: Do not run this task on a schedule

Design

Ex_3_SCD2_SFdc_Snowflake_Account

Design view showing a complex data flow map with components like SFDC_Account, exp_fields_src, src_src, src_account_dim, exp_key_dim, src_dim, src_outer, exp_calculate, trg_key, trg_dim_new_insert, trg_dim_ex_insert, and trg_dim_ex_upd.

Properties

Preview: SFDC_Account

General

Source

Connections: Salesforce_IICS (Salesforce) View... New Connection... New Parameter...

Fields

Source Type: Single Object

Partitions

Object: Account Select... Preview Data...

Serverless Spark Processing with Elasticity

The image displays two overlapping screenshots from the Informatica Cloud Data Integration (CDI) interface.

Left Screenshot: Demo_Cluster Configuration

Basic Configuration

Create or modify an at-scale configuration that you can use to run at-scale jobs.

Name: Demo_Cluster

Description:

Runtime Environment: sa_group

Cloud Platform: Amazon Web Services (AWS)

Region: US East (Ohio)

Master Instance Type: m4.2xlarge

Master Instance Profile:

Worker Instance Type: m4.2xlarge

Minimum Worker Nodes: 2

Maximum Worker Nodes: 5

Worker Instance Profile:

Availability Zones: Select Availability Zones

EBS Volume Type: SSD (gp2)

EBS Volume Size: 150 GB

Idle Timeout: 5 minutes

Staging Location: integrationat-scaledemo/staging

Log Location: integrationat-scaledemo/logs

☒ **Encrypt Data**

Right Screenshot: MyMappingTask Configuration

Created By: sabgupta@informatica.com

Updated By: sabgupta@informatica.com

Last Run: Nov 15, 2019 8:50:45 PM

Task Mapping

Name: MappingWithAnExpression - Copy 1

Location: sabira

Updated Date: Oct 15, 2019 3:55:07 PM

Advanced Session Properties

Session Property Name	Session Property Value
Session Log File Directory	asdasd
Write Backward Compatible Session Log File	No

☒ Enable cross-schema pushdown optimization

Tuning Recommendation

Property	Value
spark.driver.extraJavaOptions	some-text
spark.driver.maxResultSize	some-text
spark.driver.memory	some-text
spark.dynamicAllocation.maxExecutors	some-text
spark.executor.cores	some-text
spark.executor.extraJavaOptions	some-text

Buttons: Tuning Recommendation, Tuning by CLAIRE, Edit, Run

Cloud Data Integration

Democratize access to data

The screenshot displays the Informatica Data Marketplace interface. The top navigation bar includes 'Informatica', 'Search', 'Enterprise Catalog', 'My Items', 'Create', and 'Data Marketplace'. A search bar is present with the text 'Find Data Collections by Name, Purpose, Business Domain or Tag...'. The main content area is divided into several sections:

- My Data Collections:** A grid of data collection cards. Each card shows the collection name, status (Certified, Pending Approval, or Access Granted), a description, and a star rating. Examples include 'Fleet DOE Fleet Product', 'Mfg Parts Orders - Camry', 'Vehicle Inventory', and 'Vehicle Manufacturing'.
- My Data Collections Usage Trend:** A bar chart showing usage trends over time, with options for 'Last Week', 'This Week', and 'Weekly'.
- My Top Data Collections:** A list of top collections, including 'Mfg Parts Orders - Yaris' and 'Yaris Customer 2019'.
- Data Set Details:** A detailed view of a data set, showing its overall quality (95%), attributes (16), and a table of linked fields. The table includes columns for 'Linked Fields', 'Attribute Glossary', 'Attribute Glossary Type', 'Data Type', 'Data Length', 'Quality Ref.', and 'Quality Score'.

Linked Fields	Attribute Glossary	Attribute Glossary Type	Data Type	Data Length	Quality Ref.	Quality Score
Country Name	Country Name	ENTITY	CHAR	2	DOBULE-17	86%
Country Name	Country Name	ENTITY	VARCHAR	255	DOBULE-42	80%
Value list of KYC/KYB statuses	Value list of KYC/KYB statuses	Value List				Overall Quality: 95%

Data Marketplace

Intelligent Platform – smart recommendations

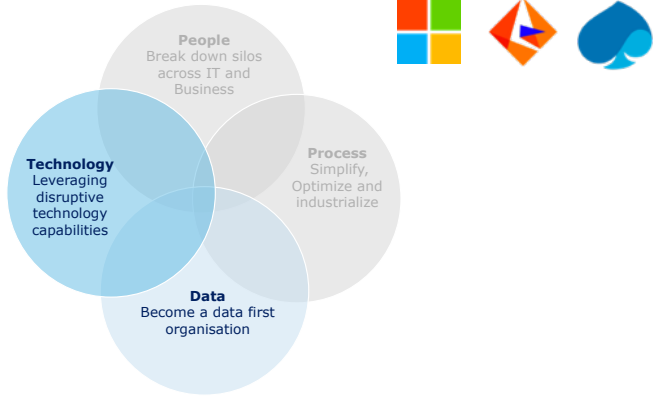
The screenshot displays the Informatica Data Integration (IICS) interface. The top navigation bar shows 'Informatica Data Integration' and 'IICS'. The left sidebar contains navigation options: 'New...', 'Home', 'Explore', 'Data Catalog', 'Bundles', 'My Jobs', 'My Import/Export Logs', 'ISD_PDF', and 'm_Workday_to_Sal...'. The main workspace shows a workflow design with a 'Source' (Wd) connected to a 'Filter' connector, which is then connected to a 'Target' (SalesforceContacts). A dropdown menu is open over the 'Filter' connector, showing options: 'Expression', 'Router', 'Lookup', 'Aggregator', 'Cleanse', and 'Data Masking'. Below the design, the 'Properties' panel for the 'Filter' connector is visible, showing 'Name: Filter' and 'Description:'. A 'Business Customer' data catalog entry is overlaid on the right, showing a 5-star rating, description 'contains most reliable customer information from dba point of view', Data Owner: 'Administrator', and tags 'Customer', 'Person', 'Address_Information', and 'EU_Citizen'. A 'Configure Tuning' dialog box is also open, providing recommendations for Spark properties. It states: 'CLAIRE Tuning runs the mapping task multiple times in succession to optimize performance. To set resource limits, configure the appropriate Spark properties before you tune the mapping task.' The 'Number of Mapping Task Runs' is set to 25. The dialog is powered by CLAIRE™ and includes 'Tune' and 'Cancel' buttons.

Business Customer
CUSTOMER
★★★★★
contains most reliable customer information from dba point of view
Data Owner: **Administrator**
Customer Person Address_Information EU_Citizen
Hermes / XE / HERMES

Configure Tuning
Get a Recommendation for a set of Spark properties that optimize the performance of the mapping task.
CLAIRE Tuning runs the mapping task multiple times in succession to optimize performance.
To set resource limits, configure the appropriate Spark properties before you tune the mapping task.
Number of Mapping Task Runs: 25
POWERED BY CLAIRE™
Tune Cancel

CLAIRE™ Powered

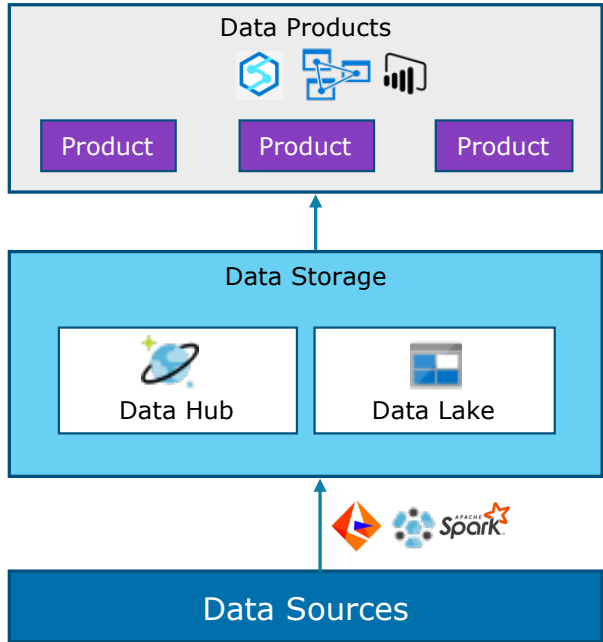
Scenario 1



Key Drivers

- Establish a Federated System Data Lake
- Data driven collaboration and realize new business value
- Provide value and capabilities to the customer ecosystem entities
- TCO reduction

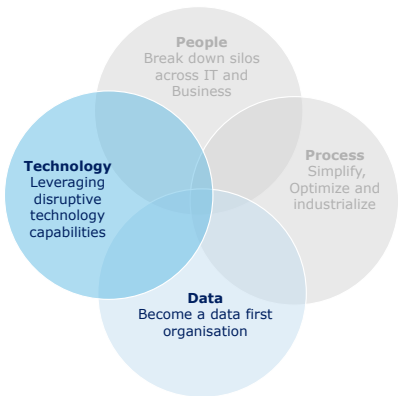
Solution



Benefits

- Rapid product development
- Faster availability of data
- On-demand scaling
- Robust data foundation
- Flexibility and cost efficiency
- Data standardization

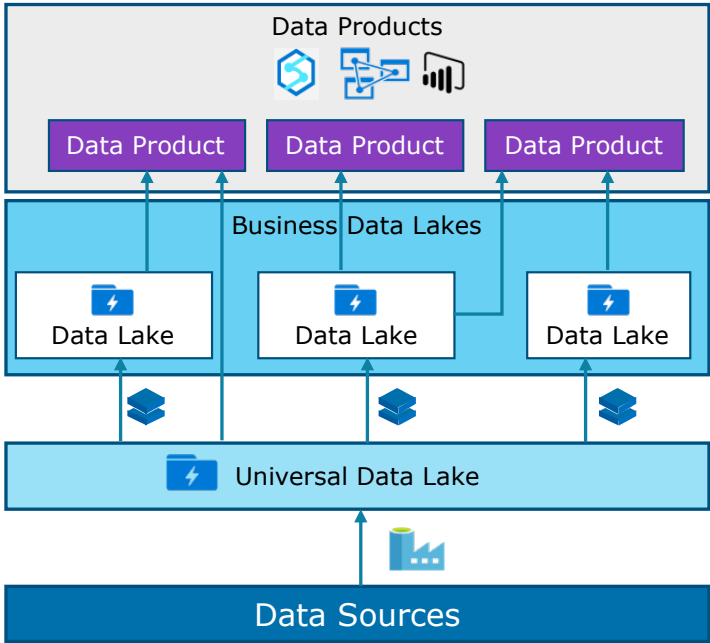
Scenario 2



Key Drivers

- Speed and agility supported by the right data
- Reduce BAU / Run costs
- Data consolidation
- On-demand analytics products

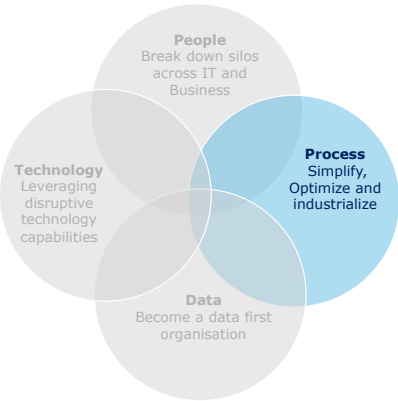
Solution



Benefits

- Improved system performance and higher speed to market
- Better cost efficiency
- On-demand analytical insights across business lines at scale
- Robust data foundation
- Improved service efficiency

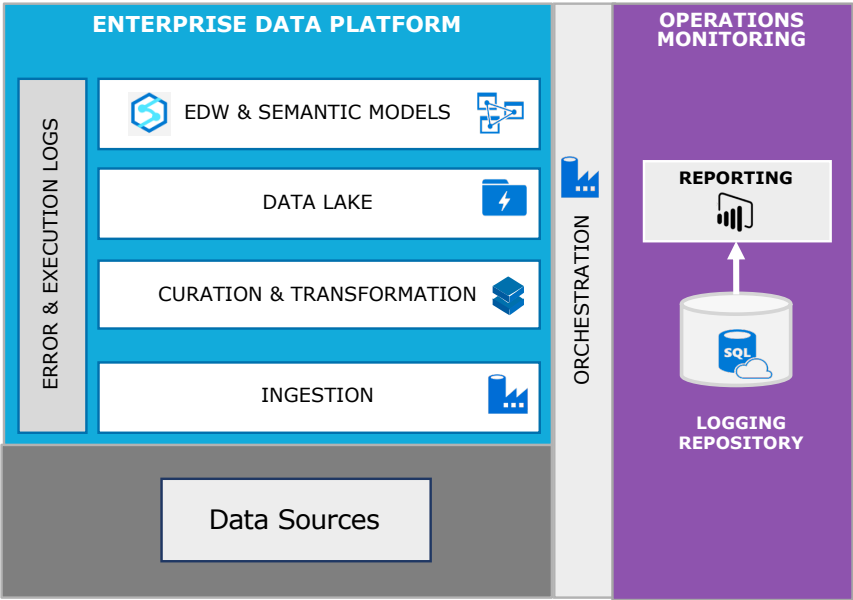
Scenario 3



Key Drivers

- Faster identification of data issues
- Improve platform operations
- Establish platform governance

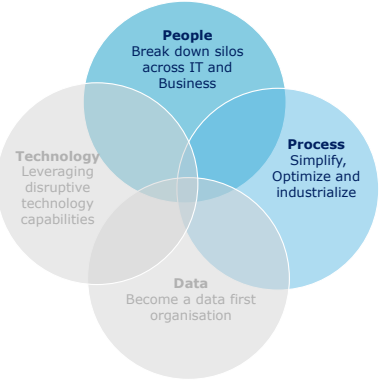
Solution



Benefits

- Efficient monitoring and management of data operations
- Auditing and faster remediation of data issues
- Better management and usage of platform resources

Scenario 4

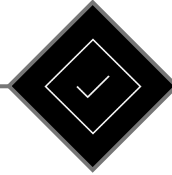


Key Drivers	Solution	Benefits
<ul style="list-style-type: none">• Business and IT productivity• Product delivery efficiency• Collaboration and interoperability between stakeholders• Quality assurance	<p>The diagram illustrates the 'Solution' architecture. At the top, a blue arrow labeled 'Continuous Integration' points to another blue arrow labeled 'Continuous Delivery'. Below this, a box labeled 'CI/CD Pipelines' contains a sequence of steps: 'VERSION CONTROL' (with a Git icon), 'ENV 1', 'ENV 2', and 'ENV N' (with a Kubernetes icon). To the left of the pipelines is a box for 'DEV TEAMS' containing 'DATA INGESTION' and 'DATA PRODUCTS'. Below the pipelines is a box for 'CHANGE MANAGEMENT' containing a sequence of steps: 'OPEN CHANGE REQUEST', 'ASSIGN RESOURCES', 'COMPLETE BUILD & TESTING', 'CHANGE APPROVAL', 'READY TO DEPLOY', and 'CLOSE REQUEST'. To the left of the change management process is a box for 'PLANNING TOOL'. To the right of the pipelines is a box for 'PLATFORM OPS' (with a server icon). Arrows indicate the flow of data and processes between these components. A vertical bar on the far left is labeled 'CHANGE PROCESS' and 'CI/CD TOOLCHAIN'.</p>	<ul style="list-style-type: none">• Reduced efforts• Improved data product quality• Collaborative and efficient planning and execution• Improved platform IT operations efficiency• Faster data ingestion

Take Aways



To become a true data driven organization you need a strategy and a plan. Capgemini can help you build that.



You need a platform to store, process and analyze data. Microsoft Azure is that platform



You need clean, trustworthy data that is well understood and democratized. Informatica delivers that.



Call to Action 1 / 2

Free Proof of Value from Capgemini + Microsoft + Informatica



	Microsoft + Capgemini	Informatica
Free Hands-On Migration Workshop	✓ Cloud Data Warehouse Modernization Workshop	✓ Cloud Data Warehouse Modernization Workshop
Free Data Catalog & ETL Migration		✓ Enterprise Data Catalog (EDC) Intelligent Cloud Services (IICS)
Free Code Migration	✓ On-prem Data Warehouse Code Migration	✓ Informatica Services + Support
Free Azure SQL DW Proof of Value	✓ Azure SQL DW Subscription (for duration of PoV) Implementation Guidance	



Call to Action 2 / 2



Interested to have a deep-dive session with us?

You can register for a deep-dive session with us at:

<https://www.capgemini.com/nl-nl/bronnen/webinar-week-data-modernization/>

We look forward to welcoming and joining you in the session.

You can also reach out to the below contacts for follow-up requests:

Pieter Nieuweboer, Head – Insights & Data, Capgemini Nederland B.V.

Email: pieter.nieuweboer@capgemini.com

LinkedIn: <https://www.linkedin.com/in/pieter-nieuweboer>

Shuvadeep Dutta, Principal Consultant, Insights & Data, Capgemini Nederland B.V.

Email: shuvadeep.dutta@capgemini.com

LinkedIn: <http://www.linkedin.com/in/shuvadeep-dutta-0b558b5>

Capgemini Expert Connect: <https://www.capgemini.com/experts/insights-data/shuvadeep-dutta/>



People matter, results count.

This presentation contains information that may be privileged or confidential and is the property of the Capgemini Group.

Copyright © 2020 Capgemini. All rights reserved.

About Capgemini

A global leader in consulting, technology services and digital transformation, Capgemini is at the forefront of innovation to address the entire breadth of clients' opportunities in the evolving world of cloud, digital and platforms. Building on its strong 50-year heritage and deep industry-specific expertise, Capgemini enables organizations to realize their business ambitions through an array of services from strategy to operations. Capgemini is driven by the conviction that the business value of technology comes from and through people. It is a multicultural company of 200,000 team members in over 40 countries. The Group reported 2017 global revenues of EUR 12.8 billion.

Learn more about us at

www.capgemini.com/PerformAI