



## Track and improve Data Quality using SAP Agile Data Preparation

Ginny Segbers, Manager Conversion and Data Readiness, Duke Energy  
Michael Eacrett, VP Product Management, SAP

Session ID 84169

# About the Speakers

## **Ginny Segbers**

- Manager Conversion and Data Readiness, Duke Energy
- Experienced technical developer and business manager, 28 years with Duke Energy, 10 years in the web space, 4<sup>th</sup> year on Customer Connect project
- Math major with art background

## **Michael Eacrett**

- VP Product Management, SAP Data Hub and EIM
- Previously lead for SAP HANA PM and was a member of the SAP NetWeaver PM. 27 years of SAP product, implementation, and customer experience
- Lack of art and math talent = a computer science degree

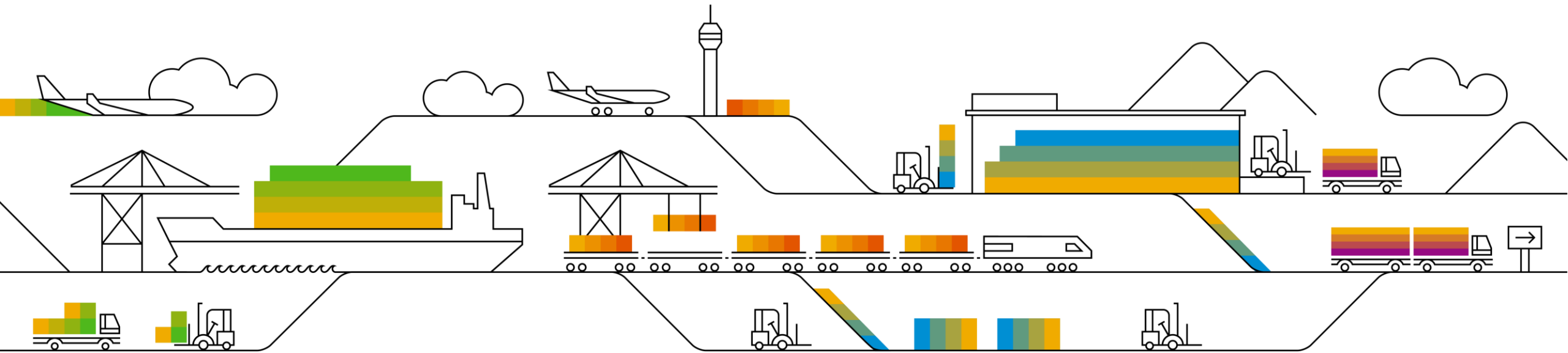
# Key Outcomes/Objectives

1. What is SAP Agile Data Preparation
2. How can it be used in my company
3. Accelerate your data trustworthiness
4. From business to action

# Agenda

- SAP Agile Data Preparation Overview
- Duke Energy Overview
- Project Presentation
- Next Steps
- Q&A

# SAP Agile Data Preparation Overview



# Multi-user-oriented, visual, interactive data prep



## Drive User Collaboration and Sharing

From IT generated data to multiple-user generated data



## Discover Data in a New, More Agile Way

From data gathering to data exploration

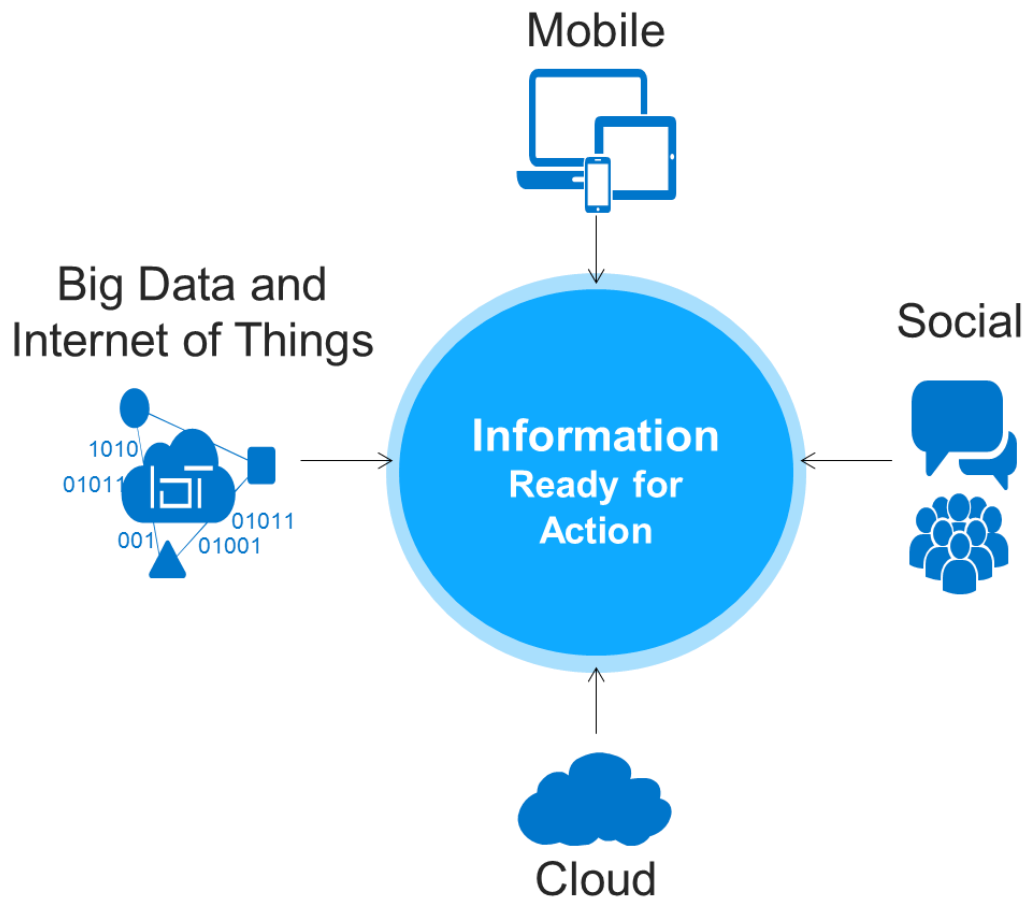


## Enable Data Curation for Better Analysis

From data nightmares to beautiful storytelling

**Driven by digital business transformation**

# Business Users Increasing Need



**10%** of organizations have adopted some form of self-service data preparation, which is estimated to grow to 30% of organizations by 2020<sup>1</sup>

Through 2020, spending on self-service visual discovery and data preparation market will grow

**2.5x** faster than traditional IT-controlled tools for similar functionality<sup>2</sup>

**20%** of global data and analytics decision-makers said the lack of end user self-service capabilities is the biggest challenge with their BI strategy<sup>1</sup>

<sup>1</sup> Gartner: Forecast Snapshot: Self-Service Data Preparation, Worldwide, 2016; <sup>2</sup> IDC: Self-Service Data Preparation Market Supply" 2016; <sup>3</sup> Forrester: "Vendor Landscape: Data Preparation Tools" 2016



# Simplifying the Way you Discover and Shape Data

**SAP Agile Data Preparation** is a self-service data preparation application providing data discovery, integration, and transformation capabilities. Quickly transform your data into actionable, easily consumable information. Built for all types of users – the software can help you drive more successful analytics, data migration, and master data management (MDM) initiatives. Simplify how you access and discover the shape of data and become far more productive and agile than you ever dreamed.





# SAP Agile Data Preparation ?

- Perform data discovery and profiling
- Enrich, blend and augment data
- Harmonize disparate data sources
- Share and collaborate prepared data sets

SAP Agile Data Preparation is a self-service data preparation application that helps multiple user roles explore, integrate and transform raw data into actionable information. It provides users with greater agility to respond to new data sources, business requirements and market opportunities.



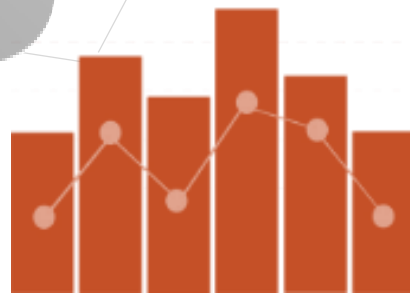
Combine data sources and find hidden patterns



Transform data with one-click fixes to ensure data trust



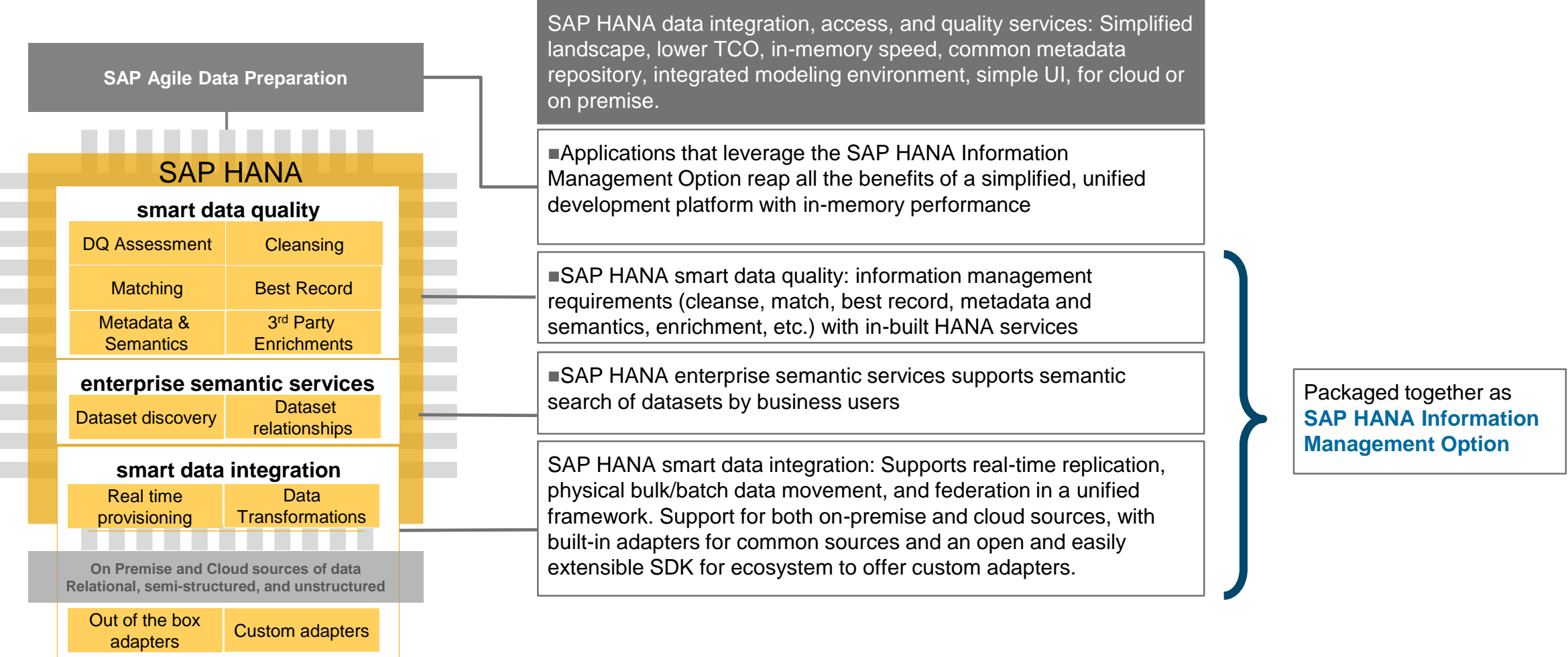
Share data discoveries and take action



Export prepared data sets to any analytical tool, including SAP Analytics & BI solutions

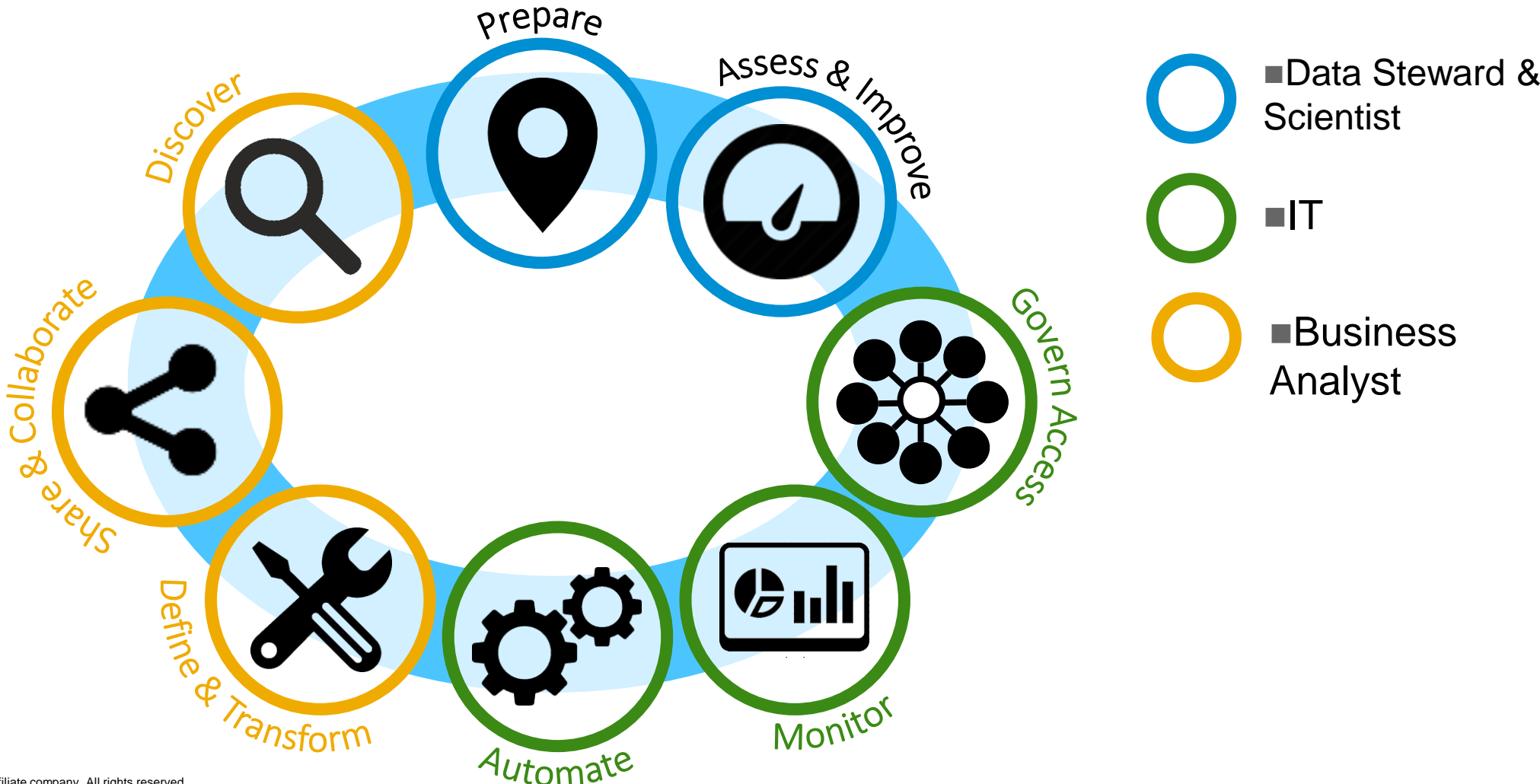
# SAP Agile Data Preparation

## Powered by SAP HANA Platform



# SAP Agile Data Preparation

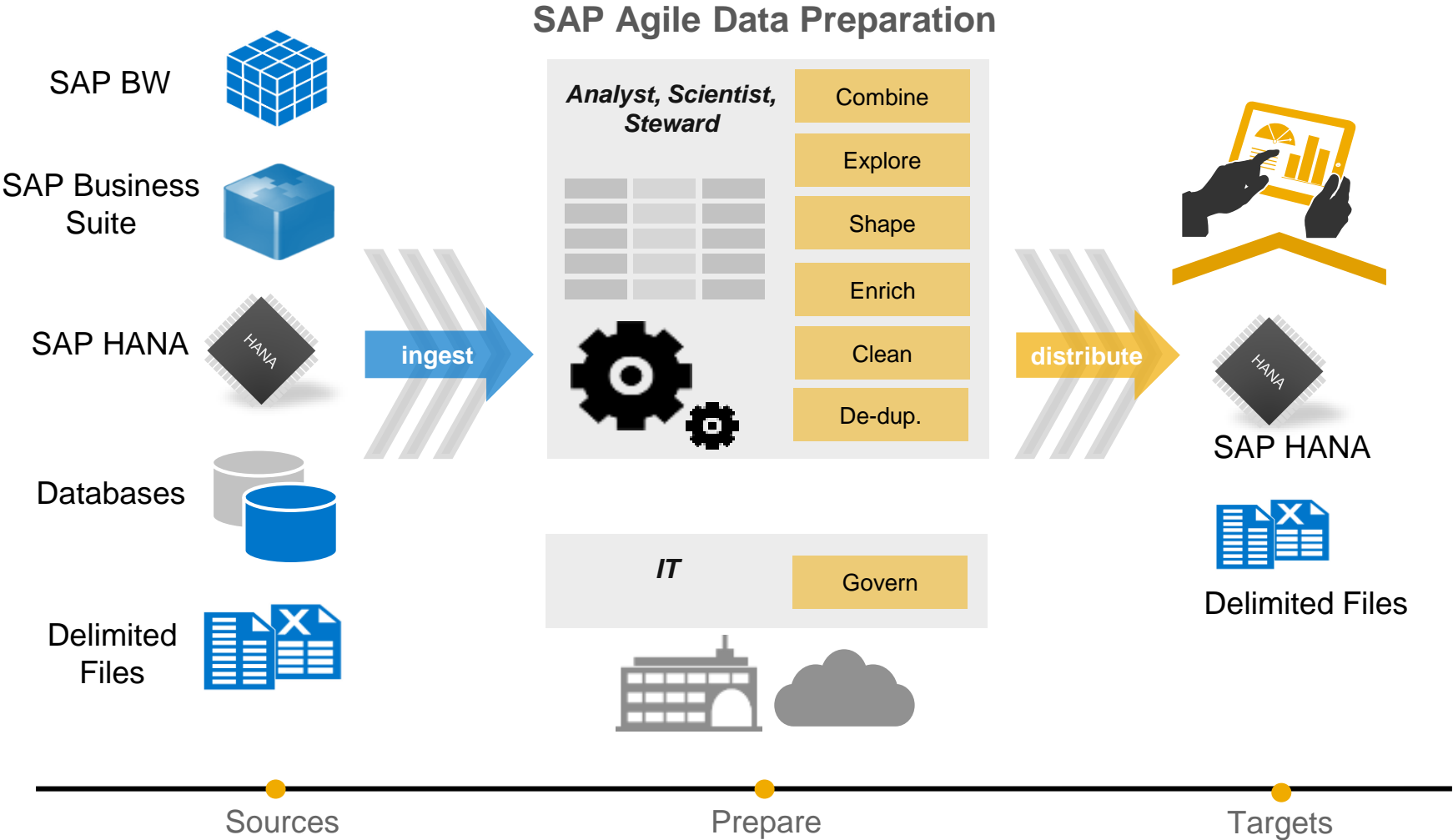
## Multiple User Roles



# SAP Agile Data Preparation

## Data Preparation Key Capabilities

- 1 Ingest data from variety of sources
- 2 Discover and profile the data
- 3 Combine, shape, enrich, or cleanse data
- 4 Output data for downstream uses
- 5 Data Steward assess and improve the data quality
- 6 IT govern, analyze and optimize user processes



# SAP Agile Data Preparation

## Data Acquisition

1 Ingest data from variety of sources

2 Discover, profile data

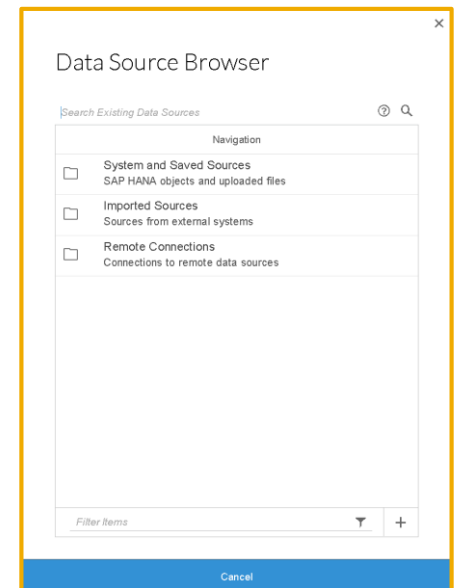
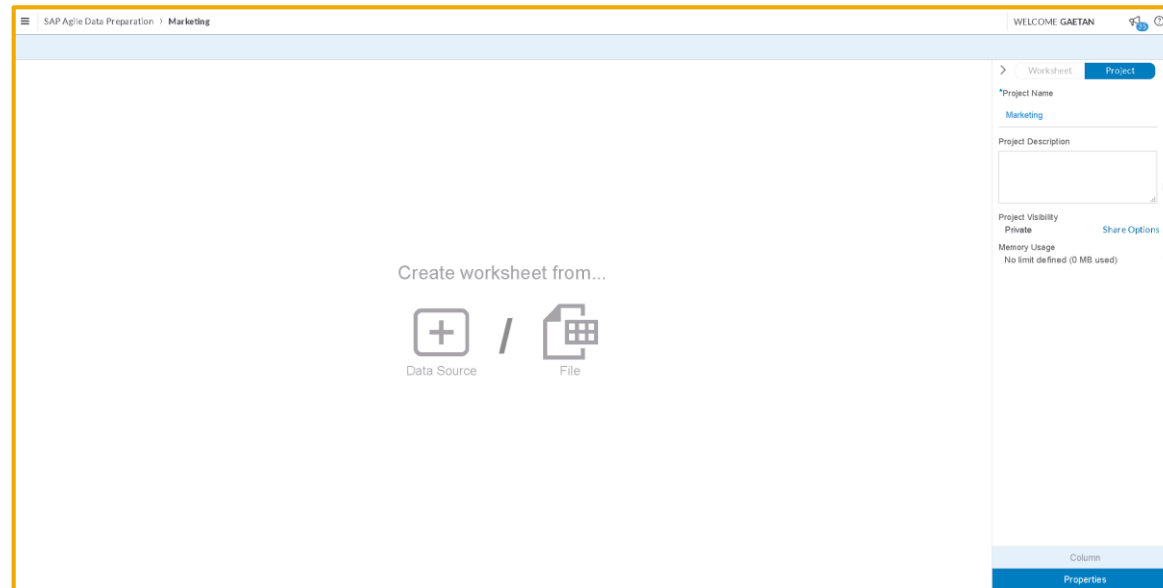
3 Combine, shape, enrich, or cleanse data

4 Output data for downstream uses

5 IT Governance team - analyze and optimize user processes

SAP Agile Data Preparation can acquire data from any remote sources that HANA Smart Data Integration can access including NoSQL databases.

It is also possible to acquire data from flat files such as csv, excel or XML files.



# SAP Agile Data Preparation

## Data Profiling

SAP Agile Data Preparation automatically detects the content type of your data (such as Addresses, Firms, Peoples, SSN...) as well as profile your data to easily understand them and highlight data quality issues you need to focus on (minimum and maximum value, % of null, blank, distinct values or patterns).

1 Ingest data from variety of sources

2 Discover, profile data

3 Combine, shape, enrich, or cleanse data

4 Output data for downstream uses

5 IT Governance team - analyze and optimize user processes

Data Quality Assessment

Please select the content types for the columns you wish to cleanse.

Column	Content Type	Minimum Value	Maximum Value	Average Length	Null %	Blank %	Zero %	Uniqueness	Distinct Value	Pattern
ID	Address Line	10145	99970	5	0.000	0.000	0.000	This column has u...	328	
Country	Country	US	US	2	0.000	0.000	0.000	This column has lo...	1	
FullName	Person	A G KEESECKER	ZORAN SINOVIC	13.600609	0.000	0.000	0.000		318	
Title	Title	ACCOUNT MANAGER	SALES TECHNICIAN	15.772727	93.293	0.000	0.000	This column is spa...	22	
Company	Firm	A H SUPPLY	ZIMMER BARTOL ASSOC	17.862962	17.683	0.000	0.000		239	
Address1	Address Line	1 MOBAY ROAD	PO BOX 955	17.304075	2.744	0.000	0.000		229	
Address2	Address Line	# 135	UNIT 70	7.714285	82.927	0.000	0.000		51	
Address3	Address Line				100.000	0.000	0.000	This column has lo...		
City	City	ABILENE	YELLOW SPRINGS	8.721875	2.439	0.000	0.000		164	
State	Region	AL	WY	1.996805	4.573	0.000	0.000		43	
Zipcode	Postcode	10014	98503	4.900392	6.707	0.000	0.000		196	
Email	Email	AKEESECKER@DOCUMENTSOLUTIONS.COM	YVETTE.THOMPSON@ATLANTAAPPRAISALS.COM	24.880434	71.951	0.000	0.000		86	
Phone	Phone	202-785-3345	WORK: 779-987-8521	12.377049	62.805	0.000	0.000		104	
Date	Date	1945-12-01	SEP. 24 2000	10.289855	78.963	0.000	0.000		68	

Save Reset Cancel

# SAP Agile Data Preparation

## Data Transformation & Data Blending

- 1 Ingest data from variety of sources
- 2 Discover, profile data
- 3 Combine, shape, enrich, cleanse data
- 4 Output data for downstream uses
- 5 IT Governance team - analyze and optimize user processes

SAP Agile Data Preparation allows to easily transform your data in a single click such as trim, replace, combine, split, data type conversions. But also allows to compute aggregates or formulas without writing a single line of code. Finally you can merge or append multiple sources in a single action.

The screenshot displays the SAP Agile Data Preparation interface. The main window is titled "Combine Worksheet" and is in the "Merge" tab. It shows a tree view with "Sales Data" and "ADP\_sample" as input sources. Below the tree, the "Define Join Conditions" section shows "Inner Join" selected. The "Select Output Columns" section shows "Sales Data" with "CUST\_ID" selected and "ADP\_sample" with "ID" selected. A "% Match" field shows "13". To the right of the dialog are two panels: "Actions" and "Change Case". The "Actions" panel lists various data manipulation actions like "Show/Hide Columns", "Add formula column", "Cleanse Worksheet", etc. The "Change Case" panel lists actions like "Trim", "Replace", "Fill", "Combine", "Merge Values", "Split", "Convert to Number", "Convert to Date", "Duplicate", "Rename", "Clear", and "Remove".



# SAP Agile Data Preparation

## Data Quality

1

Ingest data from variety of sources

2

Discover, profile data

3

Combine, shape, enrich, cleanse data

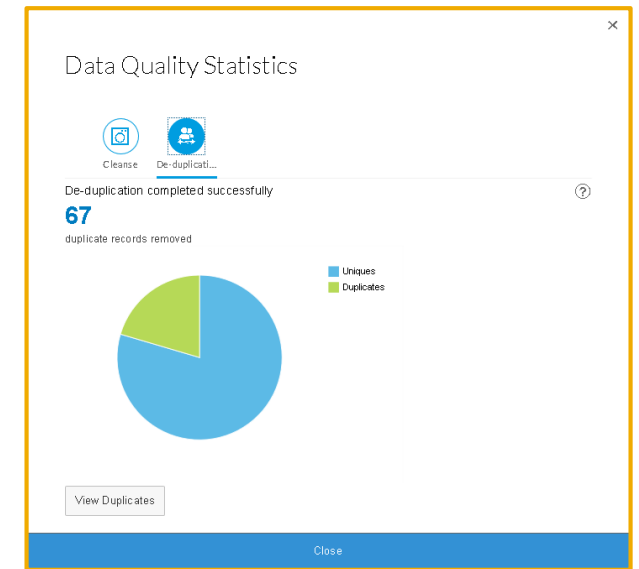
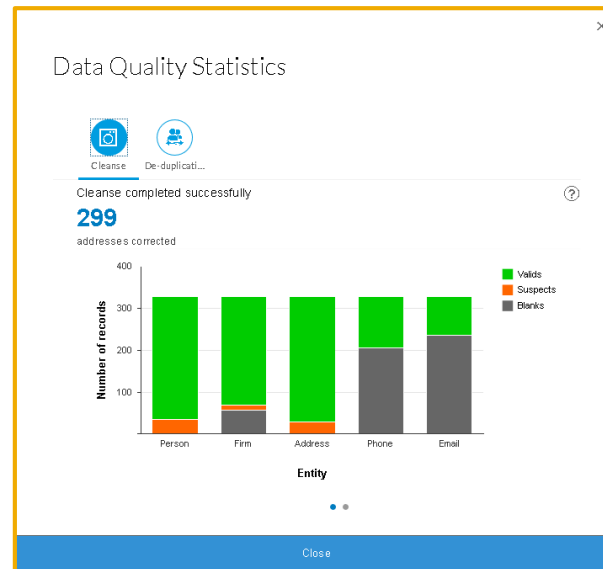
4

Output data for downstream uses

5

IT Governance team - analyze and optimize user processes

SAP Agile Data Preparation gives the ability to cleanse and remove duplicates in your data based on content types in few seconds. All cleansing and de-duplication actions comes with debriefing results that guides you on the next step that needs to be performed.



# SAP Agile Data Preparation

## Data Distribution

- 1 Ingest data from variety of sources
- 2 Discover, profile data
- 3 Combine, shape, enrich, cleanse data
- 4 Output data for downstream uses
- 5 IT Governance team - analyze and optimize user processes

SAP Agile Data Preparation provide the ability to export your results in HANA (HANA tables or HANA Calculation Views) as well as flat files (csv files or Excel files). Many functionalities also provide the capability to export the results for further analysis or action such as de-duplicate records, rule based failing records...

Select Target

Export all data

CSV

Excel

HANA Table

HANA Calculation View

Export Cancel

Removed Duplicates

Show only low confidence groups

ID	Country	FullName	Title	Company	Address1	Address2	Address3	City	State
61698	USA	GLORIA VANARSDALE		WATERMARK OF LE...	1 WILLIAMS CTR			TULSA	OK
35798	USA	GLORIA VAN ARSDALE		CROSSROADS CON...	1 WILLIAMS CTR	RM 2		TULSA	OK
51059	USA	ELIZABETH MURPHY		GARDNER HARDWA...	4501 INTELCO LOO...			OLYMPIA	WA
38694	USA	BETSY MURPHY		ALTER TRADING CO...	4501 INTELCO LOO...			OLYMPIA	WA
20091	USA	VIRGINIA S DAILY		KELLOGG SCHOOL ...	208 SOUTH AKARD ...	# 150		DALLAS	TX
34932	USA	GINNY DAILY		UNIVERSITY OF TEX...	208 SOUTH AKARD ...	SUITE 150		DALLAS	TX
91279	USA	A G KEESECKER		DOCUMENT SOLUTI...	6625 THE CORNER...	APT 260		ATLANTA	GA
61569	USA	ANDREW KEESECKER		WELLS FARGO	6625 THE CORNER...	# 260		NORCROSS	
10944	USA	ALFRED REGINALD ...		NEWPORT PACIFIC ...	17944 NE 65TH STR...			SEATTLE	WA
99876	USA	REGGIE JOHNSON		WELCOM NETWORK ...	17944 NORTH EAST...			SEATTLE	WA
45731	USA	SHERYL WOOD			1899 PALOMINO AVE			CASPER	WY
10757	USA	SHERYL THOMPSON...			1899 PALOMINO AVE			CASPER	WY

This preview is limited to 1000 entries. Use the export button to access all data.

Close

# SAP Agile Data Preparation

## IT Governance & Monitoring

1 Ingest data from variety of sources

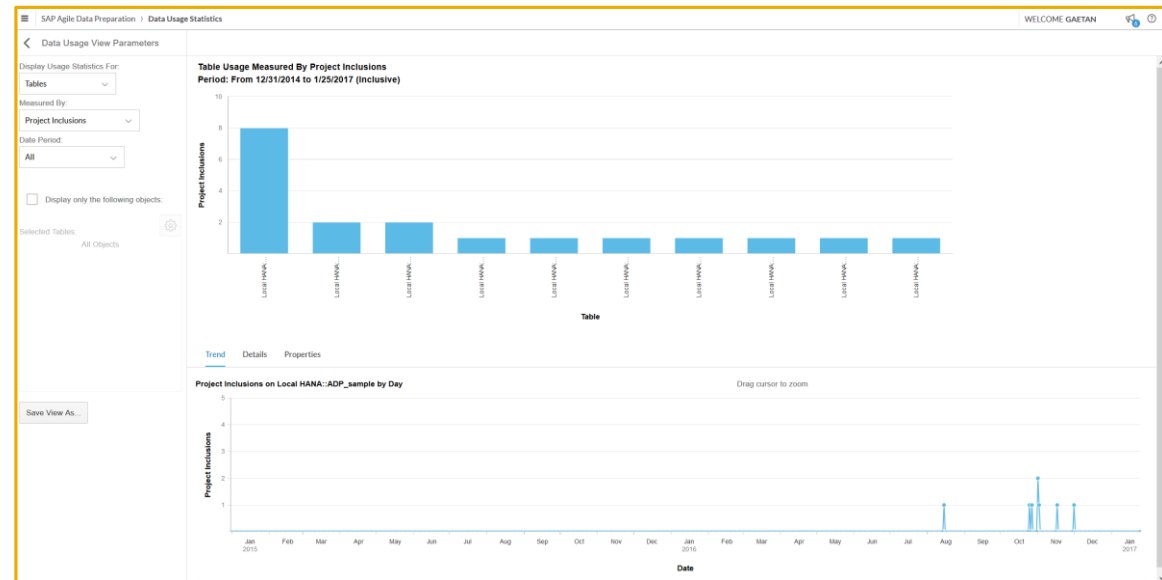
2 Discover, profile data

3 Combine, shape, enrich, cleanse data

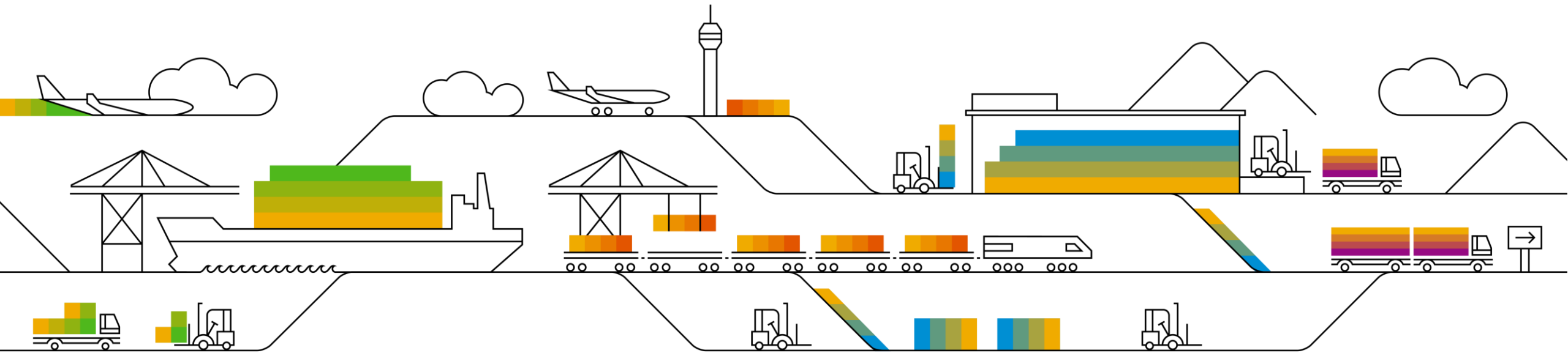
4 Output data for downstream uses

5 IT Governance team - analyze and optimize user processes

SAP Agile Data Preparation allows ITs to govern the users accesses and monitor their activities to optimize the HANA platform usage by allowing to control users, roles and data sources access and by allowing to monitor dataset usages, the memory consumption and scheduling activity.



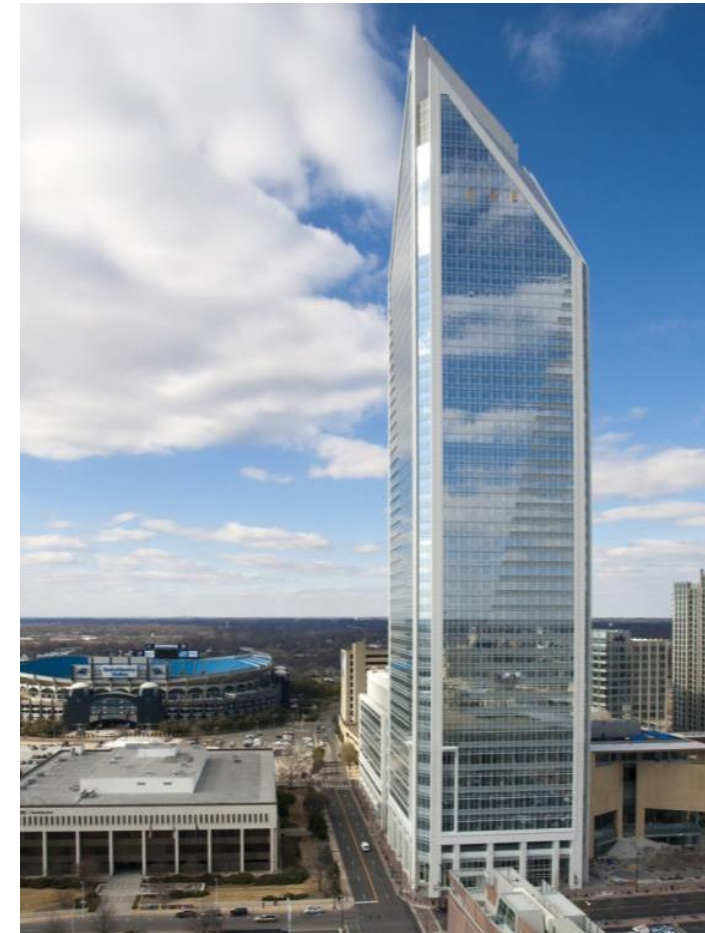
# Duke Energy Overview



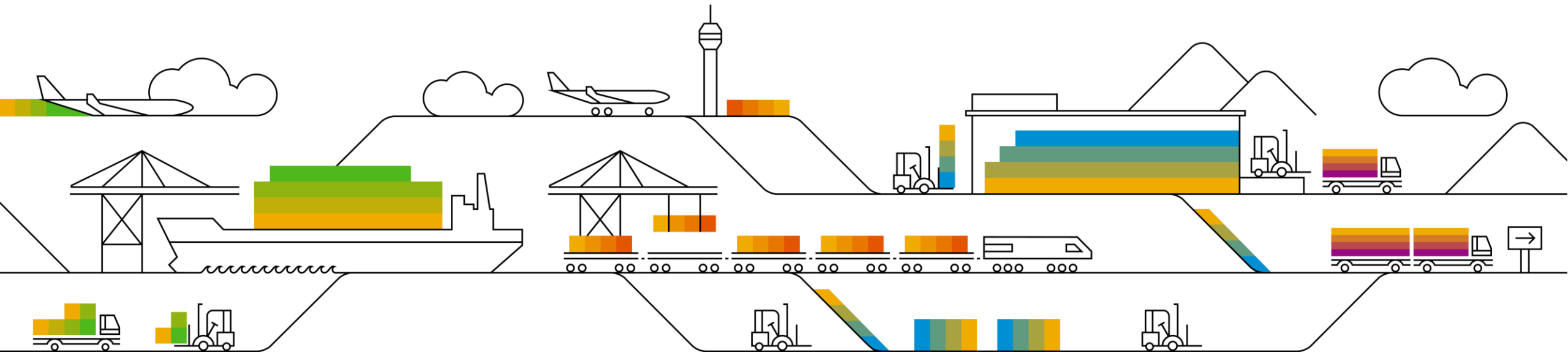
# About Duke Energy



- 150+ years of service
- 7.6 million electric customers in six states: North Carolina, South Carolina, Florida, Indiana, Ohio, Kentucky
  - Electric service area's estimated population is 24 million people
- 1.6 million natural gas customers
- Fortune 125 company
- Stock has paid cash dividends for 92 consecutive years
- Listed on:
  - Fortune's 2018 "World's Most Admired Companies" list
  - Forbes' 2018 "America's Best Employers" list
  - Dow Jones Sustainability Index



# Customer Connect



# Customer Connect will Transform the Customer Experience



Customer Connect is *foundational to transforming the customer experience:*

- Customer engagement platform that ensures we deliver universal, simple and consistent experiences across channels based on retail industry models
- Core meter-to-cash platform that evolves with the market
- Integrated operational/analytics platform that enables us to personalize experiences and serve our customers as individuals

*Customer Connect will help enable our evolution into a more customer-centric organization.*

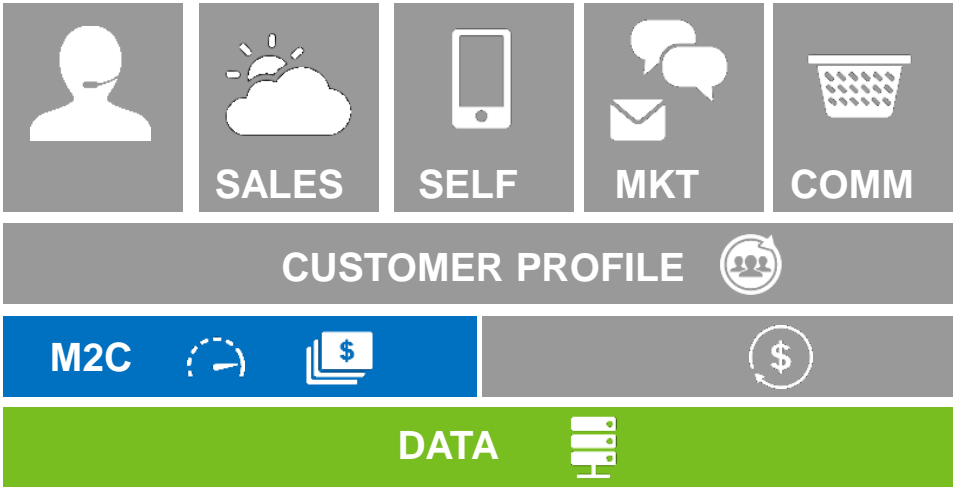





★ **Release 1**  
★ **Early Analytics**  
★ **Release 2 & 3**  
★ **Customer Engagement**  
★ **Release 4**  
★ **Universal Bill**

★ **Release 5-8**  
★ **Core Solution**



# Leveraging the power and relevancy of SAP



<p>Customer Engagement</p> 	<ul style="list-style-type: none"> <li>• Omni-channel B2B / B2C customer engagement platform</li> <li>• Capabilities – Marketing, Sales, Service, Commerce, Billing</li> </ul>
<p>Core Meter to Cash Solution</p> 	<ul style="list-style-type: none"> <li>• 785 utilities worldwide, nearly 4x SAP's nearest competitor</li> <li>• 12 utilities with customers &gt; Duke Energy, largest = 30 million</li> <li>• #1 Customer Information System ("CIS") ranking by Gartner / Utilipoint over last 10 years</li> </ul>
<p>Integrated Analytics</p> 	<ul style="list-style-type: none"> <li>• State-of-the-art in-memory transactional / analytics database</li> <li>• Dramatically improved processing speed of large data sets</li> <li>• Real-time, predictive analytics tightly integrated with solution</li> </ul>

# Conversion and Data Readiness Team

## 1. Conversion Objective

- At each Release's Go-Live Weekend, run all the conversion programs to bring the required data from legacy systems into SAP Solution. We must be very confident that the data brought over is **complete, accurate, usable**, and that our programs **finish on time**.

## 2. Controls Objective

- Demonstrate the success of conversion by **automating a set of completeness and accuracy controls** to measure if every object is converted with the right record count and every critical field is converted with the right value.

## 3. Data Cleansing Objective

- **Manage the cleansing lifecycle** of identifying, profiling, cleansing, and monitoring any data quality issues that impact conversion, solution functionality, or business outcomes.

## 4. Manual Conversion Objective

- Ensure conversion goes smoothly by tracking and **correcting any data exceptions post-conversion**. Manage the inventory of required data entry or correction tasks and coordinate the execution of every task in that inventory

## 5. Support

- **Provide other teams** with what they need to succeed, including:
  - Convert data into QA so Test Teams can run their test scripts
  - Convert data into TRN so OCM can run their training exercises
  - Generate cross reference reports so interfacing apps know how to trace accounts

## Artifacts and Deliverables

- Conversion Strategy
- Conversion RICEFW Inventory
- Mock Conversion Plan
- Key Conversion Scope Decisions
- Control Specifications
- Go-Live Conversion Controls
- Development Standards
- Designs & Data Maps
- Conversion Workplan
- Conversion Resource Plan
- Data Protection Approach for Conversion Process
- Cleansing Strategy
- Data Issue Inventory
- Data Cleansing One-Pagers
- Data Cleansing Specs
- Manual Conversion Strategy
- Manual Conversion Inventory (with controls and timings)
- Manual Conversion Specs/Guides
- Manual Conversion Workplan
- Manual Conversion Resource Plan
- Data Cleansing dev standards
- Cleansing Workplan
- Cleansing Resource Plan

# Data Cleansing Roadmap

## De-duplication

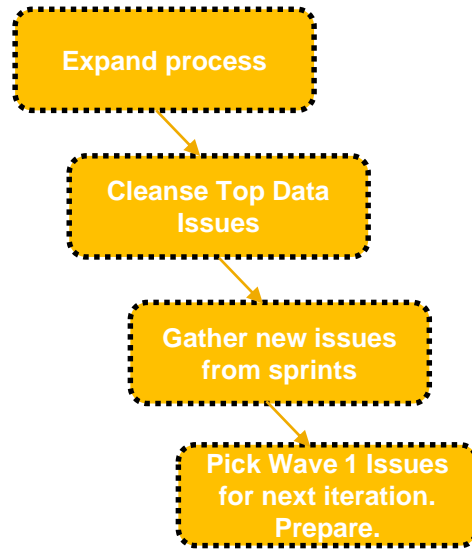
Apr - Jun 2018



Set up initial people / process / technology to tackle Customer De-Duplication.

## First 10 Data Issues

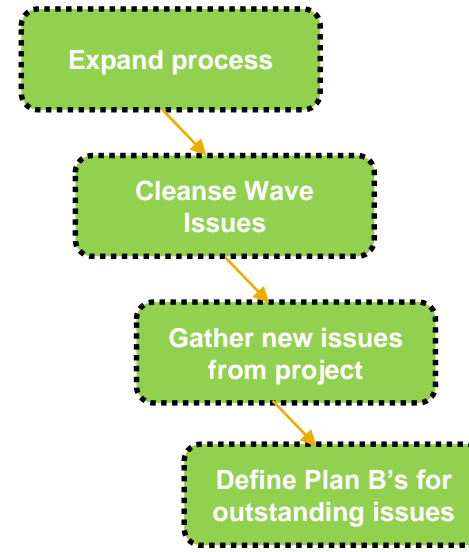
Jul 2018 - Mar 2019



Expand to tackle First Data Issues. Start production cleansing.

## 3 Waves of Issues

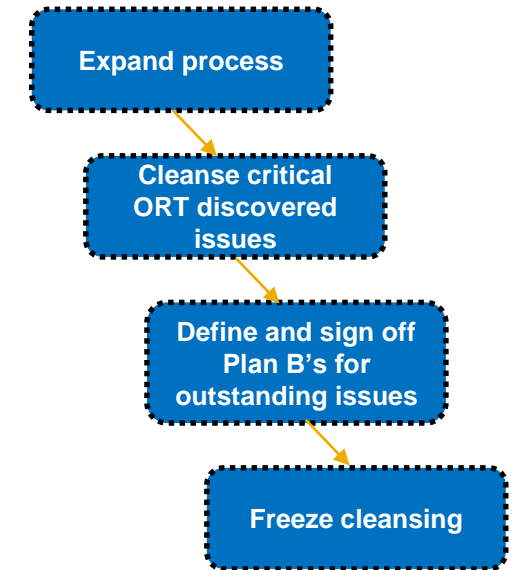
Apr 2019 - Sep 2020



Iterate to tackle inventory of project issues. Aim to optimize lifecycle processes.

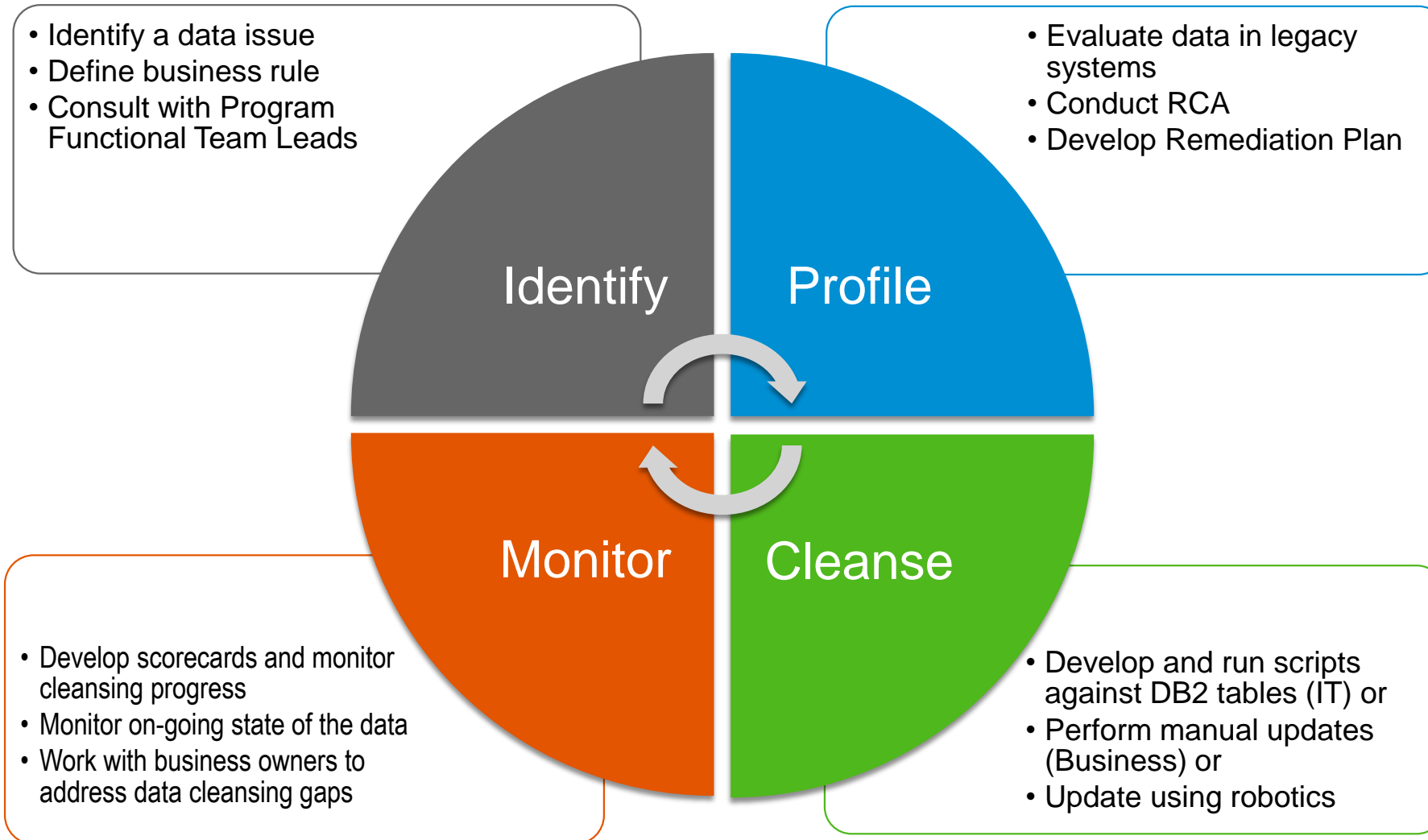
## Plan B and Freeze

Oct 2020 - Jan 2021

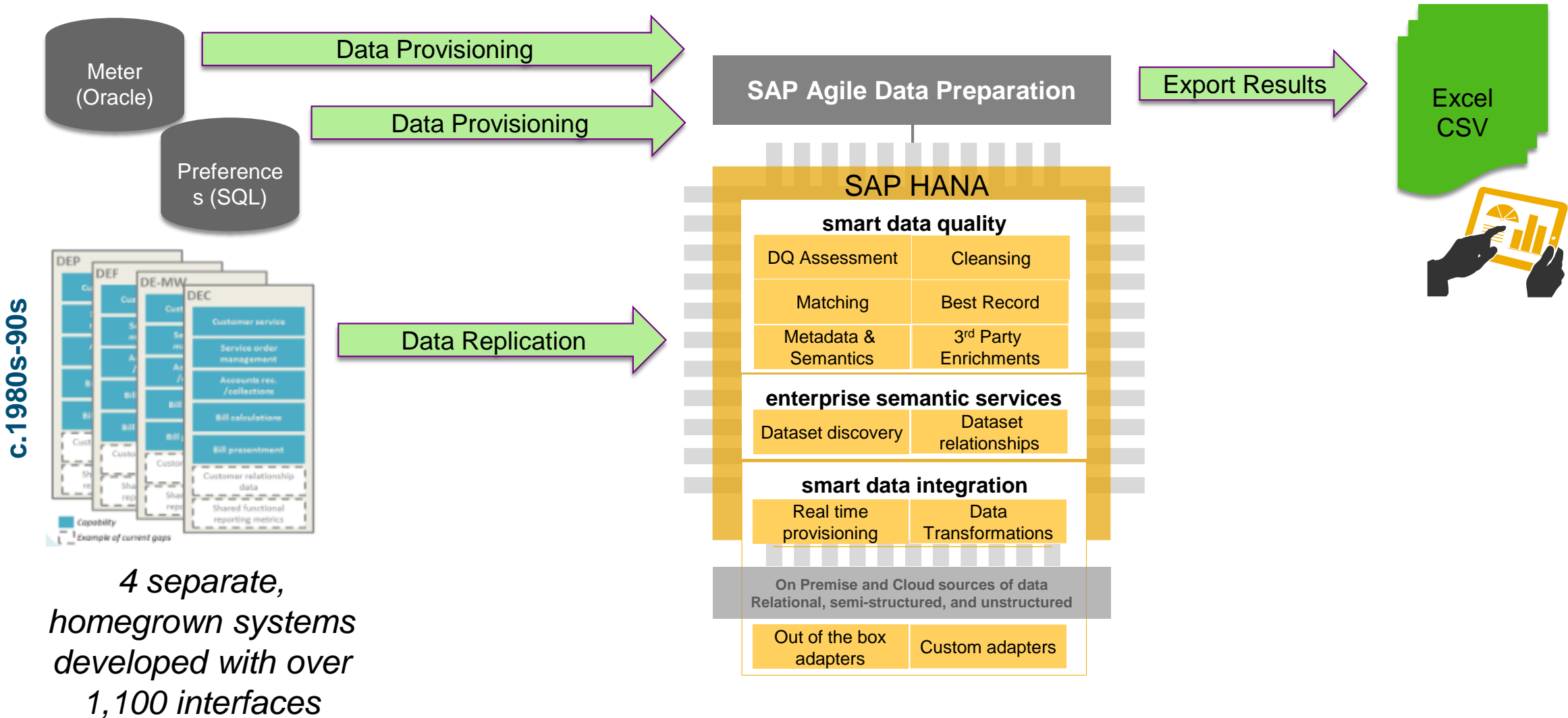


Cleanse what's critical for ORT Define and sign off Plan B's for other outstanding issues. Freeze cleansing.

# Data Cleansing Lifecycle



# Architecture



# Transforming Customer Data with SAP Agile Data Preparation

*SAP Agile Data Preparation is giving Duke Energy the ability to scale on data readiness by empowering the business to address the right data issues at the right time*

## Challenges and Opportunities

- Four separate CIS databases and other disparate systems (i.e. Meter Tracking, Customer Choice)
- Millions of records to be analyzed and evaluated
- Variety of platforms (i.e. DB2, Oracle, Hana)

## Advantages to Agile Data Preparation

- Full integration with SAP HANA
- User-friendly, self-service data management, reducing reliance on IT and streamlining data analysis
- Scalable self-service data preparation, and discovery
- Quickly identify patterns and outliers
- Create and re-use rules for analysis and scorecard development
- Easily export results in user-friendly, consumable formats

**“SAP Agile Data Preparation is improving our data governance challenges. Our data is cleaner and analysts can collaborate more easily, which improves business outcomes”** *Ginny Segbers, Manager, Conversion & Data Readiness, Duke Energy*

**~634,000**

'Business Partner records targeted for consolidation

**~4M**

Items identified for cleansing in first 6 months

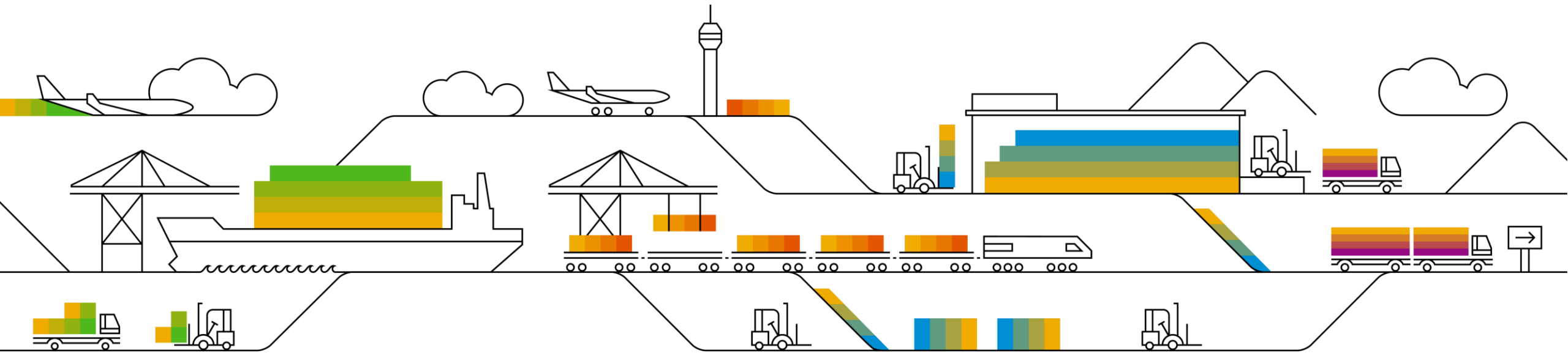
**210%**

Cleansing analysis to plan

**1.5 months**

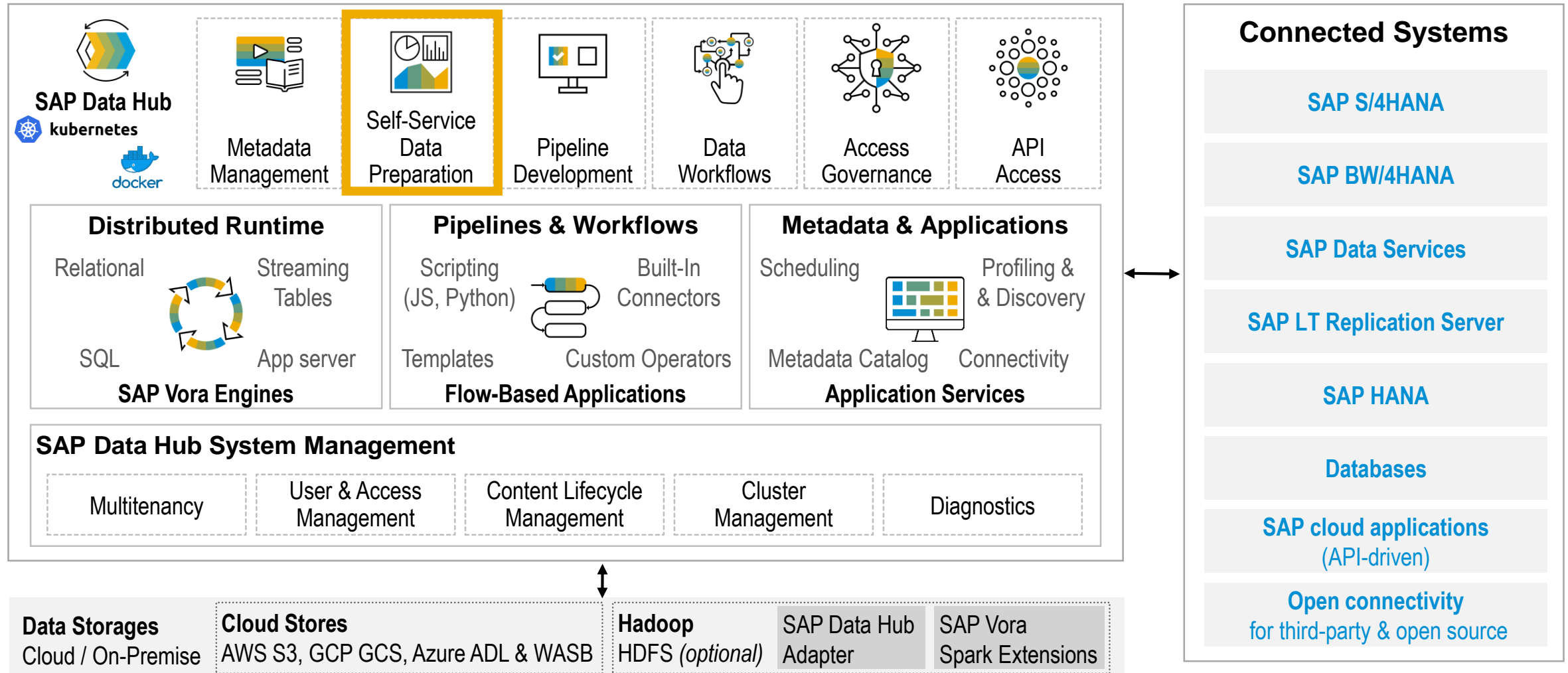
To get SAP Agile Data Preparation up and running – from development to production, deployment, and training

# Next Steps



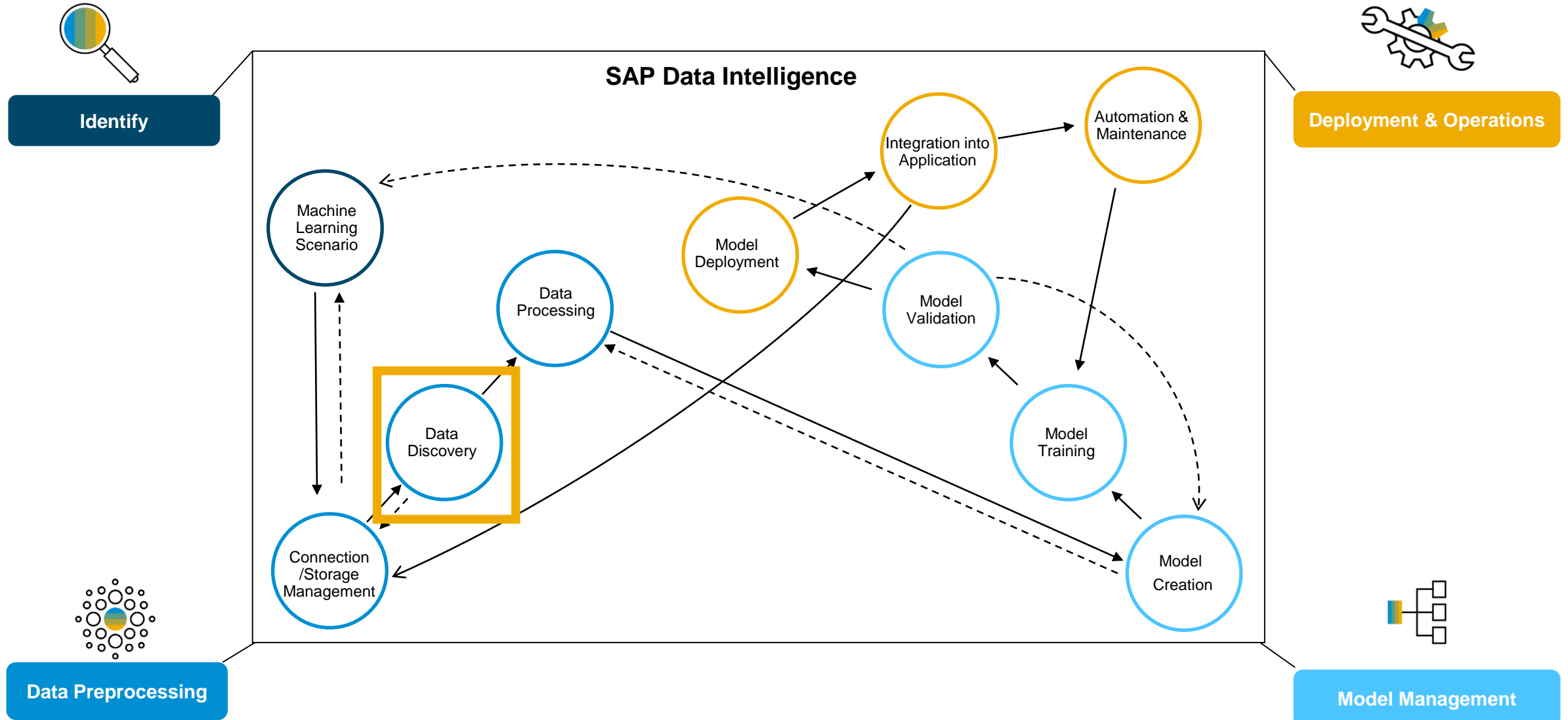


# Embedded Self Service Data Preparation within SAP Data Hub



# SAP Data Intelligence

One tool to set up, manage and automate a continuous lifecycle – using SAP Data Hub Data Preparation capabilities



# Take the Session Survey.

We want to hear from you! Be sure to complete the session evaluation on the SAPPHIRE NOW and ASUG Annual Conference mobile app.



# Presentation Materials

Access the slides from 2019 ASUG Annual Conference here:

<http://info.asug.com/2019-ac-slides>

# Q&A

For questions after this session, contact us at

[Ginny.Segbers@duke-energy.com](mailto:Ginny.Segbers@duke-energy.com)

[Michael.Eacrett@sap.com](mailto:Michael.Eacrett@sap.com)

# Let's Be Social.

Stay connected. Share your SAP experiences anytime, anywhere.

Join the ASUG conversation on social media: **@ASUG365 #ASUG**

