

データがもたらす、破壊的パワーをその手に。
デジタル変革時代のハイブリッドデータ
アーキテクチャーとは

2018年05月31日

インフォマティカ・ジャパン株式会社

データ市場における世代の変遷

Data 1.0

1960-2000

個別の業務アプリケーションでの
データ連携・統合

Data 2.0

2000-2015

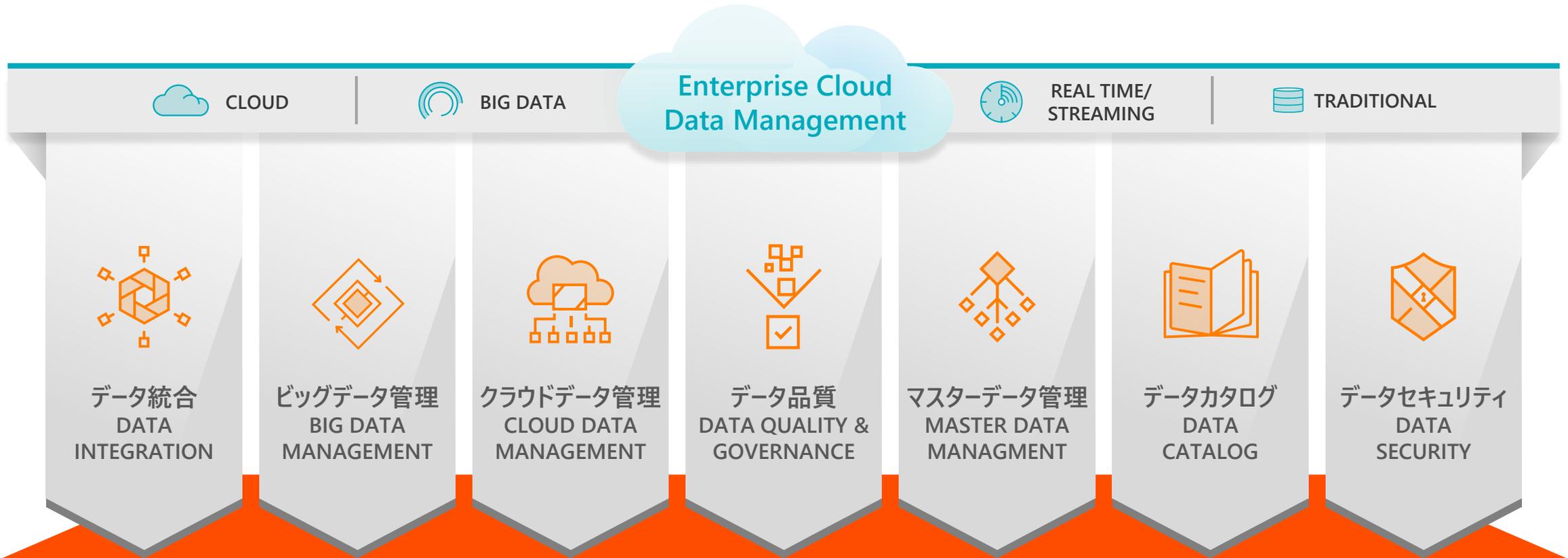
企業全体の経営資産としての
データマネジメント

Data 3.0

2015-

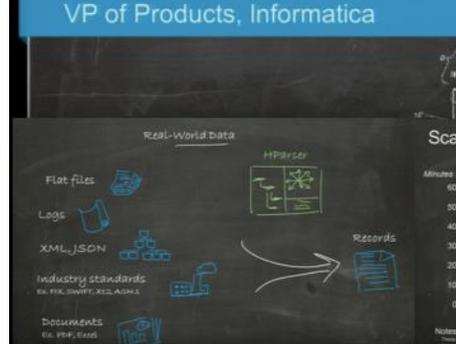
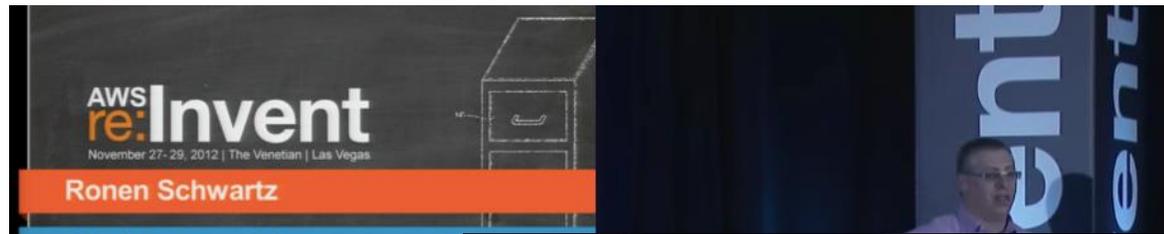
ビジネスモデルを変革する
デジタルトランスフォーメーション
のためのデータ活用

Informatica 製品ポートフォリオ



Informatica と AWS

2011年からのパートナーシップ



Informatica Products for/on AWS

Amazon S3	Amazon Redshift	Amazon Aurora	Amazon Dynamo DB	Amazon EMR
Amazon RDS MySQL	Amazon RDS Oracle	Amazon RDS SQL Server	Amazon QuickSight	

AWSとの
接続コネクタ



PowerCenter	Informatica Intelligent Cloud Services	Informatica Data Quality	Master Data Management
Enterprise Data Catalog	Big Data Management	Enterprise Data Lake	

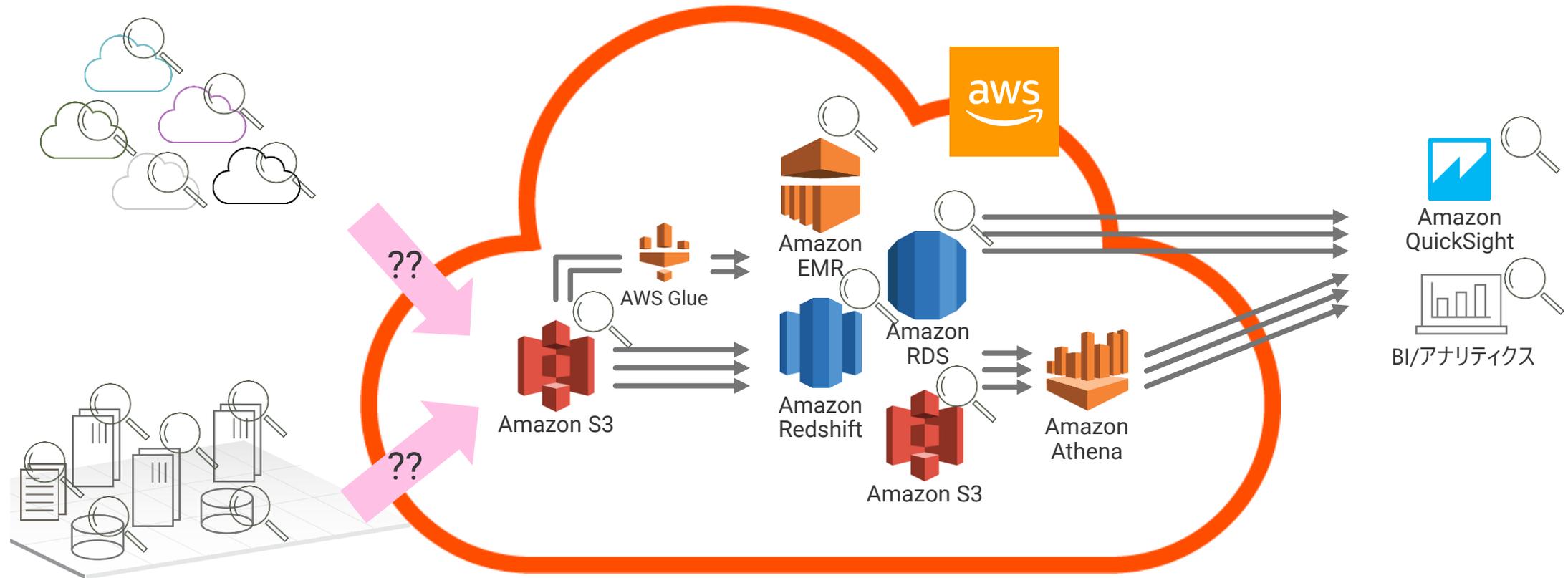
Amazon EC2上での
稼働サポート製品



PowerCenter	Informatica Intelligent Cloud Services	Informatica Data Quality	Big Data Management	Enterprise Data Catalog
-------------	--	--------------------------	---------------------	-------------------------

AWS
Marketplace

AWSにおけるInformaticaの活用

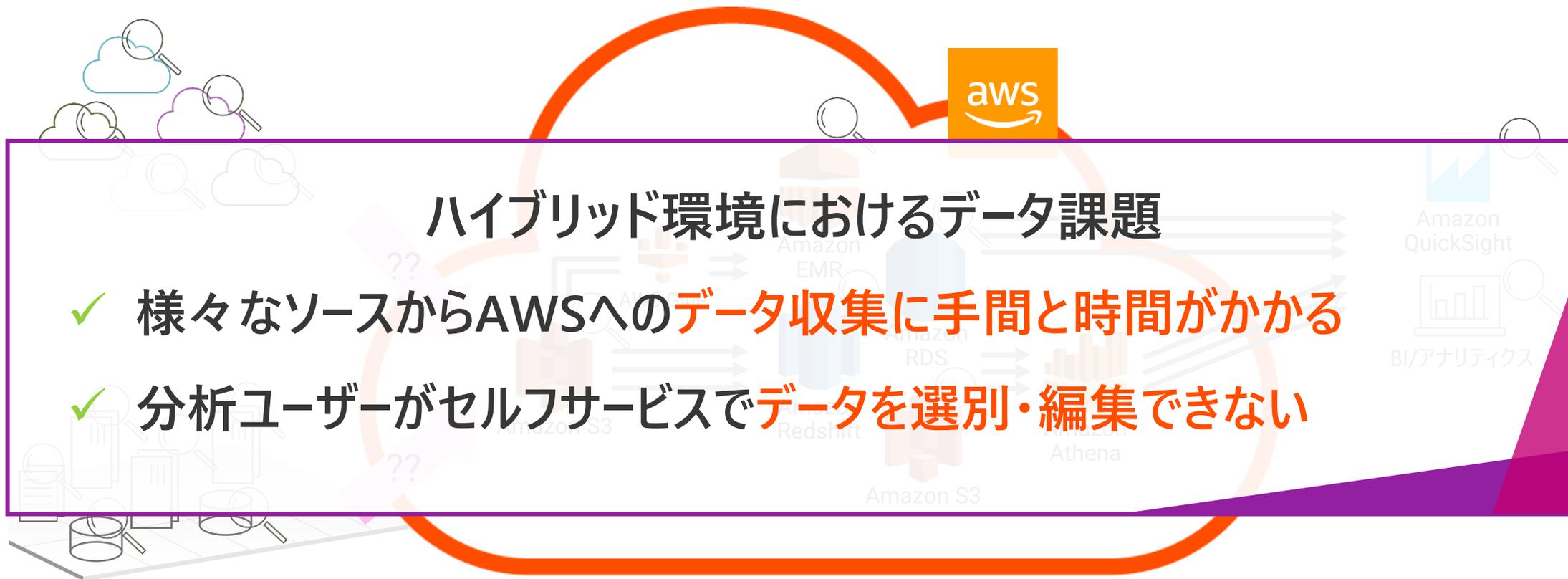


データ移行・連携処理の複雑化

求めているデータが見つからず
データの利活用が捗らない



AWSにおけるInformaticaの活用



IT/開発者

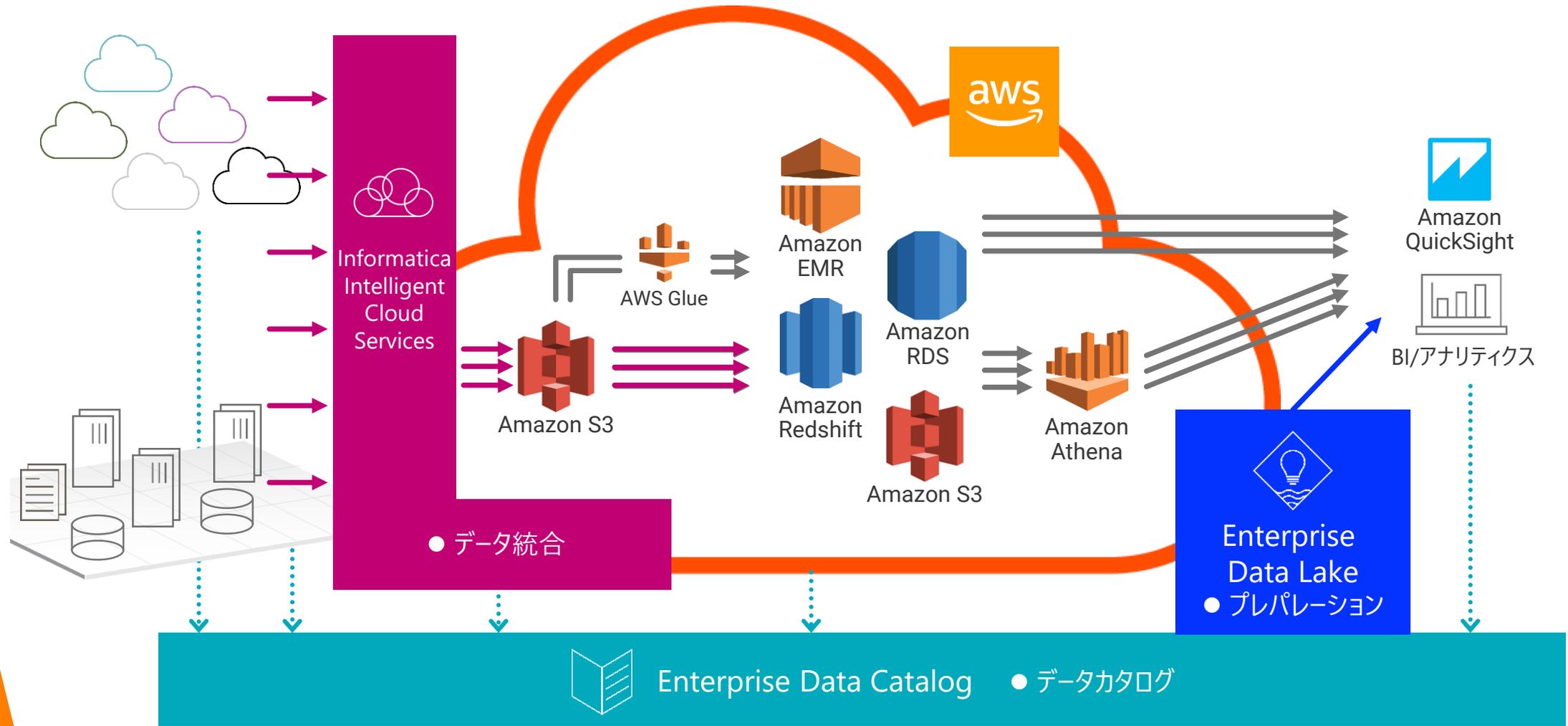
データ移行・連携処理の複雑化

求めているデータが見つからず
データの利活用が捗らない

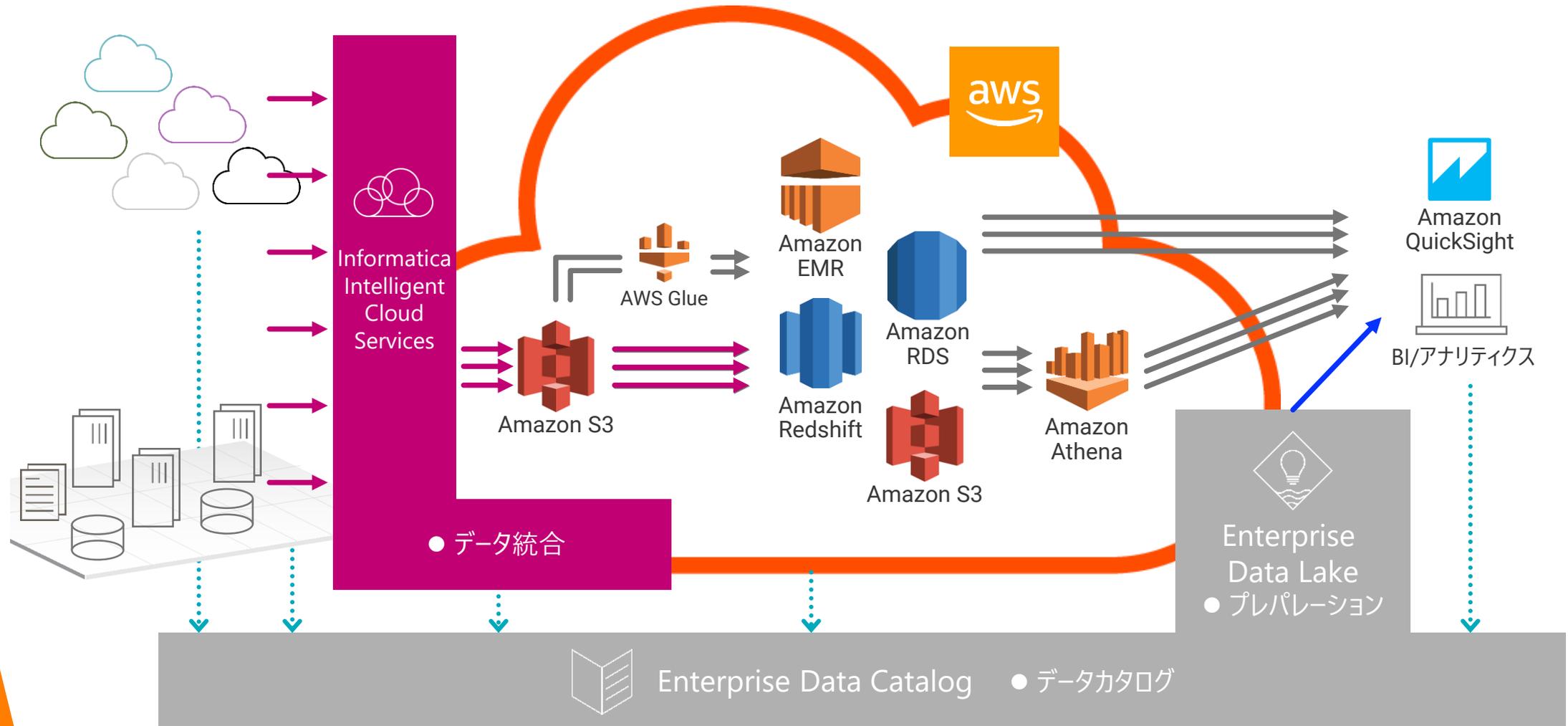


分析ユーザー

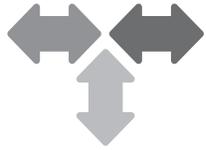
AWSにおけるInformaticaの活用



AWSにおけるInformaticaの活用



データ統合: Informatica Intelligent Cloud Services (IICS)



データ統合



プロセス統合



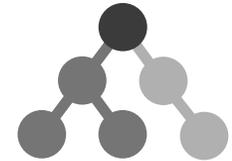
ハブ型データ統合



EDI/企業間連携



テストデータ保護



マスター統合

Informatica Intelligent Cloud Services

Cloud Data
Integration

Cloud
Application
Integration

Cloud
Integration
Hub

Cloud B2B
Gateway

Test Data
Management

Customer 360

データ統合: Informatica Intelligent Cloud Services (IICS)

データ同期/複製処理のウィザード画面イメージ

The screenshot displays the 'ds_Salesforce_Redshiftの編集' (Edit ds_Salesforce_Redshift) wizard. The progress bar at the top indicates the current step is 'ターゲット' (Target), with other steps being '定義' (Definition), 'ソース' (Source), 'データフィルタ' (Data Filter), 'フィールドマッピング' (Field Mapping), and 'スケジュール' (Schedule).

ターゲットの詳細 (Target Details):

- 接続: Amazon Redshift
- ターゲットオブジェクト: sfdc_opportunity
- 子オブジェクト: None Found
- ラベルの代わりに技術名を表示 (checked)
- ターゲットフィールドをアルファベット順に表示 (unchecked)

フィールドマッピング (Field Mapping):

ソース: Opportunity	ターゲット: sfdc_opportunity
id	Id
isdeleted	IsDeleted
accountid	AccountId
isprivate	IsPrivate
name	IIF(IsDeleted=true, CONCAT('【削除済】_',Name), Name)
description	Description
stagename	StageName
amount	Amount
probability	Probability
expectedrevenue	ExpectedRevenue
totalopportunityquantity	TotalOpportunityQuantity
closedate	CloseDate
type	Type
	NextStep
	LeadSource
	IsClosed
	...

6ステップのウィザード形式の設定でデータ同期・複製処理を実装

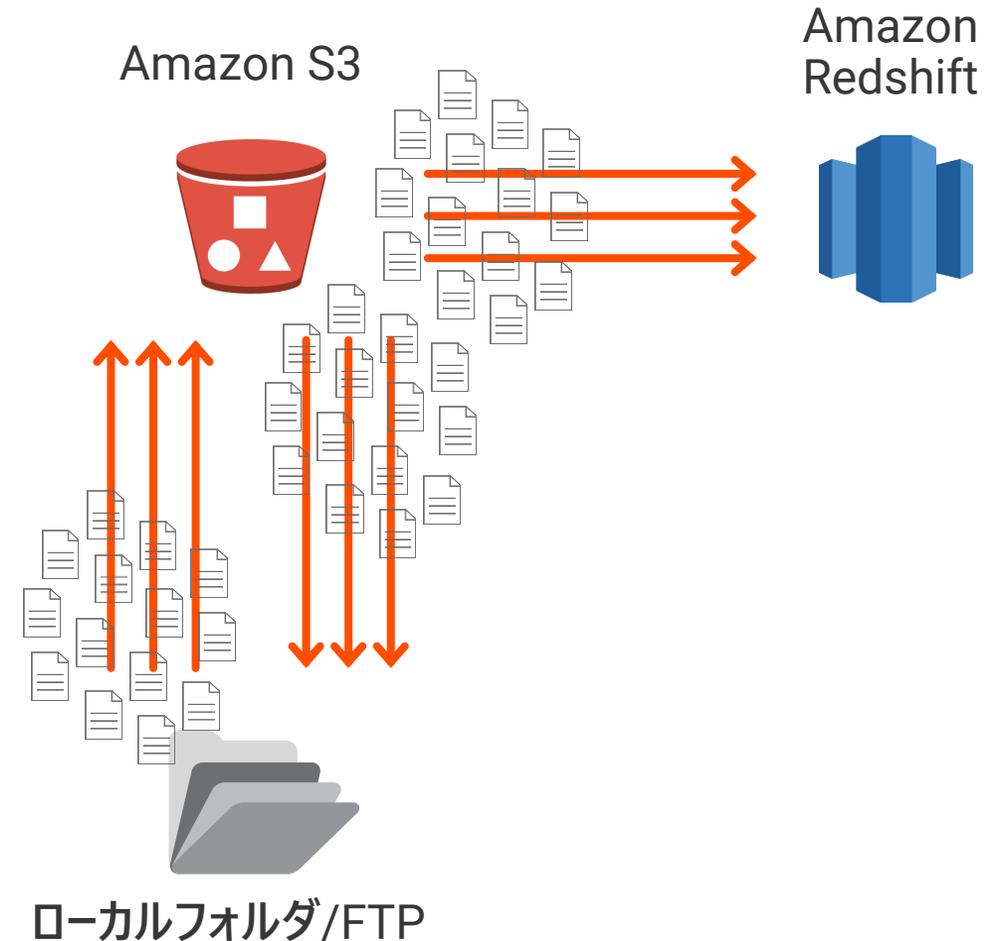
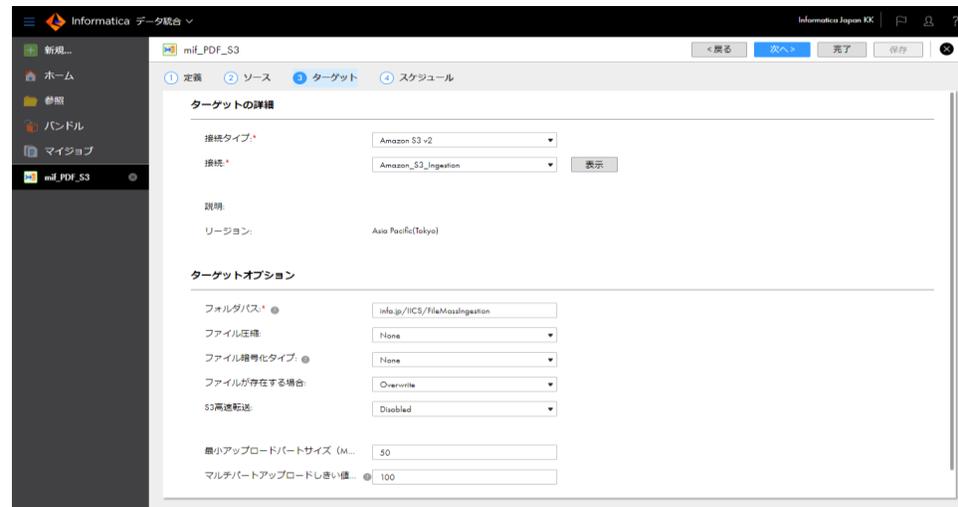
項目名が同じものは自動で項目マッピング

関数を利用した値の加工も可能

データ統合: Informatica Intelligent Cloud Services (IICS)

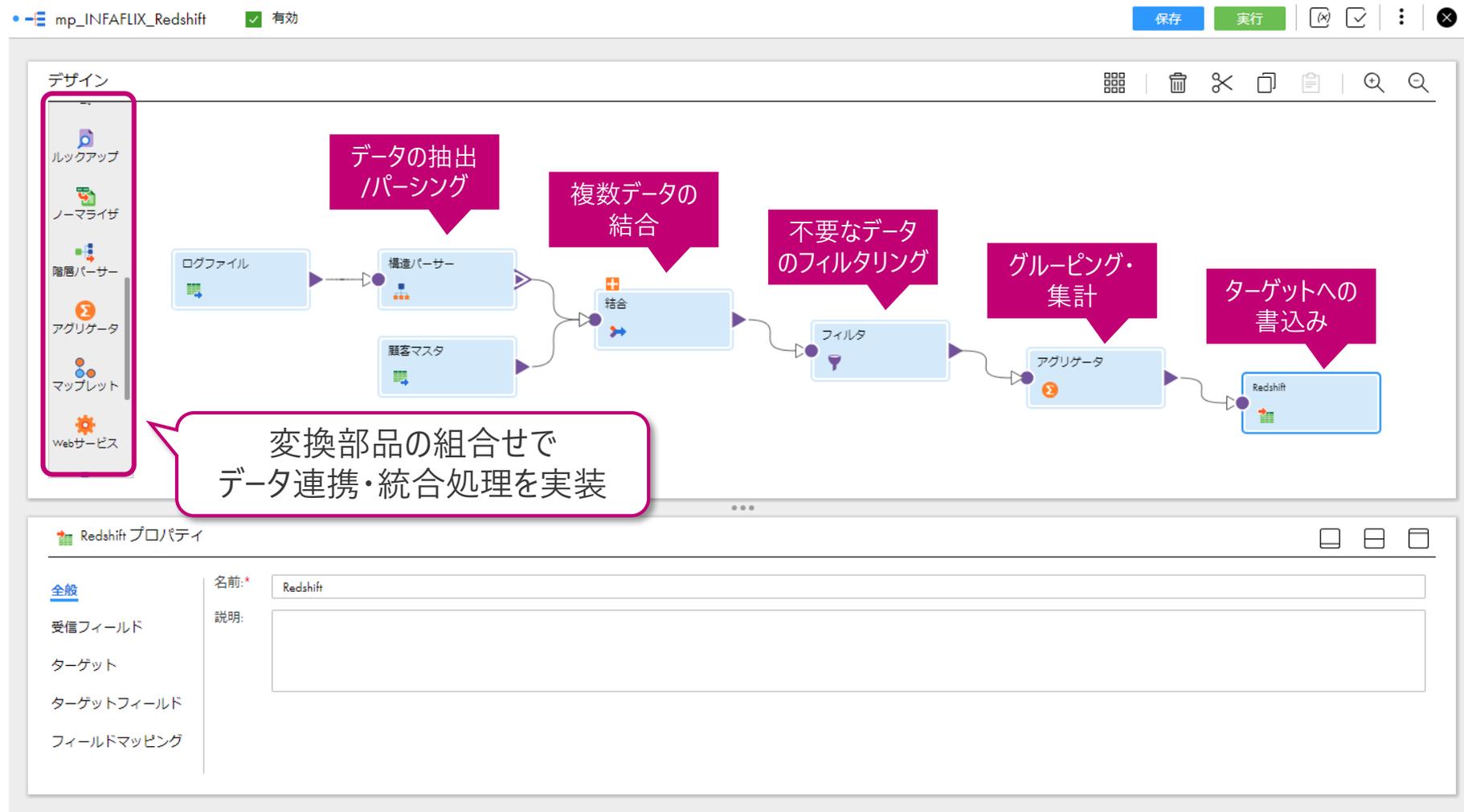
大量ファイルの一括転送: Mass Ingestion

- ウィザード形式の設定で、大量ファイルの転送を単一処理で簡単に実現
- ローカルフォルダ、FTP、Amazon S3やAmazon Redshift間の一括転送をサポート
- Amazon S3 Transfer Accelerationを利用した高速転送に対応



データ統合: Informatica Intelligent Cloud Services (IICS)

データ統合処理の開発画面イメージ





デモ

データ統合: Informatica Intelligent Cloud Services (IICS)

あらゆるデータへのアクセスを実現する接続コネクタ

Data Stores DBs, DWs, Big Data, Cloud



Cloud Apps (SaaS)



Social Apps



Middleware and Tech



Analytics



Enterprise Systems

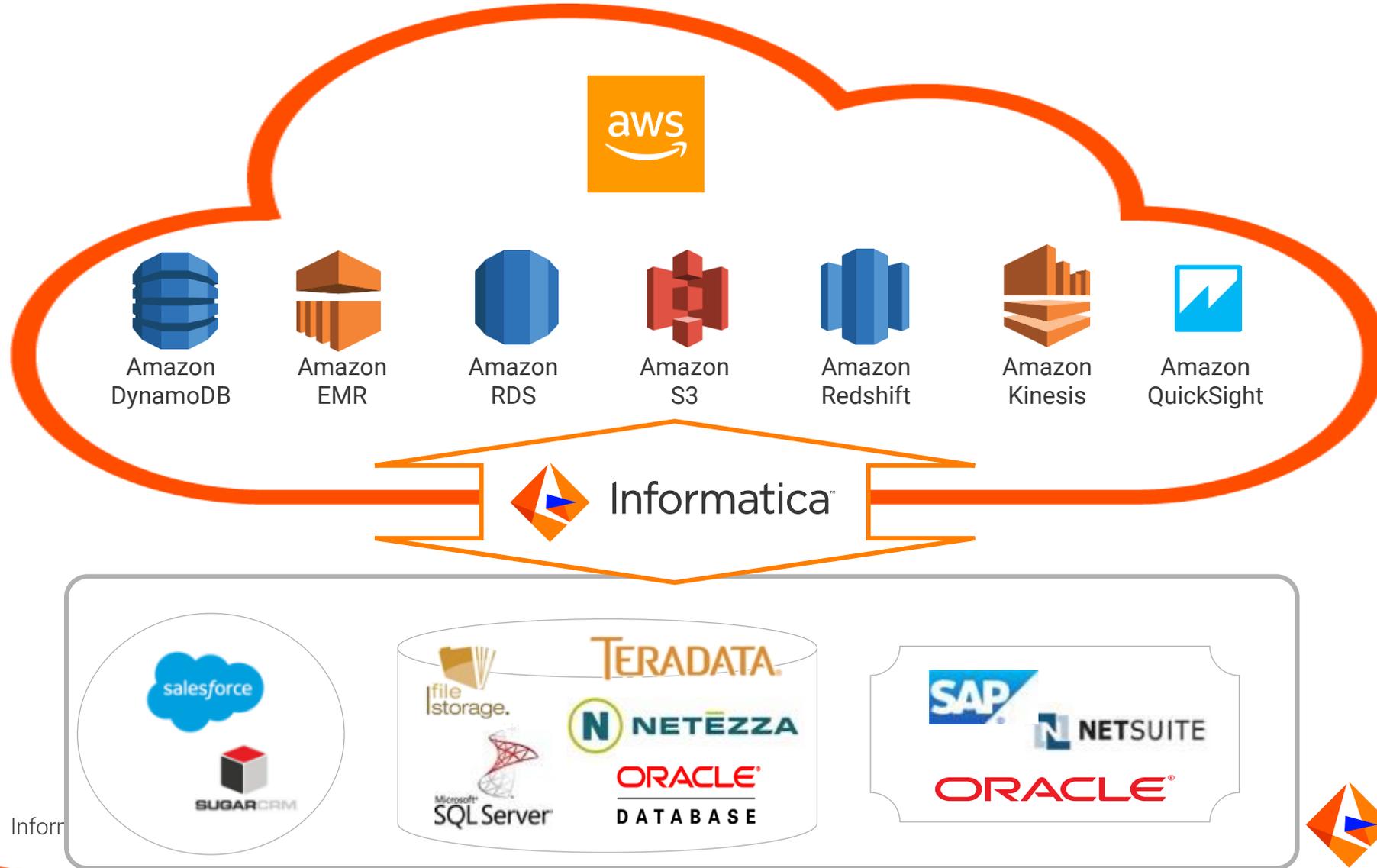


B2B



データ統合: Informatica Intelligent Cloud Services (IICS)

各種AWSサービスとのコネクティビティ



某大手通信メディアサービス業様の取り組み事例

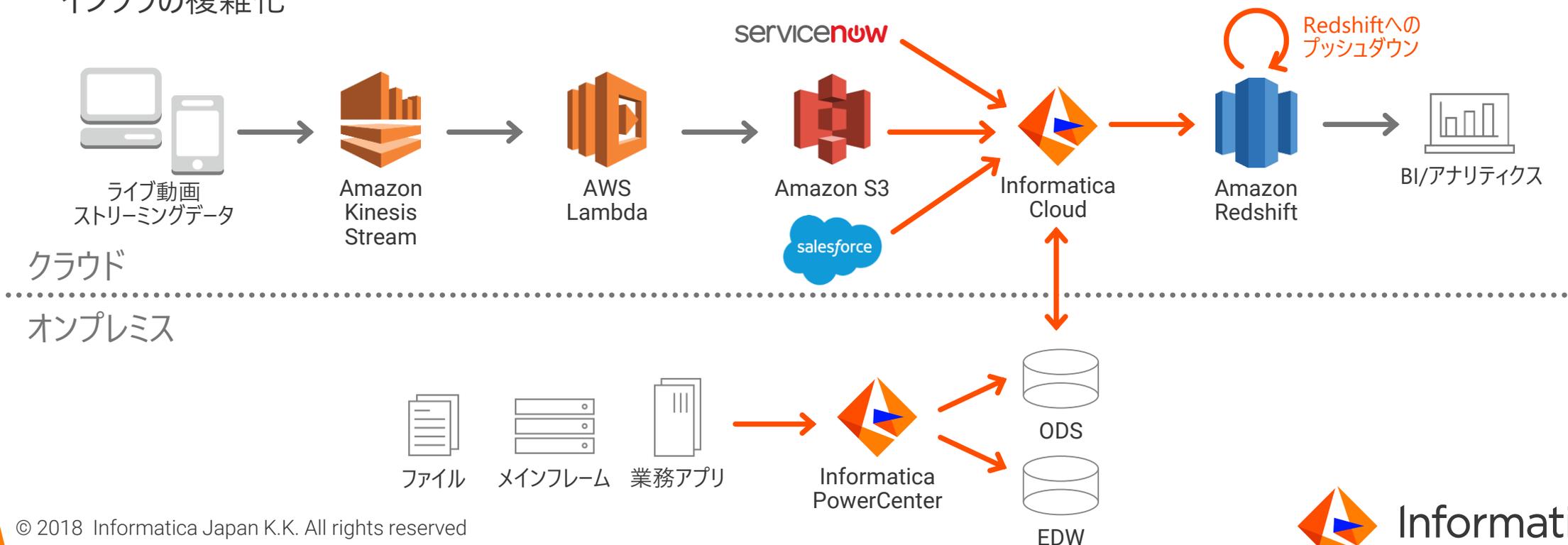
ハイブリッドデータアナリティクス環境の構築

課題

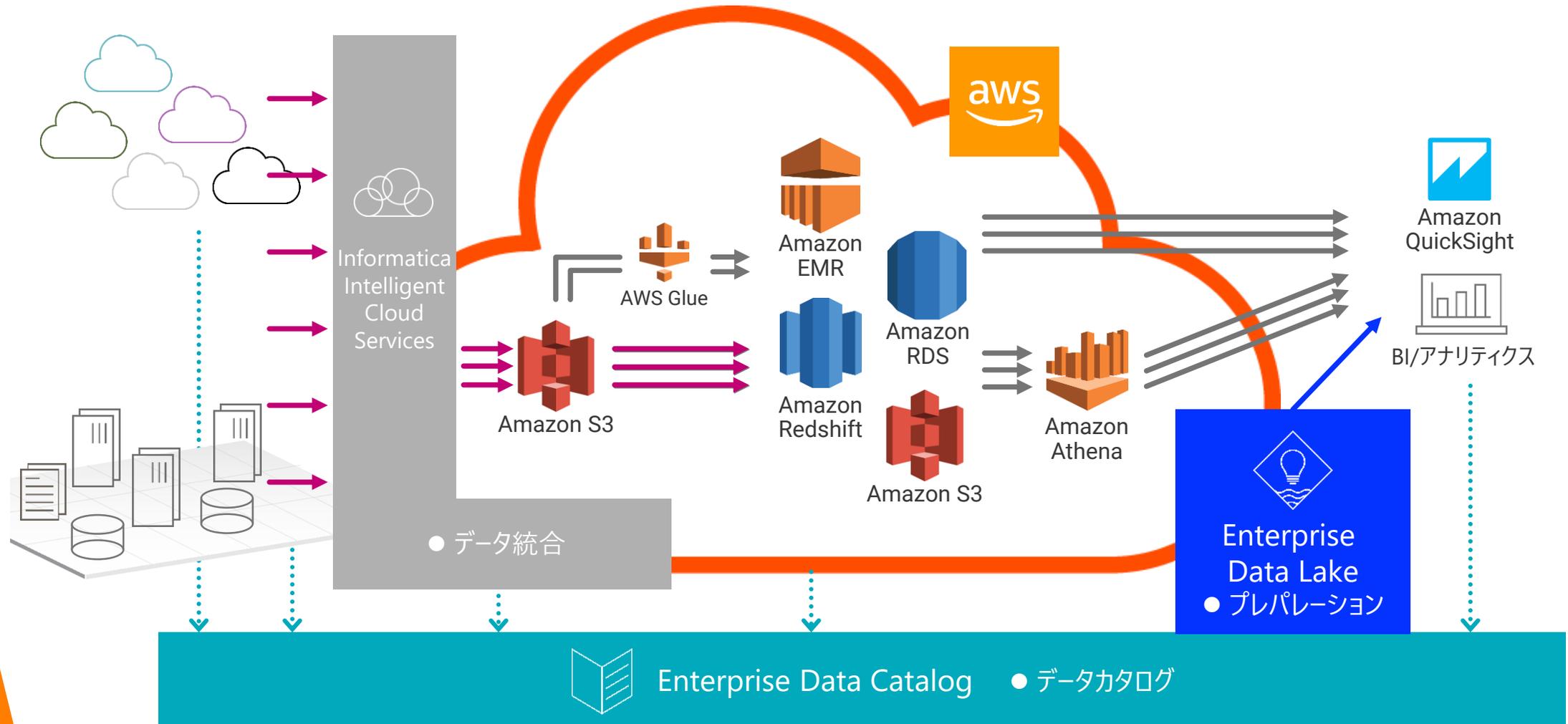
- ✓ 顧客体験(UX)の向上
- ✓ 業務ユーザによる分析ニーズの高度化
- ✓ リアルタイムデータの活用に伴うデータ増加とインフラの複雑化

効果

- クラウド/既存オンプレミスを活用した柔軟なハイブリッドインフラ
- 既存資産を活用した効率的なデータ統合によるUX向上とコスト最適化



AWSにおけるInformaticaの活用



分析プロセスとデータマネジメントの重要性

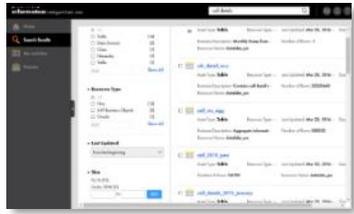


分析業務の**80%**は、分析の前の準備作業



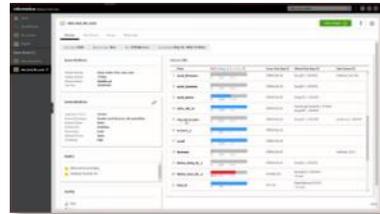
分析ユーザーのデータキュレーションプロセス

カタログサーチ



データの在り処を探す

プロファイリング



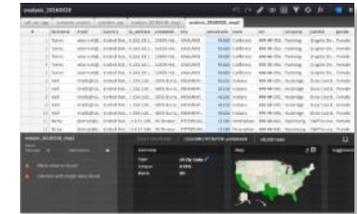
データの中身进行评估する

ビジネス用語



ビジネス用語による論理名

データプレパレーション



データを加工準備する

収集
蓄積

アクセス
Access

探索
Find

評価
Profile

エンリッチ
Enrich

ブレンド
Blend

準備
Prepare

分析
Analyze

リネージュ



データの流を把握する

リレーション



データのつながりを把握する

データドメイン

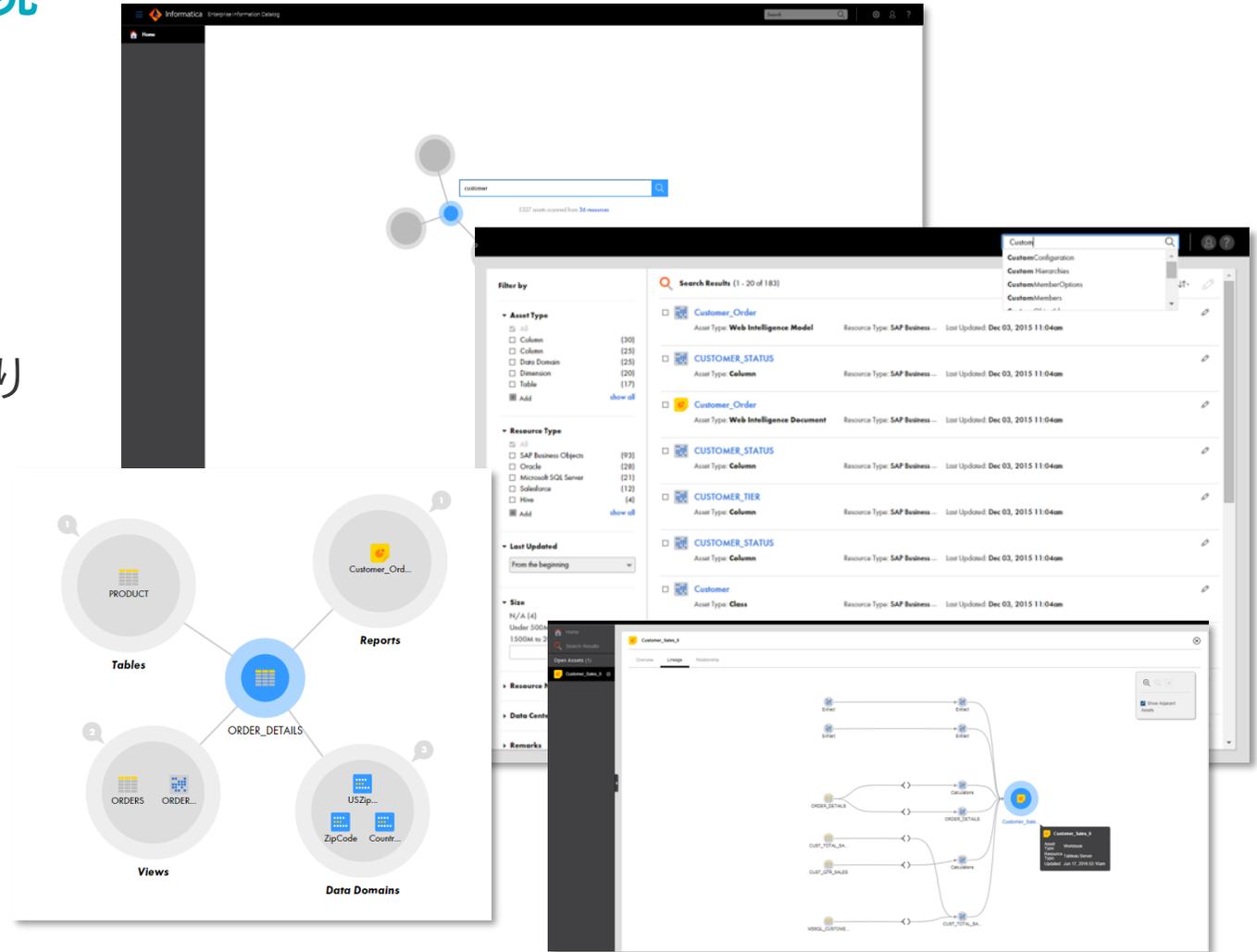
名前	ドメイン	タイプ	ドメイン	ドメイン	ドメイン
1. user_address	住所	住所	住所	住所	住所
2. user_city	住所	住所	住所	住所	住所
3. user_gender	性別	性別	性別	性別	性別
4. user_age	年齢	年齢	年齢	年齢	年齢
5. user_email	メール	メール	メール	メール	メール
6. user_phone	電話	電話	電話	電話	電話

データの意味の把握・付与

データカタログ: Enterprise Data Catalog

あらゆるデータの探索と見える化を実現

- 誰でも、容易に、データを探索
 - データ間の様々な関連性やつながり
 - データの流れ(リネージ)
 - ビジネスグロッサリー
- AIを活用したスマートタギング



データプレパレーション: Enterprise Data Lake

ユーザーによるセルフサービスでのデータ抽出 > 加工 > 提供を支援

- データカタログとのシームレスな統合により、データの探索、加工、共有と公開をサポート
- Excelライクなインターフェースで、誰でも容易にデータを編集加工
- AIを活用したデータ加工のレコメンデーション

The screenshot displays the Informatica Intelligent Data Lake interface. On the left, a 'Recipe Step' pane lists six steps: 1. Join all rows from mast_dem and call_rec_agg using 'Full Outer' on...; 2. Delete year; 3. Insert formula into new column: sum_all_charges -...; 4. Insert formula into new column: sum_all_minutes -...; 5. Insert formula into new column: spiked_charges -...; 6. Round total_intl_charge to nearest 2 digits. The main area shows a data table with columns: #, /, postalcode, state, ssn, company, and jobtitle. The table contains 23 rows of data. Below the table, a 'SHEET OVERVIEW' pane shows statistics for the 'state' column: Type: US State, Unique: 0.20%, Blank: 0%, Length Range: 2 to 20. A 'Map' pane shows a map of the United States with a green dot indicating the location of the data. A 'Suggestions' pane shows 'No suggestions right now.'

#	/	postalcode	state	ssn	company	jobtitle
13	TSBURGH	15240	Pennsylvania	###-##-8424	Topiclounge	Staff Accountant IV
14	TSBURGH	15240	Pennsylvania	###-##-8424	Topiclounge	Staff Accountant IV
15	TSBURGH	15240	Pennsylvania	###-##-8424	Topiclounge	Staff Accountant IV
16	EA	92822	California	###-##-9517	Realfire	Nurse Practitioner
17	EA	92822	California	###-##-9517	Realfire	Nurse Practitioner
18	EA	92822	California	###-##-9517	Realfire	Nurse Practitioner
19	GA	92822	California	###-##-9517	Realfire	Nurse Practitioner
20	GA	92822	California	###-##-9517	Realfire	Nurse Practitioner
21	NRICO	23228	Virginia	###-##-0429	Skidoo	Statistician I
22	NRICO	23228	Virginia	###-##-0429	Skidoo	Statistician I
23	NRICO	23228	Virginia	###-##-0429	Skidoo	Statistician I



デモ

データカタログ: 主要スキャナー

RDB/DWH



Oracle
DB2
DB2 z/OS
DB2 i5/OS
MS SQL Server
Sybase ASE
MySQL
JDBC
Teradata
Netezza
Amazon Redshift

Hadoop



Cloudera Navigator
Hortonworks Atlas
HDFS/Hive
(CDH/HW/MAPR/EMR)

ファイルシステム



Amazon S3
Microsoft SharePoint
Microsoft OneDrive
Microsoft Windows
Linux

BI



SAP BusinessObjects
Cognos
Tableau
QlikView
MicroStrategy
Oracle BIEE

アプリケーション



SAP ERP
Salesforce

モデリング



ERwin

データ統合



Informatica PowerCenter
Informatica Platform
Informatica Cloud
Informatica Axon
Informatica Business Glossary
SQL Server Integration Services

SQL解析



Oracle PL/SQL
DB2 PL/SQL
HiveQL
Teradata BTEQ

データカタログ: Amazon S3スキャナー対象ファイルフォーマット

- 圧縮ファイル

- gz, tgz, emz, sz, xz, z, 7z, ace, arj, cab, rar, gtar, zip, bz2, tbz2, boz

- Webページファイル

- chm, oth, xhtml, xht, mhtml, html, htm, ihtml

- Eメール

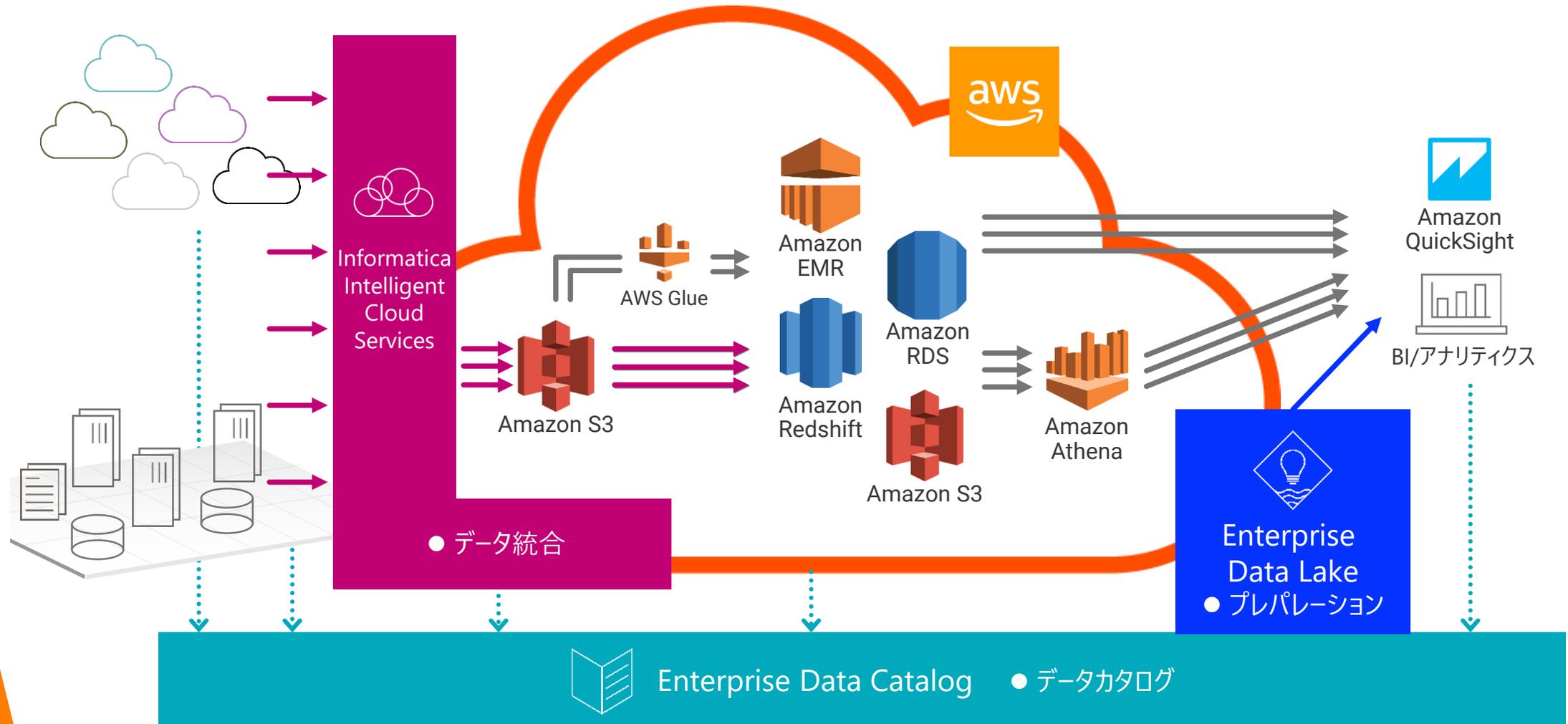
- eml, emlx, mime, mht, mhtml, msg, pst, ost, nsf, ncf

- ドキュメントファイル

- Microsoft Word (doc, docm, docx, dotm, dotx, dot)
- Microsoft Excel (xls, xlsx, xlsx, xlam, xltm, xlsx)
- Microsoft PowerPoint (pptm, potx, ppt, pot, pptx, thmx, ppsm, ppam, ppsx)
- PDF (pdf)

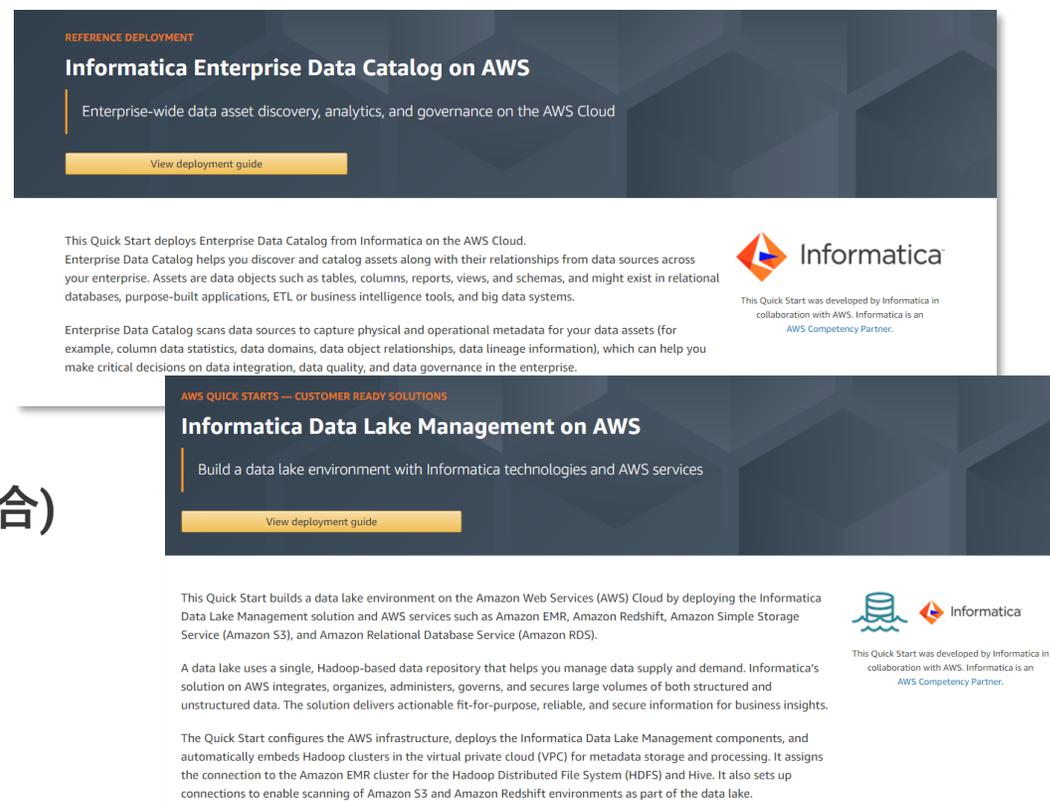


あらゆるデータをAWSへ収集し、利活用を促進するInformaticaソリューション



AWS Marketplace & AWS クイックスタート

- AWS MarketplaceやAWSクイックスタートで Informaticaソリューションを提供しています
 - Informatica PowerCenter (データ統合)
 - Informatica Data Quality (データ品質管理)
 - Informatica Big Data Management (ビッグデータ統合)
 - Informatica Intelligent Cloud Services (クラウドデータ統合)
 - Informatica Enterprise Data Catalog (データカタログ)
 - Informatica Enterprise Data Lake (データプレパレーション)



REFERENCE DEPLOYMENT

Informatica Enterprise Data Catalog on AWS

Enterprise-wide data asset discovery, analytics, and governance on the AWS Cloud

[View deployment guide](#)

This Quick Start deploys Enterprise Data Catalog from Informatica on the AWS Cloud. Enterprise Data Catalog helps you discover and catalog assets along with their relationships from data sources across your enterprise. Assets are data objects such as tables, columns, reports, views, and schemas, and might exist in relational databases, purpose-built applications, ETL or business intelligence tools, and big data systems.

Enterprise Data Catalog scans data sources to capture physical and operational metadata for your data assets (for example, column data statistics, data domains, data object relationships, data lineage information), which can help you make critical decisions on data integration, data quality, and data governance in the enterprise.

 Informatica
This Quick Start was developed by Informatica in collaboration with AWS. Informatica is an AWS Competency Partner.

AWS QUICK STARTS — CUSTOMER READY SOLUTIONS

Informatica Data Lake Management on AWS

Build a data lake environment with Informatica technologies and AWS services

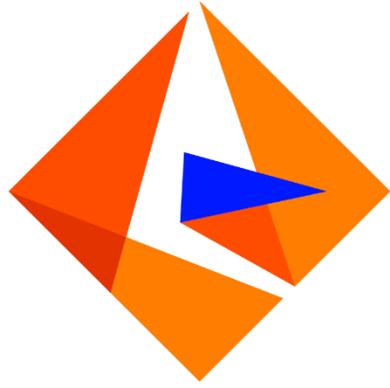
[View deployment guide](#)

This Quick Start builds a data lake environment on the Amazon Web Services (AWS) Cloud by deploying the Informatica Data Lake Management solution and AWS services such as Amazon EMR, Amazon Redshift, Amazon Simple Storage Service (Amazon S3), and Amazon Relational Database Service (Amazon RDS).

A data lake uses a single, Hadoop-based data repository that helps you manage data supply and demand. Informatica's solution on AWS integrates, organizes, administers, governs, and secures large volumes of both structured and unstructured data. The solution delivers actionable fit-for-purpose, reliable, and secure information for business insights.

The Quick Start configures the AWS infrastructure, deploys the Informatica Data Lake Management components, and automatically embeds Hadoop clusters in the virtual private cloud (VPC) for metadata storage and processing. It assigns the connection to the Amazon EMR cluster for the Hadoop Distributed File System (HDFS) and Hive. It also sets up connections to enable scanning of Amazon S3 and Amazon Redshift environments as part of the data lake.

 Informatica
This Quick Start was developed by Informatica in collaboration with AWS. Informatica is an AWS Competency Partner.



Informatica™