

DXの成功はデータの活用から！インフォマティカとAWSで実現するデータ活用基盤

インフォマティカで実現する データ資産の見える化

インフォマティカ・ジャパン株式会社

ソリューションアーキテクト & エバンジェリスト

森本 卓也



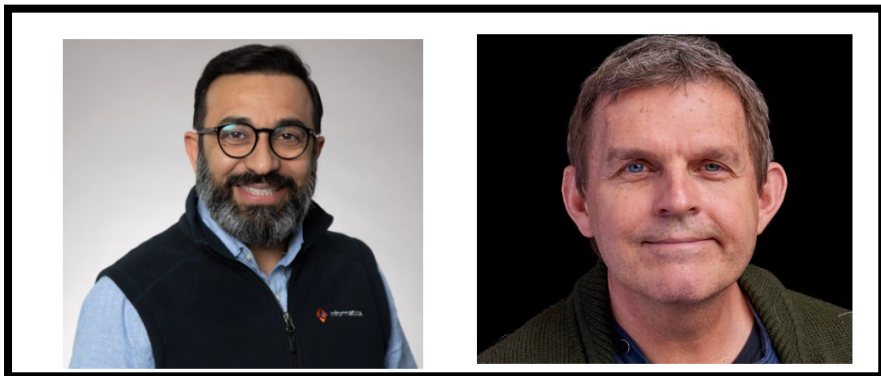
Informatica®

なぜ AWS×インフォマティカ なのか？



- ✓ 2011年から続く戦略的パートナーシップ
- ✓ エンタープライズなクラウドコンピューティングとクラウドデータマネジメントとのリーダー
- ✓ クラウドレイクハウスを迅速にスタート

AWS×インフォマティカは、強力なパートナーシップを10年以上結んでいます



Amit Walia, CEO Informatica & Charlie Bell, SVP Utility Computing Services, AWS, Keynote at Informatica World 2021



Jitesh Ghai, Chief Product Officer, Informatica & Rahul Pathak, Vice President, Analytics, AWS @ClaireVlew2020

Long Standing Partnership

10年以上にわたり、データの移行、統合、管理に関する課題を解決することで、お客様がデータ主導の変革を加速できるよう支援してきました。

Strategic Alignment

- AWSパートナーネットワーク (APN) アドバンスドテクノロジー パートナー
- データ & アナリティクス コンピテンシー
- SaaS RevRec パートナー
- Amazon RedshiftおよびAmazon RDSサービスとの連携
- AWS Marketplace ソリューション

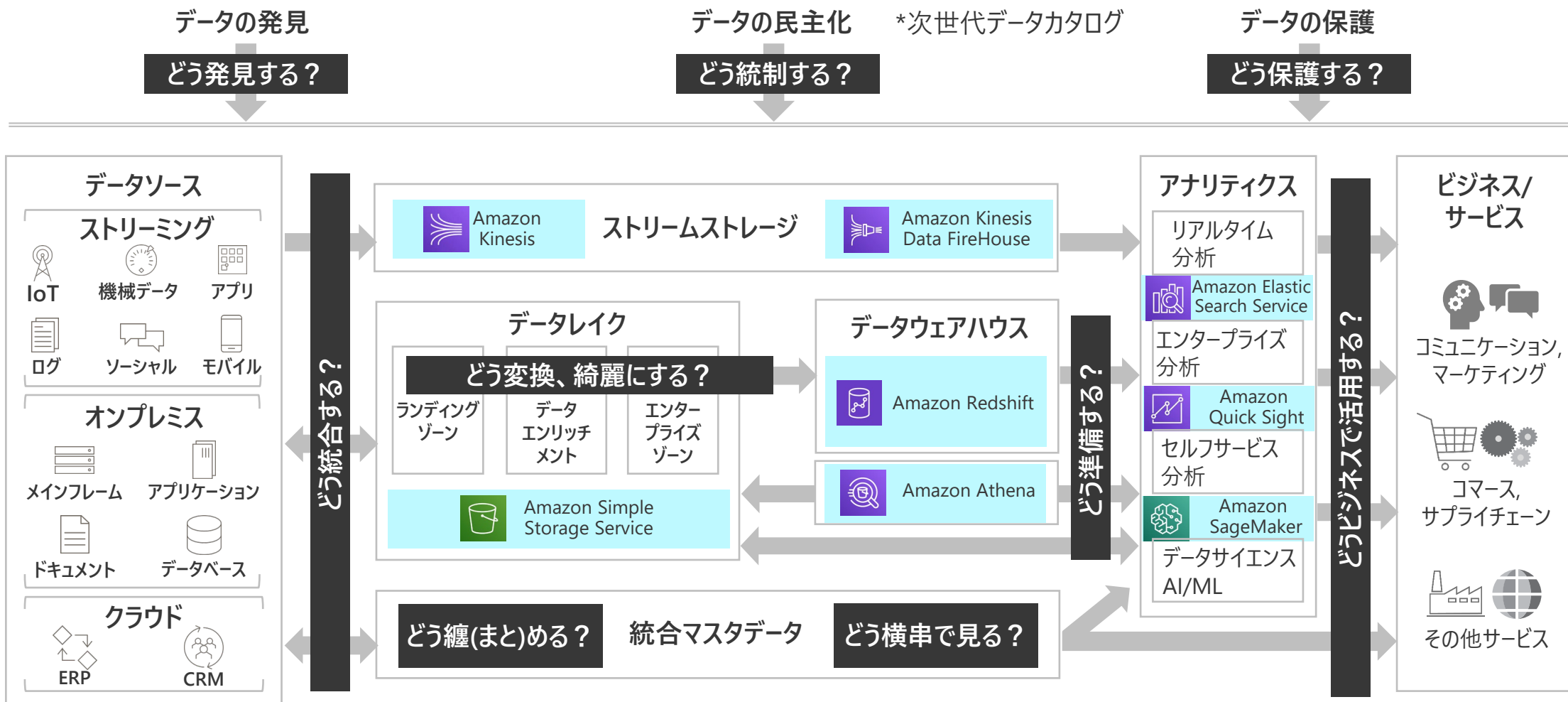
End-to-End Cloud Data Management

データの発見、カタログ化、インジェスト、クレンジング、統合、最適化、ガバナンス、セキュリティ、準備、マスタデータを実現する、AIを搭載したメタデータ駆動型のクラウド・ネイティブ・プラットフォーム

Joint Customers

Fortune 100社の85%、Global 2000社の50%以上のお客様のデータ管理を担当、あらゆる業界の1,000社以上の共同顧客

AWSのクラウドレイクハウスをとりまくデータマネジメント課題、、、



インフォマティカはAWSのレイクハウスをエンタープライズ規模で成功に導く

インフォマティカは、
これらすべての課題を
克服するのに役立つ
唯一無二の
AWS パートナーです



AWS に取り込むデータをどのように 発見 するか?
-> データカタログ

AWS にデータを 統合 するには、どうすればよいか?
-> データ統合, マスインジェスチョン

さまざまなデータを利用可能な形式にどのように 変換 すればよいか?
-> データ統合

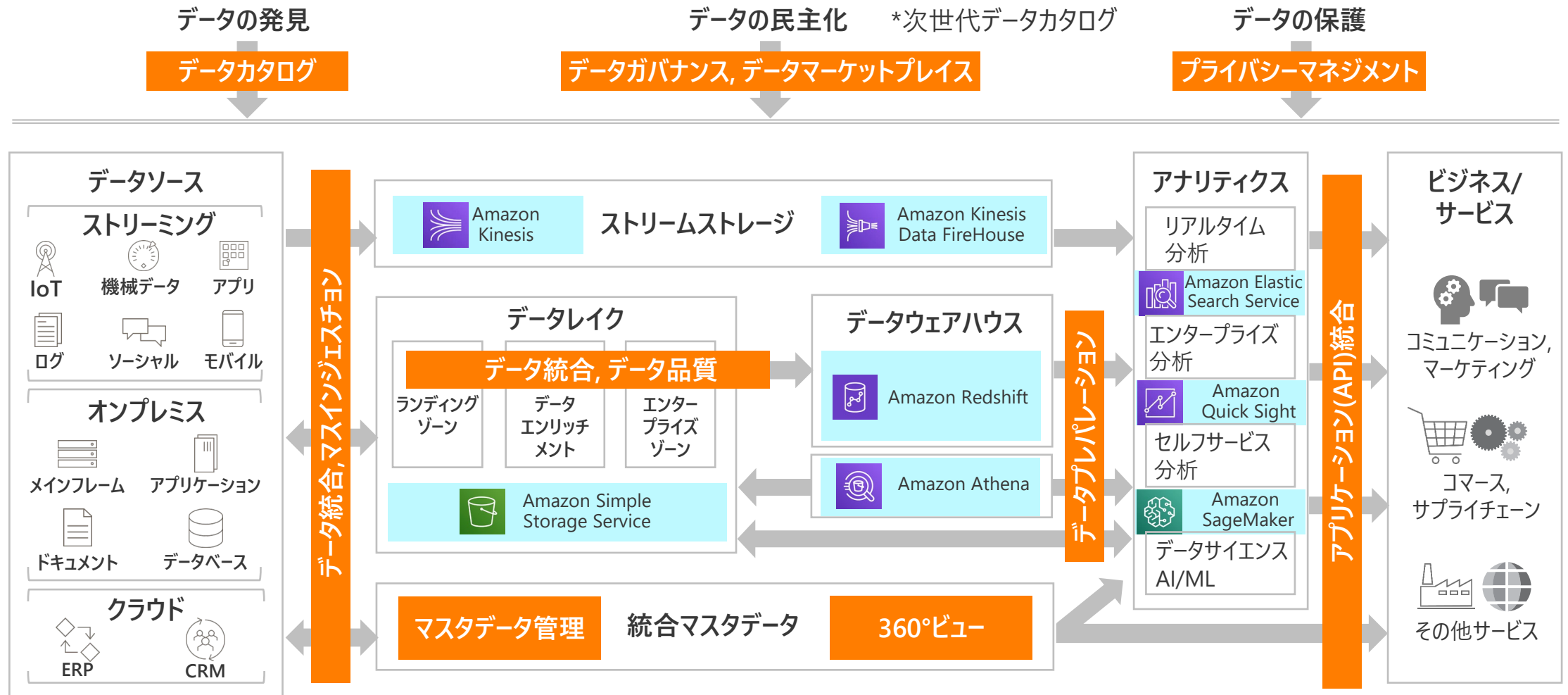
AWS に統合したデータの 品質 をどのように担保するか?
-> データ品質

サイロ化されたデータを 纏め、事業横断で 横串 分析するには、どうすればよいか?
-> マスターデータ管理, 360°ビュー

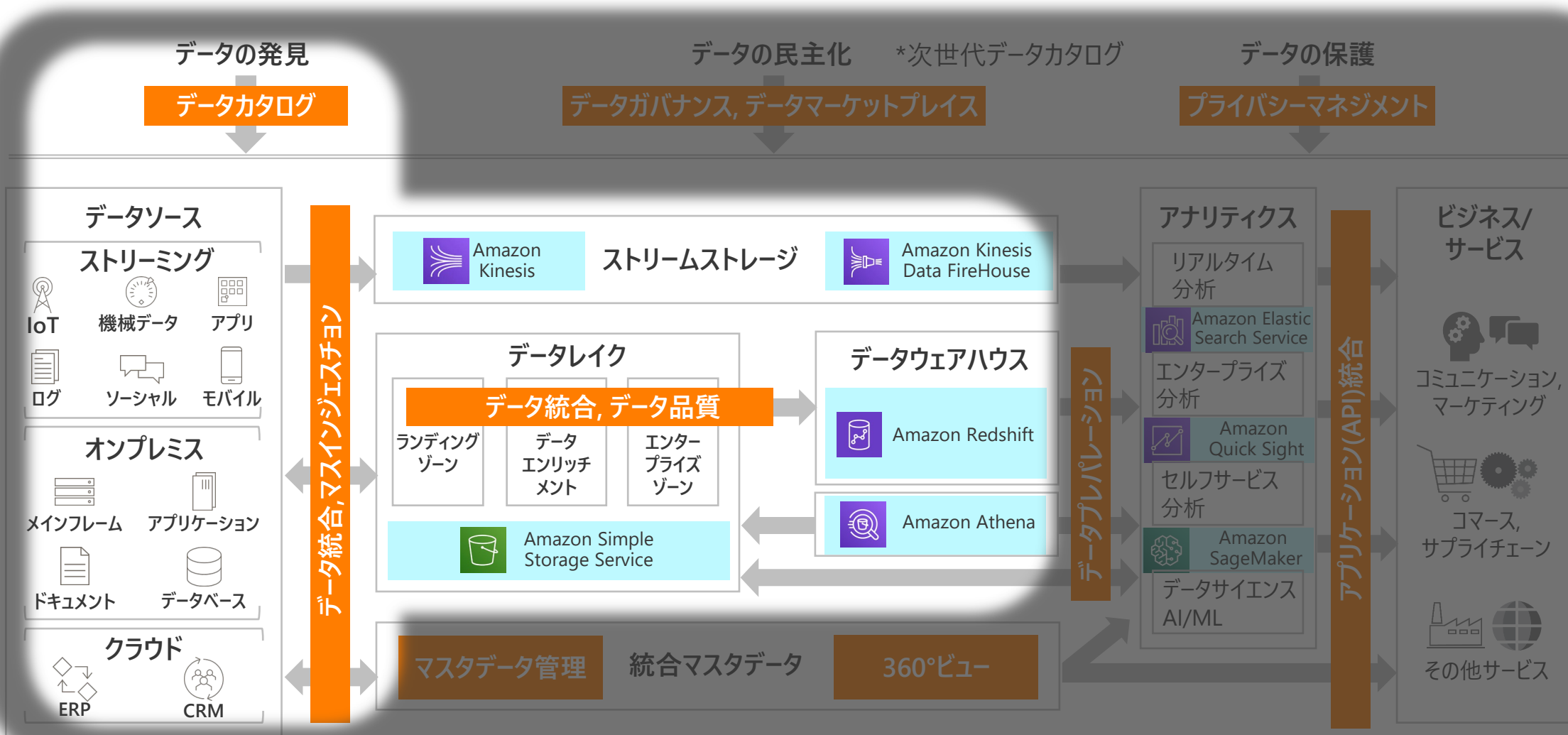
リスクの高いデータを 保護 し、企業全体のデータの 統制 を取るには、どうすればよいか?
-> プライバシーマネジメント, データガバナンス

分析に必要なデータをどのように見つけて 準備 し、すばやく 活用 するか?
-> データマーケットプレイス, データプレパレーション, アプリケーション(API)統合

AWSのレイクハウスを成功に導く クラウドデータマネジメント・アーキテクチャ



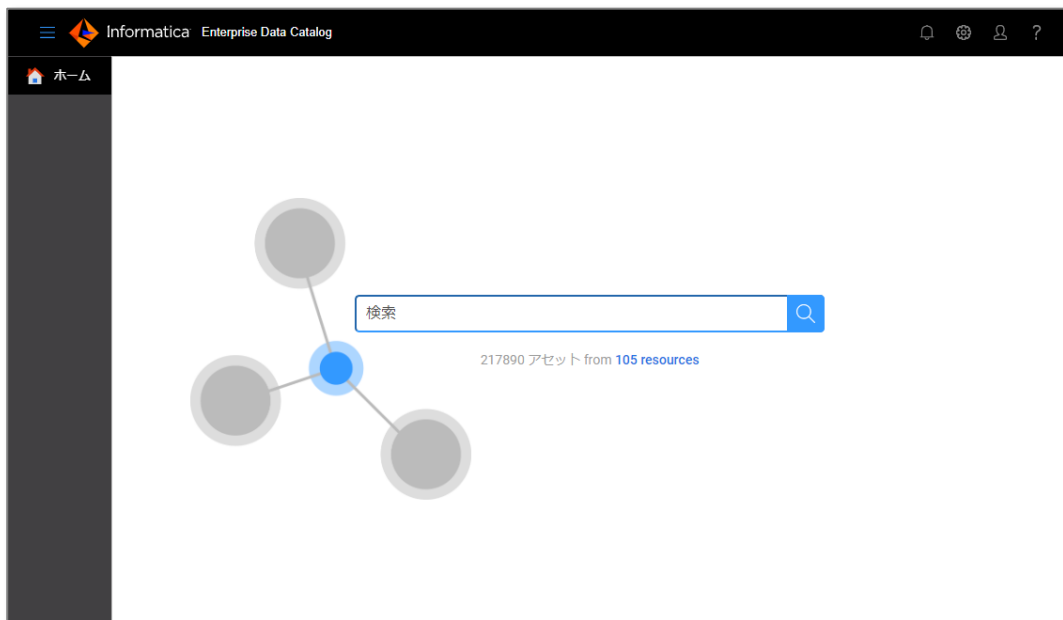
AWSのレイクハウスの迅速な立ち上げを支えるインフォマティカのソリューション



AWSに取り込むデータを発見するデータカタログ

メタデータ管理

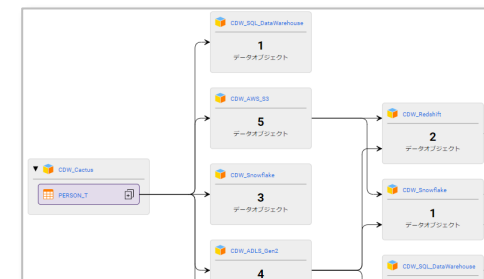
Enterprise Data Catalog



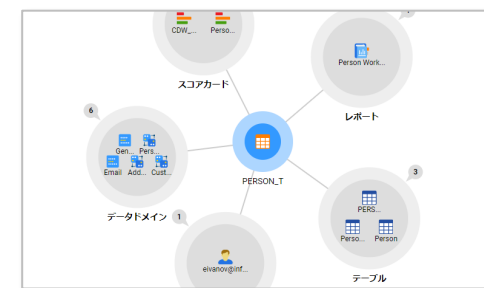
- データの傾向/特性把握
プロファイリング



- データの流れ把握
データリネージ



- データの関連性把握
データリレーション



AWSにあらゆるデータを統合する多様なデータ統合、マスインジエスジョン

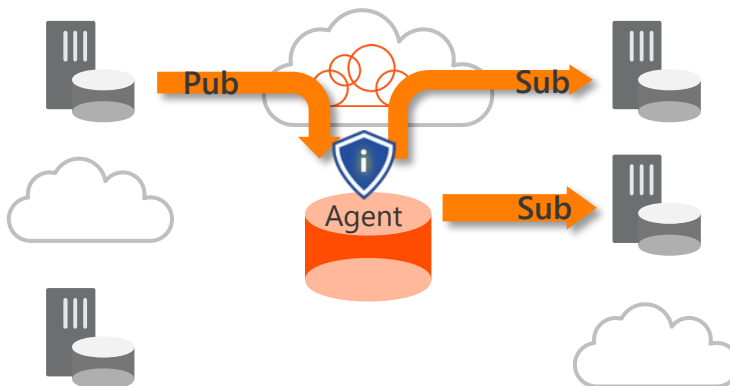
様々なシステムに対するデータの連携

P2P型データ連携 Cloud Data Integration



- Peer to Peer型のデータ連携
- ETL/ELT共にプッシュダウン可
- 変更差分同期も可

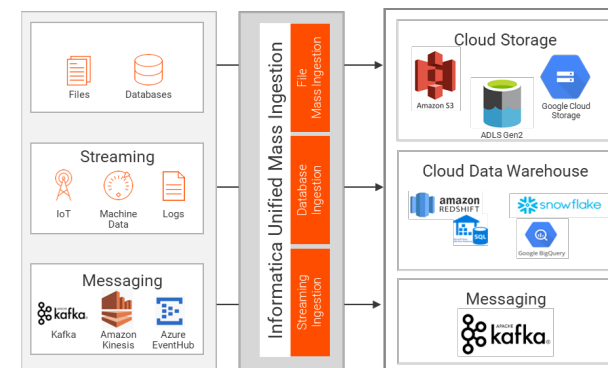
ハブ型データ連携 Cloud Integration Hub



- 送信側(Pub)と受信側(Sub)に疎結合化したハブ型のデータ連携
- データ連携の運用管理を一元化

データの一括取り込み

データインジェスジョン Cloud Mass Ingestion



- ファイル・DB・ストリーミング対応
- クラウドターゲットに一括取り込み
- ウィザード形式で実行

誰でも、あらゆるデータを利用可能な形式に変換できるデータ統合

The screenshot displays the Informatica Data Integration Designer interface for a job named 'Ora03_DeptSalTotal'. The main workspace shows a data flow diagram with the following components and annotations:

- データの抽出 (Data Extraction):** Two source nodes, '部署マスタ' (Department Master) and '社員マスタ' (Employee Master), feed into a '結合' (Join) node.
- 複数データの結合 (Join of Multiple Data):** An annotation pointing to the '結合' node.
- 不要なデータのフィルタリング (Filtering of Unnecessary Data):** A 'フィルタ' (Filter) node follows the join.
- グルーピング・集計 (Grouping and Aggregation):** A '集計' (Aggregation) node follows the filter.
- ターゲットへの書込み (Writing to Target):** The final '部署給与合計' (Department Salary Total) target node.
- AWSのコンピューティングリソースで高速変換 (High-speed transformation using AWS computing resources):** An annotation pointing to the aggregation node.
- プッシュダウン (Pushdown):** An annotation pointing to the aggregation node, with a callout box containing SQL code.

The '結合プロパティ' (Join Properties) panel is visible at the bottom, showing:

- 結合タイプ: ノーマル (Join Type: Normal)
- 結合条件: 簡易 (Join Condition: Simple)
- 結合条件 (Join Condition):

マスタ	演算子	詳細
D_DEPTNO	=	E_DEPTNO

```
INSERT INTO CIPS_DM202... ID, M_EMAIL, M_FULL_NAME, TRIPDURATION, START_TIME, COPTIME, START_STATION_ID, START_STATION_NAME, START_STATION_LATITUDE, START_STATION_LONGITUDE, END_STATION_ID, END_STATION_NAME, END_STATION_LATITUDE, END_STATION_LONGITUDE, BIKED, MEMBERSHIP_TYPE, USERTYPE, BIRTH_YEAR, GENDER) SELECT CUSTOMER_MASK.ID, CUSTOMER_MASK.EMAIL
```

AWSに統合したデータの品質を分析し改善するデータ品質

1. プロファイリング

現在のデータの品質を分析・評価

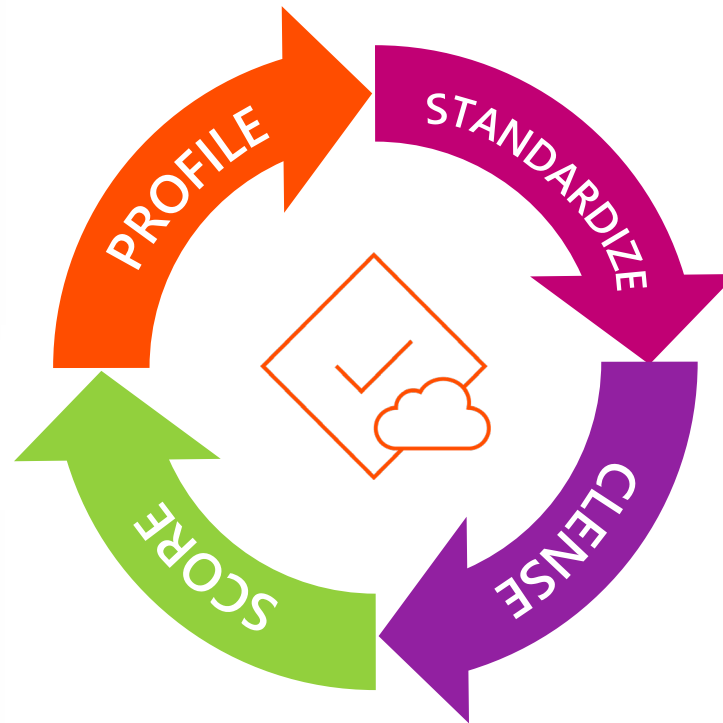


4. スコアリング

ルールに基づき定常的にデータ品質を測定

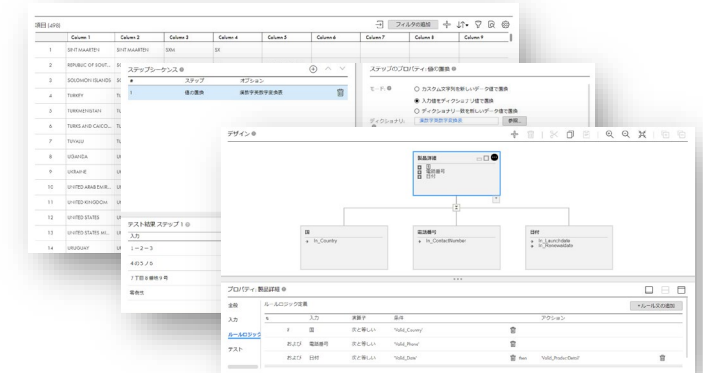
The screenshot shows a data quality scoring tool interface. On the left, there is a table with columns for 'カラム' (Column), '月の数' (Number of Months), '個別の行の数' (Number of Individual Rows), '重複のある行' (Number of Rows with Duplicates), and 'パターン' (Pattern). The table contains several rows of data, including columns like ADDRESS1, ADDRESS2, ADDRESS3, ADDRESS4, ADDRESS5, CUSTID, CUST_NAME, CUST_NAME_2, CUST_NAME_3, CUST_NAME_4, CUST_NAME_5, CUST_NAME_6, CUST_NAME_7, CUST_NAME_8, CUST_NAME_9, CUST_NAME_10, CUST_NAME_11, CUST_NAME_12, CUST_NAME_13, CUST_NAME_14, CUST_NAME_15, CUST_NAME_16, CUST_NAME_17, CUST_NAME_18, CUST_NAME_19, CUST_NAME_20, CUST_NAME_21, CUST_NAME_22, CUST_NAME_23, CUST_NAME_24, CUST_NAME_25, CUST_NAME_26, CUST_NAME_27, CUST_NAME_28, CUST_NAME_29, CUST_NAME_30, CUST_NAME_31, CUST_NAME_32, CUST_NAME_33, CUST_NAME_34, CUST_NAME_35, CUST_NAME_36, CUST_NAME_37, CUST_NAME_38, CUST_NAME_39, CUST_NAME_40, CUST_NAME_41, CUST_NAME_42, CUST_NAME_43, CUST_NAME_44, CUST_NAME_45, CUST_NAME_46, CUST_NAME_47, CUST_NAME_48, CUST_NAME_49, CUST_NAME_50, CUST_NAME_51, CUST_NAME_52, CUST_NAME_53, CUST_NAME_54, CUST_NAME_55, CUST_NAME_56, CUST_NAME_57, CUST_NAME_58, CUST_NAME_59, CUST_NAME_60, CUST_NAME_61, CUST_NAME_62, CUST_NAME_63, CUST_NAME_64, CUST_NAME_65, CUST_NAME_66, CUST_NAME_67, CUST_NAME_68, CUST_NAME_69, CUST_NAME_70, CUST_NAME_71, CUST_NAME_72, CUST_NAME_73, CUST_NAME_74, CUST_NAME_75, CUST_NAME_76, CUST_NAME_77, CUST_NAME_78, CUST_NAME_79, CUST_NAME_80, CUST_NAME_81, CUST_NAME_82, CUST_NAME_83, CUST_NAME_84, CUST_NAME_85, CUST_NAME_86, CUST_NAME_87, CUST_NAME_88, CUST_NAME_89, CUST_NAME_90, CUST_NAME_91, CUST_NAME_92, CUST_NAME_93, CUST_NAME_94, CUST_NAME_95, CUST_NAME_96, CUST_NAME_97, CUST_NAME_98, CUST_NAME_99, CUST_NAME_100. On the right, there is a detailed view of a specific rule, showing the rule name, the data type, and the rule configuration.

Data Quality



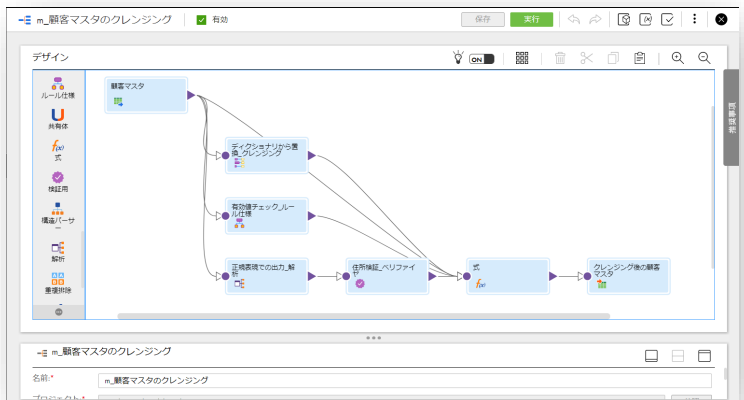
2. 標準化

ルールを定義し、ルール仕様を標準化



3. クレンジング

クレンジング部品によるデータ変換





分析環境のクラウド化とモダナイゼーション

- NAB (National Australia Bank)
- オーストラリア最大の4つの金融機関の1つ
(銀行の中で時価総額で世界21位)
- AWSのトレーニングを受けた従業員数は4000人以上
- InformaticaとAWSのソリューションを組合せ、
最新の分析環境を支えるデータアーキテクチャを実現

【参考情報】 NAB writes consumer privacy rights into its data architecture

<https://www.itnews.com.au/news/nab-models-consumer-privacy-rights-in-its-data-architecture-526804>





Informatica®