

データレイク開発実績についてのご紹介

株式会社エイエイエスティ東京

はじめに

弊社ではデータレイク技術を活用して、某大手企業グループ様の各個社が保持する様式の異なる様々なデータを収集・統合・一元管理し、グループ内の経営分析に繋げるシステム構築を手掛けてまいりました。本稿ではそのデータレイクシステムの構築実績についてご紹介いたします。

本実績の訴求ポイント

DX(デジタルトランスフォーメーション)推進に伴い、AIの利活用が叫ばれていますが、AI技術の中核である深層学習機能を効果的に正しく導くためには、多様なデータを学習データとして整備することが必須となります。弊社のデータレイク構築実績はそのデータの整備・統合化を実現する技術的基盤となります。

目次

1. データレイクの開発実績概要
2. データレイク開発全体像と課題解決策
3. 今回のデータレイクで使用している技術
4. ETLツールであるPowerCenter採用理由

1. データレイク開発実績概要

データレイク技術を活用し、某大手企業グループ様に属する各社から各種データ(売上損益データ、間接費データ、受注残データ、監査で使用するデータ等)を収集して、提出ユーザー、データ管轄部署並びに、監査法人へ集約したデータを公開します。各データを一本化することにより、各データの提出を個別に行う必要がなくなり、監査法人もデータの不備等の連絡を個別に行う必要がない。また、集約したデータ分析も容易に出来ます。(ビッグデータに該当する業務)

2. データレイク開発全体像と課題解決策

データレイクによる課題解決策

課題① 提出データの種類が多いため、提出状況の管理が難しい。

⇒(解決策) 全ての提出データがDB管理で、未提出のフォローメール等も自動で送付される。

課題② 各部署やGr会社で独自システムの為、提出データのフォーマットがばらばらで統一されておらず、項目名が微妙に違う。

⇒(解決策) 全ての提出データに対して統一フォーマット作成。

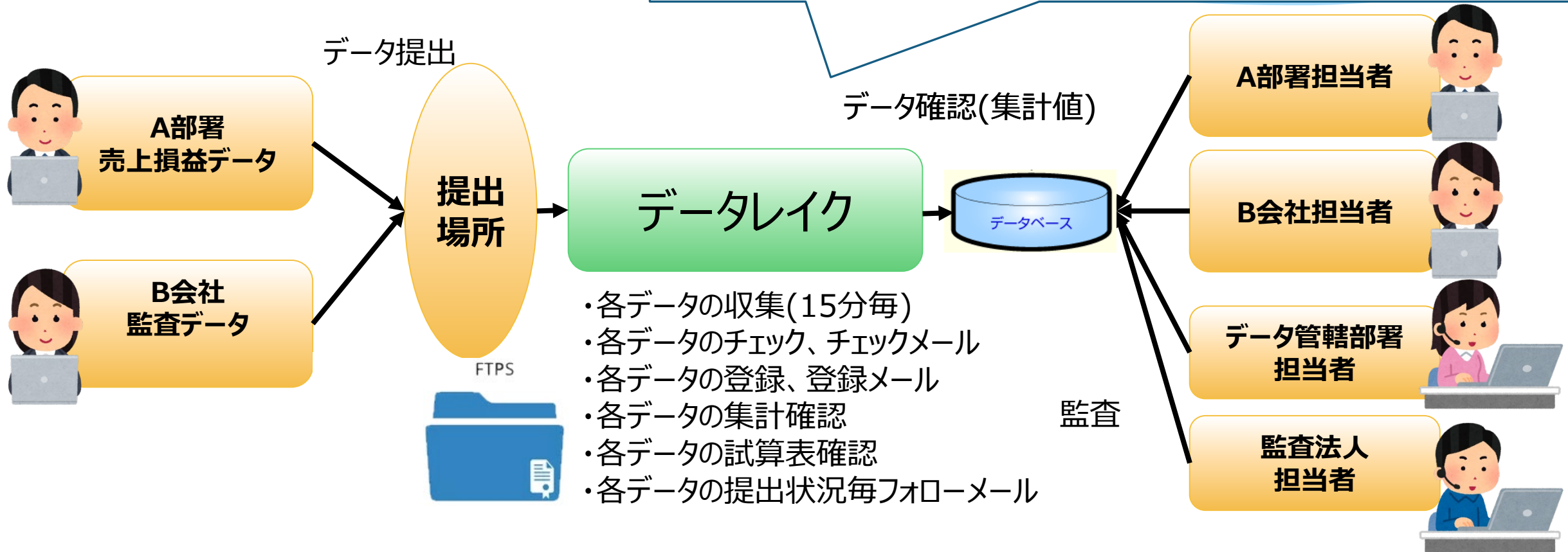
課題③ 提出データに不正データがある場合、連絡対応が煩雑。

⇒(解決策) 全ての提出データに対して自動でチェックし、DB管理した送付先に自動メールを送信。

課題④ 提出データの集計値を確認できない。

⇒(解決策) 全ての提出データを自動で取り込み、Tableau帳票で確認できる。

担当者で集計帳票の作成も可能。



3. 今回のデータレイクで使用している技術

役割	データレイクで使用しているアプリケーション名
OS	Windowsサーバー(サーバ) Windows (作業/開発環境)
DB	SAP HANA
基幹バッチ	Windows bat + Vbscript
データチェック	C#
集計データ表示(BIツール)	Tableau
Job管理	JP1
データ登録	ETLツール PowerCenter
その他	Sakura Editor, WinSCP, FFFTP, WinMargeなど

4. ETLツールであるPowerCenter採用理由

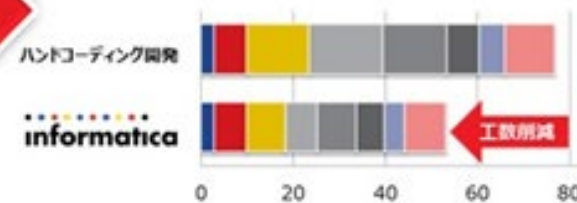
1. 接続性

あらゆるデータへの容易なアクセスを実現



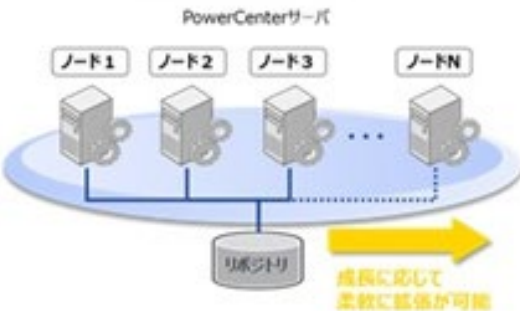
2. 生産性

GUI開発、変換処理ロジックの再利用などによる開発・運用生産性の最大化



4. 可用性・拡張性

大規模システムの安定稼動・信頼性を確保



3. パフォーマンス

マルチコアCPUの性能を最大限に活かしたデータ連携処理の最速化



出典:株式会社メトロ PowerCenterの概要

<https://www.metro.co.jp/products/information/powercenter/>

以上

* 某大手企業グループ様においてはこのデータレイクシステムの有効性が評価され、多様なデータ種の収集や外部システムからの移管等、あらゆるデータを一元的に収集するシステム拡張が行われており、エイエイエスティ東京は継続してご支援させて頂いております。

* 「PowerCenter」は、Informatica Inc. の商標または登録商標です。