IMPETUS

kyvos

SOLUTION BRIEF

Building a Single Source of Truth across the Enterprise – An Integrated Solution

From EDW modernization to self-service BI on big data

This solution brief showcases an integrated approach for modernizing your big data warehouse using automation and building an interactive, self-service BI platform at massive scale providing a 'single source of truth' across the enterprise.

IMPETUS

kyvos

Background

Data-driven decision making is changing the way businesses operate, and data warehouse is at the core of an enterprise's big data and analytics strategy. Existing data warehouses are neither easily scalable to accommodate the growing data nor analytically flexible to business users and analysts. As the traditional data warehouse fails to meet business requirements, there is an urgent need to move to cloud, on-premise or hybrid big data warehouse environment. This transformation will also eliminate the data silos that exist today and bring together the enterprise-wide data to create a comprehensive single source of truth across business teams.

Benefits of Our Solution

1

Quick, automated, no-risk transformation of data from disparate silos to modern data warehouse

2

Reduction in infrastructure and operational costs by freeing up of premium EDW capacity

3

Unified, secure, scalable enterprise-wide data platform focusing on data governance and self-service
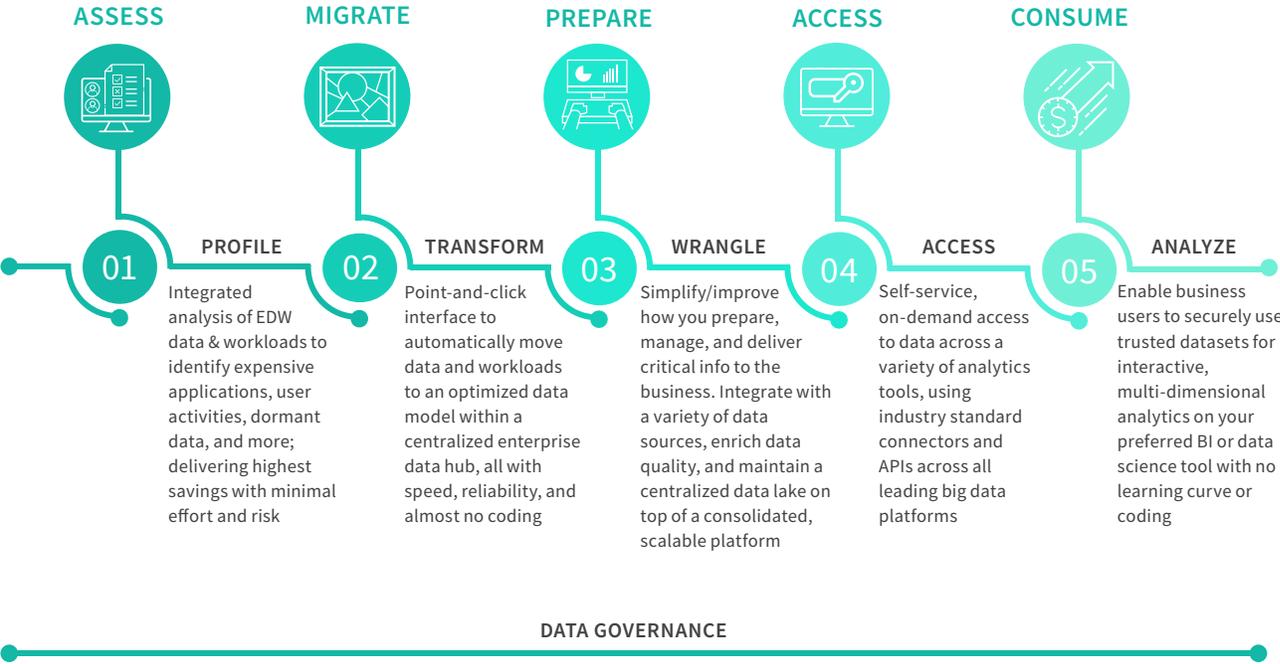
4

Powerful insights on big data enabling timely, data-driven business decisions across the enterprise

Building a Single Source of Truth across the Enterprise - An Integrated Solution

An end-to-end solution that helps you move your existing data silos to to a modern data warehouse environment or an on-premise big data warehouse and build an analytics platform on it that serves business users. The solution follows an incremental approach with five main stages – asessment, transformation, preparation, access, and consumption. Each of these stages are integrated to ensure a holistic solution that drives immediate business benefits and provides a measurable return on investment.

Automated transformation removes the risk of manual error and reduces the time for transformation. Once data from diverse and multiple data sources are brought together, an enterprise BI consumption layer is built on top of the big data warehouse to provision the data for business users by enabling fast, interactive, and self-service BI across the enterprise. The integrated solution helps in building a single source of truth for corporate-wide use with tight security protocols and appropriate governance.

A Modern Data Platform for Fast Insights on Big Data

ASSESS	MIGRATE	PREPARE	ACCESS	CONSUME
01 PROFILE	**02 TRANSFORM**	**03 WRANGLE**	**04 ACCESS**	**05 ANALYZE**
Integrated analysis of EDW data & workloads to identify expensive applications, user activities, dormant data, and more; delivering highest savings with minimal effort and risk	Point-and-click interface to automatically move data and workloads to an optimized data model within a centralized enterprise data hub, all with speed, reliability, and almost no coding	Simplify/improve how you prepare, manage, and deliver critical info to the business. Integrate with a variety of data sources, enrich data quality, and maintain a centralized data lake on top of a consolidated, scalable platform	Self-service, on-demand access to data across a variety of analytics tools, using industry standard connectors and APIs across all leading big data platforms	Enable business users to securely use trusted datasets for interactive, multi-dimensional analytics on your preferred BI or data science tool with no learning curve or coding

DATA GOVERNANCE

Assessment

This stage entails automated profiling of the existing inventory of your enterprise workloads by scanning live and offline query logs or DML/ DDL scripts. This identifies the most expensive processes, resource-intensive user activities, dormant data and more, and helps in evaluation of workloads that are ideal for transformation to the big data warehouse. The solution further provides a mechanism by which transformation pipelines can be automatically generated for incorporating the transformation candidates.

Transformation

Once the workloads are prioritized, data and logic (scripts, reports etc.) are transformed automagically. It optimizes the data structure on-the-fly for performance optimization over Hadoop.

The solution deploys an innovative and powerful automated logic translation engine that quickly translates years and decades of legacy code into big data warehouse compatible scripts. The translated scripts can be executed on any standard big data environment. The automated validation framework is used to certify transformed data, metadata and other workloads. This lays a strong foundation for your big data warehouse, by embracing the data as well as other EDW, ETL, reporting and analytical workloads.

An innovative and intelligent logic translation engine to transform decades of legacy code into Hadoop compatible scripts.

Preparation

Once the enterprise data and other workloads are available in the big data warehouse, preparation involves creation of a simplified, UI-driven data wrangling platform that enables quick preparation of data for advanced analytics. Data is wrangled across heterogeneous data repositories and formats, cleansed and enriched by using analytical transformations and custom-coded ETLs to build a fit-for-purpose dataset curating raw to analytics-ready data. This helps in delivering validated, trusted, insights-ready data in record time and with a high degree of accuracy and security.

The solution provides an end-to-end, native in-Hadoop data wrangling platform that leverages the scalability and economic benefits of Hadoop across a variety of avenues to realize a multi-platform modern data warehouse environment faster. It will help you interactively orchestrate and schedule extensible and repeatable workflows in your dedicated workspace where you can transform, aggregate, cleanse, sort, segment, and dedupe data. This stage helps in deploying enterprise-scale Hadoop, catering to self-service data sourcing, seamless data/metadata discovery, data preparation, and publishing.

Interactively orchestrates and schedules extensible and repeatable workflows in your dedicated workspace.

Access

Easy and instant access to data on Hadoop lays the foundation of building a single source of truth across the enterprise that can open up many new business possibilities. The solution deploys a scale-out architecture that delivers very high performance at any scale, regardless of the number of concurrent users. Thus enterprises can provide controlled access to a single source of data to any number of users across the enterprise. This empowers business users and analysts to get immediate access to the data they need for making decisions.

The solution supports all major security protocols and integrates with your existing enterprise-wide security model. It deploys fine-grained access control and strong authentication for safe access across the enterprise. You can provide user and group security at an object, row, and column levels to protect sensitive data.

Thus, the solution enables controlled access to data to a large number of users across the enterprise. This helps in building a data-centric enterprise where data access is not available to just a few users or teams but can be provisioned to any number of users with fine-grained access governance and control.

> *Supports all major security protocols and integrates with the existing enterprise-wide security model.*

Consumption

The final step of the solution involves building an enterprise BI consumption layer on your big data platform. Self-service consumption of big data reduces the IT team's workload and helps increase the return on your big data investments by driving enterprise-wide adoption. This layer deploys "OLAP on Big Data" technology to build OLAP cubes on massive volumes of data and provides instant response to big data queries irrespective of the size of data or the complexity of the query.

The solution integrates seamlessly with all BI tools ranging from Tableau, Excel, Qlik, Microstrategy, IBM Cognos, Tibco, Spotfire, Business Objects and more, so that the users can choose a tool of their choice or keep using existing tools. They can easily explore, slice and dice, and drill down into their big data using BI tools without being affected by any changes to the underlying data. Since there is no learning curve or coding required for business users and analysts to consume their big data, it enables them to conduct multi-dimensional analysis on their big data, both on-premise and in the cloud.

> *Deploy "OLAP on Big Data" to build OLAP cubes on massive volumes of data and provide instant response to big data queries.*

An integrated approach to accelerate your data migration results vs journey and build an insights-driven enterprise which has the power to transform your business.

Key Features

- **Rapid on-premise/cloud deployment**

  The extensible architecture enables quick and easy transformation, access and consumption of data across all leading big data storage platforms, both on-premise and in the cloud.

- **Builds an enterprise-wide data model**

  Creates a single copy of trusted source of data, made available for access and analysis for users across the enterprise. Visual tools for data preparation and analysis enable better correlation and understanding.

- **Automates processes**

  Connects disparate heterogeneous sources and automatically catalogs, curates and cleanses data for analytical consumption. Automated incremental builds update newly ingested data and truncate expired data.

- **Ensures data governance**

  Tracks the entire journey of data, enriches and enhances its quality and makes it readily discoverable. Honors the existing enterprise security models and works with big data platforms to ensure data governance.

- **Provisions self-service analytics with controls**

  Enables user-friendly analytics by integrating with existing BI tools for rich visualizations. Integrates with enterprise authorization systems and provides inbuilt granular access control with column and row level security.

- **Makes operationalization easy**

  Allows measurement of operational variables empirically and quantitatively across business use cases. Enables advanced analytics on a variety of proliferating real-time data sources in a self-serve mode.

- **Allows enterprise-wide data stewardship**

  Data steward groups can ensure availability of centralized data dictionaries, MDM, data enrichment, data preparation, pre-prepared data models and business rules. Unified metadata and services repository, data profiling/ discovery-based recommendations and curation/ collaboration enabled workflows.
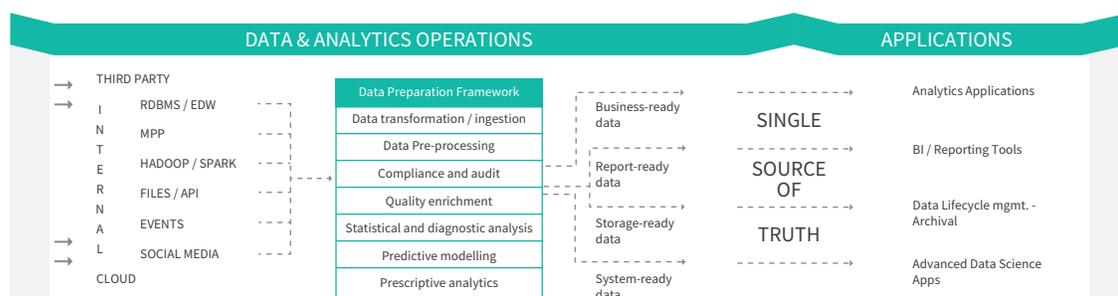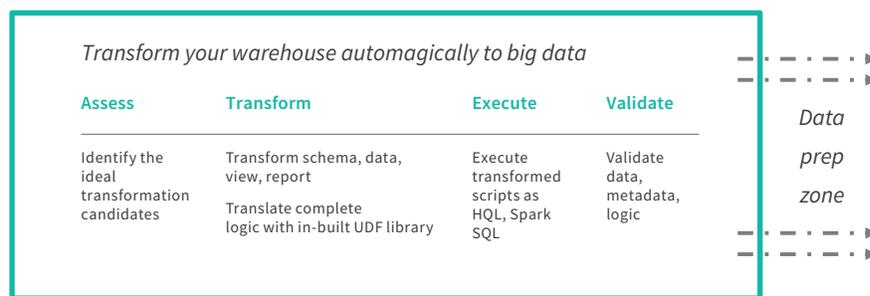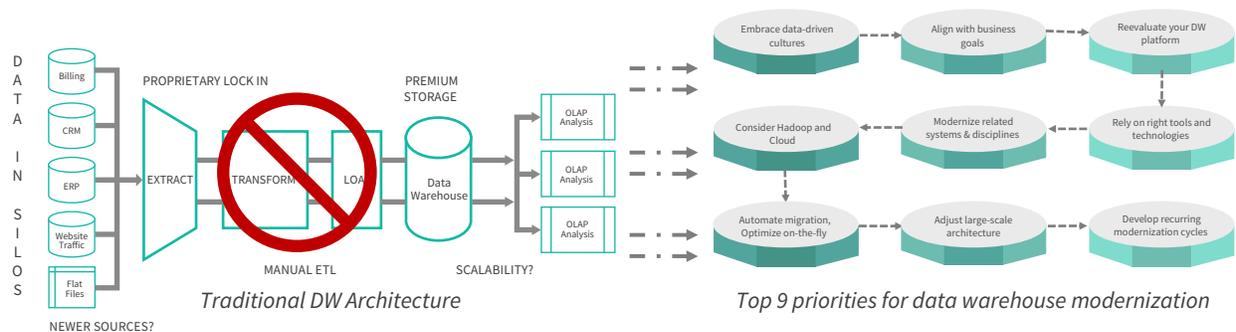
- **Phased implementation**

  Phased approach ensures proper implementation of different technologies, improves success rates, reduces time to implement and ensures user collaboration at every stage.

Key Capabilities for Conclusion- Building a Single Source of Truth

- Collaborate and enrich metadata via tagging, browse and search its definitions, and publish it across a unified metadata repository to maintain a consistent view of data and establish a single source of truth across your enterprise.

- Seamlessly integrate with a variety of security frameworks like LDAP, Kerberos, Ranger, and Sentry along with an in-built row and column level security for ease of operation.

- Establish and manifest data regulatory compliance and achieve fine-grained authorization and control.

Journey to Modern Data Warehouse Architecture

Traditional DW Architecture

Top 9 priorities for data warehouse modernization

Transform your warehouse automagically to big data

Assess	Transform	Execute	Validate
Identify the ideal transformation candidates	Transform schema, data, view, report Translate complete logic with in-built UDF library	Execute transformed scripts as HQL, Spark SQL	Validate data, metadata, logic

Data prep zone

Summary

The solution empowers enterprises to democratize their big data, and drive better and just-in-time decision making. Achieving a single source of truth not only brings enterprise-wide transparency, reaching the very end points of your organization, but also truly enhances business and operational efficiency.

kyvos

Kyvos is a disruptive Big Data solution that delivers the world's fastest BI at a massive scale. Our patent-pending OLAP technology enables Fortune 500 companies to query billions of rows of data within seconds and helps business leaders make informed decisions. We harness the true potential of data lakes in partnership with industry leaders in BI, Cloud, and Hadoop technologies. Our breakthrough technology enables analysts to interact with data in real time, using their favourite BI tools.

IMPETUS

Impetus is focused on creating big business impact through Big Data Solutions for Fortune 1000 enterprises across multiple verticals. The company brings together a unique mix of software products, consulting services, Data Science capabilities, and technology expertise. It offers full life-cycle services for Big Data implementations and real-time streaming analytics, including technology strategy, solution architecture, proof of concept, production implementation and on-going support to its clients. To learn more, visit **www.impetus.com** or write to us at **inquiry@impetus.com**.