

Mphasis Capability in Data Integration Solutions

Mphasis delivers **end-to-end Data Integration Solutions** covering ETL/ELT pipelines, real-time/streaming integration, API-based data exchange, batch ingestion, and cloud/hybrid integration. Mphasis provides enterprise-scale integration across **on-prem, cloud, multi-cloud, hybrid, polyglot** data environments.

Mphasis Data Integration – Capability Overview

1. ETL / ELT Development & Modernization

- Enterprise Data Warehouse (EDW) integration
- ETL pipeline design & refactoring
- DW consolidation & migration

2. API-based & Real-Time Data Integration

- API-driven ingestion for internal/external systems
- Real-time streaming using Kafka, DMS, Glue, Debezium etc.

3. Cloud Data Integration (AWS, Azure, GCP)

- Integrating data into modern cloud platforms like Snowflake, Databricks, Redshift, Synapse
- Cloud migration pipelines (On-prem → Cloud)

4. Data Platform & Lakehouse Integration

- Data Lake / Data Lakehouse ingestion
- Pipeline orchestration via Airflow, MWAA, DataOps, CI/CD

5. MDM/RDM Integration

- Integrating multiple source systems for golden-record creation
- Multi-domain integration (Customer / Property / Product)
-

6. Metadata, Lineage & Governance Integration

- Linking business and technical metadata for traceability
- Integrating ingestion frameworks with governance & DQ

7. Data Quality Integration

- Integration pipelines with embedded DQ rules, profiling & dashboards

8. Automation & DataOps for Integration

- Automated code generation
- CI/CD for integration pipelines
- Intelligent data pipeline validation

Key Features of Mphasis Data Integration Solutions

Feature	Description
End-to-End Data Pipeline Development	Design, build, orchestrate ingestion → transformation → delivery
API-based Data Exchange	Seamless internal/external data movement using APIs
Real-Time & Batch Integration	Kafka, DMS, streaming alongside batch ETL/ELT
Cloud & Hybrid Integration	Modern cloud ingestion into Snowflake, Databricks, Data Lake platforms
Metadata & Lineage-Driven Integration	Business + technical metadata linking; lineage capture
Integrated Data Quality Framework	Rules engine, DQ store, scorecards for pipeline quality
MDM-enabled Integration	Multi-source integration for unified master data
Automation & DataOps	Pipeline automation, CI/CD, DataOps for continuous integration

Benefits of Mphasis Data Integration Solutions

1. Faster, Unified Access to Enterprise Data

Integration enables seamless access across disparate systems, breaking silos and accelerating analytics.

2. Improved Data Quality & Trust

Embedded DQ rules, metadata governance, and lineage ensure accuracy and reliability of integrated data.

3. Accelerated Cloud Modernization

Integration pipelines simplify cloud migrations and support scalable, cloud-native architectures.

4. Reduced Operational Complexity & Cost

Automation, consolidation of pipelines, and optimized data movement reduce ETL costs and maintenance.

5. Enhanced Compliance & Audit Readiness

Lineage, metadata, DQ workflows, and governed integration support regulatory needs.

6. Faster Time-to-Insights for Business Users

Real-time ingestion and Lakehouse pipelines enable rapid analytics, AI/ML and reporting.

7. Foundation for MDM, Data Governance & AI

Clean, integrated data fuels downstream master data, governance, and AI/ML initiatives.

Case Study : Transportation Management Data Hub – Canadian Logistics Company



The Business Requirement



- Transportation Management group (TMS) sources data from multiple business applications at the beginning of business and do intra-day demand & capacity planning of intermodal train car types.
- As they work with latent data, the actual movements & loading operations of Intermodal cars, trucks at their various terminal & yards were not kept current causing under-utilization & unoptimized loading of train cars, delays in train departure & shipments resulting in revenue loss.
- TMS group needed a centralized Data Hub, a single store with low-latent schedule, car-arrival, car-departure data from multiple TMS SoR (legacy, modern apps) for intra-day planning & operations management.

Our Approach



- Build common and canonical data model that meets all the TMS demand and capacity planning, operations and intermodal dashboarding needs
- Implemented Enterprise Messaging Platform (Kafka) and Enabled critical TMS events from the SoR , publish them in canonical format .
- Developed and deployed multiple streaming jobs (Apache Spark, Azure DevOps Ci/CD) which consumes business events from various SoR and aggregated & transforms them into common data model (PostgreSQL) for TMS needs.
- Exposed Data Hub data thru Domain data Services (Java Spring Boot) for various consumes , dashboards (Angular), planning tools etc.

Key Benefits



- The planners across the terminal and yards had a good visibility on the car supply and car demand at a click of a button thru the Intermodal Dashboard, improving their productivity and work efficiency
- Due to low-latent data in TMS Data Hub, the demand & supply planning efficiency improved resulting in higher train car utilization and timely & optimal loading of train cars
- The updated & current car supply data availability resulted in optimal car-supply cut-off times resulting in on-time departure by 50% - 80%

Services Featured



- Enterprise Architecture & Data Architecture
- Data Hub – Platform Solution Architecture
- Design & Implement Enterprise Messaging Platform
- Data Pipelines – Streaming & Data Services
- DevOps for Data Pipelines
- Dashboard Development

Technologies



Cloudera | Apache Spark | Apache Kafka | PostgreSQL | Azure DevOps | ERWIN | IBM MQ | Spring Boot | Angular

Challenges



- Challenges in designing a common Data Model for Transportation Management Data Hub store and the data services
- Enabling business events from legacy application and publishing them to the messaging Platform in canonical form
- Meeting end to end latency with complex transformation and validations

Case Study : Smart Terminal – Event Processing For A Canadian Logistics Company



The Business Requirement



- The Logistics company is in the process of modernizing and automating their operations across their Intermodal terminals in CA and USA and went for implementing a new Terminal Operating System.
- For implementing TOS, one of the key requirement is to expose Car, Gate, Truck and Waybill events happening in the operational application to the terminal automation system at a very low latency
- The key challenge is that operational applications are legacy and not event enabled
- Applications are mostly black box and customization of applications stack for exposing events is not feasible

Our Approach



- Mphasis undertook a detailed assessment of the Order, Terminal and Truck management applications
- An extensive study was done and a non-intrusive mechanism for identifying events from the operational applications was strategized.
- An extensive feasibility study and PoCs were done for arriving at required events from the low-level transaction at very low latency
- Implemented Database CDC and designed and implemented stream processing required events.
- Published Car, Gate and Waybill events for the Terminal Operating System which enabled the TOS implementation

Key Benefits



- The events from operational system were exposed with low latency making it feasible for TOS to operate with maximum efficiency.
- The automation of terminal operation was achieved successfully and was able to meet the ramping, loading, unloading and departure SLAs
- Due to the automation and low-latency events, The terminals were enabled to handle addition 4 x volumes

Services Featured



- Enterprise Architecture & Data Architecture
- Data Hub – Platform Solution Architecture
- Complex Event Processing
- Data Pipelines – Streaming & Data Services
- DevOps for Data Pipelines
- API Development

Technologies



Databricks | Apache Spark | Qlik Replicate |
Apache Kafka | PostgreSQL | Azure DevOps |
ERWIN | IBM MQ | Spring Boot | Angular

Challenges



- Enabling the event with low latency from 5 secs to 50 secs from legacy applications
- Gleaning the database transactions for identifying business event and validating the same for various scenarios
- Technology challenge in processing the CDC stream and enriching the event payload

Case Study : Database Migration for a leading Logistics company in Canada



The Business Requirement



- A Leading Logistics in Canada wanted to migrate their existing on-prem Data Warehouse Appliance Netezza to Cloud
- The scope of migration included the Historical Data from Netezza and Daily incremental load from multiple Data Source types like Oracle/SQL/XML/Excel/CSV etc., feeding Data to the Warehouse
- There were around 1200+ Tables across 40+ Different DBs. They want us to migrate all the DWH to Azure Databricks and eventually to Snowflake.

Our Approach



- Study of the Warehouse Migration requirements, Infrastructure, network Capability, Storage capacity needs.
- Strategize and architected the migration process and design
- Migration Engine with Asynchronous Handshake mechanism to effectively bring down the overall Migration Execution Time
- Definite Process of categorizing the tables into buckets of T-shirts and apply slicing accordingly.
- Data Mapping across Netezza to Databricks Delta formats
- Fully Audited Process with detailed Logging in the Metadata Table

Key Benefits



- Data Migration using the Migration engine reduced the migration timeline and manual effort
- Migration of Huge Tables (Terabytes of Data) with effective execution times to run during non -batch hours
- Easy to use, Plug and play Migration engine with configurable Data Sources
- Optimal computing & storage cost and platform efficiency by utilizing Open file format
- Completed the migration in stipulated time

Services Featured



- Data Migration Engine
- Data Lake / Delta Lake / Data Pipelines
- Distributed Computing using Databricks

Technologies



- IBM Netezza
- Azure Data Factory
- Azure Databricks & Delta Tables
- Azure DevOps
- Spark SQL & PySpark
- Apache PyArrow / Parquet

Challenges



- Data Volume (1200 tables & 60 TB) and Timeline
- Automation & Orchestration of migration process with Data verification testing & Audit

Case Study : Data Analytics Platform Modernization for a Fleet Management Company



The Business Requirement



- The Fleet Management company in the process of upgrading their key ERP application to a latest version and were planning to migrate their transactional systems to Cloud.
- As part of the Cloud journey , they also wanted to move their analytical workload to Cloud Data Warehouse (Snowflake) but were facing following challenges :
 - The DW platform was grown ad-hoc & tactically over a period leveraging various technology options
 - Zero Documentation on the current data platform and it stays as a black box
 - Need to retain the external customer facing report as-is in spite of the ERP upgrade

Our Approach



- Mphasis proposed and executed an Assessment of Current Data Platform and Reverse Engineering exercise and accomplished the following :
 - Customers current ETL and Data Warehouse process was studied and documented thru multiple session with the technical owners
 - A reverse engineering exercise was performed on the SSIS packages (manual & automated) and the Store Procedures in the SQL Server database to understand the data flow, data lineage and business logic
 - Proposed high level architecture and roadmap for data platform modernization in line with the ERP upgrade
- Architected and Designed Serverless Data Pipelines on Azure for Extracting & Loading and Data Transformation on to Snowflake DW . Proved the Concept
- Currently in the process of designing the DW for self-service reporting , building and implementing the Data Pipelines

Key Benefits



- Canonical DW and Self-Service platform addressing all the reporting needs
- Serverless Data Processing , right from extraction to transformation resulting in lower TCO
- Rationalize and Optimized target Data Model resulting in data redundancy elimination, trusted , NRT and Quality data source for reporting

Services Featured



- Data Architecture & Analytics
- Assessment and Solution Architecture for Data Platform
- Proof of Concept for Solution & Technology Platform
 - CDC to DW - Serverless
 - ELT for DW stage to DM
 - Event Hub for EAI & CEP

Technologies



Azure Cloud | ADLS G2 | SQL Server CDC | Azure Data Factory | Event Hubs | Logic Apps | Tableau | Snowflake

Challenges



- Zero-Documentation
- Silo-ed DW and DM tables
- ETL & DM rationalization
 - Very complex and redundant ETLs in SSIS
 - Redundant Data Objects
 - High Data Sprawl
- Availability of SME time and low quality information

Case Study : Data Analytics Platform Modernization for a Global Consulting Company



The Business Requirement



- The leading consulting company , which has divisions across the globe , was having problems in getting reliable reports , in timely manner. The management dashboards were not rendering consistent data.
- Reporting at the country , regional and LOB levels were done by wrangling the data sourced from canned reports with high latency and ad-hoc queries, by 400+ analysts across the organization.
- The data aggregation , enrichment and cleansing were done by every analyst locally resulting in inconsistencies between the country, regional and global reports .
- Lack of single integrated view of enterprise data resulted in exorbitant redundant effort for reporting

Our Approach



- Mphasis executed a Design Thinking workshop with the client stakeholder for 7 weeks (part-time) and accomplished the following :
 - Interviewed multiple stakeholders from IT, Analyst group, end users, empathized , documented the pain areas, challenges, concerns and expectations
 - Validated the cause (architecture, technology, process etc.) for pain points and challenges in the current reporting platform and process
 - Ideated and drafted a Cloud based Data Platform architecture with Data Lake foundation, with data pipelines on cloud agnostic technology feeding the Data Lake, ODS, DW and DMs, in near-real time, powered by Dev - Sec-Ops
 - Prioritized the epics and user stories (architecture, design, platform, product etc.)
 - Executed Proof of Concept (4 weeks) to validate and built prototype for the Clients Data Analytics Platform and demonstrated.
 - Design and Development of DAP Platform on Azure with DL foundation , Talend Data Pipelines and ETL lifecycle Automation

Key Benefits



- Data Analytics Platform with data warehouse & data marts for unified analytics & accelerated insights
- Reduction in analyst effort for analytics, dashboarding & reporting since the data is cleansed transformed and managed centrally and consumed across the organization
- The Data Catalog platform helped Analyst to understand the enterprise data and the data lineage , which helps in creating new dashboards faster with right and quality data

Services Featured



- Data Architecture & Analytics
- Design Thinking and Solution Architecture for Data Platform
- Proof of Concept for Solution & Technology Platform
 - Data Lake on Cloud
 - Data Pipelines & Data Sec Ops
 - Data Security & Data Catalog

Technologies



Azure Cloud | ADLS G2 | Azure DevOps | AKS | Talend | Talend Data Catalog | Tableau

Challenges



- PoC setup took longer time than expected due to Cloud Platform Security and Data Security configurations (IAN roles and privileges)
- Challenges in building, containerizing and deploying the Talend pipelines on Azure Kubernetes Service through Azure DevOps
- Configuring the Talend Data Catalog harvester for datasets on cloud (ADLSG2) and establishing semantic lineage

Case Study : DB Purge for Major Telecom in APAC



The Business Requirement



- A Leading Telecom Giant in Australia spread across APAC and EU region needs a partner to help them in purging business data pertaining to other countries in the Siebel and WBI application databases which was cloned from a parent company's common operational application for the entire region
- Purge all other country business data from a shared application database for the entire region
- Minimal Production downtime for cutover
- Minimal Production purge impact
- 70% of space to be reclaimed at file system after purge

Our Approach



- An offline purge and data verification production process was designed
- A complete purge cycle was executed and the data anomalies sorted out
- A clone of database was then purged using the SQL scripts and the application utility in 45 days non-stop.
- The purged database was re-organized for space reclamation and synced up with production thru GoldenGate
- The entire purge had no impact on the production database as it was done off -line
- A cutover was done to make the purged database as the production with minimal down time

Key Benefits



- 3TB of space reclaimed at file system level which reduced the TCO by saving 9 TB of SAN disk space (prod , reporting and the backup volumes)
- The performance of the application improved 200% and the reports by 300%
- The reduction of administrative overheads due to reduced database size were tremendous.

Services Featured



- Data replication using Oracle Golden Gate for migration
- Data purge using SQL scripts and application utilities

Technologies



- Oracle Goldengate, Oracle 10G,
- Unix Scripting

Challenges



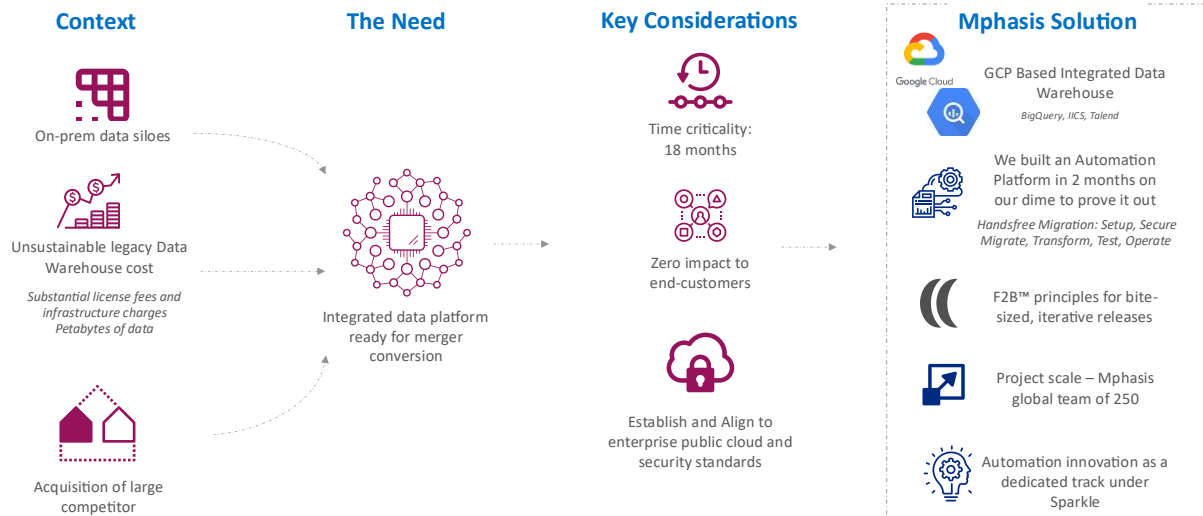
- The size of the Siebel database was 4.5TB and WBI database was 1TB where the data needs to be purge were 3 TB and 20 GB respectively.
- The purge operation should not impact the performance and the CPU bandwidth of the production system.
- The 3TB space from the Siebel database to be recovered and the performance of the database to be improved

Case Study : Data Modernization & Integration for a Wealth Manager & Custodian



Enabling Data on Cloud

Transforming data is key to delivering transformation programs successfully



Case Study : iSeries DB Migration to Snowflake on AWS



Business Requirement



- Western Union is undergoing architectural transformation from legacy to cloud platform. Transaction Database, currently on iSeries, is a part of this transformation
- Settlement is WU business-critical function on iSeries and ORMB systems. Transaction Database supports settlement processing on both these platforms. Eventually, Settlement will only happen on ORMB, which is one of the objectives of Settlement Transformation Initiative
- Transaction Load process, on iSeries, receives and processes raw transaction feeds from Unisys (POS) System and stages transactions onto Transaction DB for Settlement
- In a nutshell, Transaction DB, along with all integrated processes (e.g. Transaction Load, Transaction Enquiry & Error Correction Screens) need to be migrated to the new cloud platform

Our Approach



- Study of the business & database migration requirements, down time tolerance, SLAs, Infrastructure, network capability, storage capacity for migration.
- Overall Migration Design
- Normalize data
- Data migration using Talend ETL tool
- Data Validation / Verification Reports using MDCert Comparison tool

Key Benefits



- Alignment with Western Union 1P Standards
- Support more frequent, configurable, regional settlement cycles with increased throughput
- Provide near real-time reporting i.e., increased visibility for Agents
- Meet business needs of Agents with improved funding cycles

Data Validation



- MDCert data comparison tool used for data comparison for quantitative and qualitative validation
- Transient EMR cluster (m4.2x large – 8vcore – 32GB memory) 4 Instances and 1 master instance (m4.2x large - 8vcore – 32GB memory)
- Data compared between AWS S3 (csv) and Snowflake Database
- Volume of data ranging from 5k to 1million records
- CSV File Size vary from 1MB to 1GB
- Certify Data Verification at database and application levels

Services Featured



- Legacy On-premises to Cloud database Migration
- Data Feed from external vendor into AWS S3

Technologies



- Snowflake DB on AWS Cloud
- Okta Authentication
- Data migration using Talend ETL
- Git and Jenkins Pipeline
- Apache Airflow for Job Scheduling

Accelerators



- MDCert – Data Validation & Report

Challenges



- Migrate positional data to Snowflake normalized tables
- All data feed processed in Snowflake to be kept in sync with iSeries until a desired point-in time.

Case Study : Marketing Analytics Data Lake for Leading Logistics Company



The Business Requirement



- Client's Retail / Office IT and Customer Engagement Marketing group wanted to improve their customer experience by understanding customers behavior across various channels. This required a :
- Digital Retail Marketing Data Platform to enable Marketing Teams and Executive Management have customer behavior data correlated in a central data lake
 - Platform that enables marketing and leadership teams to extract insights from data lake
 - And to operationalize actionable insights to drive efficiency in marketing and sales and optimize customer experience

Our Approach



- Build an integrated data foundation / platform to drive advanced analytics necessary to deliver cross-channel personalized journeys by
- Extracting marketing, social and customer related insights from correlated data within EDW, CRM and Marketing systems
 - Integrate and correlate behavioral and transactional data across the funnel
 - Setup a Data Science CoE to extract insights from correlated data
 - Tangible Impact towards Efficiency, Customer Experience and Sales/retention measured and reported through KPI dashboard
 - Phased and incremental approach to mitigate risks and balance cost to value delivered over time

Key Benefits



- Ability to identify customer across channels and businesses to optimize and increase cross selling and upselling
- Ability to measure key KPIs like Click to Lead Rate, Lead Conversion Rate, Cost Per Conversion etc. to optimize and plan the advertisement and Campaign budgets across Enterprise
- Ability to predict the customer lifetime value, future potential high net worth customers etc. that play vital role in designing targeted & Customized customer journeys

Challenges



- Source systems range - legacy to modern platforms like Adobe Analytics, Sales Force Cloud etc.
- Lack of enterprise level standardized platform to enable Marketing Analytics at scale
- Lack of analytical platform that support the continuous analytical use case evangelism independent of volume, variety, velocity complexities

Services Featured



- Digital Marketing Analytics
- Data Architecture
- Data Lake – Platform Solution Architecture
- Data Lake - Design & Implementation
- AI / ML model development and deployment

Technologies



- Hortonworks Data Platform (HDP) | Spark | Hive | Sqoop | Python | One Automation

Case Study : BI Platform Consolidation & Rationalization



About the Client - A leading North American business-to-business distributor of packaging, facility solutions, print and publishing products and services. - This company is a provider of Logistics and Supply Chain Management solutions serving customers in approximately 170 operating distribution centers throughout US, Canada & Mexico	Engagement Overview - Roadmap for Datawarehouse Intégration - Enterprise Datawarehouse (EDW) Development & Implementation - Application Support & Maintenance with Major Enhancements - Implemented High Maturity model for AOS.	Key Benefits - Timely availability of critical operational and management reports for both the companies - Accuracy of company's delivery forecast. - Process Efficiency ~ 40 % of daily processing time - Productivity Gain:- Legacy programming efforts saved
Business Scenario - After the merger of 2 companies, the new entity had a challenge to have a integrated view of orders for unified company, as legacy system was unable to send the file to multiple location at the same time and using same program. - High Processing time for each locations - A deterrent data flow and data integration across different DWH systems - Large Maintenance & Programming efforts	Mphasis Solution - End-to-End engagement from assessment phase till Application Maintenance & Support. - Adoption of effective Data Modeling technique to consolidate and aggregate different data sources. - Integration across different interface systems for standardized processing. - Optimized team to Support including Automated Batch Processing	Technology Stack - Data stage 8.7 - SAP BI 4 (Business Objects) /Crystal 2013 - Oracle 11g R2 and SQL Server 2008 - MS .Net / Share Point - WebLogic, Oracle RDBMS - IBMi (Formerly known as AS400/iseries) - RPGLE, CL, DB2, MQ Series - Supply chain products (Red Parrie)

Case Study : Enterprise Datawarehouse Re -Design



About the Client Customer is a leading mutual insurance company that provides liability and property coverage, as well as related risk management services, to the utility and energy industries	Engagement Overview - RE-Architecture and consolidate Data Model and ETL to cater to the legacy as well as the New Pega system - Re-Architect the Report design for faster delivery of reports	Technology Stack - Infosphere DataStage - SSIS - Cognos 10.1.1 - SSRS - Sql Server 2012
Business Scenario - Poor Data Management; related information was not modelled appropriately - High turn-around time for new report generation - High maintenance and support - Lack of analytics insight into business operations	Mphasis Solution - A complete study of existing ADW system and its objects - Re-Architecture and design specially aimed at Consolidation and Data Model - Build a new data model that caters to the Legacy system and the new PEGA application - Reporting Architecture design to provide Faster Reporting Platform - Develop ETL framework for Error Handling, Audit Log of SSIS tasks - Build stringent Data Validation rules - Develop 105 Cognos reports as per the new data model	Key Benefits - Improved performance of reports by 20% - Policy and Claims data tightly integrated with - Single version of truth - Robust Data Warehouse to handle on-boarding of new applications/Source Systems - Scalable and Lean data warehouse where the data model can be enhanced in future

Further Information: For more information about our G-Cloud services, please contact our Public Sector Team at mphasisukps@mphasis.com

