GlobalLogic
A Hitachi Group Company

GlobalLogic Data Platform **Accelerator**

An end-to-end flexible data platform with pluggable, modular components that enables you to turn data into knowledgeable and actionable insights by gathering, transforming, and enriching data from across your enterprise.

Implement a fully functional data platform within days, not months

Although extracting value from data across an enterprise is imperative, many businesses struggle with the challenges of big data architecture, technology stacks, and hosting capabilities that are required for ramp-up.

Leveraging cloud native PaaS technologies, the Data Platform Accelerator helps businesses gather enterprise-wide data, transform and enrich it, and deliver it to analysts for extracting value and insights. The GlobalLogic Data Platform provides the option to go with the complete suite or pick and choose from the catalogue of Accelerators within the Data Platform.

Data Integration Accelerator
Provides the component library required to create data integration pipelines across different sources and targets using Apache Spark

Data Quality & Validation Accelerator
Provides the component library required to perform different validation & quality checks across source and target data sources

Data Orchestration Accelerator
Provides the sdk & library required to create orchestration pipelines on top of Apache Airflow

Data Console Accelerator
Provides the portal and management console required to configure data access, pipelines and monitor jobs
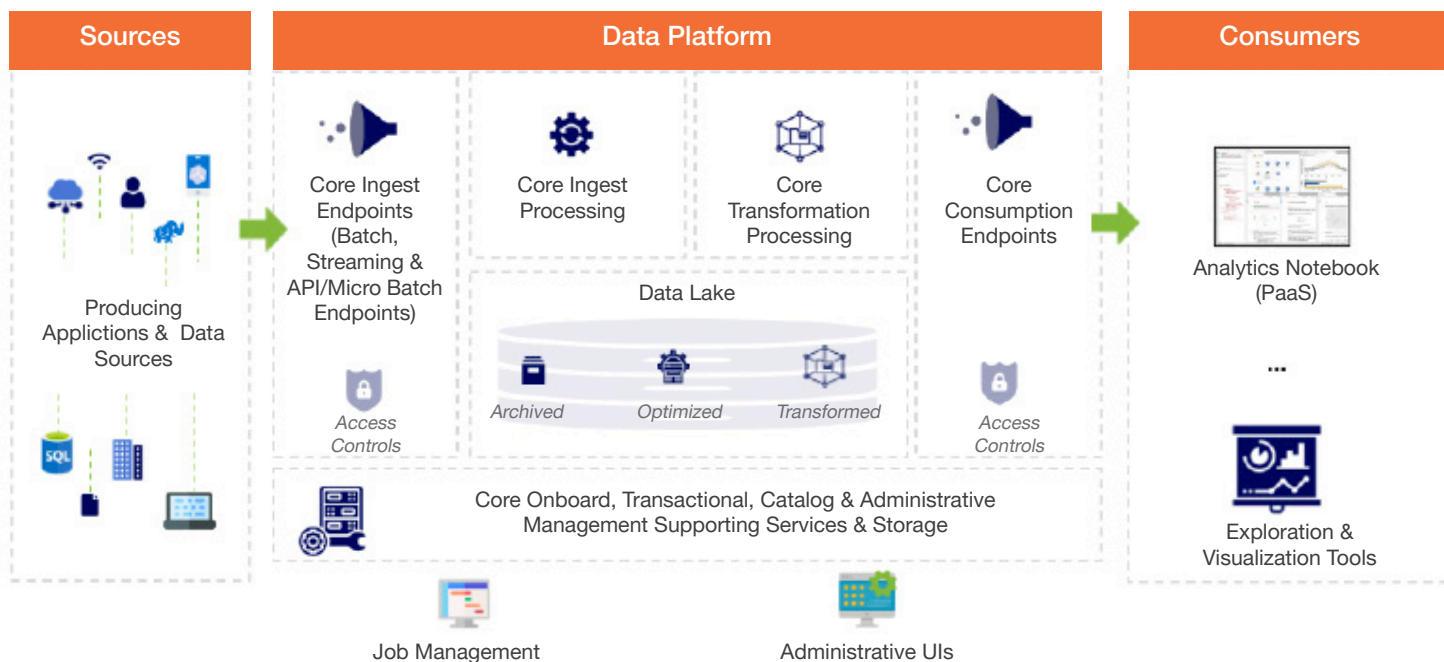
Data Plane Accelerator
Provides the component library required to perform data profiling, classification and masking on source datasets using Apache Spark.

The accelerators integrate well with each other to enable modularity, flexibility and extensibility.

GlobalLogic
GlobalLogic
A Hitachi Group Company

GlobalLogic
Data Platform
Accelerator

Sources	Data Platform				Consumers

Core Ingest Endpoints (Batch, Streaming & API/Micro Batch Endpoints)

Core Ingest Processing

Core Transformation Processing

Core Consumption Endpoints

Producing Applictions & Data Sources

Access Controls

Data Lake

Archived   Optimized   Transformed

Access Controls

Analytics Notebook (PaaS)

Exploration & Visualization Tools

Core Onboard, Transactional, Catalog & Administrative Management Supporting Services & Storage

Job Management

Administrative UIs

Add value through standardization, consistency, and resuse

The Data Platform Accelerator enables businesses to quickly deploy a centralized, next-gen platform to store and process data — even with limited technical knowledge. The accelerator adds significant value by:

- Eliminating silos that create roadblocks in generating insights
- Enabling consistency around infrastructures, technologies, pipelines and processes
- Providing transparency around available data/adaptors/pipelines to eliminate inconsistencies and duplications
- Manage tedious, non-trivial and lengthy process to provision, configure integrate with ease
- Enabling role-based access control across data, data processing, platform configurations and APIs
- Jump-starting the data science team's ability to create AI/ML models

GlobalLogic
Data Platform
Accelerator

The platform comes with the following features:

Deployment

- Automated installation
- Fully automated resource provisioning for Data Pipeline execution
- Fully automated resource provisioning to enable Consumption

Data Ingestion

- Supports structured batch-push from blob storage
- Supports data files in CSV, JSON, Avro, Parquet and ORC data formats
- Supports structured batch-pull from RDBMS / RDS and Snowflake Data warehouse

Data Integration & Transformation

- Out-of-box data integration & transformation powered by PySpark component library
- Ability to integrate with Reporting solutions and Analytical tools

Polyglot Storage

- Standard data integration zones (raw, staging, refined)
- Well-scoped hot and cold / archive zones
- Supports Blob, RDBMS and data warehouse storage

Data Quality & Validation

- PySpark based framework with pre-built component library for common validation rules
- Run validation jobs separately and persist validation report across all entities and rules

Control Plane

- Data Source Management and Schema registration
- Project Management
- User Registration

Monitoring

- Job Monitoring and Management
- High-level job tracking and monitoring
- Self-service analysis and troubleshooting

Ingress / Access endpoints

- API and JDBC/ODBC endpoints for enabling ingestion and data usage by external analytics solutions

Security

- Flexible Role Based Access Control based on Azure AD or AWS IAM
- Ownership and Access control
- Integration with AWS KMS or Azure Keyvault

Data Orchestration

- Template based pipeline orchestration accelerator based on Apache Airflow / Azure Data Factory
- Speeds up, simplifies and unifies DAGs development experience

Data Plane

- Profile the data source and generate common statistics on relevant columns
- Identify entities and columns with sensitive data based on column names and column value patterns via configurable regex patterns
- Mask or encrypt the sensitive column values using predefined masking algorithms