NTT DATA

Data and intelligence

Data management – enhancing your data analytics capabilities

"

Data intelligence is the ability to comprehend and use data by connecting the right data, insights and algorithms. It **helps deliver outcomes to the data citizen at any level by optimizing processes, increasing efficiency and driving innovation**.

We provide **frameworks, templates and, more importantly, the services to build and deploy advanced data intelligence solutions, complemented by tools, platforms and solutions that help to unlock the full potential of your data**.

services.global.ntt
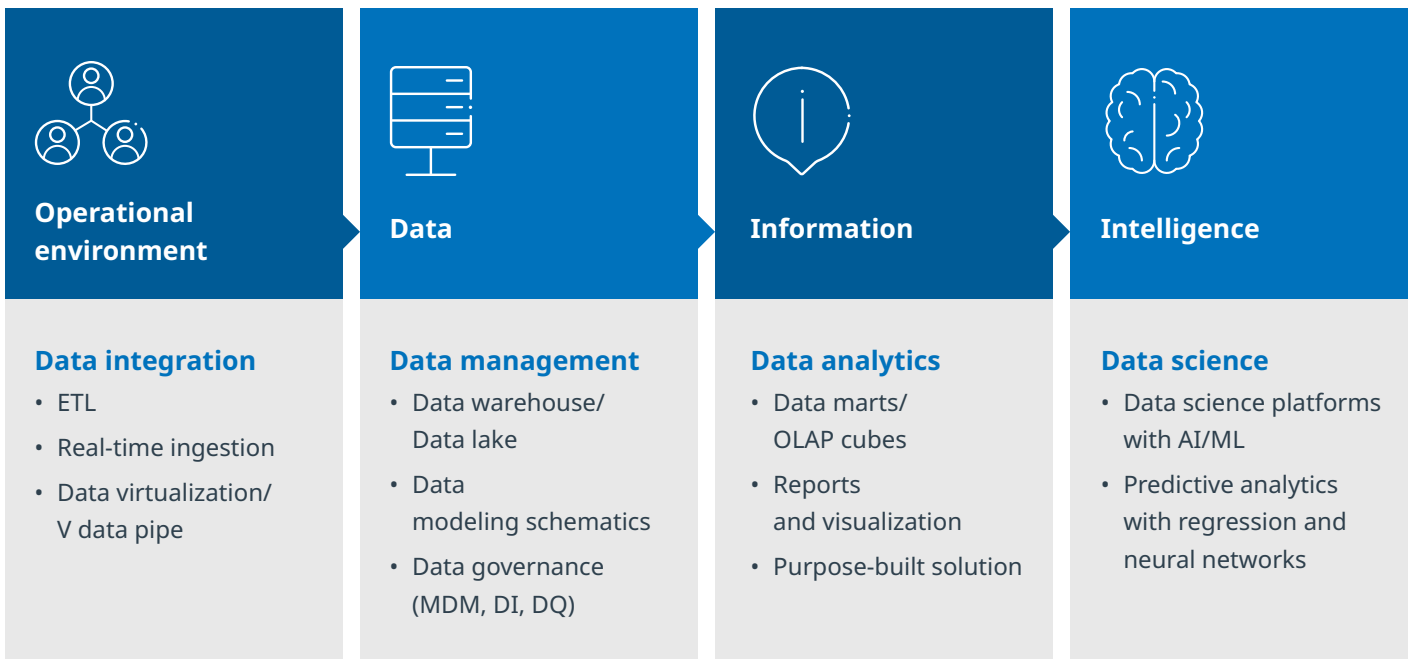
Business analytics

Business analytics is the catalyst that helps organizations set achievable goals. Over the years business analytics has helped organizations better understand business parameters and enabled them to adapt when the need arose.

We live in the age of big data. Digital systems generate large quantities of information that can be analyzed to provide insights into business operations, customer preferences and market trends. As businesses continue to grow, information has become a valuable asset. Data science and analytics tools are more relevant than ever to the growing enterprise. Businesses looking to win the race to exploit big data require sophisticated, scalable and robust business analytics solutions.

Our Digital Business Solutions deliver business analytics services to help you mine digital gold from the data you collect, offering end-to-end data processing, from extracting the data from core modules to Business Intelligence and Machine Learning.

We've helped enterprises and businesses with business analytics – levering massively parallel processing databases, big data, Online Analytical Processing (OLAP) cubes and Graphics Processing Unit (GPU)-accelerated data science platforms – in areas such as:

- Enterprise planning and budgeting
- Product pricing
- Cost allocations and transfer pricing with a focus on specialized solutions in the banking, financial services and insurance sectors
- Segmentation and profitability analytics including customer, product, market segment and region
- International Financial Reporting Standards (IFRS) and Generally Accepted Accounting Principles (GAAP) treatments

Operational environment	Data	Information	Intelligence
Data integration	**Data management**	**Data analytics**	**Data science**
• ETL • Real-time ingestion • Data virtualization/ V data pipe	• Data warehouse/ Data lake • Data modeling schematics • Data governance (MDM, DI, DQ)	• Data marts/ OLAP cubes • Reports and visualization • Purpose-built solution	• Data science platforms with AI/ML • Predictive analytics with regression and neural networks

Organizations require four basic building blocks to transform data to useful information they can use to automate their business operations, such as personalized offerings and competitive pricing, or improve clarity on key business drivers, and make informed business decisions.

The four building blocks are:

- Data integration
- Data management
- Data analytics
- Data science

1. Data integration

Data integration is the practice of consolidating data from disparate sources into a single dataset, with the goal of providing users with consistent access and delivery of data across subjects and structure types, and to meet the information needs of all applications and business processes.

Data integration

Data modeling

Data Profiling

Data cleansing

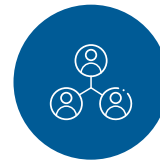**Integration of structured/ unstructured data**

Data merging migration & replication

**ETL/ELT/ETLT design and development**

Data warehousing

Legacy and 3rd party interface

1.1 Real-time data integration

In a world where getting results as quickly as possible is increasingly important for business operations, real-time data processing is a critical resource. Customers expect real-time resumption of the services once they make their subscriptions/payments for services such as their mobile or cable TV. There's also a need for customers and organizations to be able to detect fraudulent use of credit cards in online and legacy payment channels. Having

understood this need, being able to collect data associated with a payment and determine whether it matches fraud signatures in real-time – or at least within a few seconds – is vital.

Real-time data integration means that different applications and databases are connected in such a way that if data in one system changes, it's simultaneously updated everywhere else.
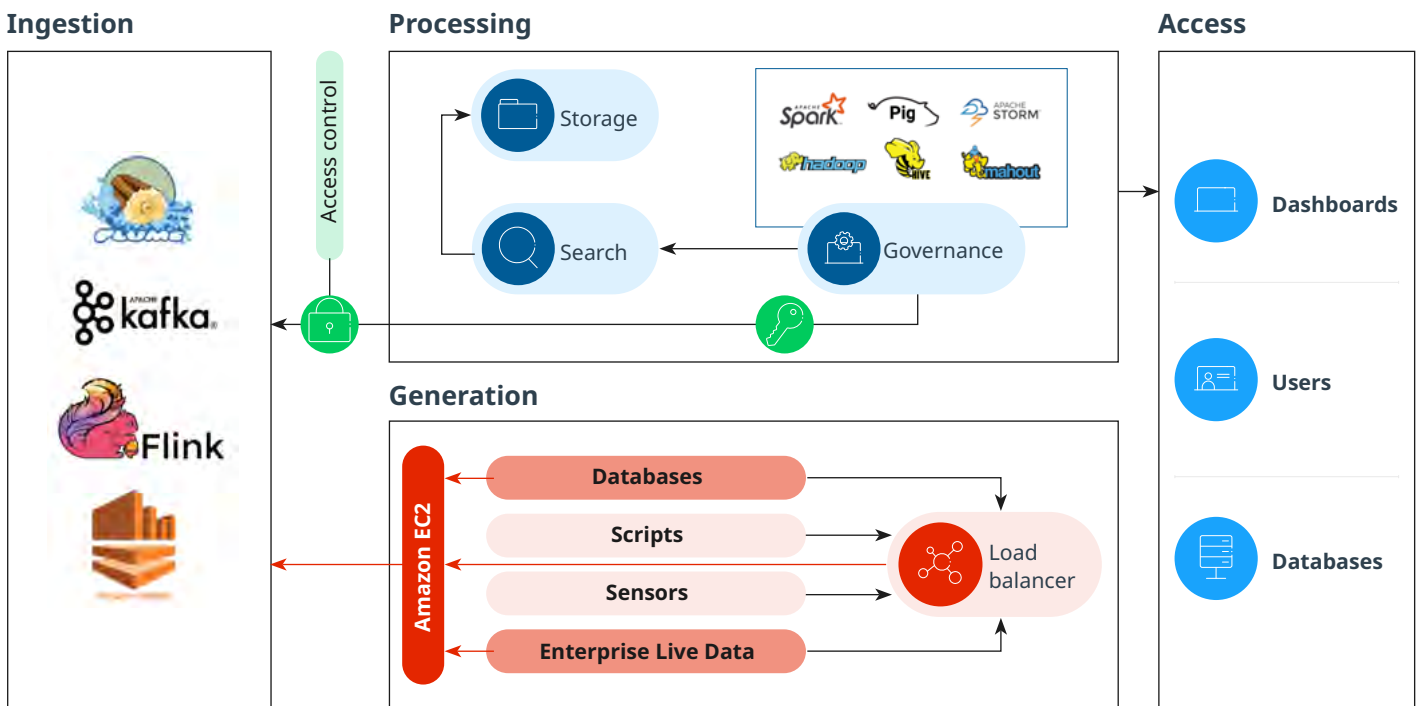
These solutions enable customers to obtain real-time solutions from customer service centers leveraging artificial intelligence and machine learning to remediate any issues faced during, for example, mobile banking onboarding.

Real-time data integration solutions are available for both on-premises and cloud models for several use cases which require specific functionalities such as:

- Legacy and relational database real-time data integration.
- Stream live updates.
- Change data capture
- IoT streaming data
- Big data ingestion
- Cloud data services

1.2 Data virtualization

Organizations have traditionally built and deployed systems to serve specific goals such as customer maintenance, product management, sales and operation, and finance and accounting. This encouraged the development of information silos. Data virtualization allows you to unlock these isolated sets of data, enabling you to view data stored in heterogeneous platforms as though they were stored in a single location.

Data virtualization allows an application to retrieve and manipulate data without requiring technical details about the data, such as how it is formatted at source or where it is physically located and can provide a single customer view (or a single view of any other entity) of the overall data.
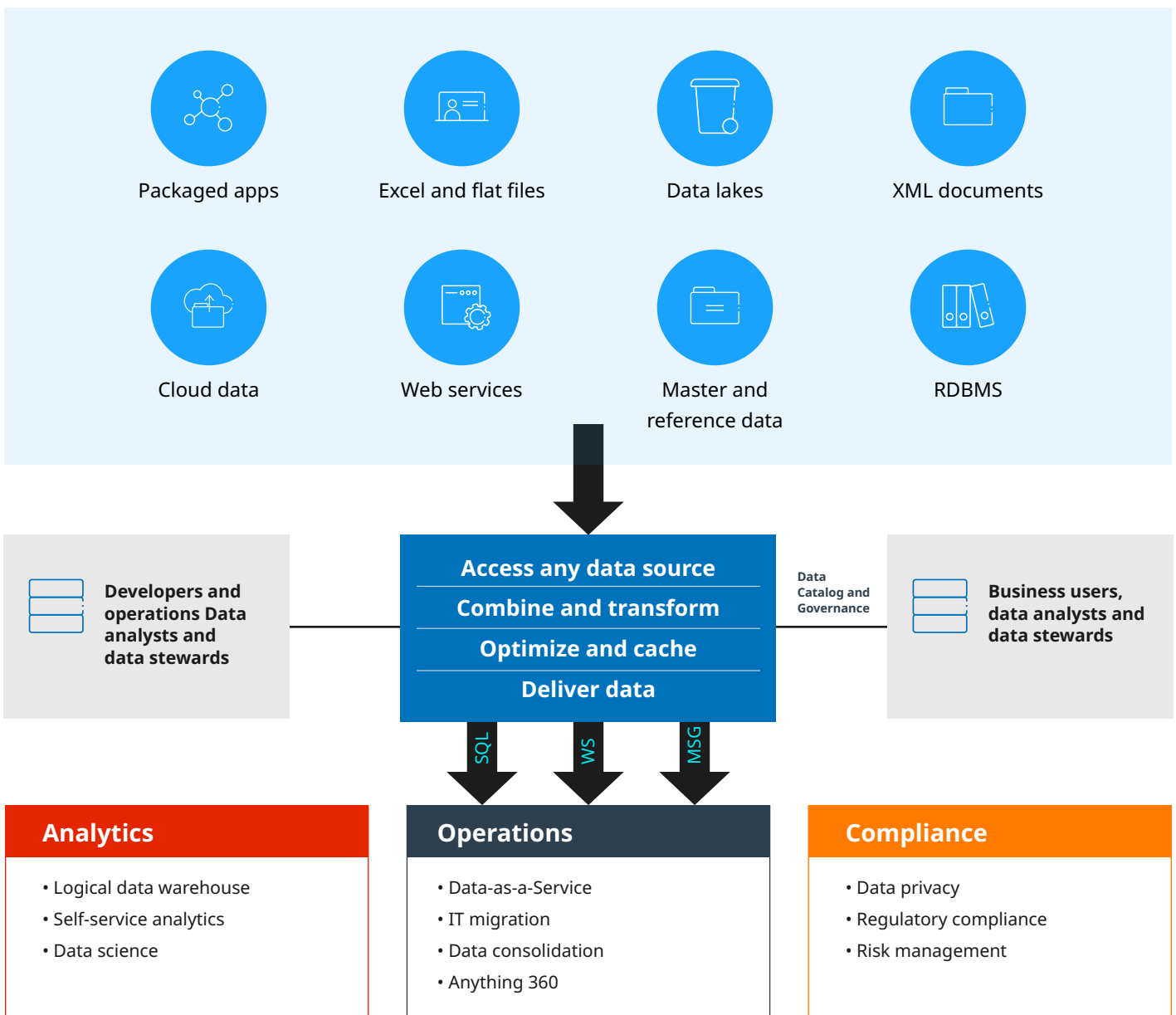
This creates a virtual data pool providing unified data services to multiple users and applications. This gives them faster access to the data, cuts down on replication, reduces costs and increases flexibility.

Data virtualization software lets you integrate data faster and more cost-effectively. It allows you to build and manage virtualized views and data services that access, transform and deliver the data your business requires to accelerate revenue, reduce cost and risk, improve compliance and more.

There are three steps to building virtual data services:

- **Connect and virtualize any source:** Quickly access disparate structured and unstructured data sources using connectors. Bring the metadata on board and create normal source views in the data virtualization layer.

- **Combine and integrate into business data views:** Integrate and transform source views into typical business views of data. This can be achieved in a GUI or scripted environment.

- **Publish and secure data services:** Turn any virtual data views into SQL views or a dozen other data formats.

How it works

Packaged apps	Excel and flat files	Data lakes	XML documents
Cloud data	Web services	Master and reference data	RDBMS

Developers and operations Data analysts and data stewards	**Access any data source**	Data Catalog and Governance	Business users, data analysts and data stewards
	Combine and transform		
	Optimize and cache		
	Deliver data		

SQL   WS   MSG

Analytics	**Operations**	**Compliance**
• Logical data warehouse	• Data-as-a-Service	• Data privacy
• Self-service analytics	• IT migration	• Regulatory compliance
• Data science	• Data consolidation	• Risk management
	• Anything 360	

- **Global metadata:** Global information search capability lets users access data through any format from anywhere in the world.

- **Hybrid query optimization:** Allows for the optimization of queries, even with 'on-demand pull and scheduled batch push data requests.

- **Integrated business information:** Data virtualization brings users integrated information while hiding the complexity of accessing varied data streams.

- **Data governance:** The data virtualization layer serves as a unified layer to present business metadata to users. Simultaneously, it helps to understand the underlying data layers through data profiling, data lineage, change impact analysis and other tools and exposes the need for data normalization/quality in underlying sources.

- **Security and service level policy:** All integrated data virtualization data views can be secured and authenticated to users, roles and groups. Additional security and access policies can manage service levels to avoid system overuse.

2. Data management

Organizations have been leveraging data to deliver business outcomes by monitoring processes including sourcing, manufacturing, inventory, sales and customer satisfaction, and innovating where necessary.

For these processes to be efficient and effective, data must be consistent, organized, complete and, most importantly, accurate. With the right data, any organization can develop and grow.

How do organizations create trust in their data, both internally and externally? And how can they begin to manage the overwhelming amount of data they now have?

By practicing good enterprise data management.

Data management is the process of ingesting, storing, organizing, and maintaining the data created and collected by an organization. Effective data management is a crucial piece of deploying the IT systems that run business applications and provide analytical information, to help drive operational decision-making and strategic planning by corporate executives, business managers and other end users.

The data management process includes a combination of different functions that collectively aim to make sure that the data in corporate systems is accurate, available and accessible. Most of the required work is done by the IT and data management teams, but business users typically also participate in some parts of the process to ensure that the data meets their needs and to get them on board with the policies governing its use.
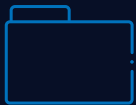
Organizations need to build a solid data management foundation to ensure that they can understand and grow their business.

**Data warehousing:
convert your data into useful information**

Our services include:

- Logical and physical data modeling
- Master data management
- Data transformation
- Data quality management

Data quality and governance: validate the accuracy and timelines of data as it flows from the source to the data warehouse or business intelligence systems.

Our services include:

- Providing scope for data governance planning and implementation.
- Implementing change management and other data quality projects.
- Ensuring that your business is following all data processing regulations.
- Ensuring that legacy systems contain updated data and follow current standards.
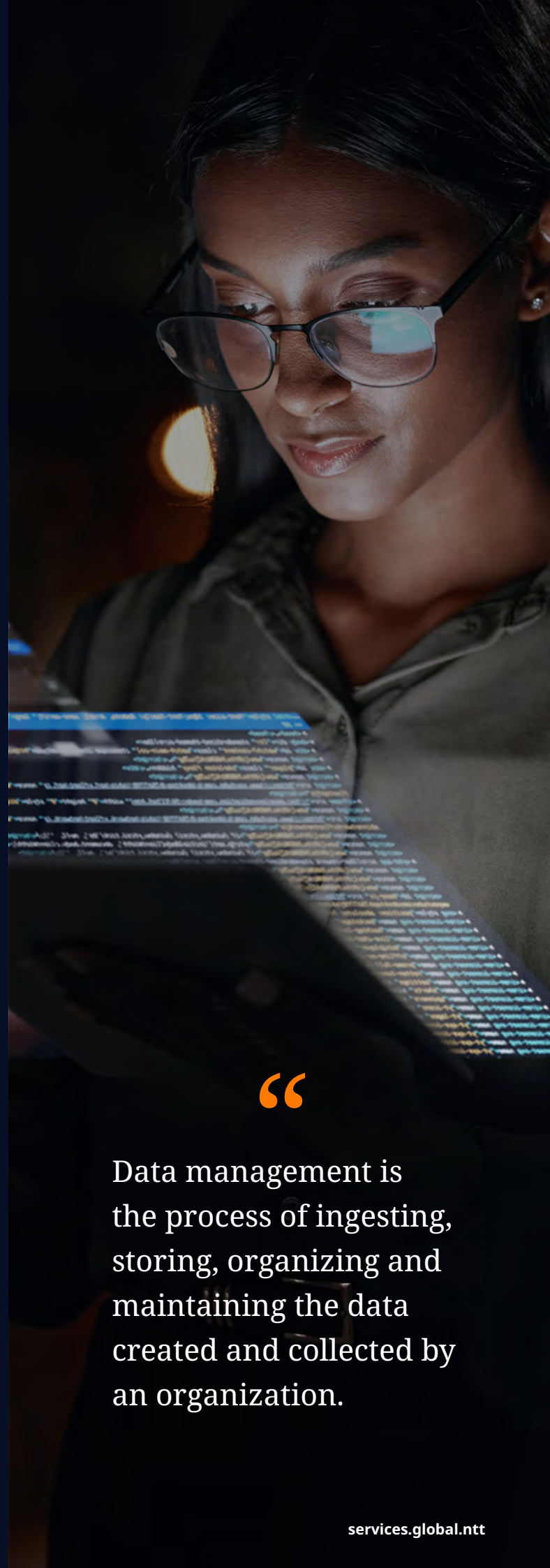
"

Data management is the process of ingesting, storing, organizing and maintaining the data created and collected by an organization.

services.global.ntt

2.1 Data warehouse

Data warehouses offer the most reliable and accurate way for organizations to store and access structured data; this, in turn, improves cross-organizational data access via reports, dashboards and analytics tools. These help businesses better monitor performance and improve decision-making because they know their data is trustworthy.

Data warehousing is a process for collecting and managing data from varied sources to provide meaningful business insights. A data warehouse is typically used to connect and analyze business data from heterogeneous sources. The data warehouse is the core of a business intelligence system built for data analysis and reporting.
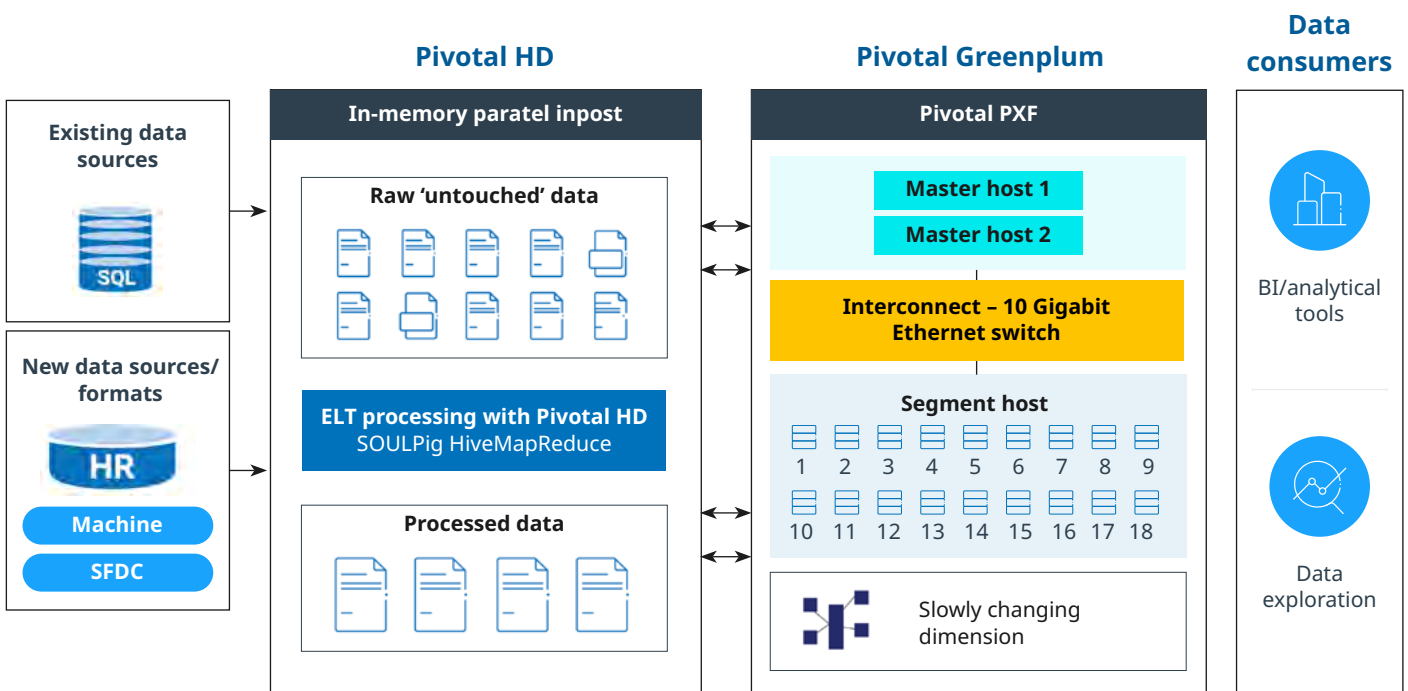
A data warehouse consists of a technical platform made up of:

- **A platform for an enterprise-wide** data warehouse with massively parallel processing capabilities.
- **A standardized data model** of a conformed data warehouse, including a system of records and aggregation.
- **A foundation** for data mart consolidation.

2.2 Big data

Enterprises usually implement and operate many purpose-built systems to deliver specific functions. Those systems in turn generate huge amounts of operational data. In the case of a financial institution, data could originate from point-of-sale terminals, merchants, credit card providers, card issuers and customer interactions related to the product or services they're buying. Usually, this data is received and stored within a sub-second range of time, resulting in millions of data points. There is a real-time necessity for the organization to determine the validity of the transaction from the angle of the customer, merchant and the financial institution itself by executing several complex fraud rules.

After the financial institution has established the validity of the transaction, they can explore potential upsell opportunities for themselves and their merchants. In the past decade, this zeal to perform real-time monitoring of transaction fraud and scanning for upsell opportunities could only be performed after the fact in a batch-processing manner. Now modern data management technologies such as big data and data science platforms can deliver those same results in real-time.

Big data is a field that analyzes and systematically extracts information from data sets that are too large or complex to be dealt with by traditional data-processing applications.

Our services include:

- Structured and unstructured data: our big data solutions framework handles both structured and unstructured data from large data source systems.
- Tailor-made solutions: our customized solutions uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful information that facilitate informed business decisions.

2.3 Data lake

A data lake is a system or repository of data stored in their raw format, usually object blobs or files. This repository provides a single location for all enterprise data including raw copies of source system data and transformed data used for tasks such as reporting, visualization, advanced analytics and machine learning.
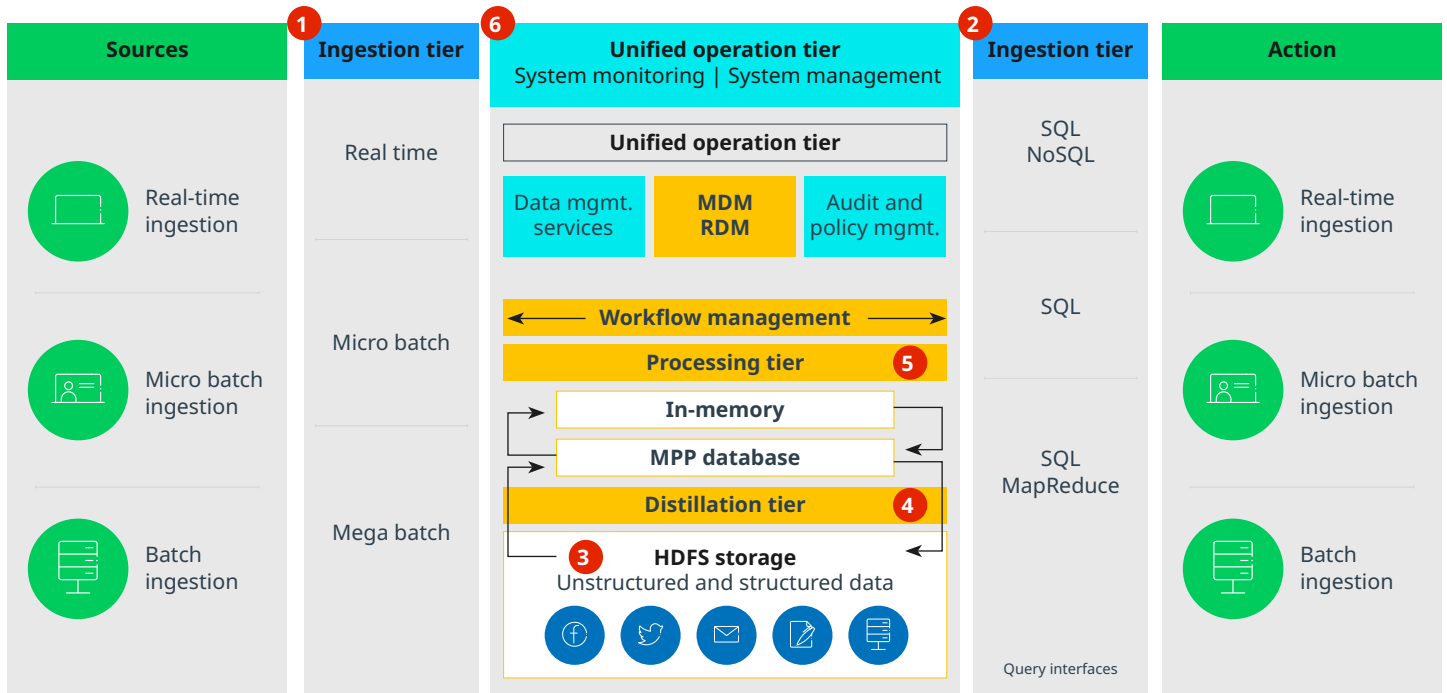
Important tiers in a data lake architecture are:

1. **Ingestion tier:** The tiers on the left side depict the data sources. This data can be loaded into the data lake in batches or in real-time.

2. **Insights tier:** The tiers on the right represent the research side where insights from the system are used. SQL, NoSQL queries or even Excel can be used for data analysis.

3. **Hadoop Distributed File System (HDFS)** is a cost-effective solution for both structured and unstructured data. It's a landing zone for all data that is at rest in the system.

4. The **distillation tier** takes data from the storage tier and converts it into structured data for easier analysis.

5. **The processing tier** runs analytical algorithms and users' queries in real-time, interactive or batches to generate structured data for easier analysis.

6. **The unified operations tier** governs system management and monitoring. It includes auditing and proficiency management, data management and workflow management.
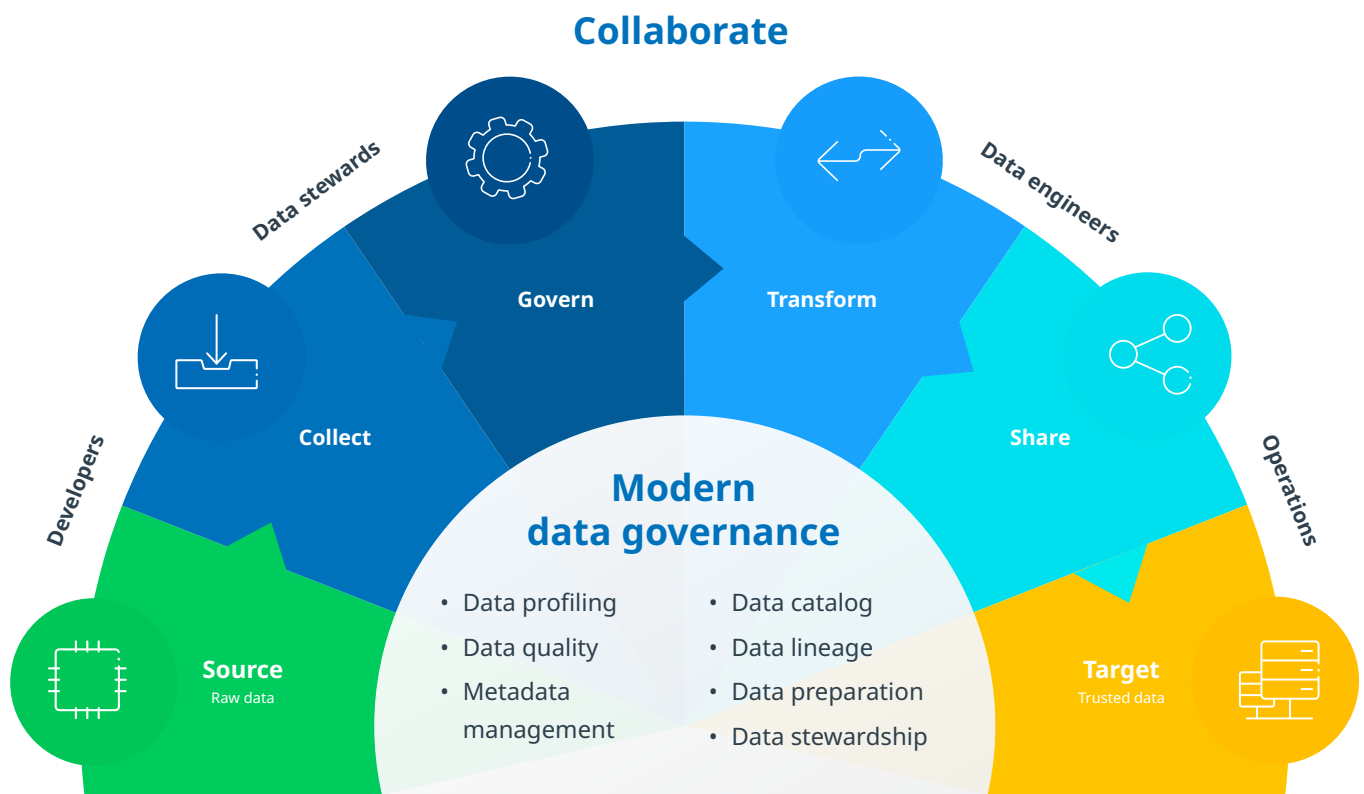
2.4 Data governance

All organizations have **multiple core systems, each with its own unique standardized information**. Data governance helps to improve understanding and lineage from different sources and improves data quality. This improves metadata management for IT managers and business glossary users when data governance is implemented well and provides a comprehensive view of all data assets. These include, but are not limited to, the standardization of data and its consistency, helping to provide:

- An enterprise-wide data catalog with consistency and accuracy
- An understanding of the entire BI and data management environment
- Traceability
- Enterprise-wide metadata governance stewardship including reference data

Business users continue to see data governance as an IT issue, but there is a need to clarify the value that high-quality and accessible data can bring to all parts of the business. Clear estimation and communication of the return on investment of a data governance program alongside tangible benefits will help drive result-oriented transformation with varying advantages for different teams such as below:

- Sales and marketing – Increase sales by using higher quality data to create more targeted campaigns.
- Procurement – Reduce costs by using governed data to optimize purchasing processes and supply chain.
- Legal and compliance – Avoid non-compliance, breaches and fines by establishing ownership and policies around data.
- Finance – Improve reporting with better governed and easily accessible data.

Collaborate

Data stewards
Data engineers
Developers
Operations

Govern
Transform
Collect
Share

Modern data governance

- Data profiling
- Data quality
- Metadata management
- Data catalog
- Data lineage
- Data preparation
- Data stewardship

Source Raw data

Target Trusted data

Our objective is to help you establish and adopt the standards and practices needed to create a data-driven organization enhancing your analytics capabilities to drive digital transformation.

- We begin with a recommendation of the structure of the data organization appropriate to your organization and confirm people in key roles (i.e., Data Steward, Data Architect, Lead Data Analyst). The team will also adopt the balanced scorecard that will enable to track and communicate progress towards sound data processes.

- Our data governance and operations framework helps you define and implement standards and practices in four areas: Business Value Management, Data Architecture Management, Analytics and Data Value Realization, and Data Solution Delivery. We introduce leading practices and standards models to the core client data team to adopt those that are fit for purpose.

- The newly defined model is validated in tabletop tests and through a data proof of concept from their existing plans.

Typical deliverables:

- Data organization model
- Data governance balanced scorecard
- Data operations and governance standards and practices
- Data governance solution
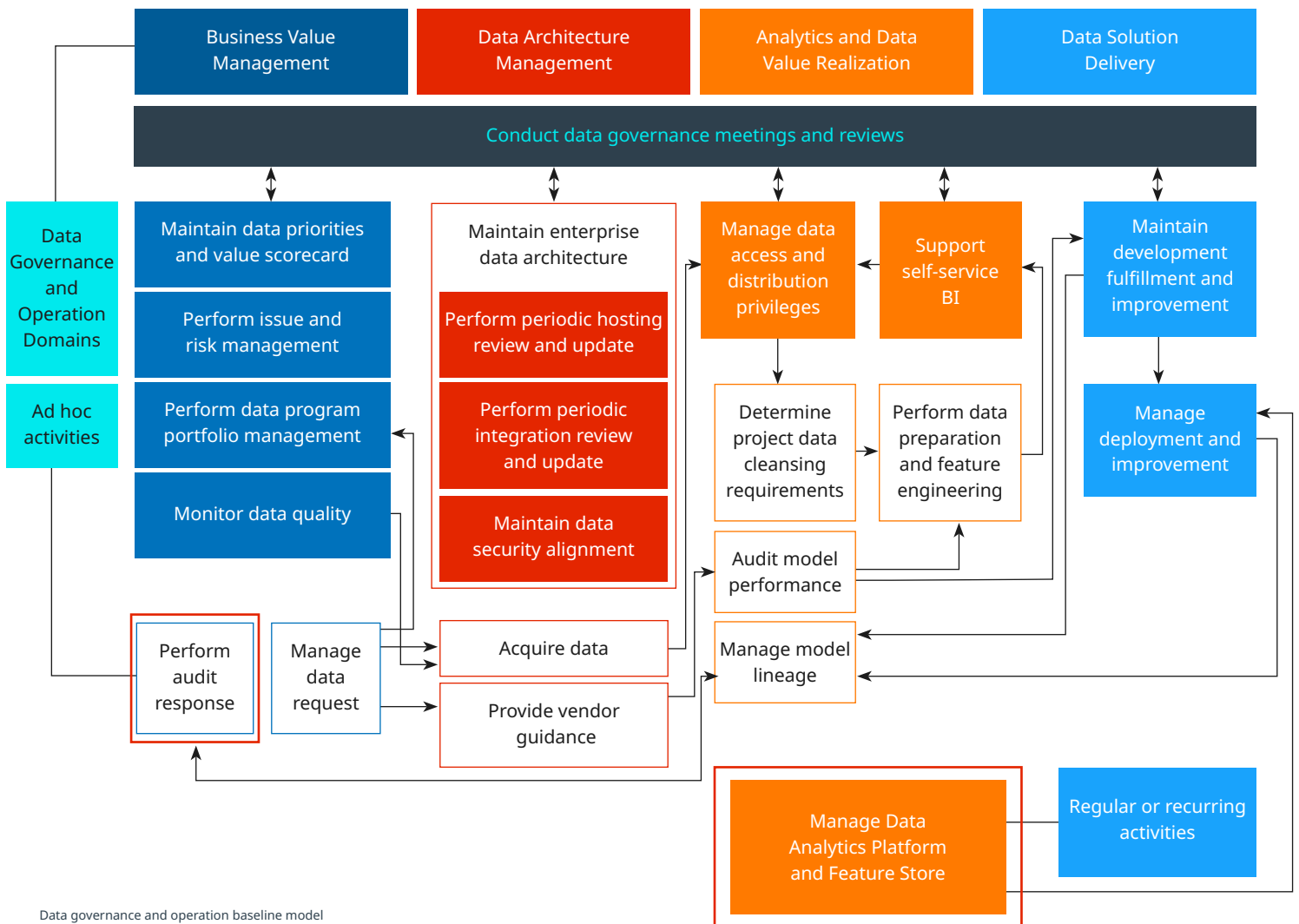
Data governance and operation baseline model

2.5 Master data management (MDM)

Organizations solve operational challenges through systems such as order management, purchasing, inventory and logistics. Each of these business functions captures and understands data pertaining to customer, products, inventory and financial performance in their own unique way resulting in the creation of silos of data each with its own trust level. In a financial organization, the home loan department has a set of customer data with attributes like email, contact address, details that aren't updated as often as they would be in the credit card system. MDM systems help to match customer information across various departments and provides a consolidated view of the data with the most recent information for key data points, creating a Golden Record.

MDM provides a data modeling and quality management solution for large enterprises. These organizations now have a single, trusted view of operational and customer information, enabling them to identify new opportunities, improve decision making and increase revenues. This process is started with extracting data from core systems and transforming them to a single view or multiple tables. Data quality is also part of the transformation, to ensure the validity and accuracy of the data.

Features:

- Multi-channel communication: supports and integrates multi-channel communication including social media.
- Industry-standard compliance data model: MDM solution that is fully compliant with standards by integrating with SOA-based infrastructure
- Customizable data models: customizable insurance data models to relate customer, policy, agent and broker information from internal and external systems.
- Seamless integration between the MDM hub and other external systems such as customer and account, credit scoring, and appraisal.

360° customer view

One of the key outcomes from the MDM solution is a 360-degree view of every customer relationship, which would have been only possible by manually piecing together several bits of information about the customer from various sources.

Social media integration

Social media provides an avenue to understand more about a customer's behavior, likes and dislikes, which enables the business to build enriched customer profiles that can be used to provide solutions more likely to be accepted on the first attempt, saving time and money.

Robust data governance

Supports robust data governance and maintains the trustworthiness of data.

MDM includes processes such as data governance rules and adjudication of the operational processes including data stewardship, data glossary definition.

Reduced cost and complexity

Standards-compliant MDM solutions reduce cost and complexity across various industry segments. These solutions have saved millions of dollars for manufacturing companies that have traditionally suffered time-to-market issues due to disparate systems which are often out of sync in the product part codes.

Improved revenue and customer satisfaction

Enhanced customer satisfaction, social selling, and increased cross- and up-sell opportunities.

3. Data analytics

Data marts and OLAP

A data mart is a problem-specific data store, designed to hold information for reporting/analysis/insight around a specific organizational function and contains a subset of data from a data warehouse.

OLAP facilitates data analysis and reports, enabling easy decision-making to move your customers forward and stay ahead of the competition. OLAP helps in merging data from different financial years. When using actual numbers for budget and calculating what-if analysis based on possibly increased or decreased costs, OLAP calculates each scenario instantaneously rather than individually. It makes easy-to-use cubes, offering flexibility and ease of use all at a low cost.
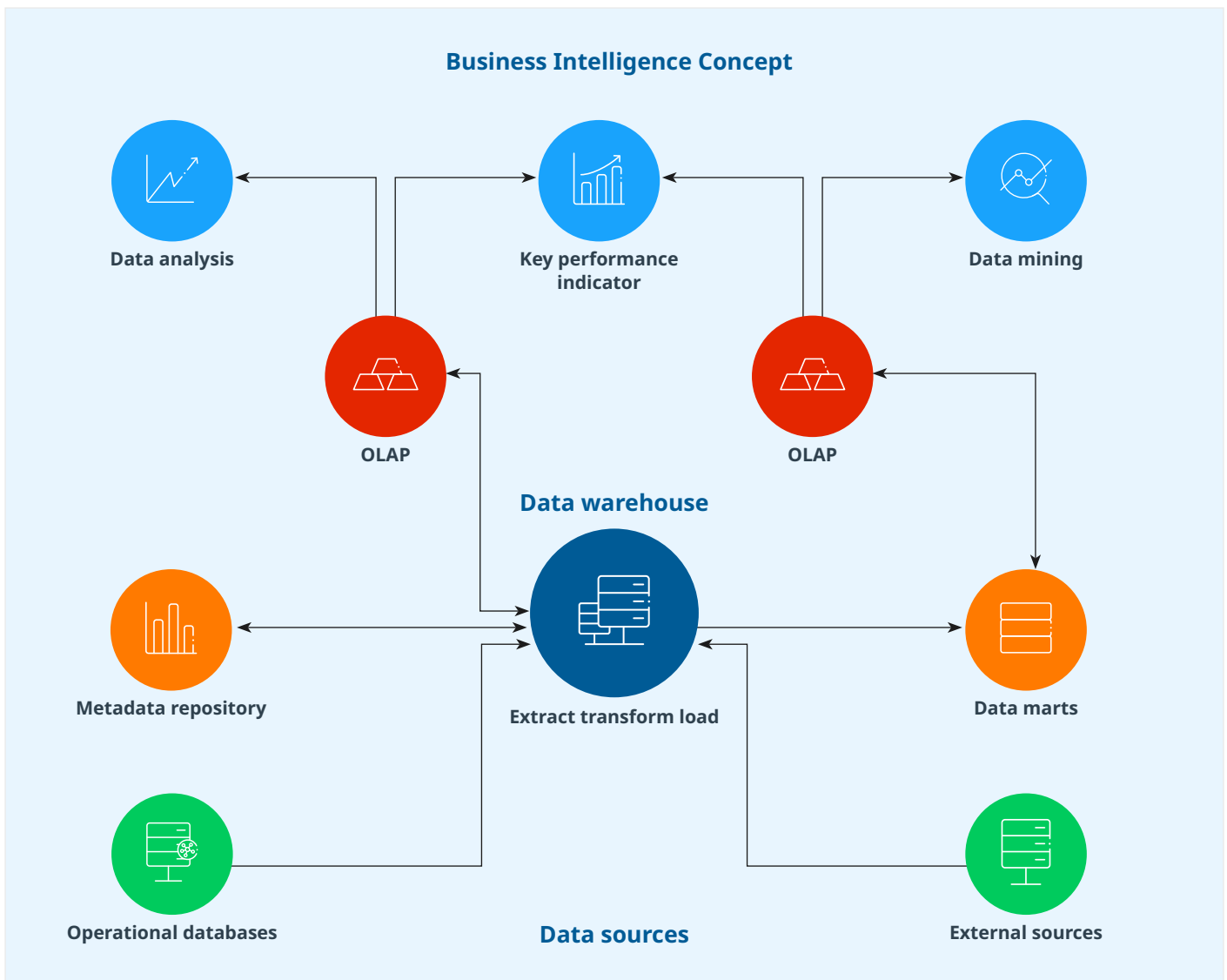
We have experience in implementing data marts for customers to solve reporting and analysis around challenges including profitability analysis, product pricing and HR expense management.

Reports and visualization

- Reporting: The process of organizing data into summaries to monitor how different areas of a business are performing. The process of exploring data and reports to extract meaningful insights, which can be used to better understand and improve business performance.

- Data visualization: Identify business patterns, trends and relationships across both historical and transactional data and structured and unstructured data through analytical reports and management dashboards.

Purpose-built solutions

- We have experience in designing and implementing specialized risk reporting, with calculators that are purpose-built to calculate credit risk, interest rate risk, etc.

- Partner solutions in the areas of fraud detection and reporting.
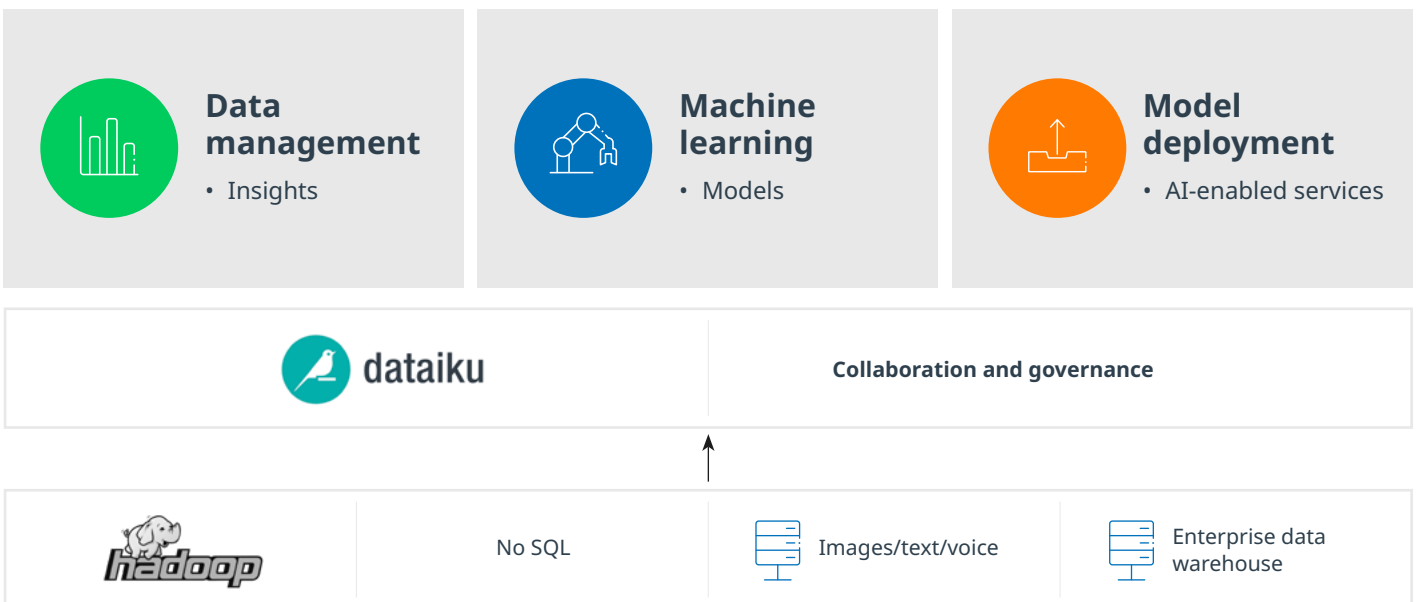

Business Intelligence Concept

Data analysis — Key performance indicator — Data mining — OLAP — OLAP — Data warehouse — Extract transform load — Metadata repository — Data marts — Operational databases — External sources — Data sources

4. Data science

Data science platforms with AI/ML

From 2012 to 2018, large technology companies implemented custom-built ML platforms for internal use (i.e., Facebook's FB Learning, Uber's Michelangelo, Twitter's Cortex, AirBnB's BigHead). Many of these platforms are based on open-source packages and are tailored for their specific use cases.

Since then, the industry has seen the evolution of enterprise-grade ML platform solutions, including those from incumbent vendors (e.g., Amazon Sagemaker, Microsoft Azure ML, Google Cloud ML, etc.) and the challengers in the space (e.g. DataRobot, H2O, BigML, Dataiku). Incumbent vendors follow an incremental strategy approach, with their ML services offering sitting on top of their existing cloud services as another application layer vs the ML native approach taken by the challengers.

We partner with clients who want to extend their internal data-warehouse or data lakes into data science platforms to harness useful intelligence. We leverage solutions like Dataiku, DataRobot and H20 to provide robust platforms for the client to build insights, models and even provide AI Services in a collaborative manner.

Data management	Machine learning	Model deployment
• Insights	• Models	• AI-enabled services

dataiku	Collaboration and governance

hadoop	No SQL	Images/text/voice	Enterprise data warehouse

Predictive analytics with regression and neural networks

We gather data from multiple sources in your organization and use our advanced and statistical techniques to predict trends, improve customer confidence, improve business performance, drive strategic decision making to maximize profit and predict behavioral patterns.

Predictive analytics encompasses a variety of statistical techniques from data mining, predictive modeling and machine learning, that analyze current and historical facts to make predictions about future or otherwise unknown events.
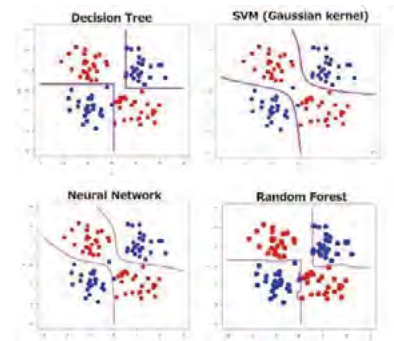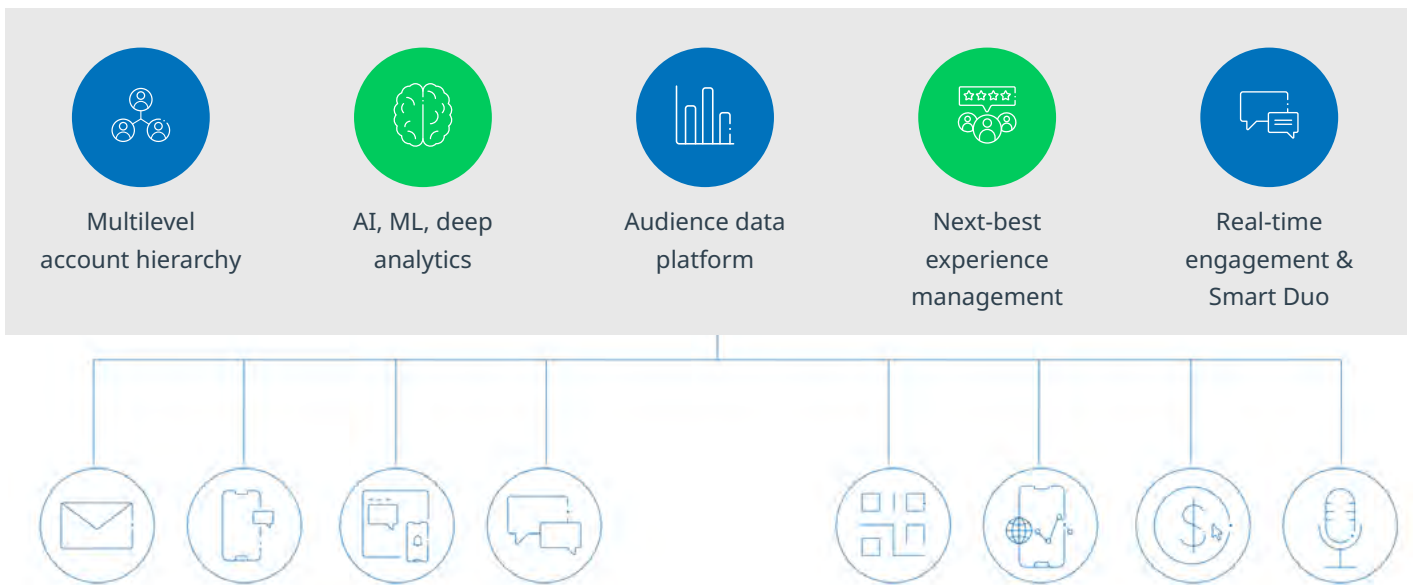
Data science-based platforms solution – Marketing automation and customer intelligence

Marketing automation enables marketers to qualify leads and pass them onto sales, resulting in higher revenue and greater growth.

This allows them to demonstrate the ROI from their campaigns. The ability to demonstrate the effectiveness of campaigns helps them move from cost centers to revenue generators. When marketing teams can prove how their campaigns impact the bottom line, they can justify increases to their budgets. With more support for their marketing campaigns, marketers can experiment, get creative and execute new lead generation campaigns that will empower lead generation.

rCloud with AI core

| Multilevel account hierarchy | AI, ML, deep analytics | Audience data platform | Next-best experience management | Real-time engagement & Smart Duo |

Marketing automation technology helps to automatically manage marketing processes and multifunctional campaigns, across multiple channels. This customer intelligence provides a guide informs and provides gives enough ammunition to digital marketers to make decisions that impact the bottom line.

- Identify audience and track results: build rewarding customer relationships by identifying your target audience and tracking campaign results.
- Manage campaigns: create and manage complex marketing campaigns using different channels with an intuitive visual interface that assists you from planning to execution.

Rapid acceleration in pace of new disruptive technologies like artificial intelligence and machine learning has transformed business functions. In particular, the financial services and insurance sectors have seen greater adoption of these technologies to improve their business processes for better customer experience through digitalization, credit lending/ underwriting, fraud detection and risk management.

Data management - enhancing your data analytics capabilities

Artificial intelligence
Engineering of making intelligent machines and programs

Machine learning
Ability to learn without being explicitly programmed

Deep learning
Learning based on deep neural network

- AI-powered with NLP for smart chatbots that provided localized services and other valuable insights. Such solutions provide a better personal banking experience and reduce the workload for the call center.

- Loan underwriting platforms with AI/ML can use thousands of additional data points and help to assess customers with no credit history.

- Quantitative trading uses large data sets able to identify the strategic trades and algorithmic trading to automate trades.

- AI/ML-based platform analyzing network network-level provides probability-based calculation in detecting suspicious and fraudulent transactions.

Service catalog

Our services	Our tools
Data warehousing and ETL	Microsoft SSIS, Oracle Warehouse Builder, Information, SAS DI Studio, Talend, Pentaho
Data quality and governance	Information Data Quality/MDM Talend Data Catalog/Data Quality/ Data Prep/Data Stewardship/Data Inventory
Data visualization	Microsoft BI, Tableau, SAS Visual Analytics, Cognos, MicroStrategy
Predictive analytics	Dataiku, Data Robot, H20
Marketing automation and customer intelligence	Resulticks Cloud/Marketing automation/ Omnichannel marketing
Big data	Hadoop, Splunk, Cloudera
AI/ML	TensorFlow, PyTorch, Scikit Learn (Python)

services.global.ntt