HITACHI
Inspire the Next

WHITE PAPER

Building the Data Factory

Streamline your data management process the same way a production plant works.

Table of Contents

03 Bridging the OT/IT Chasm

04 Data Access & Collection, like Raw Material in a Plant

05 Combine & Blend Data Like Processing and Assembly Machines Function in a Plant

05 Understand & Catalog Data the Same Way Material and Parts Are Stored in Silos or Parts Bins

06 Govern & Optimize Data in the Same Way Manufacturing Components are Molded or Prepped for Assembly

07 Analyze & Visualize information is Like the Final Packaging and Labeling Step in a Factory

Bridging the OT/IT Chasm

A gap in understanding exists between business operations and central IT organizations around overcoming data management challenges to drive digital transformation and AI analytics initiatives. A common understanding of modern data management techniques and tools is one way to foster a better working relationship between these two teams. It may also help to think of the data management process in the context of how a manufacturing assembly line works for a common analogy.

Continuous process improvement has been a cornerstone of manufacturing organizations for years. Total Quality Management (TQM) consists of organization-wide efforts to continuously improve the ability to deliver products and services that customers will find of value. Kaoru Ishikawa developed Continuous Process Improvement (CPI) (Kaizen) by translating, integrating and expanding the management concepts of W. Edwards Deming and Joseph M. Juran. CPI defines how continuous improvement can be applied to manufacturing processes when all variables are known.

DataOps is like a factory automation system in automating, upgrading, improving and assembling another valuable raw material: data. DataOps is a set of practices, processes and technologies that combines an integrated and process-oriented perspective on data with automation and methods from agile software engineering to improve quality, speed and collaboration and promote a culture of continuous improvement for data analytics.

Data Access & Collection, like Raw Material in a Plant

Data silos are some of the most destructive elements to running an effective enterprise because they hinder the ability to understand upstream and downstream information and the implications to adjacent operations, creating a more reactive versus proactive environment. Compounded with the difficulty of integrating into different systems, including varying data formats, protocol conversions and edge data management, companies find that integrating all their data sources provides an operational competitive advantage over other companies in the same industry.

Like the raw material intake to a manufacturing plant, the data integration and access function ensures that your data feed is always available and will not impede decision-making at any level of your organization, including:

- OT data integration by connecting to legacy control systems and historian databases and operations systems.
- Accessing new and expanded data sets from IoT sensors, relational databases, video, picture, audio, documents, CRM, ERP and other more traditional IT systems.
- Securely collect operations data, including storage, warehousing and exchange, inside a data lake, data warehouse or data hub.
- Storing this structured and unstructured data in the appropriate modern database, including InfluxDB, MinIO, Postgres, CouchDB and MongoDB.

Like machines process raw materials coming into a plant, Hitachi has software that integrates and stores data from all your IT and OT sources.

Like machines process raw materials coming into a plant, Hitachi has software that integrates and stores data from all your IT and OT sources.

Hitachi's Edge Data Integration Software, Powered by Pentaho:

Edge Data Integration software, powered by Pentaho, ingests OT data from industrial sources and sensors. It provides data processing, reading payloads and redirecting portions to select destinations. Data storage is provided for all data with forwarding to a destination of choice.

Hitachi's Data Integration and Analytics Software, Powered by Pentaho:

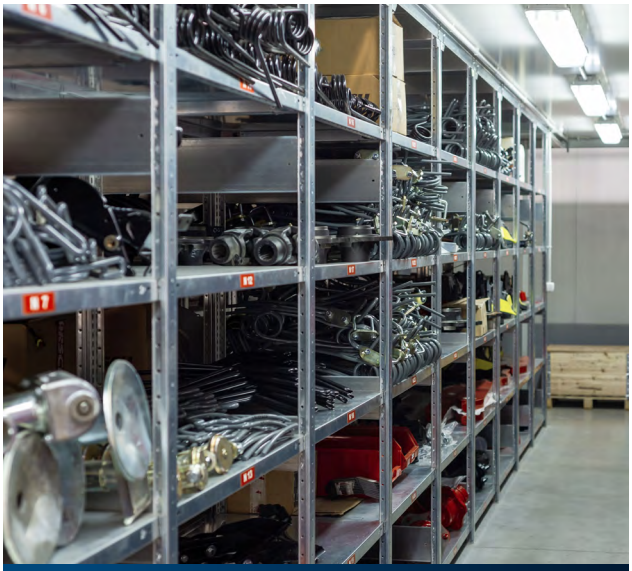
Ingest, blend, cleanse and prepare diverse data from any source in any environment — without code. Pentaho Data Integration offers broad connectivity to various data, including all popular structured, unstructured and semi-structured data sources, increasing the performance of data extraction, loading and delivery processes.

Combine & Blend Data Like Processing and Assembly Machines Function in a Plant

Data scientists spend 60% to 80% of their time on data preparation, an activity two-thirds of them dislike. These individuals are also highly paid and should be spending more time designing machine learning models and fine-tuning them for use in the production environment. Data preparation work is tedious, with many repetitive steps that can be automated out of the workflow, reducing the time to value for developing AI and ML applications.

In a factory, as parts and assemblies move down the production line, they become complete and more valuable as finished products. A DataOps system operates similarly to prep data for business users, including process tag name rationalization, time stamp alignment and data range validation, allowing quality and maintenance departments to work with data correlated to product or asset information. Data activities that can be automated include:

- Performing data contextualization and normalization close to the machine or edge.
- Scheduling (data processes) activities to determine the data transfer schedule as a collection of execution, load and transform activities, making optimal use of local resources.
- Dispatching data analysis orders. Depending on the data type needed, this may include further distribution of data queries, runs and workorders, requested by operations teams and adjusted for analysis or ML training needs.



Over 90% of industrial data is dark and unusable for broader business purposes.

Understand & Catalog Data the Same Way Material and Parts Are Stored in Silos or Parts Bins

Over 90% of industrial data is dark and unusable for broader business purposes. It is often dirty, untrusted, not standardized and lacks context. Data is also inconsistent across machinery and data streams and has little context to explain the stream, where it was from, the expected tolerances or the unit of measure. It can't be used in its raw form for critical operations decision-making. Organizing and enhancing the data for business use is often highly manual and error-prone, as well as taking up valuable time from workers who could be focusing on more meaningful work for the organization.

Manufacturing organizations grade, collage, categorize and store raw materials in the appropriate bins to be utilized in the production process. Data, too, needs to be understood and cataloged for specific use and purpose, including:

- Manage data definitions, including storage, version control and exchange with other systems of master data.
- Connect business users/analysts to data sources to explore them visually.
- Automate the discovery and classification of structured, semi and unstructured data using AI-driven unique data fingerprinting.
- Build or import a taxonomy of business terms and establish relationships using a business glossary.
- Validate data and assess conformity to business policies using business rules.
- Assess critical quality metrics across the operation with better data quality.

Hitachi's Data Catalog software:

Hitachi's Data Catalog software delivers trusted, fit-for-purpose data through automated discovery and classification. An intuitive user interface allows users to search data like an online shopping experience, providing at-a-glance views tailored to user access and data use rights. It has an easy-to-learn, award-winning, out-of-the-box visual report development environment accessible to all. It creates data marts on demand that are curated and governed.

Govern and Optimize Data in the Same Way Manufacturing Components are Molded or Prepped for Assembly

Industrial companies are finding that data privacy and regulatory compliance are becoming more critical. Information that includes patented production processes and product designs, confidential employee information, emissions compliance data and customer records must be safeguarded. Cost-effectively storing the explosion of collected data in the cloud is not feasible; some data is transient and only needed for a few hours before being summarized and placed in long-term storage.

In a factory, a manufacturing execution system (MES) or automation system keeps track of all of the parts and production processes that entail running the factory. DataOps software does the same for data trust. Trust in data requires the confidence and knowledge that your organization's data is fit for purpose and ready to act on. Data trust is about expectation and reliability. Trusting your data means knowing your data, including:

- Management of data resources: Registration, exchange and analysis of resource information, aiming to prepare and execute data operations.
- Conducting data discovery and visual analysis and reporting against prepared data.
- Data track and trace: Registration and retrieval of related information to present a complete data history, including source, combination, transformations and more.

Hitachi's Data Catalog Software:

Hitachi Vantara's Data Catalog intelligently automates profiling and discovery across your data estate of applications and data stores, edge, on-prem and multi-cloud. Make discovered data actionable by leveraging AI-driven capabilities to give business context to your data and apply tailored business rules that ensure adherence to your organization's privacy, compliance and security requirements.

Hitachi's Data Integration and Analytics Software, Powered by Pentaho:

Dataflow Studio, part of Pentaho, streamlines and speeds up data delivery to enable data self-service to collaboratively build, deploy and monitor dataflows for faster, confident business decisions. It allows your data engineers and data consumers to collaborate more effectively. It blends data across lakes, warehouses and devices while orchestrating dataflows in hybrid and multi-cloud environments with improved dataflow visibility and tools to customize your data.

Data scientists spend 60% to 80% of their time on data preparation, an activity two-thirds of them dislike.

Analyze & Visualize information is Like the Final Packaging and Labeling Step in a Factory

The final step in the manufacturing process is packaging and labeling, which makes the product valuable and consumable to the end user customer. The same is true of data analysis and visualization of information, where quickly creating an HMI or dashboard will meet a specific job role or departmental need, including:

- Build to order analytic AI/ML solution models for operations use cases.
- Embedding analytical assets into other web applications.
- Ability to train accurate and dependable ML equipment models.
- Digital Twins for data modeling and integration.
- Align with existing back-end requirements: security integration, single sign-on, multi-tenant deployment.
- Analytics dashboards for performance analysis. Displaying useful information out of the raw collected data using trained models to provide predictive and prescriptive analytics.

Pentaho Business Analytics Software:

Pentaho Business Analytics provides a spectrum of analytics for all user roles, from visual data analysis for business analysts to tailored dashboards for all audiences — from business executives to front-line workers. Users can create reports and dashboards and visualize and analyze data across multiple dimensions without dependence on IT or developers.

92% of organizations plan to deploy predictive analytics more broadly, while 50% have difficulty integrating them into existing infrastructure.

ABOUT HITACHI VANTARA

Hitachi Vantara, a wholly-owned subsidiary of Hitachi Ltd., delivers the intelligent data platforms, infrastructure systems, and digital expertise that supports more than 80% of the Fortune 100. To learn how Hitachi Vantara turns businesses from data-rich to data-driven through agile digital processes, products, and experiences, visit hitachivantara.com

Talk to an Expert →

Find out how to unlock your operations capacity with Industrial DataOps & IoT Software.

Hitachi Vantara

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact