Building and Utilizing Company-wide Data Analytics Platform, DataLab®

Kazuo MINAMI*1・Masanobu FUJIHIRA*2・Taketsugu OSAKA*2

*1 IT Planning Department
*2 Digital Innovation Technology Center, Technical Development Group

Abstract

The utilization of data is becoming increasingly important in the art of manufacturing. Kobe Steel is an early adopter of IoT and big data, which has contributed to product development and service improvement, but lacked a unified platform for efficiently utilizing data across the company. To meet this challenge, a data analysis platform, DataLab®, has been established. This platform enables efficient collection and analysis of large amounts of data and leverages advanced analytical tools to provide new insights. The purpose of DataLab® is to provide centralized data management, flexible infrastructure, advanced analytical capabilities, strict security, and user-friendly operability. This paper introduces the DataLab® initiative through material development and equipment diagnosis examples.

Introduction

To increase competitiveness in today's business environment, companies must make progress in technological areas such as telecommunications and AI. Operating and product data are particularly critical to the growth of a company in numerous areas, including new product development, product improvement, market strategy, and customer experience.[1), 2)]

The importance of IoT and big data became apparent around 2014, when Kobe Steel's businesses started using them for product development, manufacturing kaizen initiatives, and customer service improvements. At the time, however, we did not yet have a platform for company-wide data storage and analysis. As such, each of these activities occurred in a silo, leading to individualized systems for analyzing data and expertise-backed information.

This led to a collaboration between the IT Planning Department and the Technical Development Group beginning in 2016, as set out in Kobe Steel's medium-term management plan. The objective of the collaboration was to support the development of data scientists and the analysis of operating and product data. Additionally, advances in digital technology have drastically increased the processing power of computers and networks. Cloud services are another area that has seen strong development. This has facilitated the previously difficult tasks of collecting, transmitting, and processing large amounts of data.

This backdrop is what led Kobe Steel to develop DataLab® [Note 1)], a data analysis platform for data storage, preprocessing, and analysis. This platform is the foundation for the unified management of data and expertise-backed information and for the systematic use of data across the company.

DataLab® was designed to provide the entire company with a platform that can facilitate the use of large amounts of unfavorably structured data, such as R&D data and time series data of equipment. This platform makes it possible to quickly and flexibly incorporate resources as needed and collect, store, and access vast amounts of data. We also incorporated advanced data analysis tools and AI algorithms to efficiently extract insights from the data and accelerate the creation of new value.

Section 1 of this paper presents the concept of DataLab®; Section 2 details the materials development and equipment diagnosis projects comprising the initial DataLab® use cases.

1. Concept of DataLab®

A data analysis platform must have data integration and management capabilities, infrastructure that responds flexibly to changes in computing resources, and advanced analysis tools and technologies, while supporting data security and ease of operation.

Data integration and management are particularly critical to guaranteeing the accuracy, completeness, and reliability of data and to ensuring that data can be analyzed effectively. Data collected from multiple sources, such as equipment and business-related systems, may have inconsistencies or exist in multiple formats. The quality of the data must therefore be improved through data cleansing (elimination of inconsistencies and multiple formats), data conversion (conversion to a standard format), and data preservation (storage for later use by business logic and analysis models).

The next requirement is to establish a scalable infrastructure that can respond flexibly to increased

Note 1) DataLab® is a trademark of Kobe Steel.

volumes of data and computational processing needs. Cloud services achieve such scalability very well.

Advanced analytical tools and technologies enable the extraction of insights from data through targeted analytical methods such as statistical analysis, machine learning, and deep learning. Encryption and the strict control of access rights guarantee data security.

Basic data visualization tools make data analysis accessible to a larger group of people and ensure that stakeholders can quickly and intuitively comprehend the data and take action.

The DataLab® platform has two main layers to provide all functions necessary for data analysis: a database layer for storing data and a solution layer for data analysis and solution development (**Fig. 1**). The following sections provide an overview of each layer.

1.1 Database layer

One of the main challenges in data management is that data are stored in multiple locations, including individuals' machines, computers inside equipment, and shared file servers. In addition, the lack of rules for associating data means that great effort is required to convert data into a format suitable for analysis and to align data created by different people for the same issue. A further challenge is the limited storage capacity of equipment controllers, meaning that key data are lost a few months after recording.

Additionally, to support efficient analysis, raw data must be converted to a specific format suitable for the issue being studied and stored in an appropriate database.

To overcome these challenges, the database layer of DataLab® makes it possible to store and manage raw data centrally in one data aggregation area. This functionality even supports differing formats, including structured data such as Excel and CSV file types, semi-structured data such as XML, and unstructured data such as image and audio files.

The process from data collection to storage in the database layer is standardized to facilitate data collection and storage, while the data association function significantly reduces the time and effort required to prepare data for analysis. Raw data are automatically stored in unprocessed form to accommodate future analysis needs. Storing data in the cloud enables indefinite data storage.

Data formatted and stored in the database layer via the process detailed above are used by the models and tools created in the downstream solution layer.

1.2 Solution layer

Accumulated data can be used for the development of purpose-built business solutions via the solution layer. This layer also serves as the environment for developing and analyzing applications using Kobe Steel's original analysis software, third-party analysis software, and visualization tools such as BI (business intelligence) tools.

Our proprietary analysis tools can perform advanced data analysis in the database, owing to models Kobe Steel's data scientists have fine-tuned for each object of analysis using machine learning.

Visualization tools support the user's decision-

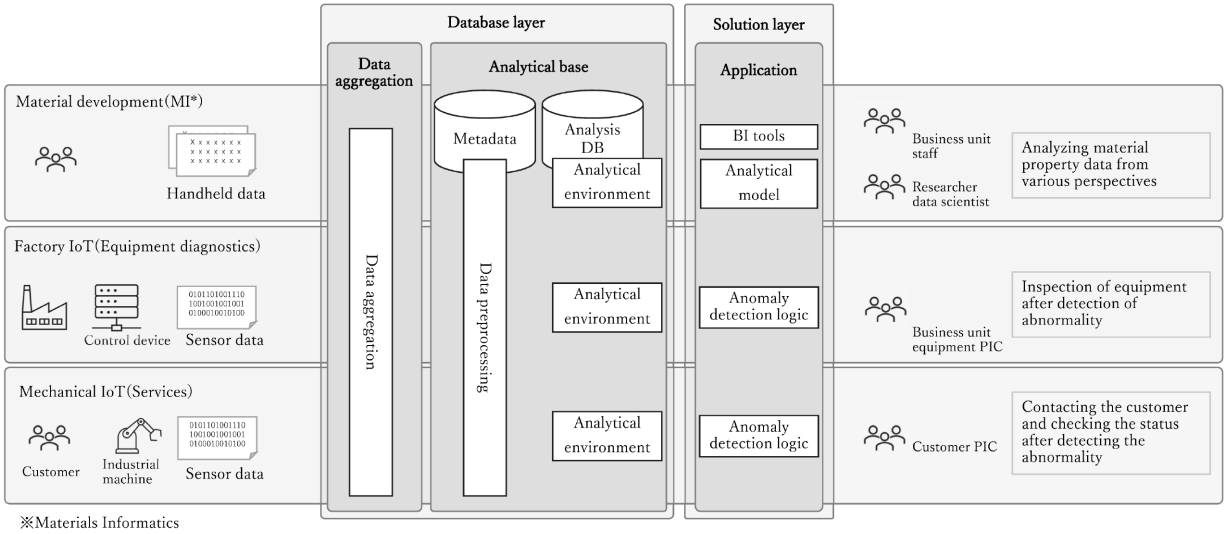

Fig. 1 Concept of DataLab®

making process by accessing the database to retrieve and visualize data.

Furthermore, the ability to securely transfer data to and from external data analysis services enables the use of third-party data analysis models. This makes it easy to perform data analysis and prediction without the need for development manpower and expertise required for AI.

2. Use cases

It is difficult to design any one system for all analysis needs. As such, the optimal configuration for each main object of analysis, such as materials development or equipment diagnosis, must be devised. Data for materials development is scant and exists in varying formats, whereas data for equipment diagnosis is abundant and has well-defined formats. These parameters lead to differences in the optimal data structures and protocols for data organization.

Materials development and equipment diagnosis served as initial use cases in developing DataLab®, as we anticipate the widespread use of such applications throughout Kobe Steel. This section provides an overview of our data analysis platform in terms of these use cases (see "MI Technology in Developing and Utilizing Materials," pp.35, and "Establishment of Condition Monitoring and Predictive Anomaly Detection System in No.2 Bloom Mill at Kakogawa Works," p.9, in this issue).

2.1 Materials development

The materials development process generally yields a broad array of research data, whether the project at hand involves developing new materials or improving existing materials. Research data include physical properties, chemical reactions, and durability test results, each of which is recorded in a different format. There is often no standard format for data regarding a given object of analysis because of differing project or team leaders, scopes, and research periods. Organizing data for analysis can be challenging because data points with the same designation may have different content, and data points with the same content may have different designations.

To organize these data, we incorporated functionality to automatically analyze accumulated data, group related data, and store data based on relationships in a database for analysis.

Research data for materials development are allocated to file formats such as CSV based on specific rules for storage in the data aggregation area. Data are then classified, tables are created, and relationships among the tables are extracted based on key fields and organized into a reverse star schema. Data are stored in a relational database, where tables can be arranged at will to extract data for analysis (**Fig. 2**).

Fit-for-purpose data tables are analyzed via commercial spreadsheet software, visualization tools such as BI tools, or our original software. Data scientists at Kobe Steel developed this software by fine-tuning machine learning output based on numerous models. To use the tool, the analyst chooses a model, enters parameters, and executes calculations to predict material properties and search for the most appropriate target material design. Cloud computing can be used for advanced calculations, negating potential speed limitations

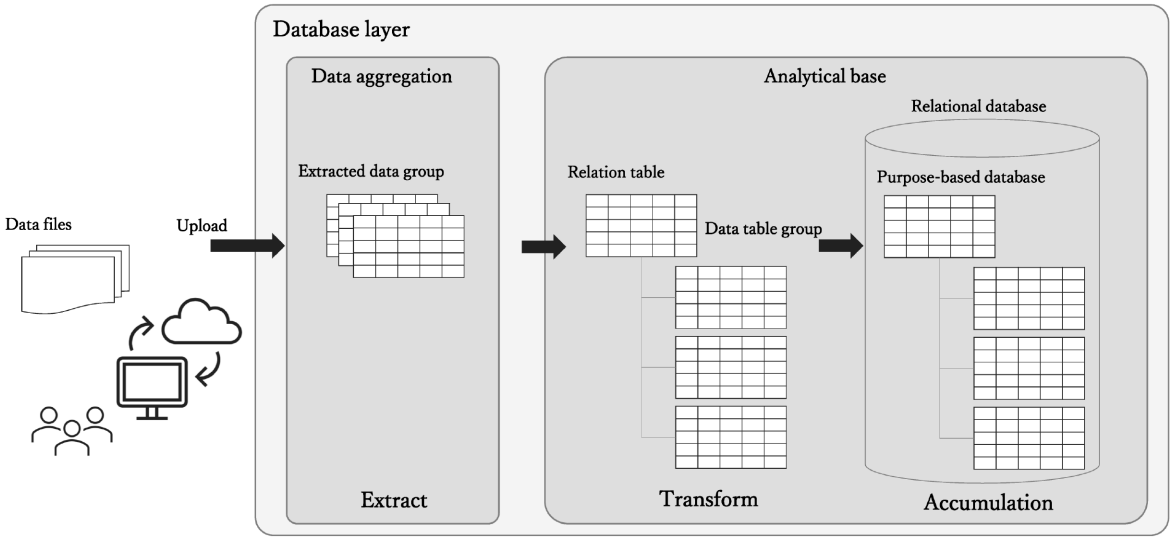

Fig. 2 Data processing image in the database

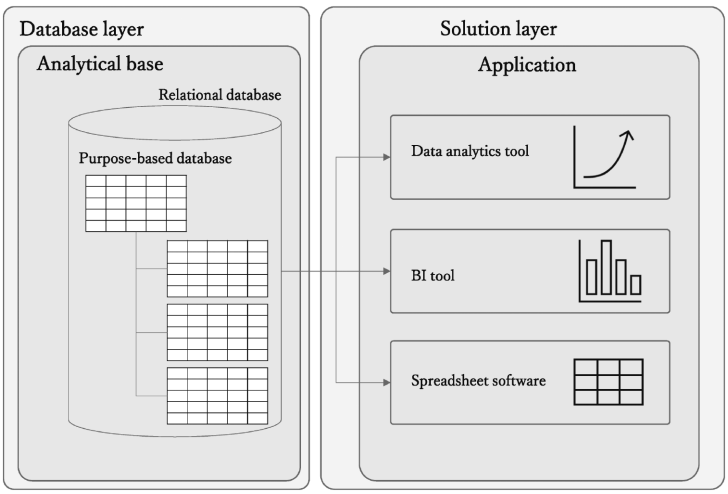

Fig. 3 System architecture of DataLab®

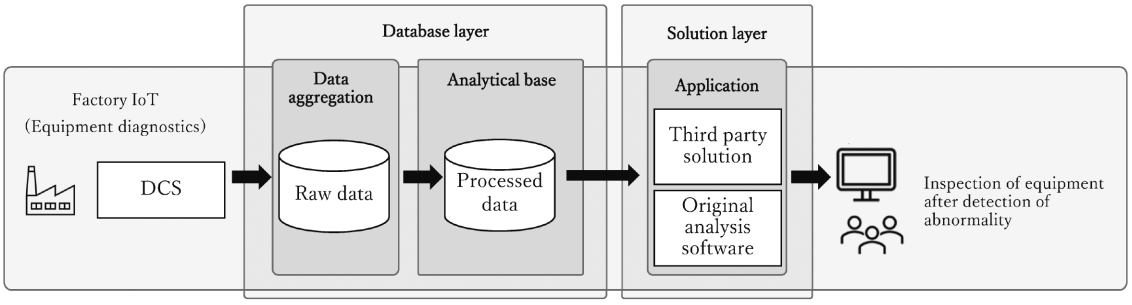

Fig. 4 Diagram of equipment diagnosis system

imposed by on-premises machines (**Fig. 3**).

2.2 Equipment diagnosis

Sensors in factory equipment provide real-time data to improve equipment availability and ensure safety; these applications fall under the field of equipment diagnosis. Data collected can include temperature, pressure, vibration levels, and equipment activity, all of which can be used to monitor equipment health and efficiency, predict faults, and set preventive maintenance intervals.

The many sensors and short acquisition cycles result in the generation of a very large amount of data. However, once a data collection mechanism is established, it can easily be deployed in other equipment since the data to be acquired is standardized. Third-party development tools for equipment diagnosis can be combined with our original software for a tailor-made analysis system.

Fig. 4 depicts all equipment diagnosis needs from data collection to analysis.

A gateway with software for transmitting data to DataLab® is used to collect the data. This device transmits data from a controller (DCS or PLC) to DataLab® for storage in the data aggregation area as raw data in a predetermined format. Data are processed and stored in a format conducive to utilization by analysis models and solutions.

In the case of equipment diagnosis, equipment sensors collect a variety of data, resulting in a large number of categories of data (i.e., columns). However, anomaly detection models only make use of data related to the anomaly being detected. In our use case, we built an analysis database by dividing time series data for equipment diagnosis at fixed intervals and storing them in a columnar database, thereby achieving both ease of analysis and high performance.

Our analysis solution uses the algorithms and functionalities of third-party tools to detect anomalies based on normal operating data. However, our data scientists can also build and implement custom analysis software using Python or programs for statistical analysis. Our system also features a data visualization tool for equipment monitoring.

3. Future prospects

Our materials development and equipment diagnosis solutions are being used only within a subset of our divisions. We plan to expand these systems to other divisions and add new

functionalities, including support for image data, which is often requested for inspection and materials development.

Additionally, it is notable that data with the potential to create value added exists in a variety of places, such as computers within equipment as well as individuals' machines. As such, we will foster data aggregation to support the sharing of data among our businesses.

We will also strive to make data analysis more accessible through the accumulation and sharing of expertise by capitalizing on the data being collected and promoting data visualization via DataLab®.

Conclusions

This paper introduces DataLab® for data utilization within an organization and presents select use cases. We believe data utilization is essential for improving manufacturing capabilities. As such, we will strengthen Kobe Steel's manufacturing capabilities by using DataLab® to accumulate and analyze data from various plants, equipment, and research projects.

References

1) N. Morita et al. Mitsubishi Electric Technical Reports. 2022, Vol.96, No.5, pp.40-43.
2) F. Taya et al. JFE Technical Report. 2022, No.50, pp.44-49.