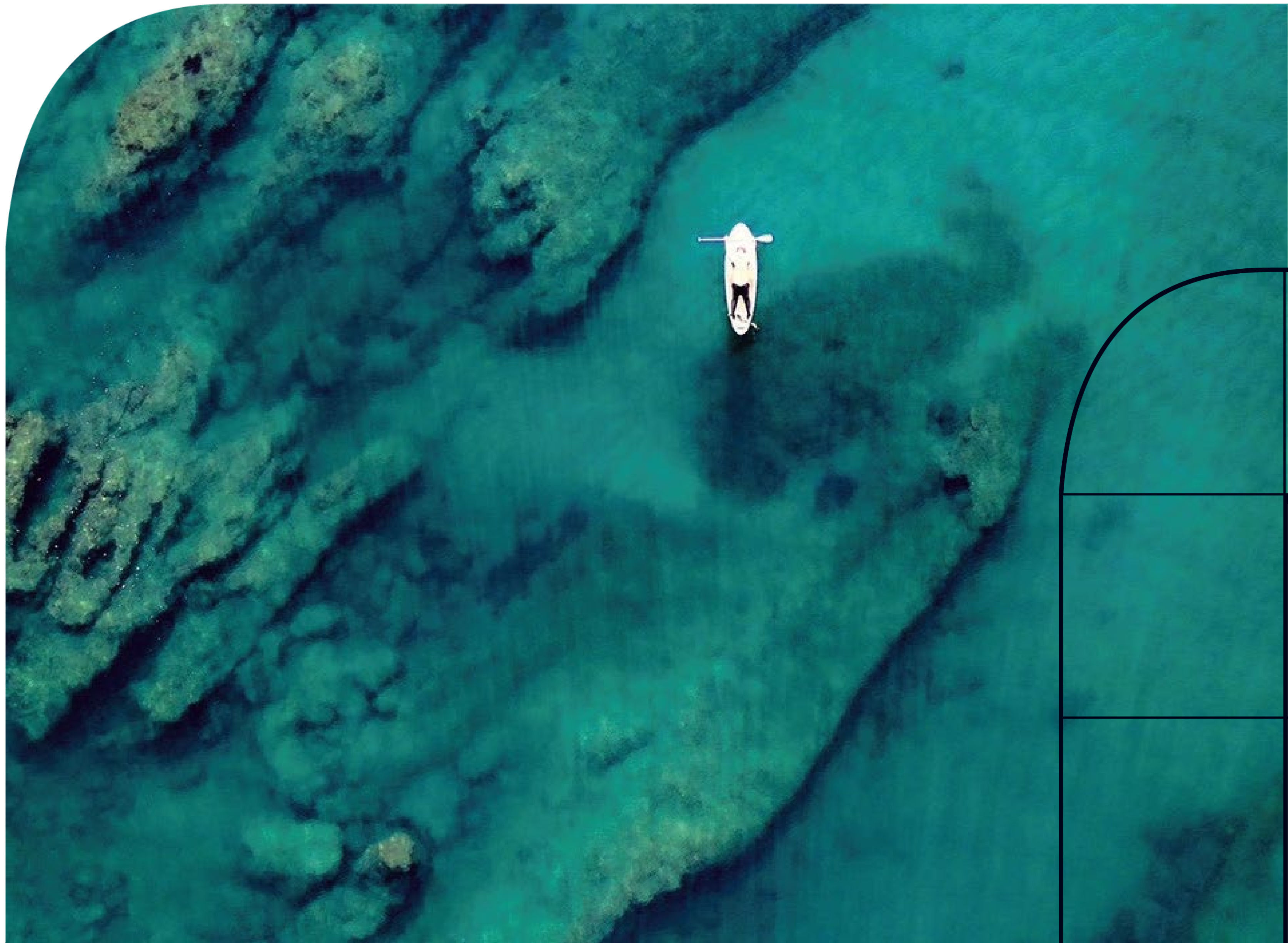
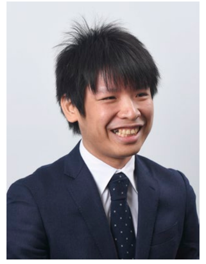
NTTDATA's Data Utilization in the Generative AI Era

LLM, Text2SQL, Semantic Layer, AI Agents for Data Utilization and Data Management Methods



NTT DATA Group Corporation
Toyosu Center Building, 3-3-3 Toyosu, Koto-ku, Tokyo 135-6033
Tel: (+81)3-5546-8051 Fax: (+81)3-5546-2405
https://www.nttdata.com/global/en/

Author



Ei Yamaguchi
Executive Data Engineer

Based on his many years of work experience in MLOps and DataOps,
he is a TechLead in the four areas of AI, Data, Software, and Infra for AI-powered services that are difficult to develop, such as applying AI to core operations.
Recently, he has been involved in supporting the introduction of AI-Agent using generated AI and in the development of offerings.

Kenta Nagatsuka
Senior Data Engineer

He provides support from the development and operation of data analysis infrastructure to data utilization using BI products and machine learning.
He also incorporates the latest trends in the field through advanced R & D using AI.

Chapter. 1

Introduction

The use of "data mining," which involves selecting necessary information from vast amounts of data, began in the 1990s. In the 2000s, the concept of "big data" was defined, and at the same time, the fields of "machine learning" and "artificial intelligence" developed. As big data technologies evolved in the 2010s, data analysis became increasingly important for business purposes, such as decision-making, customer insights , and the creation of new services. Today, many companies are introducing the latest data analysis platforms and democratizing data.

On the other hand, the democratization of data has led to many challenges.
In most large companies, the number of experts responsible for maintaining and providing data is overwhelmingly small compared to the number of business units that want to use data and create value. In many companies, engineers who provide data struggle to keep up with the rapidly increasing demand from users due to democratization of data.
Also, it is difficult for many citizen data scientists to use specialized tools, and finding the data they want is a high hurdle. However, engineering teams are often stretched thin
and it is not uncommon for them to create their own tables and reports using Excel at hand, and use data in silos.

NTT DATA has started an initiative to solve the problems of data utilization with generative AI. It uses key technologies such as LLM*1, Text2SQL*2, and semantic layer*3 to support business decision-making and promote data utilization within the company. This paper introduces the challenges of data utilization, generative AI technologies that promote data utilization, and future prospects.

*1 LLM (Large Language Model) : An advanced AI model that learns from large amounts of text data to generate and understand natural language.

*2 Text2SQL : A technology that automatically converts questions and instructions written in natural language into SQL queries.

*3 Semantic Layer : A middle layer between data and data analysts that promotes data utilization by abstracting data meanings and business logic.

Chapter. 2

Issues that slow down data utilization

This section explains specific issues related to data utilization for data providers and data analysts, who are key stakeholders in in the process.

2.1 Issues for data providers

Inability to provide data at the speed required by the business

In data utilization, it is not the analysis itself that takes the most time and effort, but actually the preparation and shaping of the data used for analysis. It is said that approximately 80% of the total effort is devoted to this process, and it is the key to increasing the agility of data utilization. On the other hand, it is difficult for many companies to provide data at the speed required by the business, and it often takes weeks to create a single data mart[*4]. To improve the speed of data provision, an increasing number of organizations are reviewing their development methods and introducing development support tools.

2.2 Challenges for Data Analysts

Inability to find the data needed for analysis

In the early stages of data analysis, there is often a lack of knowledge about the data needed for analysis. Data catalogs are used to find out what data is available for analysis, but most data catalog products use n-grams (word-based) searches and require that you know the exact table and column names in advance. Many analysts do not always know this information, which is a barrier to data exploration. In this situation, spending a lot of time exploring the data, making it difficult to focus on effective analysis. To improve the efficiency of data analysis, it is important to make sure to have fast access to the necessary data.

Business users are less proficient with the tools and fewer people can analyze the data

Moreover, many business users do not have sufficient knowledge of the BI tools[*5] and analysis tools used in data analysis, and they feel the psychological and technical burden of analyzing data. Analytical tools such as BI tools, SQL, and Jupyter Notebooks[*6] provide advanced functionality, but a certain level of proficiency is required to effectively use them. To reduce the hurdles of analysis, an environment must be created where business users can perform data analysis autonomously with minimal effort.

*4 Data Mart : A table or view that is formatted according to the purpose and purpose of a specific business department or project.

*5 BI tools : tools for analyzing and visualizing data and using it for business.

*6 Jupyter notebooks : interactive execution environment for developing and sharing software in a browser.

2.3 Issues common to data providers and data analysts

Significant effort in managing indicator values

In large organizations, managing indicator values such as KPIs is often challenging. As services start to end, and management policies are reviewed annually, the definition of these indicator values needs to be constantly reviewed and maintaining them is a very labor-intensive task. In addition, the information required by each organization for the same KPI is subtly different, and it is a major challenge how and where to absorb these differences.

Common examples of indicator value management

① Developers Provide Views on the Data Warehouse (DWH) Side

This is the most common solution, but the more users analyze the data, the greater the burden on the data provider. Even for a simple index value, the number of aggregation axes to focus on varies as many users (Want to see sales by region, month-on-month, product category, etc.), so it is impractical to create a view that supports all combinations of aggregation axes in advance.

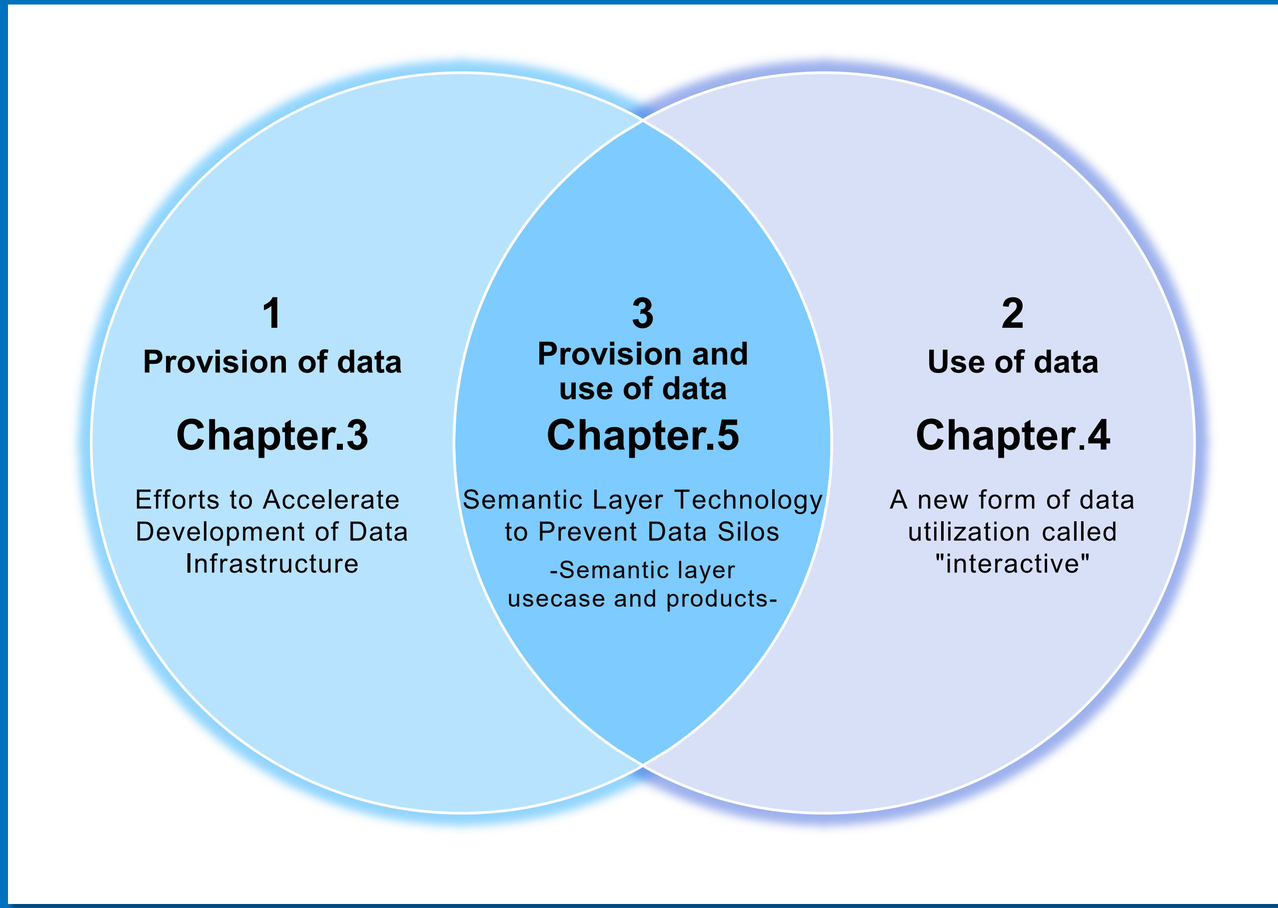
② Users Define Index Values in Various BI Tools

When data engineers cannot keep up with the demand for data marts, data analysts sometimes create their own data marts in the BI tools they manage. In DWH, it seems to be a good balance to create a huge daifukucho (a comprehensive dataset) and then perform aggregation and analysis from an arbitrary angle in the user department. However, especially in a large company where many departments are aggregating and analyzing data, the problem arises that business indicators such as KPIs are calculated using their own logic, and the index numbers are not consistent between departments.

2.4 Structure of this article

Regarding the issues mentioned above, this article introduces measures for data providers in Chapter 3, measures for data analysts in Chapter 4, and measures common to both the provision and use of data in Chapter 5. Chapter 6 describes NTT DATA's view of the future data utilization world based on the latest trends.

Figure 2.4 Issues that slow down data utilization



Chapter. 3

Efforts to Accelerate Development of Data Infrastructure

NTT DATA supports the improvement of productivity in data mart development by automatically generating SQL by utilizing generative AI to address the issues of data providers mentioned in the previous chapter.

3.1 Utilizing generative AI in data mart development

There are various approaches in improving development efficiency using generative AI, but the most effective is to improve the efficiency of coding work. Various products provide SQL generation functions, and if you search, you can see demo videos of each product. However, it seems that many of the demos generate SQL for light users based on simple conditions such as "sales by product in 2024."On the other hand, a data mart created in commercial development is a mass of complex business requirements, and giving each of these requirements to a prompt rather hurts productivity.

Introduction of SQL generation for DWH products

Databricks Assistant

Databricks Assistant, provided by Databricks, helps generate, debug, optimize, and explain code and queries simply by querying in natural language. Specifically, in response to questions or directives about SQL or Python code, Databricks Assistant generates optimal queries and provides explanations based on Unity Catalog metadata, automatically corrects errors, and suggests inline code. It is also possible to analyze, visualize, and filter data using Databricks Assistant from the dashboard.

※ Formed a capital and business alliance with NTT DATA.

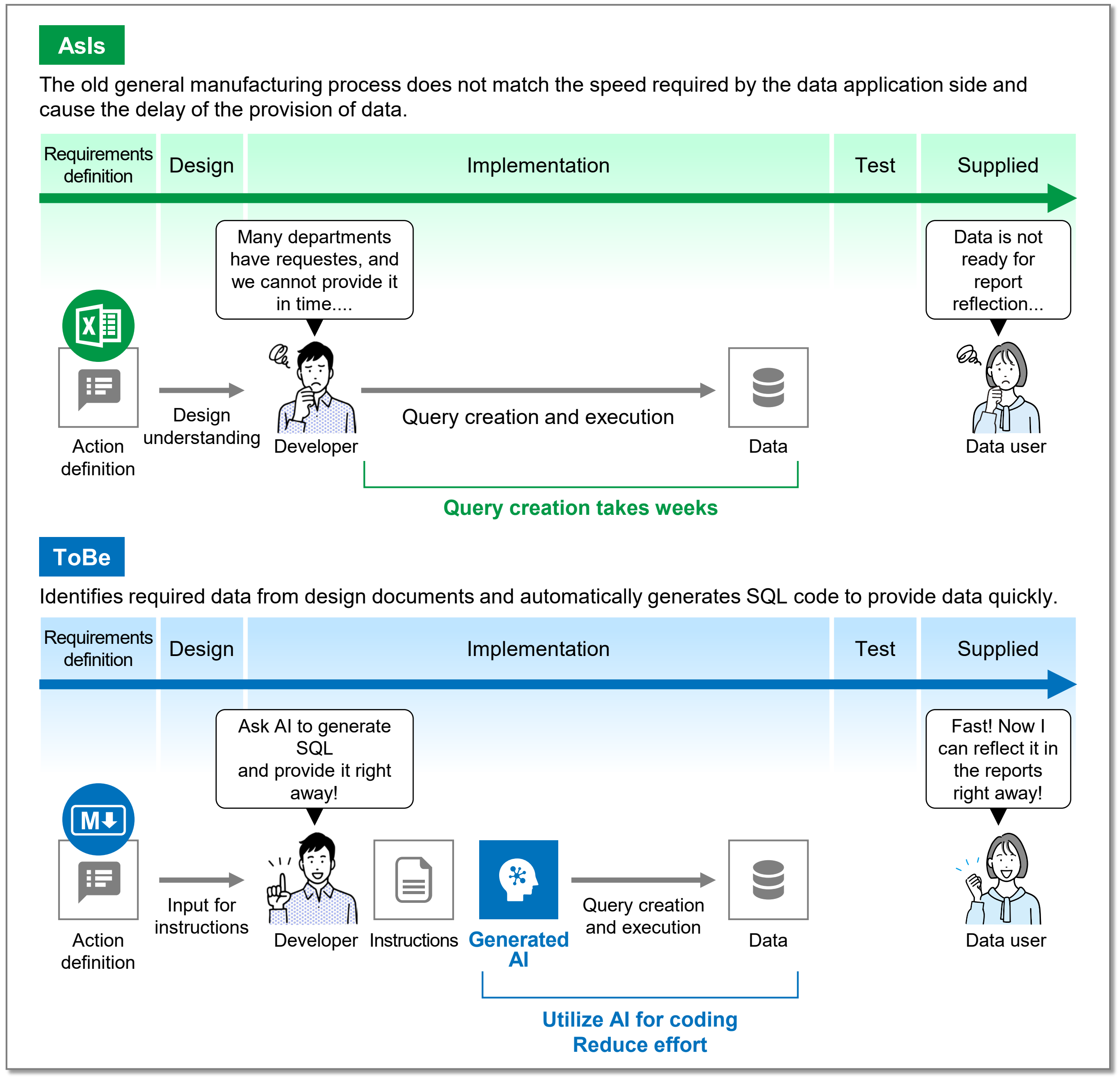
Snowflake Copilot

Snowflake Copilot, also provided by Snowflake, is a natural language queryable AI assistant that supports SQL query generation and optimization. Snowflake Copilot can generate and optimize SQL queries and fix problems by understanding the relationships between tables based on table schema information in response to SQL query questions and directives. It can also answer questions about Snowflake functionality based on Snowflake documentation.

※ NTT DATA is Snowflake's Elite Partner, a top tier partnership.

At our company, we are improving development efficiency by giving design documents directly to the SQL generation function prompt. There is always a design process before the manufacturing process, and by using design documents as they are, we can introduce AI without changing the development flow. In this case, if the design document is written in Excel, etc., it cannot be given as a prompt, so it must be written down in Markdown, etc. (It is also a good idea to write the design document in Markdown from the beginning). In addition, rather than simply giving the design document as a prompt, if you write the instructions in Few-Shot[*7] format, it is possible to improve the accuracy with detailed ingenuity.

Figure 3.1 Image of SQL Generation by Generated AI



*7 Few-Shot : A method in which a small number of examples are presented in a prompt and the AI learns to respond to a new task.

3.2 Importance of Design Documents

Now that it is possible to generate SQL applicable to commercial development by using design documents, the quality of design documents may become a bottleneck in future development using AI. Even at our company, we have often seen cases where the quality of generated SQL is poor, and even after trial and error with the prompts, the design documents were written incorrectly. In addition, as coding work becomes faster with the assistance of AI, it is expected that there will be an increase in cases where the speed of creating design documents cannot keep up with coding work.
In light of this trend, NTT DATA is working to support the creation of design documents by AI. In this way, AI supports the creation of design documents based on the output of the upstream requirement definition phase. When AI is introduced into general application development, the output of the upstream process is basically used as input to AI. On the other hand, there is a different perspective in the development of data marts and data pipelines at DWH, which will be introduced as future trends in Chapter 6.

Chapter. 4

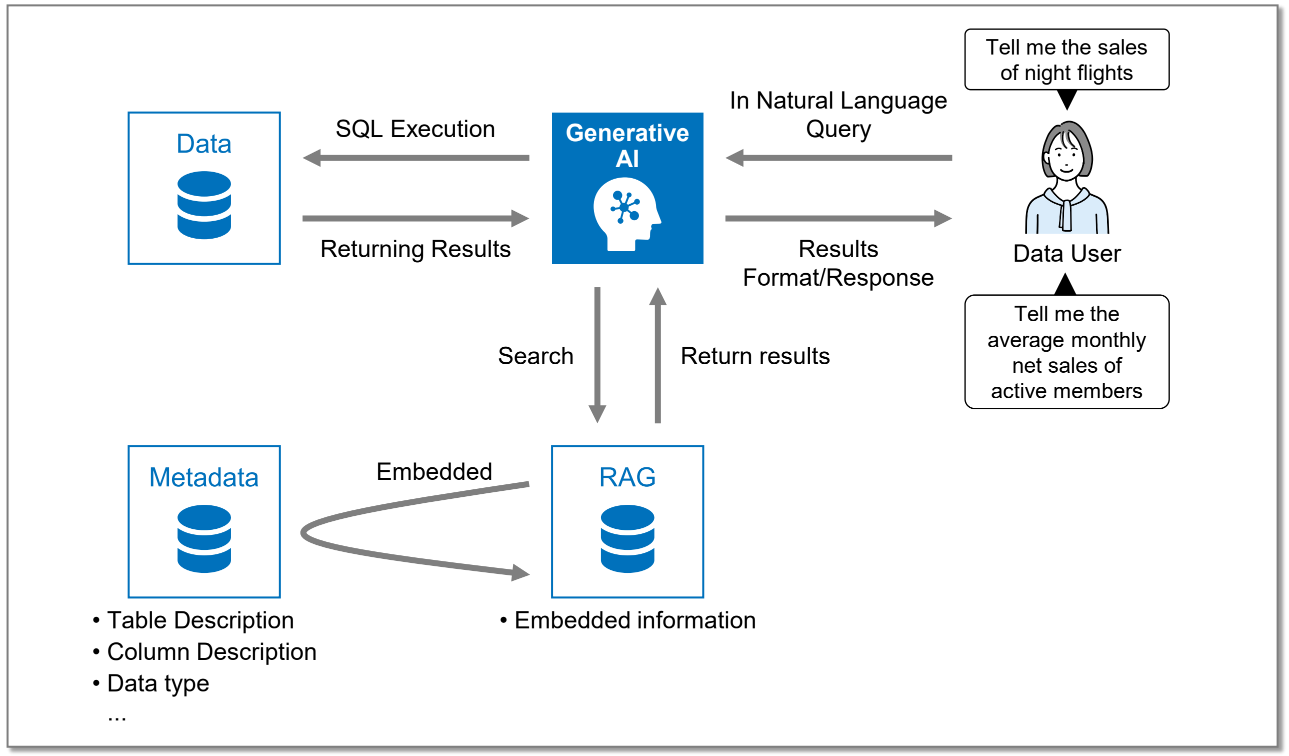
A new form of data utilization called "interactive"

This chapter explains how to solve the challenges of data analysts using generative AI. It explains how to support data exploration, aggregation, and analysis by AI, and how to prepare metadata to achieve this.

4.1 Interactive data utilization supporting data democratization

A major feature of data democratization is that non-experts play a leading role. As an interface that can be used by people who do not necessarily have high technical skills, chat-based forms of interactive data utilization are attracting attention.
Even in such interactive data utilization, the accuracy of answers is directly linked to what information is provided to the generation AI. Unlike the manufacturing phase, in the analysis phase, detailed specifications for processing such as design documents are rarely arranged beforehand. Instead, it is effective to provide the generation AI with information about the data itself: metadata.

Figure 4.1 Interactive Data Utilization Concept



Data Exploration

Semantic search using LLM can be implemented by embedding metadata and constructing an RAG[*8]. Semantic search is a method of providing more relevant search results by understanding the meaning and context of words.
For example, a natural language query such as "table storing sales data for 2024" or "table of customer purchase history" can quickly find an appropriate data set without depending on table or column names. Semantic search using LLM can reduce search time and greatly improve the efficiency of data analysis.

Data aggregation and analysis

In the same way, natural language instructions can be used to perform aggregation and analysis on the data found by searching. In this case, AI can generate SQL that understands the unique business conditions of the organization by having proper access to metadata.
For example, in order to perform a task such as "totaling the number of active members," it is necessary to understand the definition of each term in the organization, and the ability to generate SQL based on this information is required for AI application in this phase. This system makes it easy for business users who are not familiar with data to engage in data analysis and promotes data utilization throughout the organization.

What is metadata?

Metadata is the collateral information of main data. It is the idea of organizing information such as the use, constraints, and scope of application of data. It is necessary to manage metadata to extract the necessary information from the huge amount of data.
The following three types of metadata are used for data analysis.
In particular, business metadata that explains the meaning of the target data is important when using generative AI in the analysis phase.

Business metadata

- Business information, such as business rules and definitions and descriptions of tables and columns
- Meaning of category values and interpretation of special values such as 0

Technical metadata

- Information about IT, such as column properties and database object properties

Operational metadata

- Information generated during system operation, such as job execution logs for batch programs and history of data extraction and results

*8 RAG : Abbreviation for Retrieval Augmented Generation, a method that utilizes the latest information, such as a large amount of business documents, regulations, and other internal information accumulated in the company, and has a large-scale language model (LLM) answer questions based on this information.

4.2 Importance of maintaining metadata

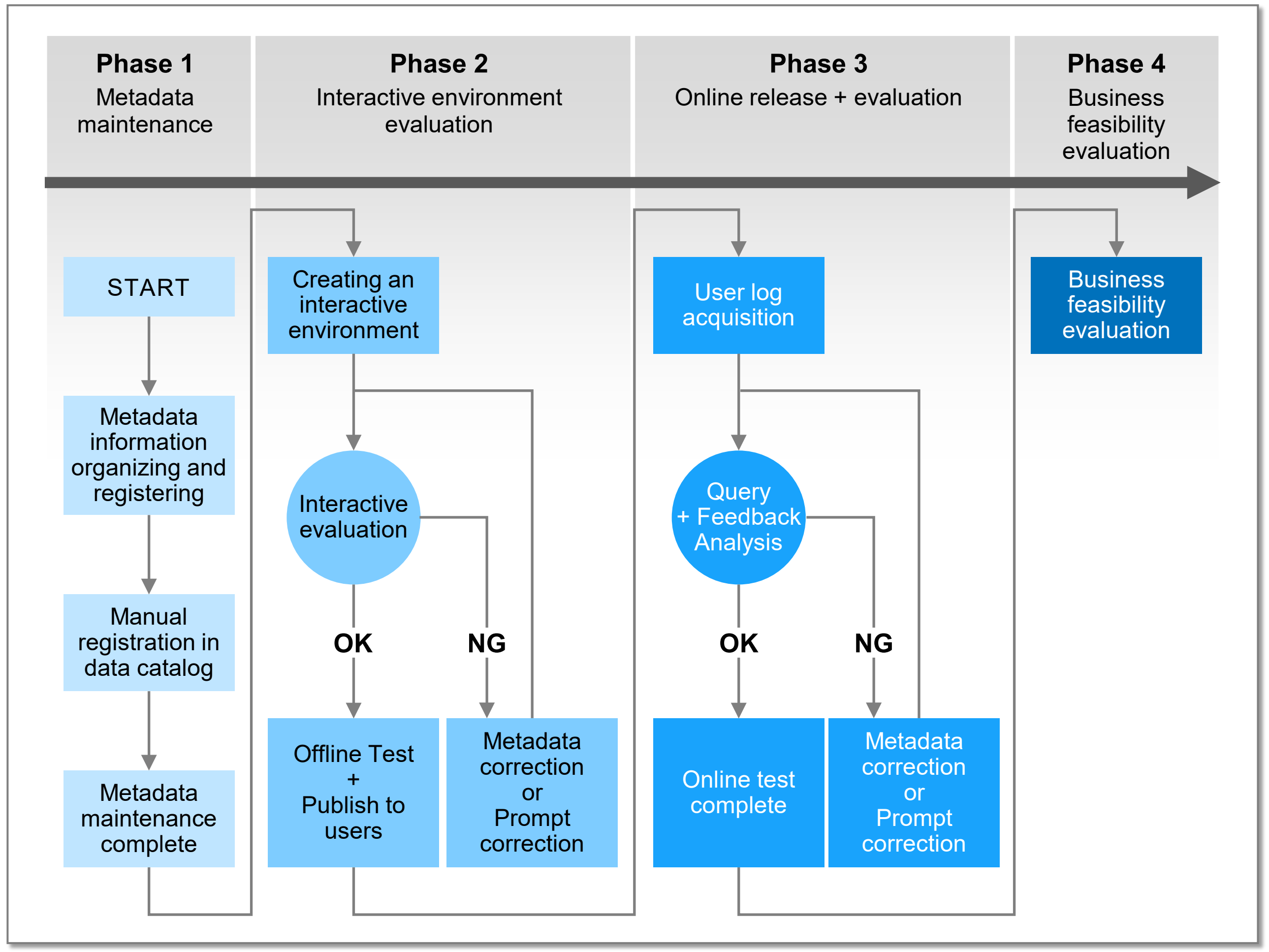
Generation AI is by no means a silver bullet (technology or practice that immediately improves development productivity), but metadata is the key. However, metadata is not a silver bullet either. At present, few companies have sufficiently maintained metadata that can be used for AI. While the minimum information, such as the logical names of columns and the frequency of table updates, is often included in the catalog as part of data mart development, detailed column descriptions often take a backseat. This is not only a low priority during development, but also because it is difficult to know what information to include in advance because it is impossible to assume 100% in advance how the data mart will be used.
In particular, business metadata, which is important for AI generation in the analysis phase, must be actively nurtured in the business processes of both data providers and data analysts. In contrast to the traditional passive and static approach, this approach is called Active Metadata Management*9 and has become a new trend in metadata management in recent years.

4.3 Flow of development and deployment

The major difference between development and deployment of an interactive analysis environment and regular SI development is that the quality of the metadata greatly influences the achievement of functional requirements. Therefore, it is very important to cycle through the addition and modification of metadata based on the test results of answer accuracy. In addition to offline testing by the developers themselves, it is also desirable to conduct online testing by users because it is difficult to know in advance how interactive analysis will be used. In addition to registering metadata in a general data catalog, it is also possible to register a set of questions and answers directly in the form of FAQs. Especially in the phase of improvement based on user feedback, continuous improvement of accuracy can be expected by having users evaluate the appropriate answers to the questions and registering the ones with high evaluation in FAQs.

*9 Gartner Research "Quick Answer: What is Active Metadata?": https://www.gartner.com/en/documents/4556899

Figure 4.3 AI × Data Utilization : Developing an Interactive Analysis Environment



Techniques for Improving Accuracy of Text 2 SQL

To improve response accuracy, it is useful to add the following metadata

		(Example)
①Data modeling	A) Relationships between entities	Composite Primary Key and Foreign Key Relationships for Tables A and B
	B) Proper data type	Date format should be date type ∵ To make date-related calculations easier for the AI
②Metadata	C) Domain Definition	Business meaning of each individual code value with respect to the code value
	D) Domain-specific information	Business metadata and company-specific terminology definitions
	E) Filter Conditions	Filter conditions based on domain knowledge, such as the date of the most recent inventory stock take, etc.
③Dialogue examples	F) Query example	Frequent queries in an analytical environment
④Complex table generation	G) Materialized Views	Temporary view generated from complex logic ∵ To prevent hallucination, do not allow AI to generate it
	H) User-defined table functions(UDTF)	Complex aggregate indicator definition ∵ To prevent hallucination, do not allow AI to generate it

AI/BI Genie Enables Data Analysis in Natural Language

Databricks' AI/BI Genie enables business users to perform self-service data analysis and visualization using natural language.
Specifically, it translates business domain questions and directives into analytical queries based on Unity Catalog metadata and visually answers them using tables and figures. It also continuously updates the metadata in response to data changes and new questions. It can also tune for wrong answers, providing users with more accurate analytical information.
In addition, benchmarking can be used to create test questions to evaluate answer accuracy for questions that users frequently ask.

Chapter. 5

Semantic Layer Technology to Prevent Data Silos

Semantic layer technology has been attracting attention in recent years to solve the problem of misalignment of indicators, a common issue for both data providers and data analysts. This chapter provides an overview of the semantic layer along with its history and also confirms that it can be applied to interactive analysis in the previous chapter. On the other hand, since it is a relatively new technology field, the current issues are also summarized in this chapter.

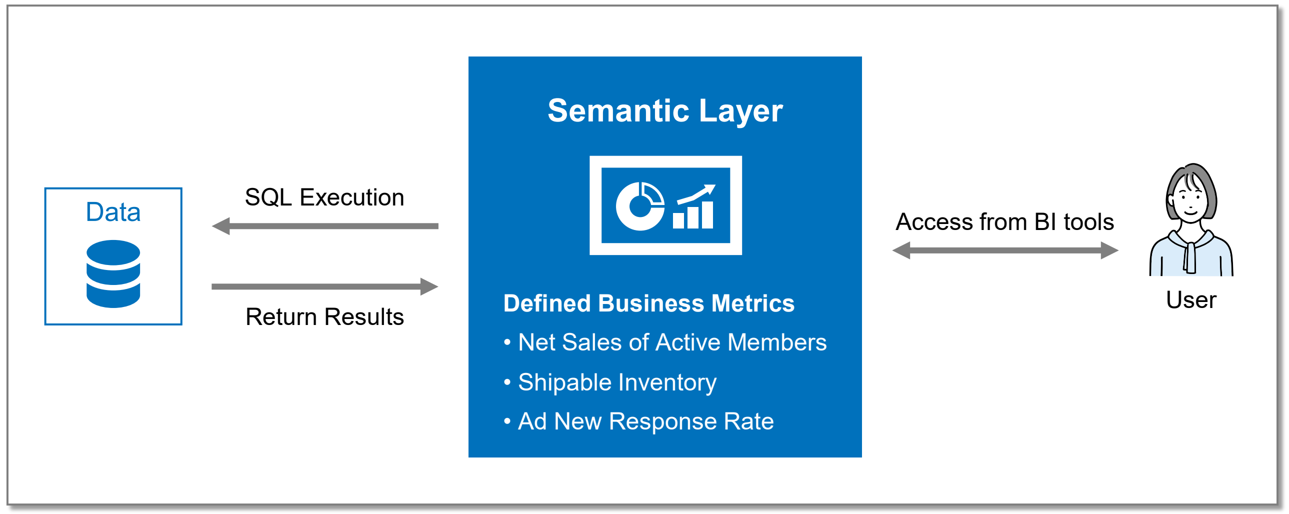
5.1 History of Semantic Layer

The term “Semantic Layer" may sound like an entirely new technology area, but it has its roots in BI tools and is a familiar concept to those who use BI regularly. Similar functions are called "semantic models" in Power BI and “Semantic Layers" in Looker.
In BI, before creating a dashboard, you first define relationships between tables and create new columns (often called "measures" in BI) to calculate metrics. By defining the semantics of this data in advance, you can ensure that different dashboards display the same metric numbers consistently.
The semantic layer, which has been gaining attention in recent years, is technically called the (Universal) semantic layer. It is a (local) semantic layer that was originally provided as a closed function within each BI tool, but now exists independently, allowing the various tools to access it in common.
In recent years, with the democratization of data and the resulting silos, the semantic layer has gained attention. In 2024, the 4th Semantic Layer Summit was held, and the father of DWH, Bill Inmon, attended.

5.2 What is the Semantic Layer?

The (Universal) semantic layer is a middle layer that sits between data and data analysts, and translates complex data meanings and business logic into understandable business concepts.

Figure 5.2 Conceptual Diagram of the Semantic Layer



As with BI tools, relationships between tables are defined, and the logic of business metrics such as KPIs is also defined so that they can be reused. For example, by defining logic such as "active members" and "net sales," or the combined "active member net sales," sales departments can aggregate monthly data by store, or marketing departments can aggregate quarterly data by product category. Traditionally, developers have often created a new data mart for each analysis from these different perspectives, but this increases the burden on developers. By centrally managing the core logic of the metrics to focus on at the semantic layer, and allowing users to choose the analysis points flexibly, the agility of data utilization can be enhanced while maintaining effective data governance. In this way, the Semantic Layer is a technology area that can kill two birds with one stone, which makes both developers and analysis users happy.

5.3 Semantic Layer × Generated AI

The semantic layer defines the meaning (semantics) of data and provides consistent logic for data analysis and usage. This is extremely valuable metadata, and as discussed in the previous chapter, interactive analysis using this metadata is increasingly provided as a function of Semantic Layer products.

Product Examples

Cortex Analyst

Snowflake's Cortex Analyst[*10] is a data analysis application development support function that enables business users to ask questions in natural language and reply highly accurate and reliable answers. Specifically, it converts business domain questions and instructions into highly accurate SQL queries based on measure and dimension information predefined in a semantic model and provides highly reliable answers.
In addition to seamless integration with existing business tools through APIs, data is securely processed within Snowflake's security and governance framework.

Ask dbt

Ask dbt[*11] provided by dbt is an application development support function that combines the semantic layer functionality with LLM to enable querying of data in natural language. dbt Semantic Layer allows defining semantic models and metrics (business indicators) from tables. Ask dbt selects and executes metrics and dimensions defined in the dbt semantic layer based on natural language questions and directives, and outputs results. This allows for highly accurate responses to complex metric capture using a centralized set of metrics.

*10 Cortex Analyst : https://docs.snowflake.com/en/user-guide/snowflake-cortex/cortex-analyst
*11 Ask dbt : https://docs.getdbt.com/docs/cloud-integrations/snowflake-native-app

5.4 Challenges for widespread use

This is a promising technology that not only prevents data from becoming siloed but also supports data analysis in conjunction with AI. However, a major challenge for widespread use is the narrow scope of information. Currently, both Ask dbt and Cortex Analyst can only reference metadata defined for the semantic layer. Therefore, even if you add a comment to an object such as a table or column in a DDL statement, you cannot use that information. To utilize such metadata, you need to copy that information into the metadata definition file for the semantic layer, which is double-managed and time-consuming. In some cases, support for the creation of definition files is provided, but even so, metadata is fundamentally distributed, which increases the operational burden in the field.
In the next chapter, we will introduce the latest trends and NTT DATA's vision.

Chapter. 6

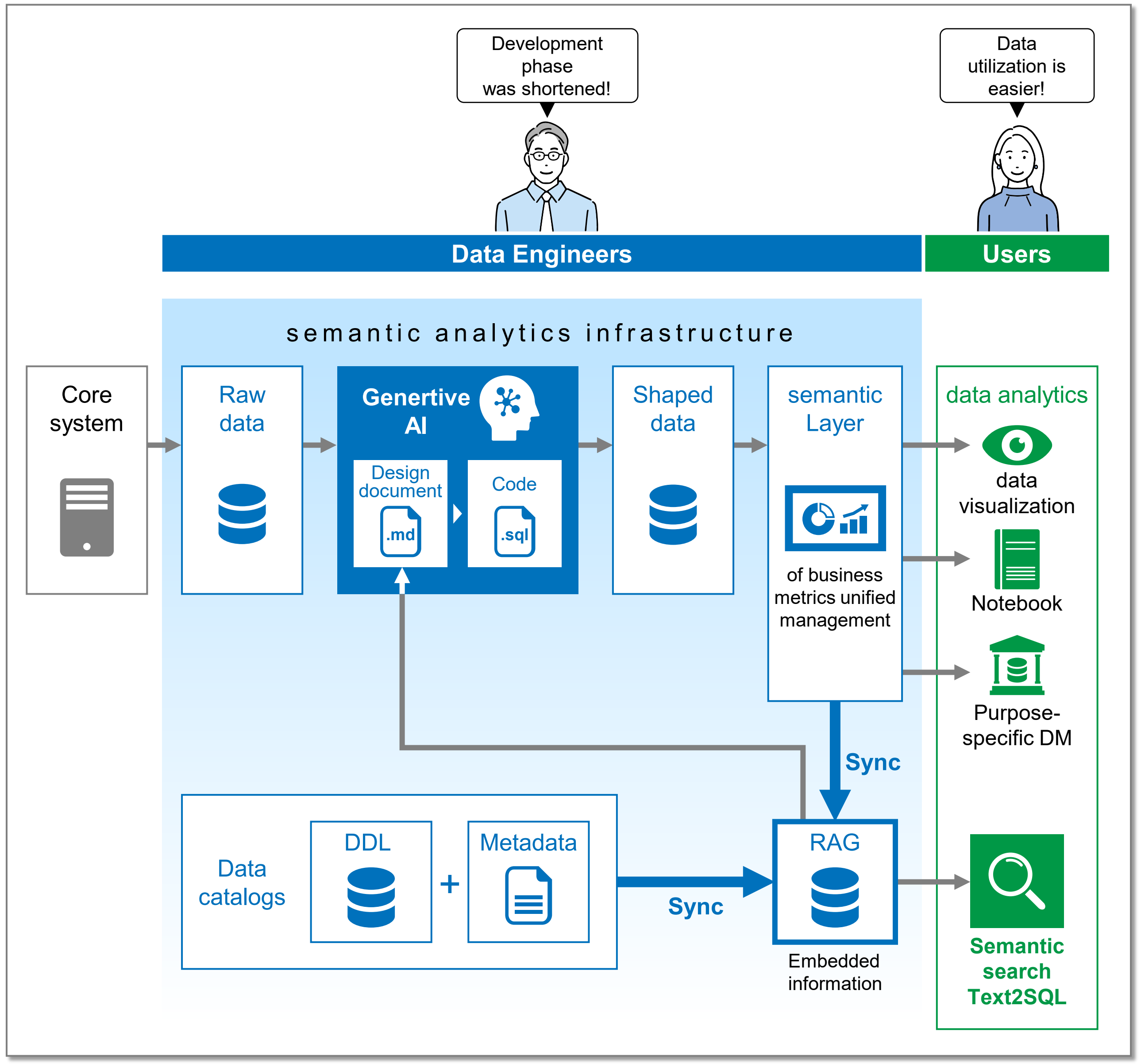
NTT DATA's Vision of the Future

In order to use generative AI in the data utilization area, data management assumes the use of generative AI is necessary.
In addition to the existing metadata management, we will introduce new metadata, such as design documents for ETL development in the data provision business and semantic layers for data analysis business.
In addition, we will introduce an AI-Agent specialized in data utilization that can use this metadata to improve the efficiency of operations in the data utilization area, thereby realizing autonomous execution of operations and accelerating and enhancing the overall process of data utilization.

6.1 Integration of Metadata

In the previous chapters, we explained that it is important to provide sufficient background information to AI when using generated AI. Going a step further, in particular, in the application of AI in DWH, it is important to provide information related to the data itself, which is the core of the application of AI. In other words, how to provide sufficient metadata to AI is a major theme of the application of AI in DWH in recent years.
Since the current context (storage area in human terms) of generative AI is finite, RAG, which is a technology to pass only the necessary number of metadata to AI each time, is rapidly attracting attention. However, what is important here is the range of information that can be referenced by the RAG infrastructure. Even if RAG is an excellent technology and has raw data information sent from, for example, the core accounting system, it is not hard to imagine that it would be difficult to search for and aggregate data if the sales department did not know about the data mart.
In Chapter 5, I mentioned the problem that, while the semantic layer is a useful metadata maintenance tool and provides interactive analysis functions that take advantage of it, metadata from other products such as data catalogs cannot be referenced. In fact, such a problem can be seen in various tools. In BI, data catalogs, and document management tools, various functions have been released one after another to keep up with the generative AI boom of the past few years. However, the metadata used by these generative AI functions is basically closed within each product, and those that are designed to work with other products have not yet become widespread. In the future, however, our company believes that generative AI services with RAGs, which provide access to all metadata in an organization across product boundaries, will be able to deliver the truly "right" metadata to AI and will become widespread.

Figure 6.1 Ideal diagram of metadata within an organization aggregated into a single RAG



Some products are actually moving in this direction. For example, Databricks announced Unity Catalog Metrics at the Data+AI Summit 2024. This is a new semantic layer function added to Unity Catalog, the data catalog function of Databricks. It also aims to link with metadata defined by popular semantic layer products such as dbt and cube[*12]. We are seeing a movement to integrate metadata across products and technology domains.

In the past 10 years, the leading role of data utilization was played by DWH, which manages data. In the generation AI era, the leading role of data utilization will be products that manage metadata. At our company, we are researching and verifying the latest trends in this area and creating value in the business field with our customers.

*12 Cube : A service that consolidates logic across tools through a semantic layer that exists independently between DWH and each BI tool and has been provided as a closed function within each BI tool.

6.2 Design Document and SQL Generation Become More Accurate with Metadata

In Chapter 3, we introduced that the quality of design documents is important for AI-driven SQL generation, and based on this trend, our company is also working to support the creation of design documents by AI. On the other hand, in the development of data marts and data pipelines at DWH, it may become a trend in the future not only to input the products of the upstream requirement definition process, but also to use information from the integrated metadata infrastructure in the form of RAG as input. Considering that even in conventional development without AI, design was conducted by investigating information (column names, data types, interpretation of null values, frequency and timing of updates, impact on downstream dashboards, etc.) in existing tables, this information is an essential element in the design process, and it is expected to play an important role in improving the accuracy of AI-driven design document creation support. As AI technology develops, the ability to maintain this information in a form that can be used by AI will greatly influence the development productivity of organizations.

6.3 Integrated Metadata Infrastructure Supports Offensive and Defensive Data Use

In the context of data democratization, Chapter 4 introduced interactive analysis, which enables the use of data without technology, and Chapter 5 introduced the semantic layer, which prevents silos of data and logic. In the past, if data scientists in specialized departments performed sophisticated analysis, and therefore metadata was understood by members of those specialized departments, it was not a big problem. However, in order for each person in the business and sales departments to find value from data in their daily operations, metadata needs to be widely open, not closed, and integrated without silos. Not only can we expect an aggressive effect of faster and more accurate cross-departmental data analysis, but also in the sense of defense, which is to understand data marts and dashboards within an organization and to optimize across organizations while preventing reinvention of the wheel, demand for an integrated metadata platform is expected to increase in the future.

6.4 Automated development and analysis in general

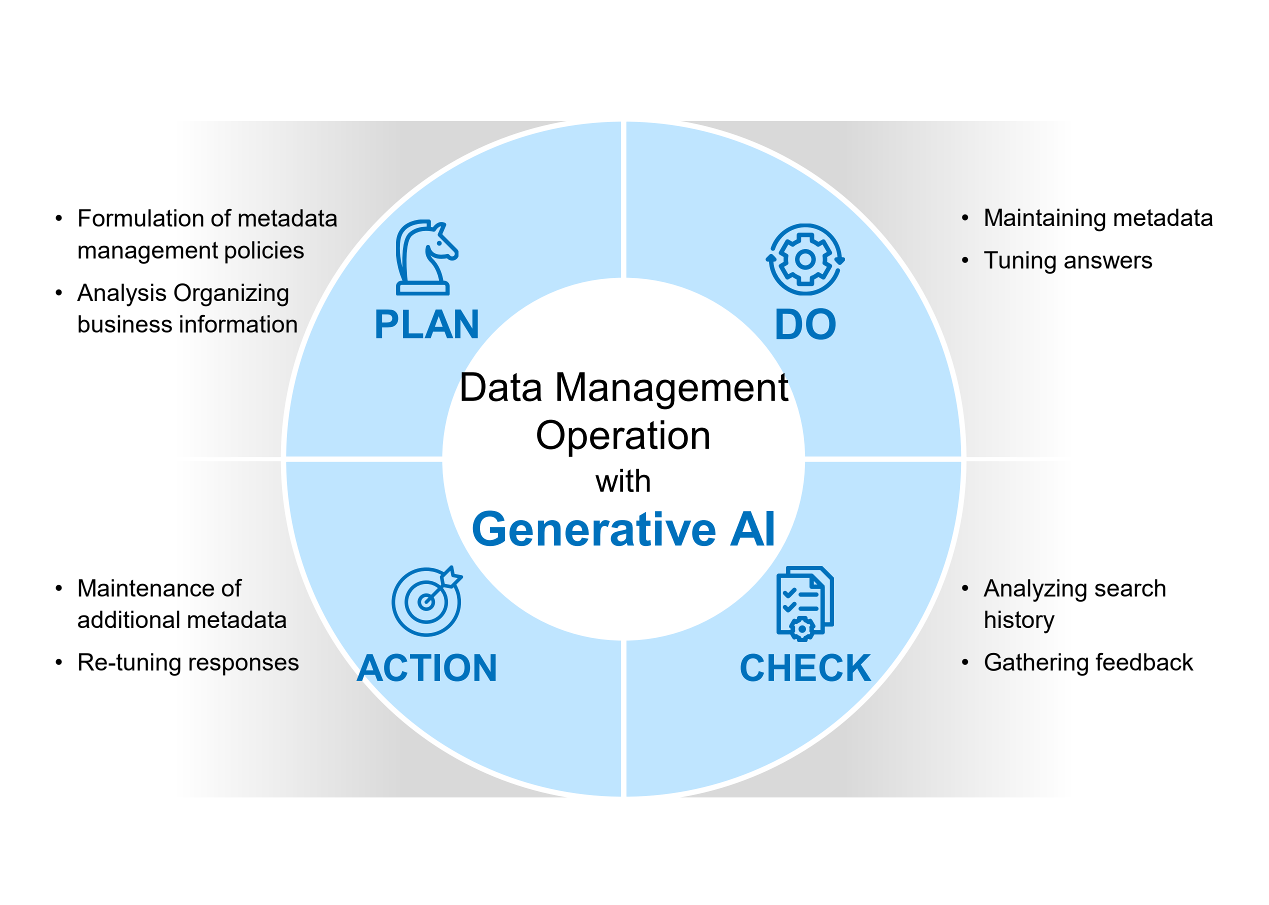
Even though the use of generative AI has become widespread, there are still many tasks in development and analysis that require human intervention, such as hearings with relevant parties, creation of design documents, review, revision, and hypothesis verification, and overall productivity does not improve as expected.
Our hypothesis for solving these challenges is to automate development and analysis in general by implementing an AI-Agent for data utilization based on "Smart AI Agent™," a concept that encompasses the AI-Agent provided by NTT Data, announced in October 2024.
For example, in data utilization, for ETL development, which is the main task of data providers, we are envisioning a future where design documents can be automatically created, reviewed, and displayed in BI tools after testing based on source code samples and specification information.
In data analysis, repeating the PDCA cycle of hypothesis examination and analysis involves a lot of tacit knowledge and is difficult to achieve with the accuracy of conventional AI.
For such tasks, we believe that by formalizing human tacit knowledge and reflecting it in the design of an AI-Agent in a form that can be independently executed, we will be able to analyze and report with insights.
However, even as AI agent technology evolves, in the absence and dispersion of metadata, humans have no choice but to respond. Once metadata can be organized and aggregated, AI-Agents in the data utilization domain will be able to demonstrate a great deal of self-reliance in business execution. This will accelerate fundamental changes in the entire data utilization process.

6.5 The Future of Data Management

This article has consistently emphasized the importance of metadata. Metadata maintenance and management is a theme that is often discussed as part of the larger framework of data management. Data management is an extremely important initiative that determines the success or failure of a company's data utilization. However, our company often receives inquiries about how to proceed with data management, and many companies feel that there is a disconnect between data management as a concept and how to proceed within their organization.
If you refer to DMBOK[*13], which provides a systematic overview of data management, you will find chapters such as metadata management, data quality, and document and content management. While these topics are extremely important, they are difficult to measure quantitatively. This difficulty is one of the reasons why it is difficult to proceed with data management in the field, obtain budgets and promote results within the organization.

However, recent technological developments have enabled both data providers and data analysts to maximize the value of metadata by using generative AI and use it directly and quickly in their own operations. This strong incentive serves as a great banner for promoting data management efforts within organizations and can also function as a health meter that quantitatively evaluates a company's data management through AI response accuracy. Using this health meter as a compass, our company believes that promoting down-to-earth data management through the PDCA cycle is what data management looks like in the generative AI era.

Figure 6.5 Metadata Operation with Generated AI Assumption



Our company provides support for customer data management. Based on the latest trends, including generative AI, we support the success of customer data utilization, from assessment of the current situation to consulting for the planning of consulting measures, and support for the implementation of planned measures.

*13 DMBOK : A book and framework that systematically summarizes knowledge about data management. DMBOK defines 3 types of metadata: business metadata, technical metadata, and operational metadata.

Chapter. 7

Conclusion

In this white paper, we have presented general issues in data utilization and introduced generative AI and semantic layers as solutions, and data management based on generative AI.

NTT DATA has supported generative AI and semantic layers, and has signed partner agreements with Snowflake, Databricks, and dbt Labs. In addition, by applying Smart AI Agent™ to the data utilization area, we aim to improve the agility of customer data utilization.

If you have any problems or questions regarding data utilization, please contact us.

Note : The product names, company names, and organization names used herein are trademarks or registered trademarks of their respective companies.

Contact

Apps & Data Technology Department,
NTT Data Group Technology Innovation Headquarters

Mail: dioffering-contact@kits.nttdata.co.jp

Contact

Hironobu Hamakata General Manager

He leverages his extensive knowledge of data collection and storage from data source systems, large-scale data distributed processing, data prep, data provision, data management, data governance, and a wide range of other areas from consulting to implementation to lead the clients' businesses.

Yoshimitsu Yagi Manager

After working on large-scale renewal and development of a subsidiary of a major banking corporation, he worked as a development team leader in the R&D department.
Currently, he is leading many data utilization projects with his expertise in data science, agile, and AI governance.