



Discovery User Manual

Easy and quick connection, analysis, and visualization of
Big and complicated data



metadiffInDays =
formatDate = MM/DD/YYYY
dtExpiryDate =
diffInDaysmillisec=0;
differenceInDays=0;
discountEligibility = parse
discountsecondEligibility
validateDifference(ISaleDate
sendEmailAlertforDateDiff()
sendEmailAlertforError();
difference between two
diffInDays = diffInMillise
if(count != 0){instObject.s
ageMailAlert=true}if(Age
Numeric getDateDiffIn
format formatDate =
format dtExpiryDate =
diffInDaysmillisec=0;
differenceInDays=0;
discountEligibility = pars
secondEligibility
difference(ISaleDate
sendEmailAlertforError());
differenceInDays=0;



metatron 2.0

Part 1

Discovery Outline

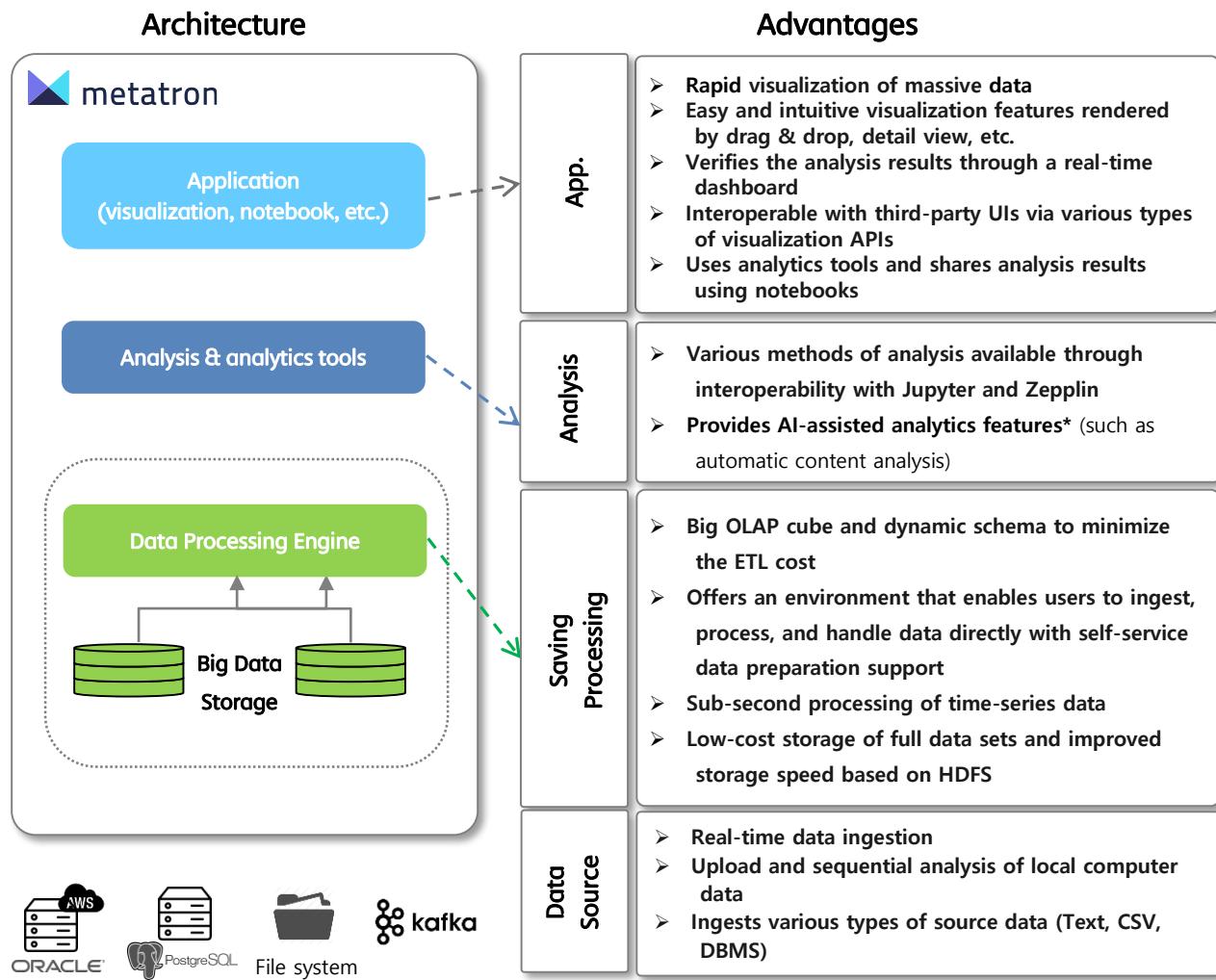
Contents

1. Advantages of metatron Discovery.....	3
2. Structure of metatron Discovery	5
3. Relational OLAP vs. Multidimensional OLAP.....	7
3.1 Characteristics and limits of the relational OLAP system.....	7
3.2 Multidimensional OLAP system and druid engine.....	9

1. Advantages of metatron Discovery

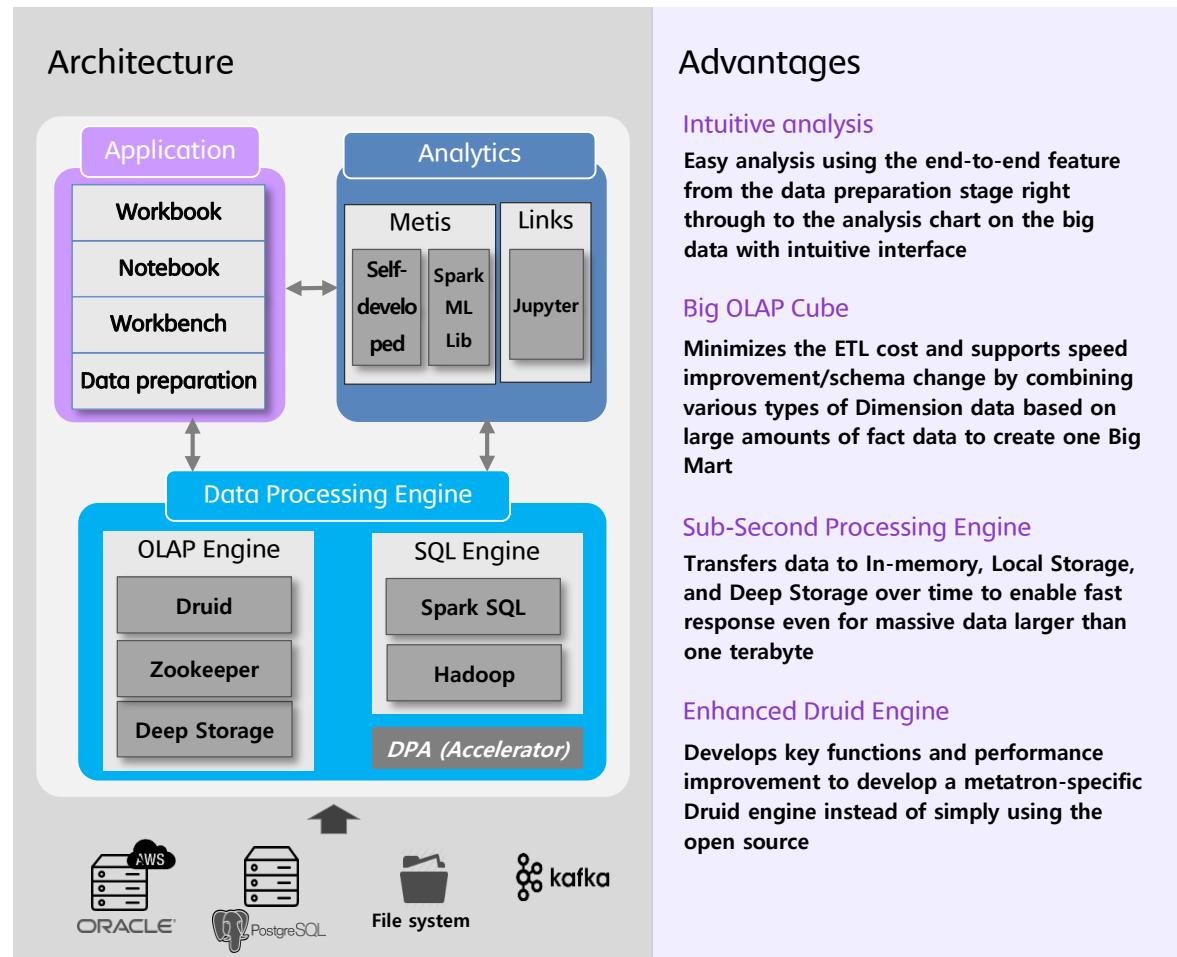
metatron Discovery is a product for Single Solution Data Discovery, supporting end-to-end functions for massive data from preparing to searching data based on visualization and advanced analysis. The following figure summarizes the architecture and main advantages of metatron.

Architecture and advantages of metatron-Based Technology



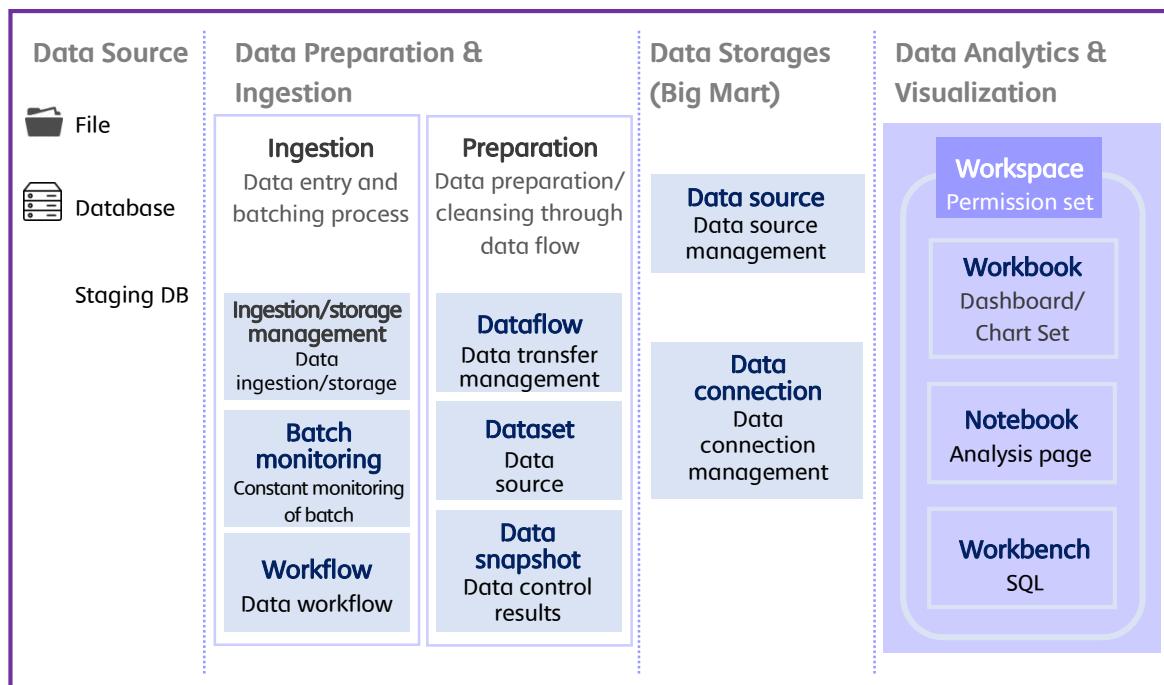
* The features offered are under implementation.

Architecture and advantages of the metatron Discovery module



2. Structure of metatron Discovery

metatron Discovery imports data from a data source ingested in the metatron operation server or other external data sources, analyzes it using various advanced analysis functions, and outputs the result as various charts and reports. To utilize this module, you must understand the overall structure shown below.



Source Data

Data imported when creating the data source. The current metatron version supports file type, databases, and staging DB type.

Data Preparation & Ingestion

Refines and processes imports from source data, and then ingests them to metatron's Hive storage or exports them into local files.

Data Storage

Saves data to the Druid engine of metatron based on multi-dimensional OLAP. Data saved as this type facilitates easy visualization and analysis. (See Section 3.2 for more information.)

Data Analytics & Visualization

Searches, analyzes and visualizes data stored in data storage.

- **Workspace:** Space to create and manage workbooks, workbenches, and notebooks.
- **Workbook:** Analyzes multi-dimensional, time-stamped data sources stored in the Druid engine and visualizes and reports the results.
- **Notebook:** Able to perform advanced analysis such as deep learning or machine learning by linking external analytics tools including R, Python, and Scala.
- **Workbench:** Performs an SQL-based analysis.

Data monitoring

Monitors various data logs generated in metatron's Staging DB (internal Hive DB) and workbench (external DB).

User Permission and Account

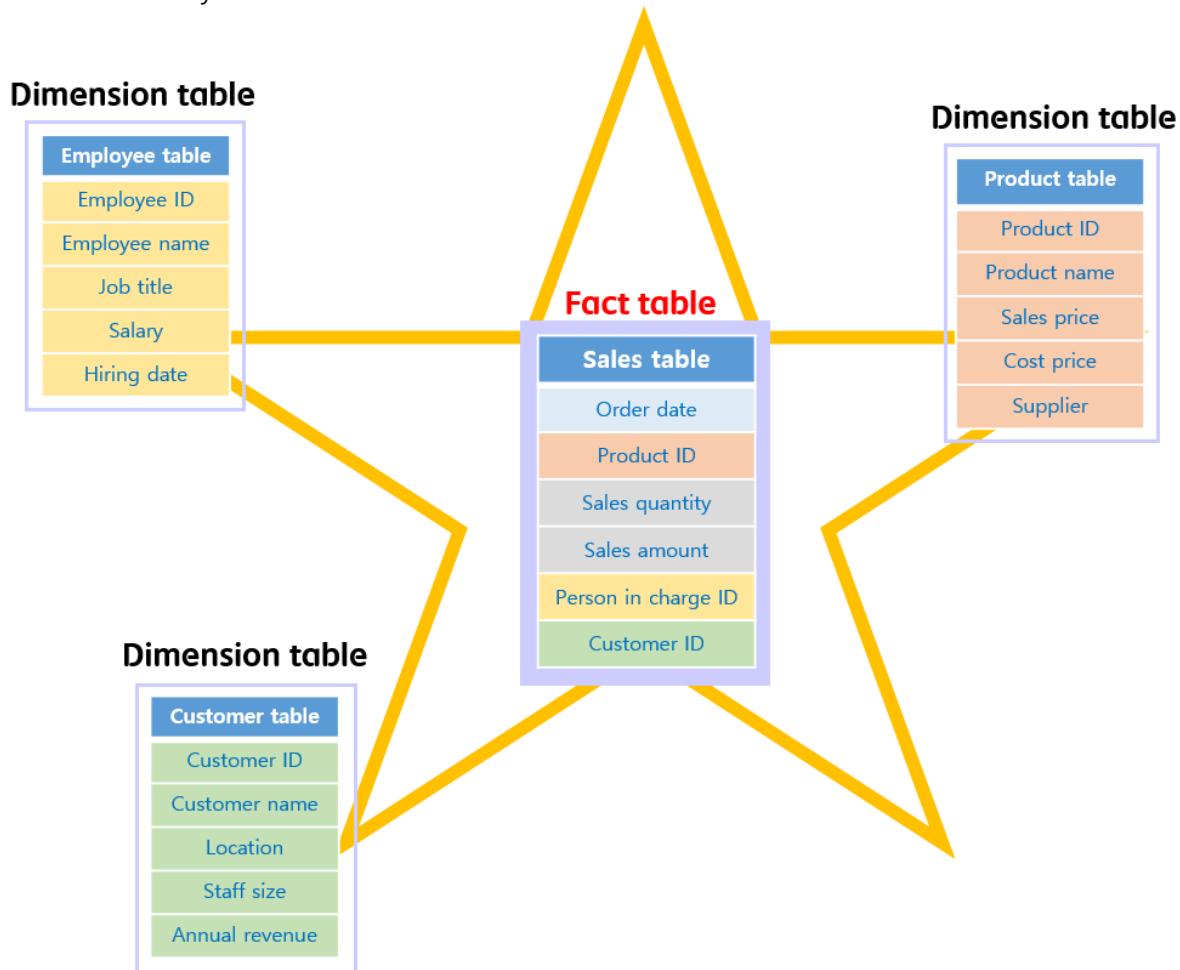
Adds and deletes metatron Discovery users, or manages user permission.

3. Relational OLAP vs. Multidimensional OLAP

Various cutting-edge systems and methods for big data analysis are applied in metatron Discovery. Among them, you must understand the concept of the "multi-dimensional OLAP system" using the Druid engine to use Discovery properly. This section will describe this concept. This section briefly describes the characteristics and limits of the relational OLAP system, which has been the most popular system for data collection and management in the past, and explains how Discovery's multidimensional OLAP system resolves these limits.

3.1 Characteristics and limits of the relational OLAP system

A relational OLAP system is the most popular method of collecting and managing data. The figure below indicates an example of the star schema structure, the most frequently used structure in a relational OLAP system.



As shown in the figure, a basic star schema is composed of many dimension tables and a small number of fact tables.

- Detailed data on each dimension of the schema is stored in the **dimension table**. In the example shown in the figure above, the three dimension tables each contain data relating to the product, the employee, and the customer. Each of these is called a "dimension" because you can use them as a standard for data classification (e.g., quantity sold **by each employee** and quantity purchased **by each customer**).

- In the **fact table**, a record containing the details of the event is added one at a time whenever a new event occurs. In this case, a connection is made to each dimension table using the reference ID as mediator.

The figure below summarizes an example of simple table based on the schema structure above to aid in understanding.

Dimension table (product)					Dimension table (person in charge)					Dimension table (customer)				
ID	Product name	Sales price	Cost price	Supplier	ID	Employee name	Job title	Salary (\$)	Hiring date	ID	Customer name	Location	Staff size	Annual revenue
P1	Pencil	200	100	Supplier 1	E1	Employee 1	Manager	3,500	2009.09.01	C1	Corp. A	Paris	100–499 persons	\$500B–\$999B
P2	Scissors	1,000	700	Supplier 2	E2	Employee 2	General manager	4,000	2006.10.10.	C2	Corp. B	London	50–99 persons	\$100B–\$199B
P3	Eraser	100	50	Supplier 1	E3	Employee 3	Assistant manager	2,500	2016.07.08	C3	Corp. C	Seoul	100–499 persons	\$500B–\$999B
P4	Ballpoint	500	300	Supplier 3	E4	Employee 4	Director	6,500	1999.08.01	C4	Corp. D	New York	5–49 persons	\$50B–\$99B

Fact table

Order date	Product ID	Sales quantity	Sales amount	Person in charge ID	Customer ID
1/2/2015	P2	8	8,000	E4	C4
1/31/2015	P2	20	20,000	E1	C2
2/3/2015	P2	16	16,000	E3	C2
2/22/2015	P4	19	9,500	E1	C4
2/28/2015	P3	3	300	E1	C2
2/28/2015	P3	8	800	E2	C3
3/2/2015	P2	4	4,000	E4	C3
3/3/2015	P3	20	2,000	E3	C2
3/5/2015	P3	8	800	E1	C2
4/4/2015	P3	16	1,600	E3	C3
4/22/2015	P3	10	1,000	E4	C4
5/5/2015	P4	9	4,500	E2	C1
6/9/2015	P2	3	3,000	E3	C2
8/11/2015	P3	5	500	E2	C3
9/21/2015	P3	9	900	E1	C1
10/30/2015	P4	3	1,500	E4	C4
11/11/2015	P4	15	7,500	E3	C4
...
10/19/2018	P2	12	12,000	E4	C3
11/2/2018	P3	20	2,000	E2	C4

Although the products sold, the sales staff, and the customer making the purchase are recorded as a simple ID in each record of the fact table, each of these IDs can be used for searching detailed information if necessary, since they are connected with a unique dimension table. For example, if we search the product table, the P2 product in the first record indicates that the applicable product type is pencil, the unit price for sale is 200 KRW, the unit price for supply is 100 KRW, and the supplier is Supplier 1. Using such data collection method offers the advantage of being able to use the storage area effectively by storing only the minimum amount of data in the fact table.

When trying to analyze data stacked this way, however, the reference ID of the relational OLAP actually becomes the cause of decreasing data accessibility and speed. This is because the original data connected to all of the reference IDs included in each record used for analysis in the fact table must be loaded one by one into each dimension table. If there are 100,000 records for analysis, and 4 reference IDs per record, this must go through a data loading process of 100,000 x 4 times, causing that amount of load on the data processing device. Moreover, there are only a limited number of skilled people who understand the schema structure of the database properly, meaning only they are able to use such a data load query.

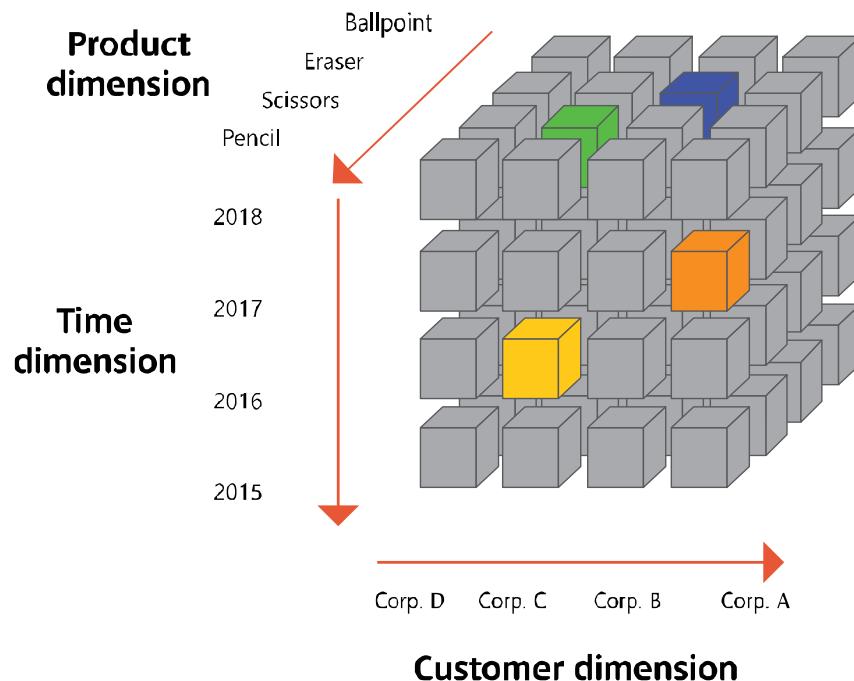
The multi-dimensional OLAP system described in the next section overcomes the limitations of such a legacy structure by saving all data into one or a small number of tables.

In a Druid engine-based multidimensional table, a new record is added each time an event occurs in the same way as with the fact table of the star schema system. The time is also recorded as a time stamp. As shown in the example above, however, all data items are saved to one table at once without using the IDs that refer to specific dimension tables. Since the time stamp is also considered to be a dimension (time dimension), the data mart above has a total of 4 dimensions.

Although data capacity is increased since duplicated data is increased if this storage method is selected (for example, the same product name is mentioned repeatedly in many records), it enables faster and more effective operation when it comes to analysis for the following reasons:

- Since all data is stored in one or a small number of tables, and can be extracted by simply selecting only the columns required for analysis at the time, there is far less data load compared to the relational structure.
- The cost required for understanding reference relations between tables and filling out/executing the query is reduced accordingly.
- In particular, since it is easy to store and process the data mart by distributing it over many nodes (based on Hadoop cluster in the case of Druid engines), data can be processed and analyzed in terabyte units very promptly and effectively.

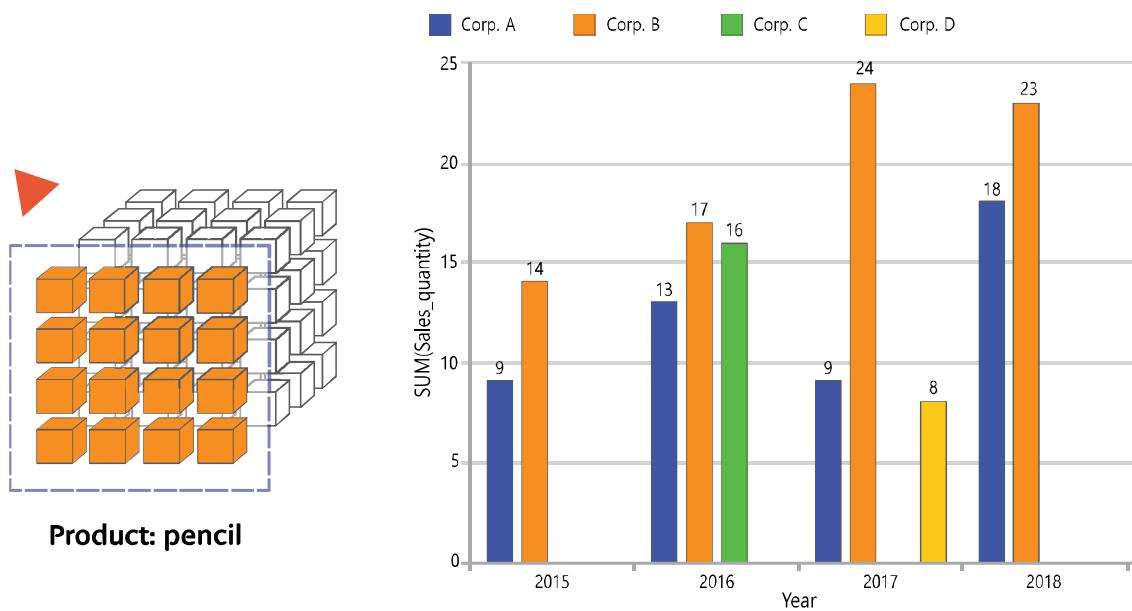
We are now going to provide you with more information on how data analysis takes place in a multidimensional OLAP system. A three-dimension cube model is sometimes used for understanding multi-dimensional data mart analysis.



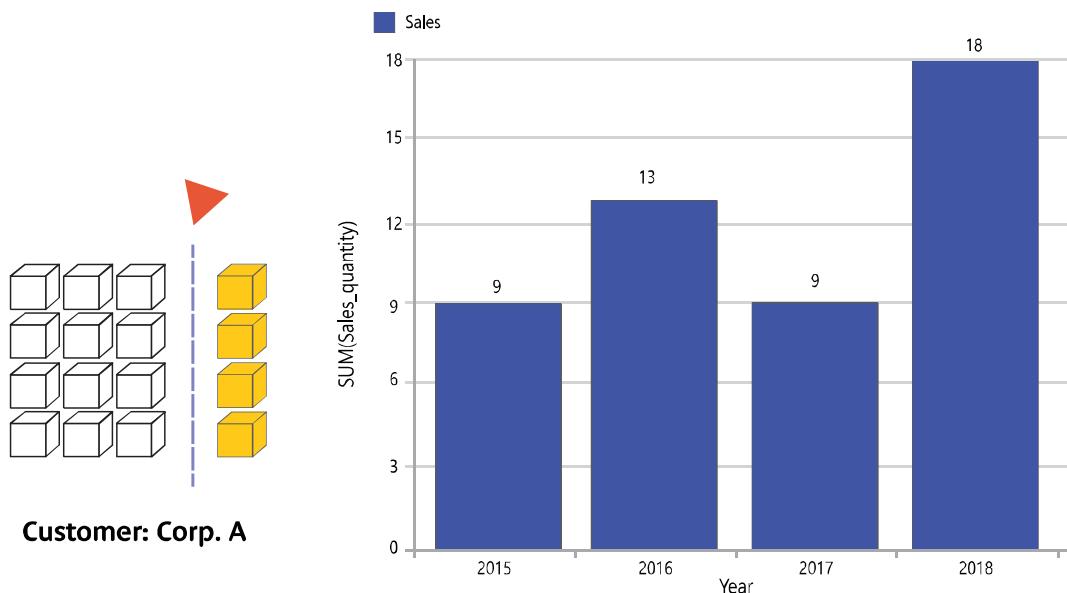
Each axis of the cube represents each dimension of the OLAP-type data mart. As cases in which there is a limit of three dimensions are rare in the actual data mart, such a cube model is merely for illustrating the process of analyzing the data mart simplified into three dimensions; it cannot express the data mart from the reality of four dimensions or more.

Although the data mart used as an example in this section is also composed of a total of four dimensions, we will use only three dimensions by ignoring the "employee" dimension for convenience in order to explain the fundamentals of the analysis process using such a three-dimensional cube model. The following is an analysis process for checking the sales of a specific product for a specific customer during a specific period.

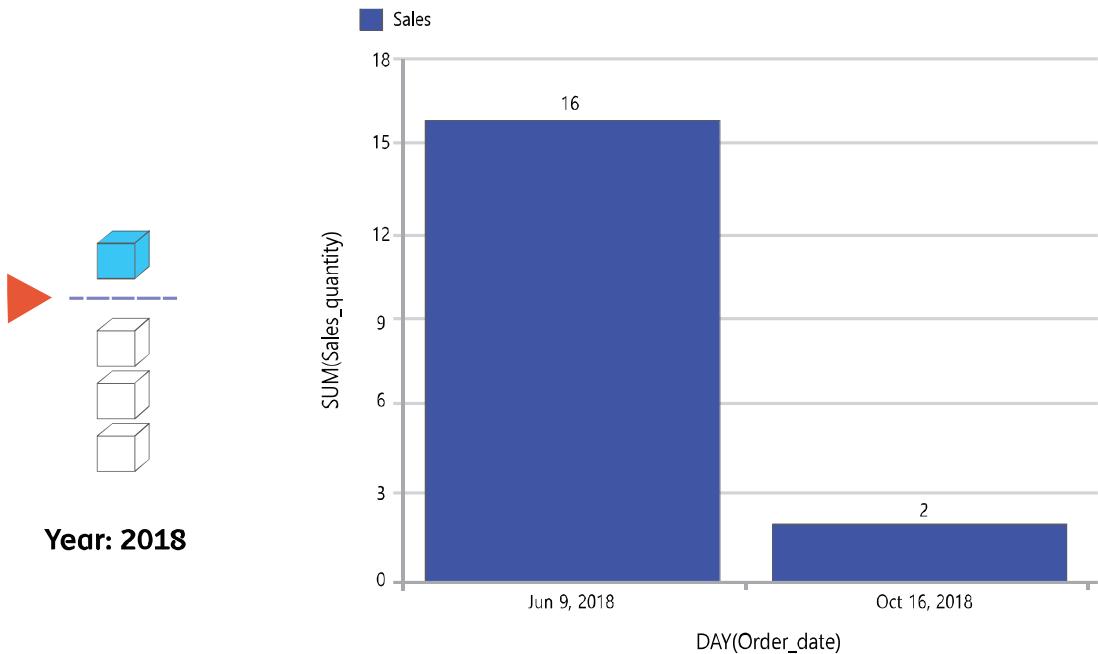
1. First, pick out only the data with product type "pencil" from the data mart.



2. Then select the purchases of Corp. A.



3. Finally, select data in 2018. The following example indicates Corp. A purchased 16 units of Product A on June 9, 2018, and 2 units on October 16, 2018.



metatron Discovery provides many functions for analysis of multi-dimensional data with an intuitive GUI.



metatron 2.0

Part 2

Management

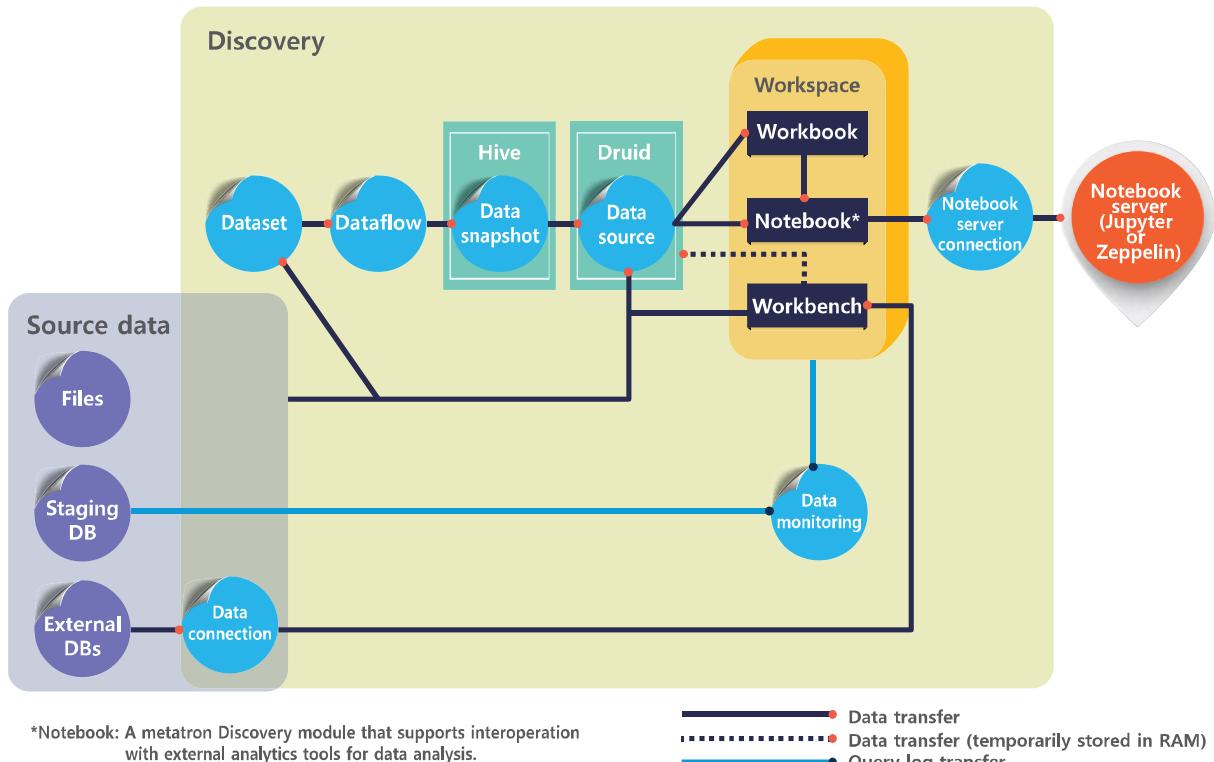


Contents

1. Management Outline	4
2. Data Storage Management.....	5
2.1 Data source.....	5
2.1.1 Data source management home screen.....	5
2.1.2 Data source details	7
2.1.3 Create a data source.....	15
2.2 Data connection	28
2.2.1 Data connection management home.....	28
2.2.2 Create a new data connection.....	30
3. Data Preparation Management.....	33
3.1 Dataset.....	33
3.1.1 Dataset management home	34
3.1.2 Dataset details.....	35
3.1.3 Create a new dataset.....	36
3.2 Dataflow	44
3.2.1 Dataflow management home screen.....	44
3.2.2 Create a new dataflow	45
3.2.3 Dataflow details.....	47
3.2.4 Edit a Wrangled dataset.....	50
3.2.5 Create a data snapshot.....	52
3.3 Data snapshot	55
3.3.1 Data snapshot management home.....	55
3.3.2 Data snapshot details	56
4. Notebook Management	57
4.1 Notebook server.....	57
4.1.1 Search notebook server list.....	57
4.1.2 Register a new notebook server.....	58

5. Data Monitoring Management.....	59
5.1 Log analysis	59
5.1.1 Log analysis home screen.....	59
5.2 Job log.....	61
5.2.1 Job log home	61
5.2.2 Job log details.....	62
5.3 Data lineage	63
5.3.1 Data lineage management home screen.....	63
5.3.2 Data lineage details.....	64

1. Management Outline



As shown above, data used by the 3 metatron Discovery analysis modules (workbook, notebook and workbench) is imported from various types of source data, engines and storage systems. Therefore, such dataflow needs to be standardized and managed, and different types of source data need to be linked.

The management menu is used to manage the above processes in black. These processes are divided into the 4 sub-menus as follows.

Chapter 2 Data Storage (Data Source and Data Connection): Describes the procedure for

managing data sources and the database connection required for data analysis and visualization.

Chapter 3 Data Preparation (Dataset, Dataflow, and Data Snapshot): Describes the

procedure for preparing the imported original data.

Chapter 4 Notebook management: To use the notebook module, the server must be linked with Jupyter or Zeppelin. This chapter describes the procedure to set and manage these servers.

Chapter 5 Data Monitoring (Log Analysis, Job Log and Data Lineage): Describes the

procedure to monitor and track data usage.

2. Data Storage Management

Data to be used in metatron Discovery can be linked in the two following ways.

- Loads the original data in the internal engine 'Druid', and stores it by 'data source' unit. (See Paragraph 2.1.)
- Links the SQL database directly. (See Paragraph 2.2.)

Data storage provides functions to link such data to metatron and manage it.

2.1 Data source

In metatron Discovery, 'data source' means the unit of data collected by the Druid engine. Each data source is saved in the Druid database as a table. These data sources are used in 'workbook' or 'notebook' for analysis and visualization.

2.1.1 Data source management home screen

In this screen, a user can register, edit or search data sources.

The screenshot shows the 'Data Storage' management screen. At the top, there are tabs for 'Data Source' and 'Data Connection'. Below the tabs are several filter options: 'Data type' (set to 'All'), 'Ingestion type' (set to 'All'), 'Status' (radio buttons for 'All', 'Enabled', 'Disabled', 'Preparing'), and 'Allowance' (checkbox for 'open data only view'). There are also time-related filters ('Time' dropdown set to 'Created', 'Updated' radio button, 'All' button, 'Today' button, 'Last 7 days' button, date range input fields for 'yyyy-MM-dd hh:mm' to 'yyyy-MM-dd hh:mm', and an 'Apply' button). A search bar labeled 'Search by name of datasource' is followed by a message 'There are 50 lists' and a button '+ Create a new datasource'. The main area displays a table of data sources:

Datasource	Data type	Ingestion type	Status	Created
cell_test	File	Ingested data	Enabled	2018-05-08 20:27 by P...
test	File	Ingested data	Enabled	2018-05-04 16:13 by P...
test_index_sales1	Database	Ingested data	Enabled	2018-04-30 21:18 by 관...
winemag -테스트용입니다. [Open data]	File	Ingested data	Enabled	2018-04-23 14:18 by P...
[Open data]	File	Ingested data	Enabled	2018-04-22 23:33 by 관...
[Open data]	File	Ingested data	Enabled	2018-04-22 22:59 by 관...
[Open data]	File	Ingested data	Failed	2018-04-22 21:01 by 관...
aaaaaaaaaa [Open data]	Database	Ingested data	Enabled	2018-04-20 17:12 by 관...

① **Data type:** Search data sources by type of original data.

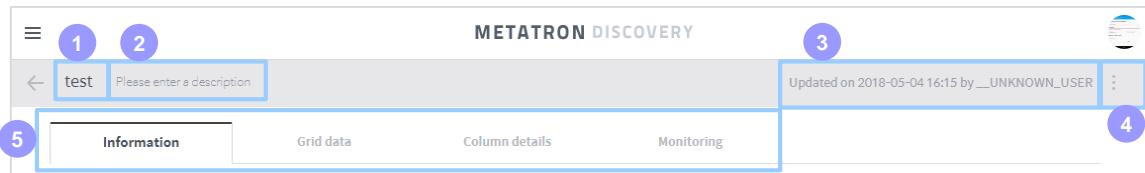
- All: Displays all data sources regardless of original data type.
- File: Displays data sources created by the file imported from a local PC.
- Database: Displays data sources created by the data imported from a database.
- Staging DB: Displays data sources created by data imported from metatron's internal Hive database.

- ② Ingestion type:** Search by the data source's type of ingesting data.
- All: Displays data sources regardless of ingestion type.
 - Ingested data: Displays data sources ingested by saving data in the metatron server directly.
 - Liked data: Displays data sources that import data from the linked database whenever necessary.
- ③ Status:** Searches by availability of the data source saved in the active data storage.
- All: Displays all data sources regardless of their availability.
 - Enabled: Displays data sources that passed data ingestion and are available in a workbook or workbench.
 - Disabled: Displays data sources that passed data ingestion but are unavailable because of a problem in a certain Druid process.
 - Preparing: Displays relatively new data sources in the data ingestion process.
- ④ View only open data:** Searches only data sources allowed in all workspaces.
- ⑤ Time:** Standard time applied when searching a data source. Select created date or updated date, and select All/Today/Last 7 days/Specific day as a time range.
- ⑥ Search by data source name:** Searches a registered data source by its name.
- ⑦ Number of data:** Displays the number of data sources searched from the current list.
- ⑧ Create a new data source:** Click to create a new data source.
- ⑨ Data source list:** Displays data sources that meet the established criteria. Click a data source to see its details. (See Paragraph 2.1.2.)
- ⑩ Delete:** Hover the mouse over the data source to display the trash icon. Click the icon to delete the data source.

2.1.2 Data source details

Click a data source listed on the data source management home to view various attributes of that data source. When viewing the following areas, each data source is saved in metatron's internal Druid database as a table, and includes a timestamp column due to the time-series attribute of Druid.

A. Common Top Area



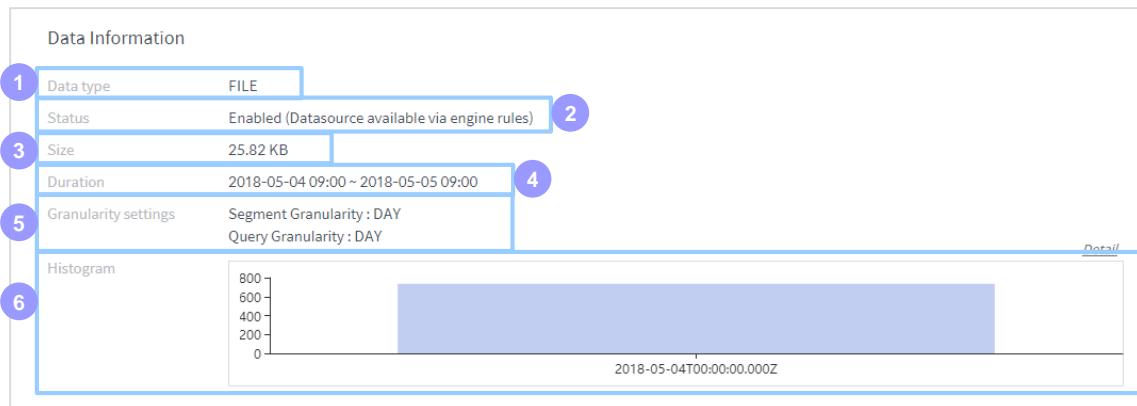
- ① **Name:** Name of the data source. Click to edit it.
- ② **Description:** Description of the data source. Click to edit it.
- ③ **Information last updated by:** Displays the user who last updated the data source.
- ④ **Delete:** Click this icon to display the menu to delete the data source.
- ⑤ **Tab selection area:** Each tab displays a specific attribute group of the data source.
According to the type of data source, not all of the 4 tabs may be displayed tab data sources. For more information on each tab, refer to the relevant paragraph below.

B. Information tab

In the information tab, a user can view details of the data source and edit basic matters. This tab is composed of 'Data information', 'Permission' and 'Ingestion information'.

Data information area

The area displays basic information of the data source.



- ① **Data type:** Type of original data imported when creating the data source.
- ② **Status:** Displays the availability of the data source.
- ③ **Size:** Displays the size of the data source.
- ④ **Duration:** Displays the time range of the timestamp included in the data source.
- ⑤ **Granularity settings:** Displays the Granularity interval defined when creating the data source.
 - Segment Granularity: Defines the unit of divided data saving required to utilize Druid operating in the distributed environment.
 - Query Granularity: Defines the unit of minimum time for performance in the analysis. It is required to obtain responses faster by generating the result of a minimum unit in advance.
- ⑥ **Histogram:** A graph displaying the capacity to save data at each time by Kbyte. This histogram is enabled because the Druid engine must record a timestamp for each record.

Permission area

In this area, a user can check and set the workspace where the data source can be used.

The screenshot shows a section titled "Permission" with three numbered options:

- 1 **Allow all workspaces to use this datasource**
- 2 [Edit](#)
- 3 **0 workspaces**

- 1 **Allow all workspaces to use this data source check box:** Click this check box to use the data source in all workspaces.
- 2 **Edit:** Used to designate a specific workspace to allow the data source. This button will disappear if the data source is set as open data.
- 3 **Number of shared workspaces:** Displays the number of workspaces where the data source can be used.

Ingestion information area

This area displays ingestion information on the master data, and origin of the relevant data source.

The screenshot shows the "Ingestion information" section with the following details:

Master data	Type	JDBC
Database / Table	polaris / salesqa	

Ingestion settings

Ingest once
• Scope of Ingesting data : Gets the entire data

C. Grid data tab

The grid data tab displays details of data in the data source.

The screenshot shows a data grid with the following columns: ab_dt, ab_yn, and current_datetime. The data consists of 16 rows, each containing a date from 20161229 to 20170116, a value 'N' or 'Y' in the ab_yn column, and a timestamp '2018-05-04 00:00:00' in the current_datetime column. Above the grid is a toolbar with the following controls:

- Search data (1)
- Role (2) with radio buttons for All, Dimension, and Measure
- Type (3) with a dropdown menu set to All
- Row count (4) showing 100
- Download CSV (5)

ab_dt	ab_yn	current_datetime
20161229	N	2018-05-04 00:00:00
20161230	N	2018-05-04 00:00:00
20161231	Y	2018-05-04 00:00:00
20170101	Y	2018-05-04 00:00:00
20170102	N	2018-05-04 00:00:00
20170103	N	2018-05-04 00:00:00
20170104	N	2018-05-04 00:00:00
20170105	N	2018-05-04 00:00:00
20170106	N	2018-05-04 00:00:00
20170107	Y	2018-05-04 00:00:00
20170108	Y	2018-05-04 00:00:00
20170109	N	2018-05-04 00:00:00
20170110	N	2018-05-04 00:00:00
20170111	N	2018-05-04 00:00:00
20170112	N	2018-05-04 00:00:00
20170113	N	2018-05-04 00:00:00
20170114	Y	2018-05-04 00:00:00
20170115	Y	2018-05-04 00:00:00
20170116	N	2018-05-04 00:00:00

- 1 **Search data:** Searches by the contents of a data table.
- 2 **Role:** Searches a column of the data table by All/Dimension/Measure.
- 3 **Type:** Searches a column of the data table by field type.
- 4 **Row:** Displays the number of records registered in the data table.
- 5 **Download CSV:** Download data displayed on the screen as a CSV file.

D. Column detail tab

In the column detail tab, a user can view details of each column of the data source table.

Column search/setting

In the top of the column detail tab, a user can use the UI to filter columns by certain criteria. Columns that meet the criteria will be displayed on the left. Also, a user can edit column settings.

The screenshot shows the Column detail tab interface. At the top, there is a search bar labeled "Search data" (1) and three filtering options: "Role" (2), "Type" (3), and "View all" (4). A "Configure schema" button (5) is also present. Below these, the data source name "cell_test" (6) is displayed, followed by a list of columns (7). The first column in the list is "occ_r_dth". The main panel displays two sections: "Schema informations" and "Summary". The "Schema informations" section shows details like Name: occ_r_dth, Role: TIMESTAMP, Logical name: occ_r_dth, Logical type: TIMESTAMP, Format: yyyyMMddhh, and Description. The "Summary" section shows statistics such as Column name: occ_r_dth, Row count: 6, Valid: 6 (100%), Unique: 0 (0%), Outliers: 0 (0%), and Missing: 0 (0%).

- 1 **Search data:** Searches by column name.
- 2 **Role:** Searches a column of the data table by All/Dimension/Measure.
- 3 **Type:** Searches a column of the data table by field type.
- 4 **View all:** Cancels all search criteria set in the Search data, Role and Type options, and returns to View All Columns.
- 5 **Configure schema:** Click to display a window to edit the settings of the current column.
- 6 **Data source name:** Displays the name of the data source (relevant 'table').
- 7 **Column list:** Lists table columns.

Column basic information area

This area displays the various basic information of the selected column.

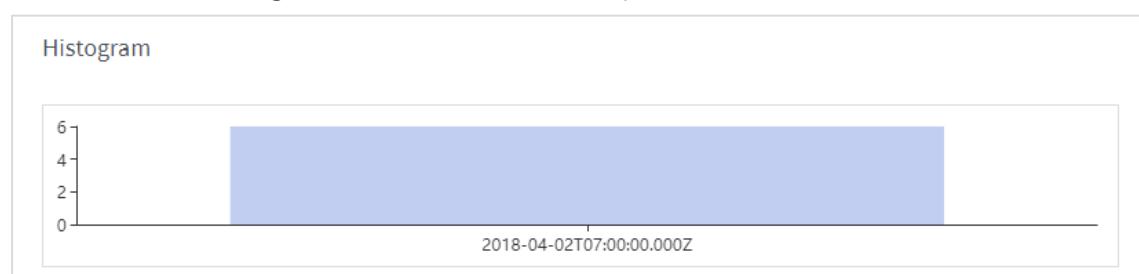
The diagram illustrates the four main sections of the Column basic information area:

- 1 Schema information:** Displays the attributes of the selected column. It includes fields: Name (occr_dth), Role (TIMESTAMP), Logical name (occr_dth), Logical type (TIMESTAMP), Format (yyyyMMddhh), and Description.
- 2 Display setting:** Displays the meta data information of the selected column. It includes fields: Missing (Do not set) and Do not set.
- 3 Summary:** Displays the value configuration information of the selected column. It includes fields: Column name (occr_dth), Row count (6), Valid (6 (100%)), Unique (0 (0%)), Outliers (0 (0%)), and Missing (0 (0%)).
- 4 Statistic:** Displays the minimum value and the maximum value among the values entered in the selected column. It includes fields: Minimum (2018-04-02T07:00:00.000Z) and Maximum (2018-04-02T08:00:00.000Z).

- 1 Schema information:** Displays the attributes of the selected column.
- 2 Display setting:** Displays the meta data information of the selected column.
- 3 Summary:** Displays the value configuration information of the selected column.
- 4 Statistics:** Displays the minimum value and the maximum value among the values entered in the selected column.

Histogram area

A graph displaying the capacity to save data at each time by Kbyte. This histogram is enabled because the Druid engine must record a timestamp for each record.

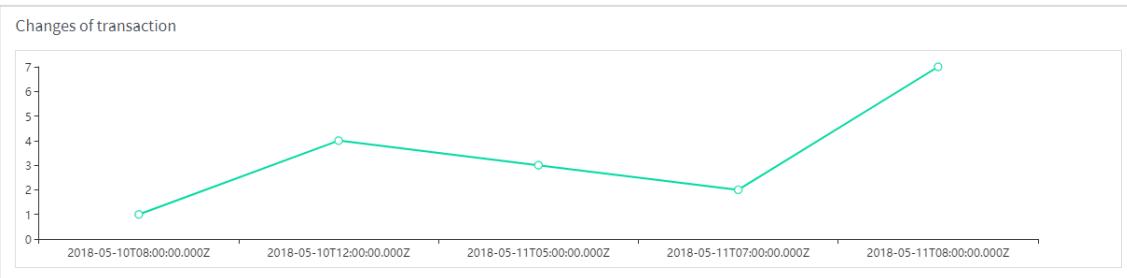


E. Monitoring tab

Displays the log where the data source is used.

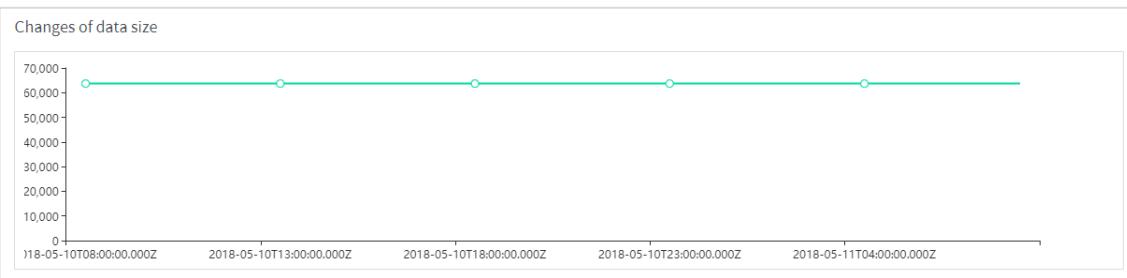
Transaction change area

Displays the trend of data source transactions over time.



Data size change area

Displays the trend of data source capacity over time.



Query distribution area

Displays information on the query performed for the data source.



- ① **Query distribution by user (during the last week):** Displays a graph of query distribution by the user who performed the query in the past week.
- ② **Query distribution by elapsed time (during the last week):** Displays a graph of query distribution by elapsed time for query performance in the past week.

Query log area

Used to view a detailed history of each performed query.

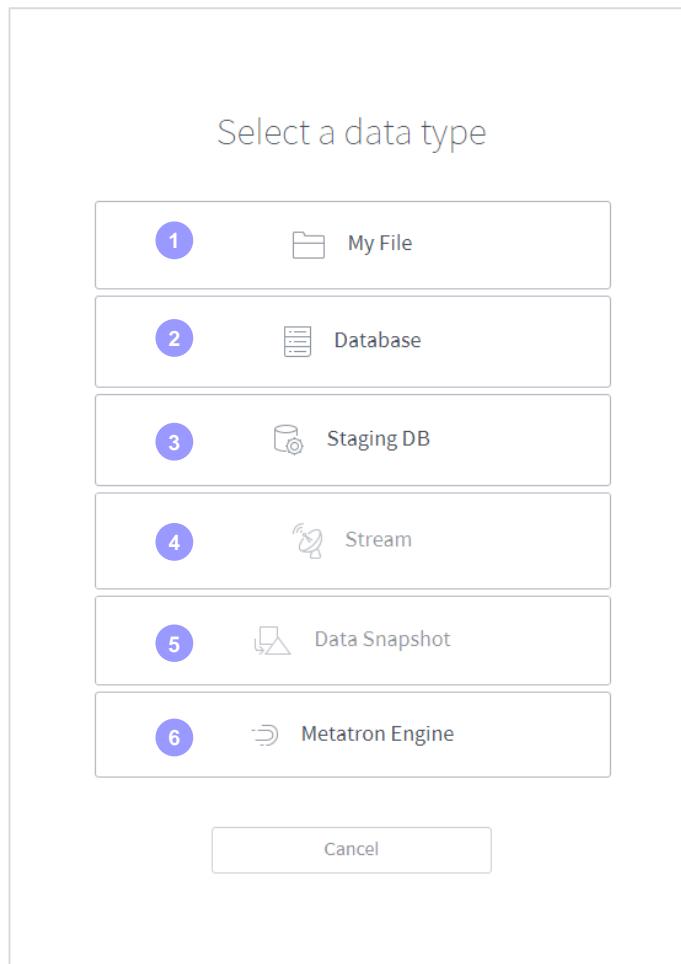
The screenshot shows a table titled 'Query log' with the following columns: No., Query date, Query type, User, Elapsed time, and Result. The table contains five rows of data. To the right of the table, there is a vertical column of five blue rectangular boxes, each containing the text 'Detail>'. The interface includes various filters at the top: 'Query date' (All, Today, Last 7 days, Date range), 'Query type' (All), and 'Result' (All). There is also a 'More' button at the bottom of the table.

No.	Query date	Query type	User	Elapsed time	Result
1	2018-05-10 21:17	SUMMARY		85ms	Success
2	2018-05-11 16:41	SUMMARY		78ms	Success
3	2018-05-10 21:17	SEARCH		78ms	Success
4	2018-05-10 21:17	SUMMARY		76ms	Success
5	2018-05-11 17:30	SUMMARY		64ms	Success

- ① **Query date:** Set the running time zone for queries to be confirmed.
- ② **Query type:** Prints the performed queries by type.
- ③ **Result:** Prints succeeded or failed query results.
- ④ **Query list:** Lists queries that meet the established criteria.
- ⑤ **More:** Click to view the query statement.

2.1.3 Create a data source

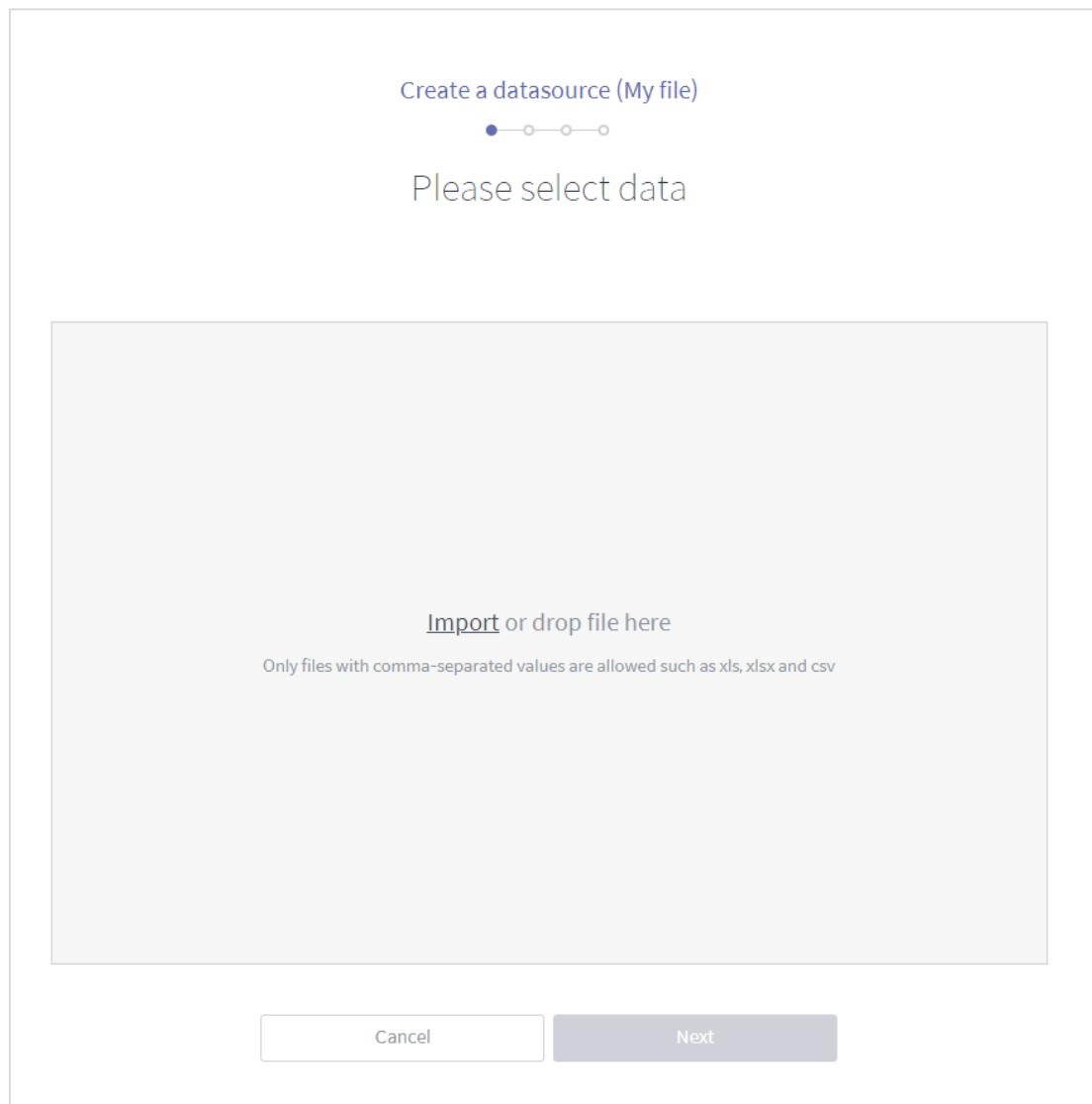
Click the '+Create a new data source' button in the top-right of the data source management screen to create a new data source. First, select the type of original data.



- ① **File:** Imports a file saved on the user's local PC to create a data source. Only files separated by a comma such as xls, xlsx and csv can be imported (for more information, see Paragraph A).
- ② **Database:** Imports data from an external database to create a data source. Currently, metatron Discovery supports Oracle, MySQL, Hive, presto and TIBERO (for more information, see Paragraph B).
- ③ **Staging DB:** Creates a data source based on the data imported from the metatron internal Hive database.
- ④ **Real-time:** This function is not currently supported.
- ⑤ **Data snapshot:** This function is not currently supported.
- ⑥ **metatron engine:** Migrate a data source saved in metatron version 1.0.

A. Create a data source with file

1. In the screen to select the type of original data, select 'File'.
2. This imports the file to be used as a data source from the user's local PC. Click the Import button and select the file, or drag a file to the screen. Once the file is imported, click 'Next'.



3. From the file, select the sheet to be included in the data source.

The screenshot shows the 'Create a datasource (My file)' interface. At the top, it says 'Please select data'. Below that, there's a file list with 'sale_20180130.xls' and 'Import or drop file here'. Underneath, there's a list of sheets: 'sale.txt (2)' (selected) and 'sale.txt (1)'. To the right, there are summary statistics: 7739392 byte, 28 Columns, 100 / 9994 Row, and 1 Types. A checkbox labeled 'Use the first row as the head column. (If not checked, a new row is created and used as the head column)' is checked. At the bottom are 'Cancel' and 'Next' buttons.

ab OrderDate	ab Category	ab City	ab Country	ab CustomerNa...	ab OrderID
2011-01-04T00:...	Office Supplies	Houston	United States	Darren Powers	CA-2011
2011-01-05T00:...	Office Supplies	Naperville	United States	Phillina Ober	CA-2011
2011-01-05T00:...	Office Supplies	Naperville	United States	Phillina Ober	CA-2011
2011-01-05T00:...	Office Supplies	Naperville	United States	Phillina Ober	CA-2011
2011-01-06T00:...	Office Supplies	Philadelp...	United States	Mick Brown	CA-2011
2011-01-07T00:...	Office Supplies	Athens	United States	Jack OBriant	CA-2011
2011-01-07T00:...	Office Supplies	Los Angeles	United States	Lycoris Saunders	CA-2011
2011-01-07T00:...	Furniture	Henderson	United States	Maria Etezadi	CA-2011

- 1 **File name:** Name of the imported file. The user can import another file.
- 2 **File sheet list:** Displays sheets included in the imported file. Select the sheet to be made as data.
- 3 **File sheet name:** Name of the currently selected sheet.
- 4 **Capacity:** Capacity of the imported file.
- 5 **Column:** Number of columns in the imported file.
- 6 **Row:** Number of rows in the imported file. Input the number of rows to be displayed on the screen.
- 7 **Type:** Displays the number of data types recognized from each column. Data type by column can be edited in a later screen.
- 8 **Use the first row as the column name check box:** Select it to use the contents of the first row of the file as the column name. If you don't select it, a new row to enter the column name will be created.

- Set schema to be realized in the data source.

Search columns that meet the criteria

The screenshot shows the 'Create a datasource (My file)' interface at step 4: 'Please configure the schema'. The interface has the following components:

- Search by column name:** A search bar labeled '1'.
- Filter Options:** 'Role' dropdown (set to 'All') with 'Dimension' and 'Measure' options, 'Recommendation filter' checkbox, and 'Type' dropdown (set to 'All').
- Column List Area:** A list of columns labeled '5' (highlighted with a blue border). It includes a header 'Column' and a list of rows with columns 'Role' (e.g., Dimension, Measure), 'Value' (e.g., ab 2011-01-04T00:00:00Z, Office Supplies), and 'Description' (e.g., Houston, United States).
- Individual Column Setting Area:** A panel for '2011-01-04T00:00:00Z' labeled '6' (highlighted with a blue border). It shows 'Data' (100 Row) and 'Setting' (Role set to Dimension). The 'Setting' section includes 'Type' (String), 'Recommendation filter' (checkbox checked), 'Apply as a priority filter' (checkbox checked), 'Missing' (Replace with, Discard, Do not set), and 'Time-type column' (checkbox checked).
- Timestamp Setting:** A note at the bottom labeled '7' stating 'One of the time-type columns or current time must be specified as a Timestamp' with options 'Current time' (radio button checked), 'Time-type column' (checkbox checked), and 'No time-type column' (checkbox checked).
- Navigation:** 'Previous' and 'Next' buttons at the bottom.

- Search by column name:** Search a column in the imported file by its name.
- Role:** Searches a column in the imported file by All/Dimension/Measure.
- Recommended filter:** Search columns where the priority filter is applied.
- Type:** Searches a column in the imported file by field type.
- Column list area:** Displays columns that meet the established criteria (for more information, see 'Edit and delete columns in a batch' below).
- Individual column setting area:** This area is used to set the attributes of the column selected from the column list (for more information, see 'Edit attributes of individual column' below).
- Timestamp setting:** Sets the timestamp that must exist in the data source. The metatron engine is a time-series engine that requires a time value to save the data source. Therefore, a user can designate a time-type column of the existing data as a timestamp, or create a time-type column based on the current time value, and designate it as a timestamp.

Edit and delete columns in a batch

In the data column list area, a user can edit or delete columns in a batch. Select the check boxes of columns to be deleted or which need to change type, set the operation you want referencing the following description, and click 'Apply'.

The screenshot shows a list of columns in a table. The first column has a checkbox header. The second column is labeled 'Column'. The third column contains the data values. A 'Dimension' checkbox is checked for the first item, '2011-01-04T00:00:00Z'. Below the table is a toolbar with three numbered steps:

- 1 Unselect all
- 2 1 selections
- 3 apply

Below the toolbar are three dropdown menus:

- Change Type (selected)
- Dimension
- String

	Column	
<input checked="" type="checkbox"/>	Dimension	ab 2011-01-04T00:00:00Z
<input type="checkbox"/>	Dimension	ab Office Supplies
<input type="checkbox"/>	Dimension	ab Houston
<input type="checkbox"/>	Dimension	ab United States
<input type="checkbox"/>	Dimension	ab Darren Powers
<input type="checkbox"/>	Dimension	ab CA-2011-103800
<input type="checkbox"/>	Dimension	ab 77095
<input type="checkbox"/>	Dimension	ab Message Book, Wirebound, Four 5 1/2 X 4 Forms...
<input type="checkbox"/>	Dimension	ab 2
<input type="checkbox"/>	Dimension	ab Central

- 1 Change/delete type:** Select the operation to be applied to the selected columns.
- 2 Dimension/measure:** Change the role of the columns as dimension or measure.
 - If you select dimension, you can change the data type to character/Boolean/Integer/Decimal/Date/Time/Latitude/Longitude.
 - If you select measure, you can change data type to Integer/Decimal.
- 3 Column type:** Select the new type of columns.

Edit attributes of an individual column

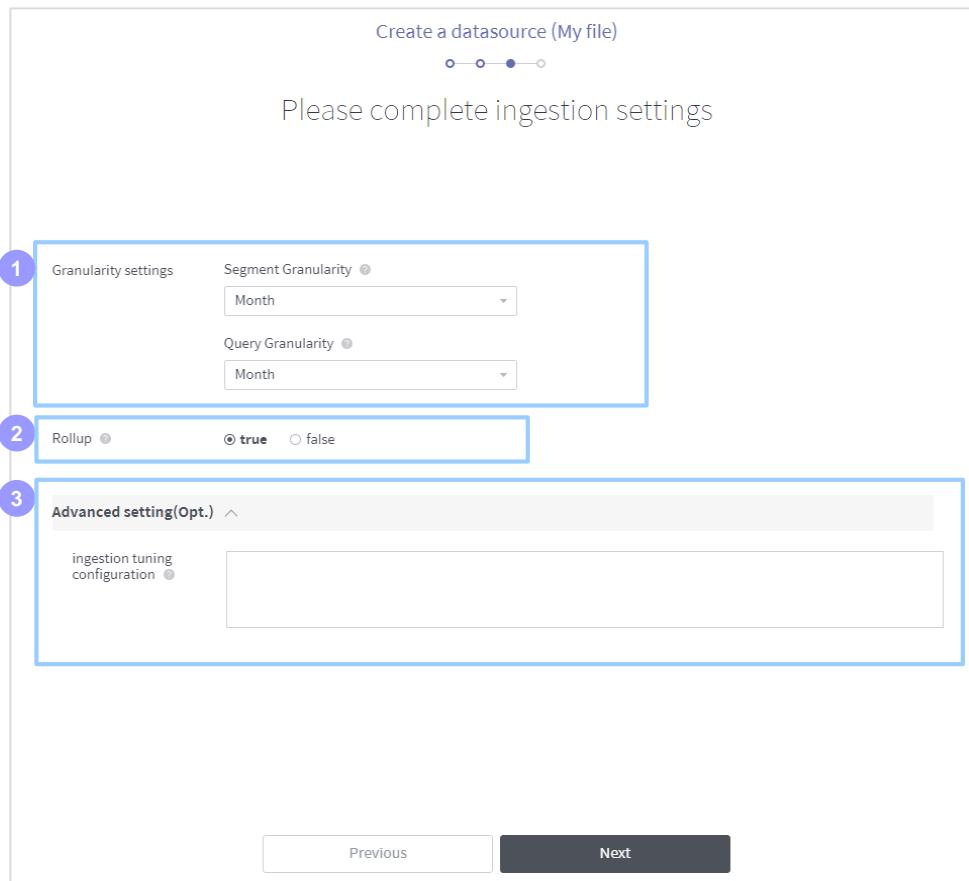
This area is used to edit attributes of the columns selected from the column list.

The screenshot shows a user interface for editing column attributes. On the left, there is a list of data rows. On the right, there are several configuration sections:

- Setting**:
 - Role**:
 - Dimension
 - Measure
 - Type**:
 - String
 - Recommendation filter**:
 - Apply as a priority filter
 - [Edit order](#)
 - Missing**:
 - Replace with
 - Discard
 - Do not set

- ① **Name:** Name of the selected column.
- ② **Number of data rows:** Indicates the number data rows displayed on the screen.
- ③ **Role:** Displays the dimension/measure (role applied to the column). A user can change it from this menu.
- ④ **Type:** Displays the data type of the column. A user can change it to another data type. Available data types are different based on the selected role (dimension/measure).
- ⑤ **Recommended filter:** If there is a column containing a massive amount of data, a timeout error may be generated due to a long loading time. The data administrator can register such a column as a recommended filter to facilitate the configuration of the dashboard or chart.
 - **Apply as a priority filter:** Select whether to apply the selected column as a priority filter.
 - A user can select only a single item. The time is generated when the column is applied as a priority filter. Check if the data is too big.
 - **Edit order:** Click to the change priority ranking of columns where the current recommended filter is applied.
- ⑥ **Missing:** Set how to handle Null values in the column.
 - **Replace with:** Replaces a Null value with the value inputted here.
 - **Discard:** Discards a Null value.
 - **Not set:** Displays a Null value as is. However, a Null value of a data source's timestamp will be discarded.

5. Set data source ingestion and click Next.



1 Granularity settings: Set two types of Granularity.

- Segment granularity: Defines the time unit of divided data saving required to utilize Druid operating in the distributed node environment.
- Query granularity: Defines the unit of minimum time for performance in the analysis. It is required to obtain responses faster by generating the result of a minimum unit in advance.

2 Rollup: Rollup includes a summarization of data based on dimension. A summarization rule can be the calculation of a sum following hierarchy or application of a formula group such as 'profit=sales=expenses'. In short, rollup is an effective option for ingestion.

3 Advanced setting (Opt.): Configures the ingestion tuning. Ingest tuning configuration is an effective option for ingestion. Input JSON format syntax in the text box. Example:

```
{maxRowsInMemory : 75000,  
 maxOccupationInMemory : -1,  
 maxShardLength : -2147483648,  
 leaveIntermediate : false,  
 cleanupOnFailure : true,  
 overwriteFiles : false,  
 ignoreInvalidRows : false,  
 assumeTimeSorted : false}
```

6. Confirm the information in the data set from the imported file, enter the name and description, and click Done to create a data source. It may take a few seconds or minutes according to the amount of data, because data is ingested from the original data to the metatron internal engine (Druid).

Create a datasource (My file)

● ● ● ● ●

Please complete datasource creation

File	sale.txt.csv
Granularity settings	<input checked="" type="radio"/> Segment Granularity: Month <input checked="" type="radio"/> Query Granularity: Month
Rollup	true

Name

Please enter a name

Description

Please enter a description

[Previous](#)
Done

7. Move to the data source management home to check the new data source from the screen. During ingestion, the status will be displayed as Preparing. Once it is finished, the status will be changed to Enabled. After that, a user can use the data source.

METATRON DISCOVERY					
Data Storage					
Data Source		Data Connection			
Data type	<input type="button" value="All"/>	Ingestion type	<input type="button" value="All"/>	Status	<input checked="" type="radio"/> All <input type="radio"/> Enabled <input type="radio"/> Disabled <input type="radio"/> Preparing <input type="checkbox"/> open data only view <input type="button" value="Refresh"/>
Time	<input checked="" type="radio"/> Created <input type="radio"/> Updated <input type="button" value="All"/>	Today	Last 7 days	yyyy-MM-dd hh:mm	~ yyyy-MM-dd hh:mm
					<input type="button" value="Apply"/>
Search by name of datasource There are 51 lists <input type="button" value="Create a new datasource"/>					
Datasource	Data type	Ingestion type	Status	Created	
sale	File	Ingested data	Preparing	2018-05-11 19:45 by 관리자	

B. Create a data source using a database or Staging DB

1. From the original data type selection screen, select 'database' or 'Staging DB'.
2. If you select 'database', load an existing data connection or enter connection information for a new database, in the data connection setting screen as follows. This step is skipped if you select 'Staging DB'.

Create a datasource (DB)

Please set data connection

1 Load a data connection

2 DB Type

ORACLE DATABASE	MySQL
HIVE	presto
TIBERO	

3 Ingestion type
 Ingested data Linked data

4 Host 5 Port
Host Port

6 SID
Please insert database name

7 User name 8 Password
Please enter user name
9 Test
Host is required

10 Save as a new data connection

Cancel Next

- 1 **Load a data connection:** Used to select a saved data connection. Select it to automatically load access information on the database connected to the relevant data connection. However, you must verify the connection by clicking the 'Test' button.
- 2 **DB type:** Select the type of database to be connected. Currently 5 database types are supported. (Oracle, MySQL, Hive, presto, TIBERO)
- 3 **Ingestion type:** Select the data source's method to ingest data.
 - Ingested data: Displays data sources ingested by saving data in the metatron server directly.
 - Liked data: Displays data sources that import data from the linked database whenever necessary.
- 4 **Host:** Select the value of the host to be connected.
- 5 **Port:** Enter the number of the port to be connected.

- 6 SID/Catalog:** For Oracle and TIBERO, enter SID. For presto, enter Catalog value.
- 7 Username:** Enter the username of the database.
- 8 Password:** Enter the password of the database.
- 9 Test:** If you enter all fields, the Test button will be activated. Click it to view the connection test result at the bottom of the button. 'Valid connection' means it is normal, and 'Invalid connection' means it is abnormal.
- 10 Save as a new data connection:** To link a new database rather than using an existing data connection, you must save connection information as a new data connection. Enter the name to be used in this step.

3. Select data. You can select a table from the account of linked databases, or write a query statement by yourself. You can only designate a table when creating a data source using Staging DB.

Table

Select the database and table name, confirm the data to be saved and click 'Next'.

Create a datasource (DB)

Please select data

1 Table 2 Query

OrderDate	Category	City	Country	CustomerName	OrderId	PostalCode	ProductName
2011-01-01 00:00:00	Office Supplies	Houston	United States	Darren Powers	CA-2011-19143	77095	Message Block
2011-01-02 00:00:00	Office Supplies	Philadelphia	United States	Mick Brown	CA-2011-19143	19143	Avery Hi-Lif
2011-01-03 00:00:00	Furniture	Huntsville	United States	Vivek Sundaresam	CA-2011-19143	77340	Howard Mil
2011-01-04 00:00:00	Office Supplies	Huntsville	United States	Vivek Sundaresam	CA-2011-19143	77340	Acco Four F
2011-01-05 00:00:00	Furniture	Springfield	United States	Anthony Jacobs	CA-2011-19143	22153	Howard Mil
2011-01-06 00:00:00	Office Supplies	Springfield	United States	Anthony Jacobs	CA-2011-19143	22153	Avery 482
2011-01-07 00:00:00	Furniture	Philadelphia	United States	Brendan Sweed	CA-2011-19143	19140	Sauder Cor
2011-01-08 00:00:00	Furniture	Philadelphia	United States	Delfina Latchford	CA-2011-19143	19134	DAX Black I
2011-01-09 00:00:00	Office Supplies	Philadelphia	United States	Delfina Latchford	CA-2011-19143	19134	GBC Linen
2011-01-10 00:00:00	Office Supplies	Philadelphia	United States	Delfina Latchford	CA-2011-19143	19134	Round Spe
2011-01-11 00:00:00	Technology	Philadelphia	United States	Delfina Latchford	CA-2011-19143	19134	AT&T 8410
2011-01-12 00:00:00	Office Supplies	Springfield	United States	David Wiener	CA-2011-19143	97477	Fellowes M

Previous Next

① **Select database:** Select a database linked with the selected data connection.

② **Select schema:** Select a table in the selected database.

Query

Write a query statement to import the data you want, and click 'Run' to display data at the bottom. Confirm the data and click 'Next'.

The screenshot shows a user interface for querying a database. At the top, there is a progress bar labeled "Create a data source(DB)" with several steps completed. Below it, a message says "Please select data". There are two tabs: "Table" and "Query", with "Query" selected. A SQL query "select * from test.emp_table" is entered in the query editor. A "Run" button is located to the right of the query. Below the query editor is a table with 7 columns and 7 rows of data. The columns are: contract_..., contract_id, product_..., product_..., product_..., product_..., customer..., and detail_st... . The data in the table is as follows:

contract_...	contract_id	product_...	product_...	product_...	product_...	customer...	detail_st...
2011-11-20	cid000000...	01	01	004	0002	uid007409...	0004
2011-02-20	cid000000...	02	05	004	0048	uid004774...	0308
2012-02-03	cid000000...	04	04	004	0030	uid008105...	0243
2014-12-01	cid000000...	03	04	004	0050	uid005363...	0012
2015-11-11	cid000000...	03	05	002	0015	uid010054...	0157
2010-04-28	cid000000...	01	01	004	0032	uid018579...	0224
2016-05-19	cid000000...	01	02	003	0037	uid014980...	0110

At the bottom of the interface are "Previous" and "Next" buttons.

4. The subsequent procedures are the same as the procedures to import data for the file.

See Step 3 and onwards of Paragraph 2.1.3 A. However, when creating a data source from a database, you must configure additional settings to set ingestion.

The screenshot shows the "Ingestion settings" screen. It has two radio button options: "Ingest Once" (selected) and "Ingest periodically". Below these are two numbered steps: 1. Ingestion settings: Set data ingestion. 2. Scope of Ingesting data: Set the scope of data ingestion. Step 2 is highlighted with a blue box. Inside the box, there are two radio button options: "Gets the entire data" (selected) and "Gets only the first [10000] rows".

① Ingestion settings: Set data ingestion.

- Ingest once: Select to save data only once.
- Ingest periodically: Save data on a regular basis.

② Scope of data ingestion: Set the scope of data ingestion.

- Retrieves the entire data: Ingests all data when running data ingestion on a regular basis.
- Receives only the first: Sets the amount of data to be inserted at the start of data.

C. metatron engine (v1.0) data source migration

1. From the original data type selection screen, select 'database' or 'metatron engine'.
2. When data sources created by metatron V1.0 are listed on the left as follows, select the check box of the data sources to be migrated to V2.0.

Create a datasource (Druid)

Please select data table

<input type="checkbox"/> 0 Selections	_aiinb		
<input type="checkbox"/> _aiinb >	event_t... ▾	Field 1 ▾	Field 2 ▾
<input type="checkbox"/> _bfzew	2017-12-26 00...	1.0	a
<input type="checkbox"/> _cjada	2017-12-26 00...	10.0	j
<input type="checkbox"/> _ddfkcb	2017-12-26 00...	2.0	b
<input type="checkbox"/> _euuaa	2017-12-26 00...	3.0	c
<input type="checkbox"/> _hihgv	2017-12-26 00...	4.0	d
<input type="checkbox"/> _itjly	2017-12-26 00...	5.0	e
<input type="checkbox"/> _kiyky	2017-12-26 00...	6.0	f
<input type="checkbox"/> _mlwxb	2017-12-26 00...	7.0	g
<input type="checkbox"/> _mowrq	2017-12-26 00...	8.0	h
	2017-12-26 00...	9.0	i

Cancel

Done

3. Click 'Done' to migrate the selected data sources.

2.2 Data connection

metatron Discovery can link an external database directly. To link an external database, you must create and manage a data connection containing access information for that database. Register such a data connection to avoid entering access information again for a new database. The purpose of a data connection is divided as follows.

- For general purpose: Linked to the database when creating a new data source based on a database (see 2.1.3 Paragraph B).
- For workbench: Linked to the database to be used in workbench (see 'Part 6 Workbench').

2.2.1 Data connection management home

In this screen, a user can register, edit or search a database connection.

The screenshot shows the 'Data Storage' section of the Metatron Discovery interface. At the top, there are two tabs: 'Data Source' and 'Data Connection'. The 'Data Connection' tab is selected and highlighted with a blue border. Below the tabs are several search and filter options:

- 1 Type:** Radio buttons for 'All' (selected), 'General', and 'Workbench'.
- 2 DB Type:** A dropdown menu currently set to 'All'.
- 3 Time:** Radio buttons for 'Created' (selected) and 'Updated'. Below these are buttons for 'All', 'Today', 'Last 7 days', and date range inputs 'yyyy-MM-dd hh:mm' and 'yyyy-MM-dd hh:mm' with an 'Apply' button.
- 4 Search by name of data connection:** An input field with a magnifying glass icon.
- 5 Number of data connections:** A text box stating 'There are 10 lists'.
- 6 Create a new Data Connection:** A button with a plus sign and the text 'Create a new Data Connection'.
- 7 Data connection list:** A table listing 10 data connections. The columns are 'Type', 'Data connection', 'DB Type', 'Host', 'Port', and 'Created'. Each row includes a small trash can icon in the last column.
- 8 Delete button:** A blue button with a trash can icon located at the bottom right of the data connection list area.

The table data is as follows:

Type	Data connection	DB Type	Host	Port	Created
General	Tibero-exntu	TIBERO	exntu.kr	8629	2018-05-08 13:39 by 관리자
Workbench	Tibero_Exntu	TIBERO	exntu.kr	8629	2018-05-08 09:06 by 관리자
Workbench	local_mysql	MYSQL	metatron-poc-a01	3306	2018-04-10 17:19 by 관리자
Workbench	azure-mysql-test	MYSQL	metatron-poc-h04	3306	2018-03-22 11:05 by 관리자
General	Test_Connection	MYSQL	metatron-poc-h04	3306	2018-02-09 16:24 by 관리자
Workbench	metatron-azure	MYSQL	metatron-poc-a01	3306	2018-01-25 11:07 by 관리자

1 Type: Search a data connection by its use.

- All: Displays all data connections regardless of their use.
- General: Displays general purpose data connections. Used to create a new data source.
- Workbench: Displays data connections for workbench. Used to create a new workbench.

2 DB type: Search a data connection by database type

(Oracle/MySQL/Hive/presto/Tibero).

3 Time: Standard time applied when searching a data connection. Select created date or updated date, and select All/Today/Last 7 days/Specific day as a time range.

4 Search by data connection name: Searches a registered data connection by its name.

5 Number of data connections: Displays the number of data connections searched from the current list.

- 6 Create a new data connection:** Click to create a new data connection.
- 7 Data connection list:** Displays data connections that meet the established criteria. Click a data connection to edit its setting.
- 8 Delete:** Hover the mouse over the data connection to display the trash icon. Click the icon to delete the data connection.

2.2.2 Create a new data connection

A. Create a new general purpose data connection

A general purpose data connection is used to create a new data source from metatron Discovery. Enter the following fields to create a new one.

Create a data connection

Please set required items and complete data connection creation

1 Type General Workbench

2 DB Type

ORACLE DATABASE	MySQL
HIVE	presto
TIBERO	

3 Host

4 Port

5 SID

Please insert database name

6 User name

Please enter user name

7 Password

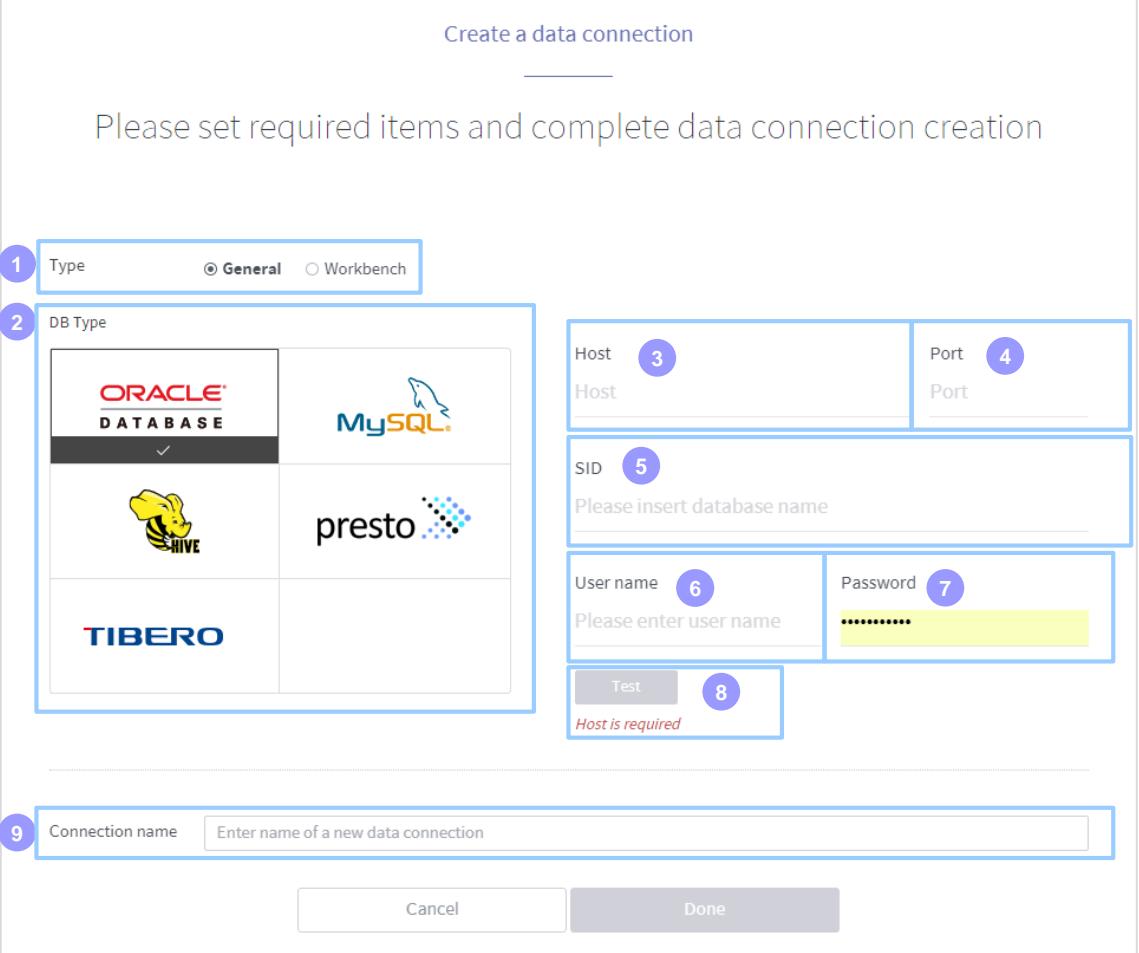
.....

8 Test

Host is required

9 Connection name

Cancel Done



- 1 **Type:** Purpose of the data connection to be created. To create a general purpose data connection, select 'General'.
- 2 **DB type:** Currently 5 database types are supported. (Oracle, MySQL, Hive, presto, TIBERO)
- 3 **Host:** Select the value of the host to be connected.
- 4 **Port:** Enter the number of the port to be connected.
- 5 **SID/Catalog:** For Oracle and TIBERO, enter SID. For presto, enter Catalog value.
- 6 **Username:** Enter the username of the database.
- 7 **Password:** Enter the password of the database.
- 8 **Test:** If you enter all fields, the Test button will be activated. Click it to view the connection test result at the bottom of the button. 'Valid connection' means it is normal, and 'Invalid connection' means it is abnormal.
- 9 **Name of data connection:** Enter the name of the data connection to be created.

B. Create a new data connection for workbench

Data connection for workbench is used to link a database to be used in workbench. Enter the following fields to create a new one.

Create a data connection

Please set required items and complete data connection creation

1 Type General Workbench

2 DB Type

ORACLE DATABASE	MySQL
HIVE	presto
TIBERO	

3 Host 4 Port

5 SID Please insert database name

6 Account input method
Enter by manager

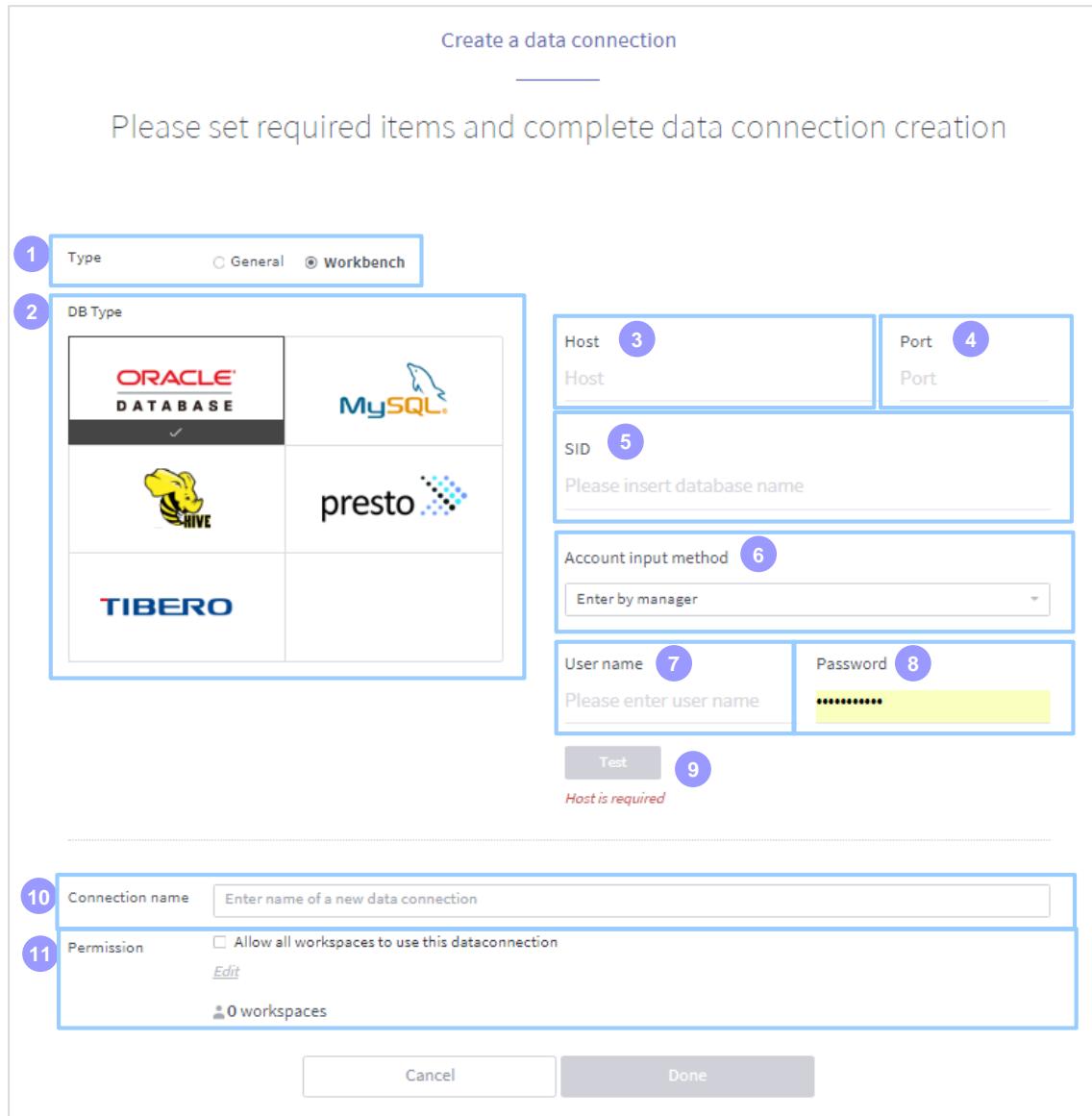
7 User name 8 Password
Please enter user name

9 Test Host is required

10 Connection name Enter name of a new data connection

11 Permission Allow all workspaces to use this dataconnection
[Edit](#)
0 workspaces

[Cancel](#) [Done](#)



① **Type:** Purpose of the data connection to be created. To create a data connection for workbench, select 'Workbench'.

② **DB type:** Currently 5 database types are supported. (Oracle, MySQL, Hive, presto, TIBERO)

③ **Host:** Select the value of the host to be connected.

④ **Port:** Enter the number of the port to be connected.

⑤ **SID/Catalog:** For Oracle and TIBERO, enter SID. For presto, enter Catalog value.

⑥ **Account input method:** Set how to log in to use the connection in workbench.

- Enter by manager: Log in using information entered by the user when creating a data connection.

- Use user account: Log in using user account information registered in metatron Discovery.
- Input on connection: Log in by entering information whenever connecting to the workbench.

7 Username: Enter the username of the database.

8 Password: Enter the password of the database.

9 Test: If you enter all fields, the Test button will be activated. Click it to view the connection test result at the bottom of the button. 'Valid connection' means it is normal, and 'Invalid connection' means it is abnormal.

10 Name of data connection: Enter the name of the data connection to be created.

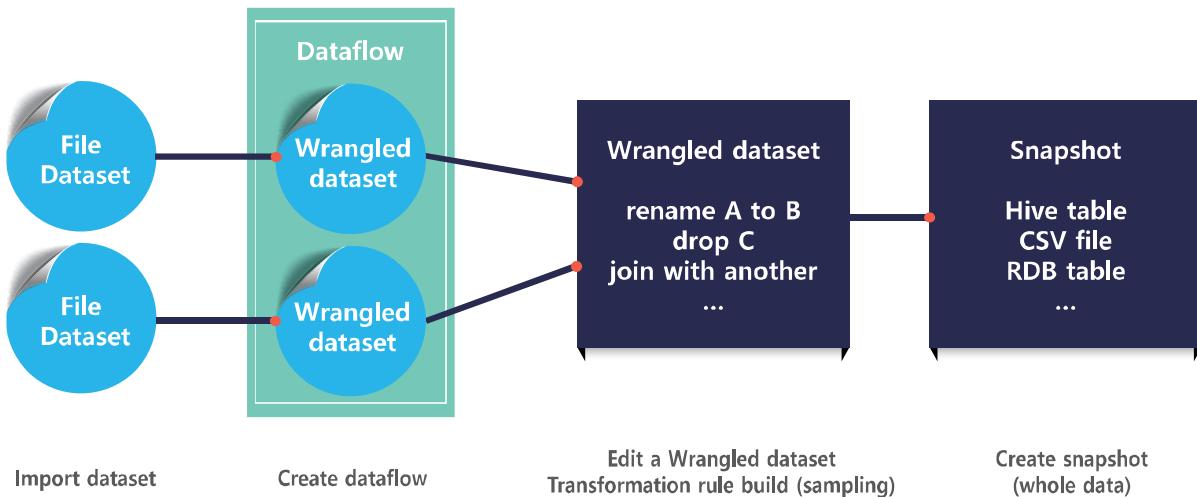
11 Permission: Designates a workspace to allow the use of the new data connection.

- Allow all workspaces to use this data source check box: Click this check box to use the data connection in all workspaces.
- Edit: Used to designate a specific workspace to allow the data connection. This button will disappear if the data connection is set as open data.
- Number of shared workspaces: Displays the number of workspaces where the data connection can be used.

3. Data Preparation Management

Data preparation is a process to prepare (refine, integrate, arrange and transform) the original data before creating a data source for analysis. This is not a process to refine the existing data, but a process to define preparation procedures for records to be added later, in order to automatically refine every new piece of data according to the defined rules. The menu is composed as follows.

- **Dataset:** Manages the connection with the original data. (See Paragraph 3.1.)
- **Dataflow:** Defines the procedure to prepare a dataset. (See Paragraph 3.2.)
- **Save snapshot:** Identifies data inputted and prepared during the dataflow and saves it as a file or database table. (See Paragraph 3.3.)



3.1 Dataset

Dataset means the original data for data preparation. A dataset is divided into two types.

- **Imported dataset:** A dataset containing the original data without defining the preparation procedure.
- **Wrangled dataset:** A dataset with a defined preparation procedure. The preparation procedure can be defined in the dataflow screen.

3.1.1 Dataset management home

In this screen, a user can register, edit or search a dataset.

The screenshot shows the 'Dataset' tab selected in the navigation bar. A search bar at the top allows filtering by dataset name. Below the search bar is a table listing datasets. Each dataset entry includes its name, used count, source type, creation date, and a trash icon for deletion. A summary at the top right indicates there are 105 lists and a button to generate a new dataset.

Dataset	Used In	Source	Created
rrr [W]	1	rrr	2018-05-02 11:35 by admin
rrr	1	HIVE	2018-05-02 11:35 by admin
조민정_테스트1 [W]	1	ddd	2018-05-02 11:23 by admin
○○○ [W]	1	○○○	2018-05-02 11:02 by admin
○○○	1	FILE	2018-05-02 11:02 by admin
winemag-v2 [W]	1	winemag-v2 dataflow	2018-04-23 14:25 by polaris
winemag-v2 [W]	1	winemag-v2 dataflow	2018-04-23 14:22 by polaris
winemag-v2	1	FILE	2018-04-23 14:22 by polaris
1222 [W]	1	33	2018-04-22 23:23 by admin
1222	1	FILE	2018-04-22 23:22 by admin

① **Type:** Search a dataset by its type.

- All: Displays datasets regardless of their type.
- Imported dataset: Displays Imported datasets.
- Wrangled dataset: Displays Wrangled datasets.

② **Search by dataset name:** Search a registered dataset by its name.

③ **Number of data:** Displays the number of datasets searched from the current list.

④ **Create a new dataset:** Click to register a new dataset.

⑤ **Dataset list:** Displays datasets that meet the established criteria. Click a data source to see its details. (See Paragraph 3.1.2.)

⑥ **Delete:** Hover the mouse over the dataset to display the trash icon. Click the icon to delete the dataset.

3.1.2 Dataset details

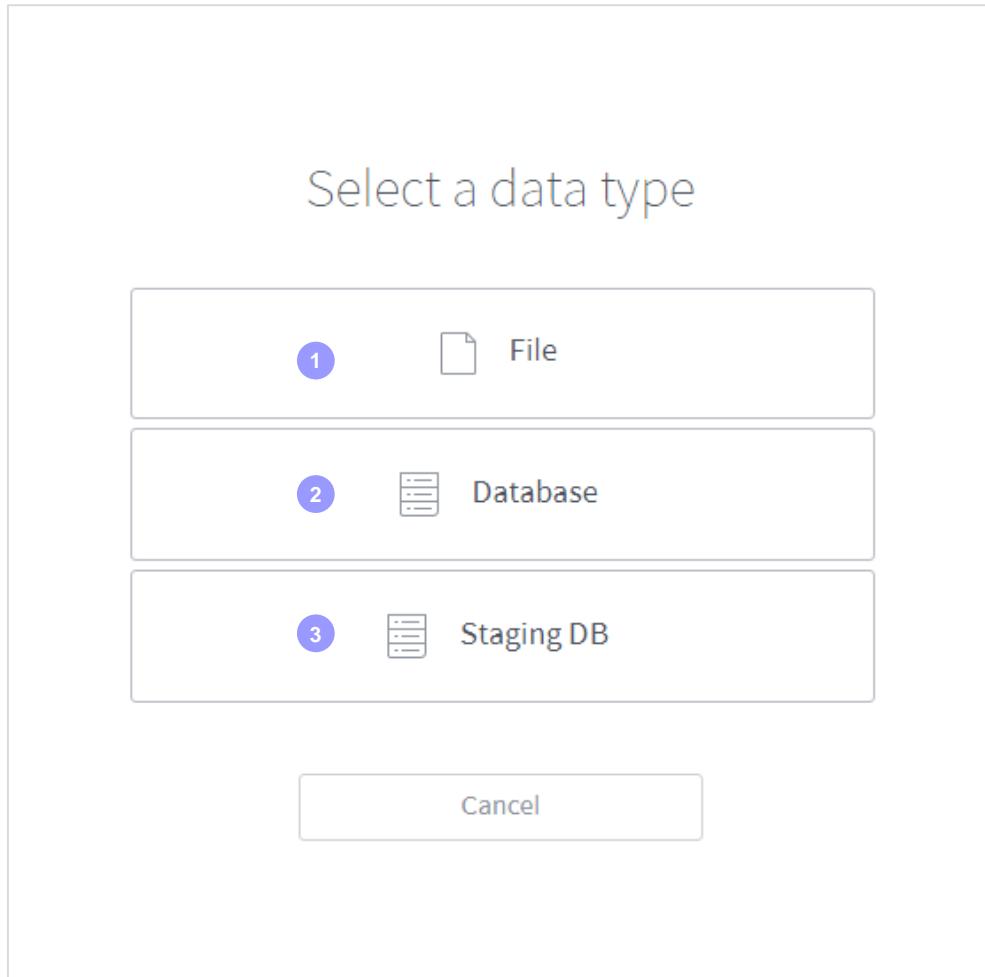
Click a dataset listed in the dataset management home to view various attributes of that dataset.

The screenshot shows the dataset details page for 'winemag-v2 [W]'. The top navigation bar includes a back arrow, the dataset name, a search bar placeholder 'Please enter a description', and a user status 'Updated on 2018-05-01 15:13 by admin'. Below the header, the dataset type is listed as 'WRANGLED'. A summary box indicates '10,000 Rows/ 14 Columns'. The main content area displays a table of data with columns 'ab_no', 'ab_country', and 'ab_description'. The first three rows of the table are visible, showing sample data. At the bottom of the page, a 'Used in' section lists a single entry: 'Created in winemag-v2 dataflow - winemag-v2 dataflow' with a timestamp 'Updated on 2018-04-23 14:25 by polaris'. Numbered callouts (1-6) point to specific elements: 1 points to the dataset name; 2 points to the description input field; 3 points to the 'WRANGLED' type; 4 points to the summary statistics; 5 points to the data table; and 6 points to the 'Used in' section.

- 1 **Name:** Name of the dataset. Click to edit it.
- 2 **Description:** Description of the dataset. Click to edit it.
- 3 **Type:** Displays the type (Imported/Wrangled) of the dataset.
- 4 **Summary:** Number of records and columns of the original data table linked to the dataset.
- 5 **Data:** Information of columns of the original data table linked to the dataset.
- 6 **Used in:** Displays the name of the dataflow that uses the dataset, number of Imported and Wrangled datasets used in that dataflow, time and User ID of the last update of that dataflow. The details screen of the Import type dataset displays all dataflows linked to that dataset.

3.1.3 Create a new dataset

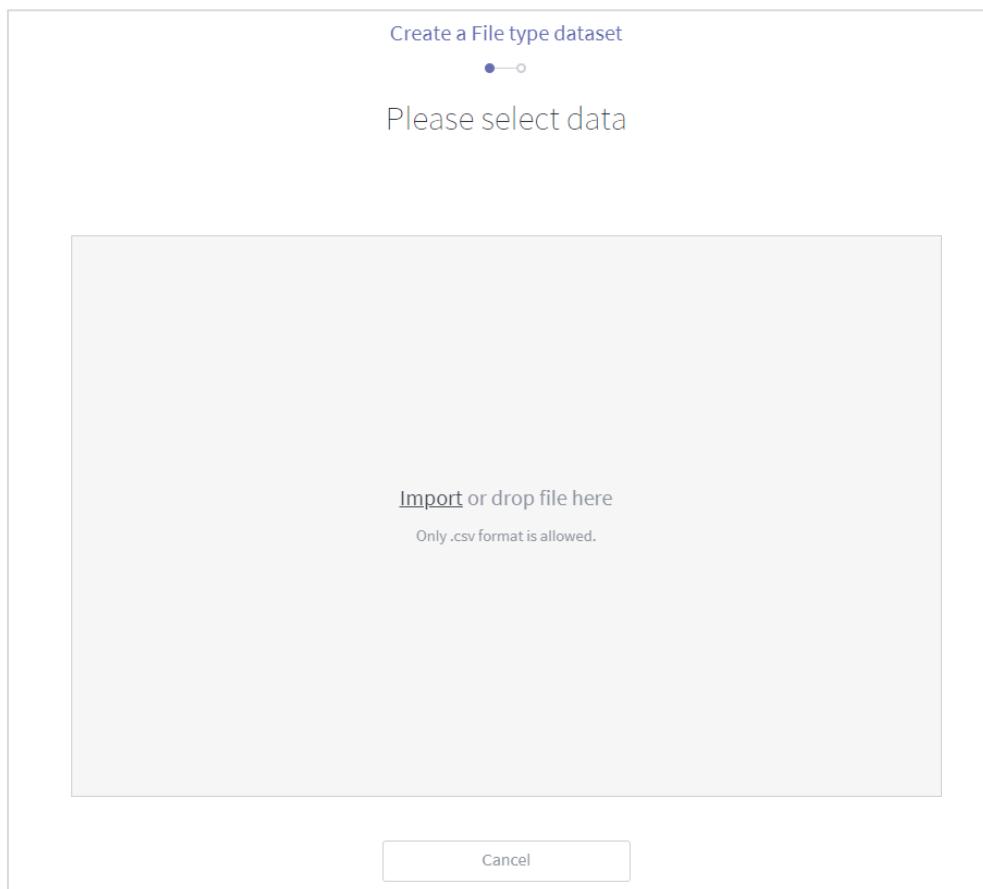
Click the '+ Create a new dataset' button in the top-right of the dataset management home to create a new dataset and register data to be used for preparation.



- ① **File:** Imports files saved on the user's local PC and registers them as a dataset. Only files separated by a comma such as xls, xlsx and csv can be imported.
- ② **Database:** Imports files from an external database and registers them as a dataset.
- ③ **Staging DB:** Registers data imported from the metatron internal Hive database as a dataset.

A. Create a dataset using a file

1. In the screen to select the type of original data, select 'File'.
2. Imports the file to be used as a dataset from the user's local PC. Click the Import button and select the file, or drag a file to the screen. Once the file is imported, click 'Next'.



3. Data in the imported file will be displayed as follows. Confirm that the data is what you want, and click 'Next'. If it is not what you want, import the file again.

Create a File type dataset

Please select data

1 2 Import or drop file here 3

```
"OrderDate","Category","City","Country","CustomerName","OrderID","PostalCode","ProductName","Quantity","Region","Segment","ShipDate","ShipMode","State","Sub-Category","ShipStatus","OrderProfitable","SalesaboveTarget","Latitude","Longitude","Discount","Profit","Sales","DaystoShipActual","SalesForecast","DaystoShipScheduled","SalesperCustomer","ProfitRatio"
2011-01-04T00:00:00Z,Office Supplies,Houston,United States,Darren Powers,CA-2011-103800,77095,"Message Book, Wirebound, Four 5 1/2 X 4 Forms/Pg., 200 Dupl. Sets/Book",2,Central,Consumer,2011. 1. 8.,Standard Class,Texas,Paper,Shipped Early,null,null,29.8941,-95.6481,0.2,6.0,16.0,4.0,22.0,6.0,16.45,33.8
2011-01-05T00:00:00Z,Office Supplies,Naperville,United States,Phillina Ober,CA-2011-112326,60540,GBC Standard Plastic Binding Systems Combs,2,Central,Home Office,2011. 1. 9.,Standard Class,Illinois,Binders,Shipped Early,null,null,41.7662,-88.141,0.8,-5.0,4.0,4.0,5.0,6.0,3.54,-155.0
2011-01-05T00:00:00Z,Office Supplies,Naperville,United States,Phillina Ober,CA-2011-112326,60540,Avery 508,3,Central,Home Office,2011. 1. 9.,Standard Class,Illinois,Labels,Shipped Early,null,null,41.7662,-88.141,0.2,4.0,12.0,4.0,15.0,6.0,11.78,36.3
2011-01-05T00:00:00Z,Office Supplies,Naperville,United States,Phillina Ober,CA-2011-112326,60540,SAFCO Boltless Steel Shelving,3,Central,Home Office,2011. 1. 9.,Standard Class,Illinois,Storage,Shipped Early,null,null,41.7662,-88.141,0.2,-65.0,273.0,4.0,357.0,6.0,272.74,-23.8
2011-01-06T00:00:00Z,Office Supplies,Philadelphia,United States,Mick Brown,CA-2011-141817,19143,"Avery Hi-Liter EverBold Pen Style Fluorescent Highlighters, 4/Pack",3,East,Consumer,2011. 1. 13.,Standard Class,Pennsylvania,Art,Shipped Late,null,null,39.94
```

Column delimiter

4

- 1 **Name:** Name of the file.
- 2 **Import or drop file here:** Used to import a new file.
- 3 **Data preview:** Used to preview the data of the file.
- 4 **Column delimiter:** Enter the delimiter of the column. The default value is a period (.).

4. Confirm information on the file type data imported to create a dataset, enter the name and description, and click 'Done' to create a new dataset.

Create a File type dataset

• •

Please complete dataset creation

File sale_20171212.csv

Name
test

Description
test

Previous Done

The screenshot shows a 'Create a File type dataset' wizard. It's step 2 of 2, indicated by the progress bar at the top. The main message says 'Please complete dataset creation'. Below it, a preview shows a CSV file named 'sale_20171212.csv'. There are input fields for 'Name' (containing 'test') and 'Description' (containing 'test'). At the bottom are 'Previous' and 'Done' buttons, with 'Done' being the primary action button.

B. Create a dataset using a database or Staging DB

1. From the original data type selection screen, select 'database' or 'Staging DB'.
2. If you select 'database', load an existing data connection or enter connection information for a new database, in the data connection setting screen as follows. This step is skipped if you select 'Staging DB'.

Create a DB type dataset

Please set data connection

The screenshot shows the 'Create a DB type dataset' interface. At the top, it says 'Please set data connection'. Below is a grid of database icons:

DB Type	ORACLE DATABASE	MySQL
PostgreSQL		
presto		
TIBERO		

Below the grid, numbered steps 1 through 9 guide the user through the connection setup:

- 1 Load a data connection: A dropdown menu showing 'Hive(1.2)'.
- 2 DB Type: A grid of database logos.
- 3 Host: A text input field labeled 'Host'.
- 4 Port: A text input field labeled 'Port'.
- 5 SID: A text input field labeled 'SID' with placeholder 'Please enter SID'.
- 6 User name: A text input field labeled 'User name' with placeholder 'Please enter user name'.
- 7 Password: A password input field with placeholder '*****'.
- 8 Test: A button labeled 'Test' with a red error message 'Host is required'.
- 9 Save as a new data connection: A button to save the connection with an input field 'Enter name of a new data connectic'.

At the bottom are 'Cancel' and 'Next' buttons.

- 1 **Load a data connection:** Used to select a saved data connection. Select it to automatically load access information on the database connected to the relevant data connection. However, you must verify the connection by clicking the 'Test' button.
- 2 **DB Type:** Select the type of database to be connected. Currently 7 database types are supported. (Oracle, MySQL, PostgreSQL, Hive, presto, Phoenix, TIBERO)
- 3 **Host:** Select the value of the host to be connected.
- 4 **Port:** Enter the number of the port to be connected.
- 5 **SID/Catalog/DB name:** Enter SID for Oracle and TIBERO. Enter Catalog for presto. Enter DB name for PostgreSQL.
- 6 **Username:** Enter the username of the database.
- 7 **Password:** Enter the password of the database.

- 8 Test:** If you enter all fields, the Test button will be activated. Click it to view the connection test result at the bottom of the button. 'Valid connection' means it is normal, and 'Invalid connection' means it is abnormal.
- 9 Save as a new data connection:** To link a new database rather than using an existing data connection, you must save connection information as a new data connection. Enter the name to be used in this step.
3. Select data. You can select a table from the account of linked databases, or write a query statement by yourself. You can only designate a table when creating a data source using Staging DB.

Table

Select the database and table name, confirm the data to be saved and click 'Next'.

Create a DB type dataset

Please select data

ab contract_id	ab product_01_code	ab product_02_code	ab product_03_code	ab product_04_code	ab custom
cid000004744	01	02	001	0014	uid007
cid000006368	01	03	003	0040	uid007
cid000007306	04	04	004	0016	uid015
cid000007406	04	03	001	0030	uid006
cid000007799	02	04	002	0045	uid005
cid000008881	04	04	004	0034	uid011
cid000011827	03	03	002	0028	uid019
cid000016114	02	04	001	0043	uid009
cid000018663	01	05	004	0018	uid019
cid000019207	01	02	004	0019	uid008
cid000021427	02	04	001	0019	uid011
cid000028012	02	02	002	0015	uid014

Previous Next

- 1 Select database:** Select a database linked with the selected data connection.
- 2 Select schema:** Select a table in the selected database.

Query

Write a query statement to import the data you want, and click 'Run' to display data at the bottom. Confirm the data and click 'Next'.

Table Query

select * from product

✓ Valid ConnectionRun

ab_product_01_code	ab_product_02_code	ab_product_03_code	ab_product_04_code	ab_product_name
01	01	001	0001	pn00001
01	01	001	0002	pn00002
01	01	001	0003	pn00003
01	01	001	0004	pn00004
01	01	001	0005	pn00005
01	01	001	0006	pn00006
01	01	001	0007	pn00007
01	01	001	0008	pn00008
01	01	001	0009	pn00009
01	01	001	0010	pn00010

4. Finally, confirm the data configured so far, enter the name and description of the dataset to be created, and click 'Done' to create the dataset. This dataset will be an Imported dataset.

Create a DB type dataset

Please complete dataset creation

Type	HIVE
Host	metatron-poc-h03
Port	10000
Database / Table	default / contract_part
Query	SELECT * FROM default.contract_part

Name
Please enter a name

Description
Please enter a description

Previous Done

3.2 Dataflow

Dataflow is a process of preparing a dataset to secure data quality and make it into a type optimized for analysis or visualization.

3.2.1 Dataflow management home screen

In this screen, a user can manage and prepare the original data.

The screenshot shows the 'Data Preparation' section of the METATRON DISCOVERY interface. At the top, there is a search bar labeled 'Search by name of dataflows' (1). To the right of the search bar, it says 'There are 28 lists' (2) and has a button to 'Add a dataflow' (3). Below the search bar is a table listing 14 dataflows (4). Each row in the table contains the dataflow name, the number of imported datasets ('Dataset') and wrangled datasets ('Wrangled'), and the creation date ('Created'). A trash icon (5) is located at the end of each table row. The dataflows listed are: rrr, ddd, o o o, winemag-v2 dataflow, 33, 연도 구분 포함 플로우, Inpatient_Prospective_Payment_System, eee - eee, TEST, hong, 빈플로우, 테스트, and 인구통계.

Dataflow	Dataset	Wrangled	Created
rrr	1	1	2018-05-02 11:35 by ad-
ddd	1	1	2018-05-02 11:23 by ad-
o o o	1	1	2018-05-02 11:02 by ad-
winemag-v2 dataflow	1	2	2018-04-23 14:22 by pol-
33	1	1	2018-04-22 23:23 by ad-
연도 구분 포함 플로우	1	1	2018-04-22 20:48 by ad-
Inpatient_Prospective_Payment_System	1	1	2018-04-20 14:17 by ad-
eee - eee	1	1	2018-04-19 11:00 by ad-
TEST	1	2	2018-04-17 11:04 by ad-
hong	2	3	2018-04-17 09:50 by ad-
빈플로우			2018-04-13 10:27 by ad-
테스트	3	3	2018-04-11 14:40 by ad-
인구통계	3	3	2018-04-10 13:47 by ad-

- ① **Search:** Search the saved dataflows by name.
- ② **Number of data:** Displays the number of dataflows searched from the current list.
- ③ **Add a new flow:** Click to add a new dataflow.
- ④ **Dataflow list:** Displays dataflows that meet the established sort criteria. Each number next to the icon assigned to each dataflow in the 'Dataset' column represents the number of Imported datasets and Wrangled datasets used in the dataflow, respectively. Click a data source to see its details. (See Paragraph 3.2.3.)
- ⑤ **Delete:** Hover the mouse over the dataflow to display the trash icon. Click the icon to delete the dataflow.

3.2.2 Create a new dataflow

Click the 'Add a dataflow' button in the dataflow management home to display the screen to create a dataflow.

1. Select datasets from the list to be used to create a dataflow.

The screenshot shows a 'Create a dataflow' interface. At the top, a progress bar indicates the task is 100% complete. Below it, a message says 'Please select a dataset'. A search bar labeled 'Search by name of datasets' is followed by a table of datasets. The table has columns for selection (checkbox), dataset name, type, and last updated date. The datasets listed are: rrr (Imported dataset(HIVE)), ooo (Imported dataset(FILE)), winemag-v2 (Imported dataset(FILE)), 1222 (Imported dataset(FILE)), 연도 구분 포함 (Imported dataset(FILE)), Inpatient Prospective Pay... (Imported dataset(FILE)), LLLL (Imported dataset(FILE)), HONG_TEST (Imported dataset(FILE)), and TEST (Imported dataset(FILE)). At the bottom, there are buttons for 'Cancel', 'Next', and '+ Create a dataset'.

	Dataset	Type	Last updated
<input type="checkbox"/>	rrr	Imported dataset(HIVE)	2018-05-02 11:44
<input type="checkbox"/>	ooo	Imported dataset(FILE)	2018-05-02 11:02
<input type="checkbox"/>	winemag-v2	Imported dataset(FILE)	2018-04-23 14:22
<input type="checkbox"/>	1222	Imported dataset(FILE)	2018-04-22 23:22
<input type="checkbox"/>	연도 구분 포함	Imported dataset(FILE)	2018-04-22 20:48
<input type="checkbox"/>	Inpatient Prospective Pay...	Imported dataset(FILE)	2018-04-20 14:16
<input type="checkbox"/>	LLLL	Imported dataset(FILE)	2018-04-19 14:50
<input type="checkbox"/>	HONG_TEST	Imported dataset(FILE)	2018-04-17 14:31
<input type="checkbox"/>	TEST	Imported dataset(FILE)	2018-04-17 11:04

- 1 **Search by dataset name:** Search a registered dataset by its name.
- 2 **Dataset list:** Displays registered datasets. If you search by dataset name, only datasets containing the string in their name will be searched.
- 3 **Create a dataset:** Click this button to display a dialog window to move to the screen to create a new dataset.

2. Confirm the selected datasets, enter the name and description of the dataflow to be created, and click 'Done' to create the dataset.

Create a dataflow

Please complete dataflow creation

Datasets	winemag-v2
----------	------------

Name
Please enter a name

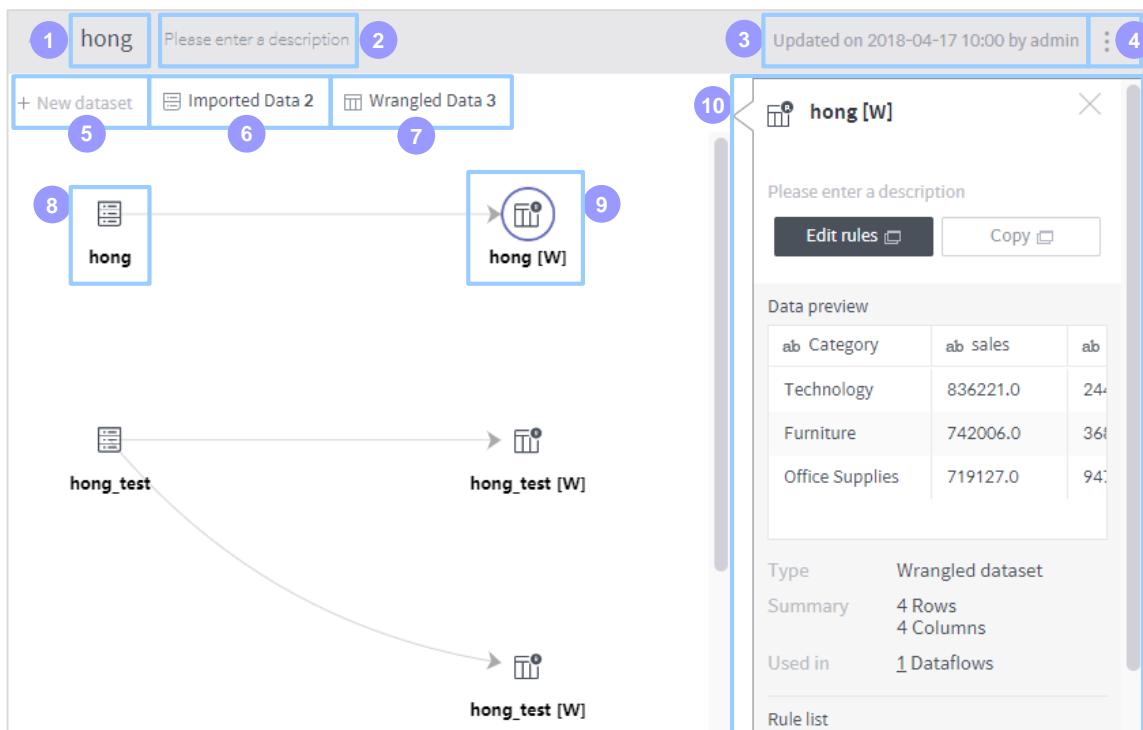
Description
Please enter a description

Previous Done

3.2.3 Dataflow details

Displays the relation between all datasets in the dataflow, and displays each dataset for editing. The relation of influence between datasets is displayed as a figure using functions such as JOIN/UNION.

Outline of the dataflow details screen



- 1 Name:** Name of the dataflow.
- 2 Description:** Description of the dataflow.
- 3 Information last updated by:** Displays the user who performed the last dataflow update.
- 4 Delete:** Click this icon to display the menu to delete the dataflow.
- 5 Add a new dataset:** Click to display a new window to add a new dataset to the dataflow. Click the dataset to be added and click 'Done' to add that dataset.
- 6 Imported dataset:** Displays the number of Imported datasets included in the dataflow.
- 7 Wrangled dataset:** Displays the number of Wrangled datasets included in the dataflow.
- 8 :** Represents each Imported dataset included in the dataflow.
- 9 :** Represents each Wrangled dataset created by wrangling each Imported dataset. 'Wrangling' is a process used to transform an unrefined Imported dataset as easy-to-use data using various tools (see Paragraph 3.2.4). A Wrangled dataset is registered as a data source, and can be used for data analysis and visualization.

- 10 Information on the selected dataset:** Click a dataset icon on the screen to display an overview of that dataset in this area. In this area, a user can confirm, copy or delete the selected dataset, and edit the rules applied to this dataset. For more information on this area, see the following description.

Information outputted when an Imported dataset is selected

- 1 **Name:** Name of the selected dataset.
- 2 **Description:** Description of the selected dataset.
- 3 **Create a new dataset:** Click to create one more Wrangled dataset included in the selected Imported dataset.
- 4 **Data preview:** Displays basic information on the selected dataset.
- 5 **Created on:** Displays the creation time of the selected dataset and the creator.
- 6 **Updated on:** Displays the user who last updated the selected dataset.
- 7 **Delete this data:** Delete the selected dataset.

country 1

Please enter a description 2

Generate new dataset 3

Data preview 4

#	r_regionkey	ab_r_name
0		AFRICA
1		AMERICA
2		ASIA
3		EUROPE

Type HIVE
Summary 5 Rows 3 Columns
Used in 1 Dataflows

Created on 2018-05-02 11:35 by admin
Updated on 2018-05-15 19:31 by admin

[Delete this data](#) 7

Information outputted when a Wrangled dataset is selected

- 1 **Name:** Name of the selected dataset.
- 2 **Description:** Description of the selected dataset.
- 3 **Edit rules:** Click to open a window to edit rules for the selected dataset (for more information on this window, see the next paragraph).
- 4 **Copy:** Click to create another Wrangled dataset by copying the selected dataset.
- 5 **Data preview:** Displays basic information on the selected dataset.
- 6 **Rule list:** Displays rules currently applied to the selected dataset.
- 7 **Created on:** Displays the creation time of the selected dataset and the creator.
- 8 **Updated on:** Displays the user who last updated the selected dataset.
- 9 **Delete this data:** Delete the selected dataset.

hong [W] 1

Please enter a description 2

Edit rules 3 Copy 4

Data preview 5

ab Category	ab sales	ab sur
Technology	836221.0	244.3%
Furniture	742006.0	368.8%
Office Supplies	719127.0	947.8%

Type Wrangled dataset
Summary 4 Rows 4 Columns
Used in 1 Dataflows

Rule list 6

- (H) header header rownum: 1
- (Rm) rename rename col: sum_
_Sales to: 'sales'

2 Rules

Created on 2018-04-17 09:50 by admin
Updated on 2018-04-23 14:21 by polaris

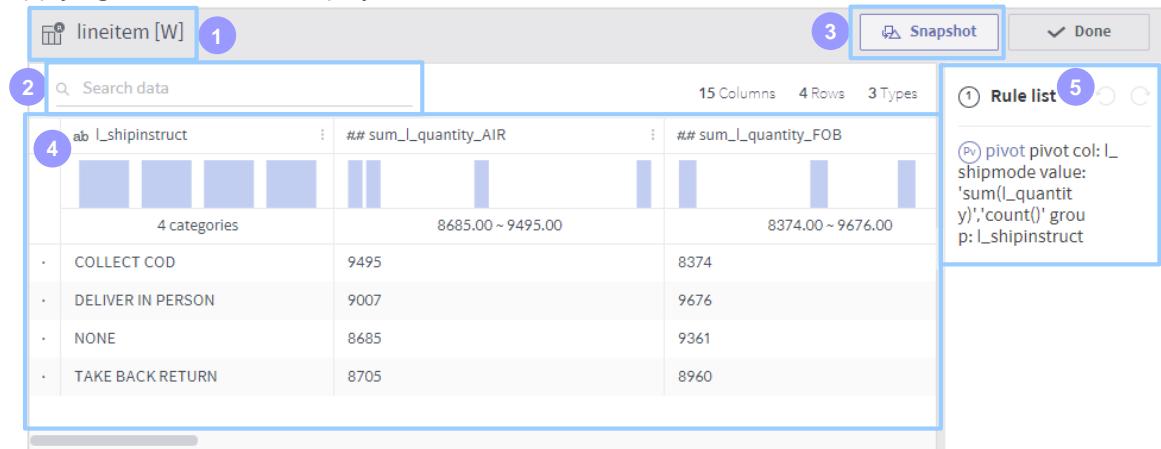
[Delete this data](#) 9

3.2.4 Edit a Wrangled dataset

Editing a Wrangled dataset is the most basic work of data preparation. A user can define the rules to transform a dataset, apply the rules and, based on the result, append a new rule, update, delete, undo or redo the work to make rules to transform data quickly.

Dataset transformation rules and result of application

The currently applied transformation rules are listed on the top-right, and the result of applying those rules is displayed on the left.



- ① **Name:** Name of the selected Wrangled dataset.
- ② **Search:** Search data in the selected Imported dataset to display only data containing the searched string.
- ③ **Snapshot:** Saves all data collected and prepared based on the result of the selected Wrangled dataset as a file or database table, to use as a data source (see Paragraph 3.3).
- ④ **Wrangling result data:** Displays Wrangling result data by applying rules to the dataset.
- ⑤ **Rule list:** Displays rules applied to the dataset.

Edit and add a dataset transformation rule

At the bottom of the screen, there is an area to edit or add a dataset transformation rule.



Add Rule [switch to editor](#)

Command

Choose Rule Function

Cancel Add

The following is a description of each command's function.

Command	Function
rename	Changes the column name
drop	Removes the column
delete	Deletes rows that meet the conditional expression
keep	Deletes rows except those that meet the conditional expression
set	Sets the result of the given conditional expression as a value of the column
derive	Similar with the above 'set', but leaves the original column and creates a new column
header	Sets column values in the designated row as the name of each column at once
replace	Applies a conversion equation to the column values at once
settype	Changes the type of column (Integer, Float, String, etc.)
extract	Finds the pattern and makes it a new column
split	Splits one column into multiple columns
merge	Merges multiple columns into one column
join	Transforms into a result joined with another dataset
union	Merges contents of different datasets that share the same schema
sort	Sorts by the selected column value
move	Moves a column up or down
aggregation	Creates a new column based on the result of the grouping operation
flatten	Creates each element as a new column when the arrangement column is designated
pivot	Creates columns whose name is each value of the designated column, and performs the grouping operation
unpivot	Designates column names as column values, and converts them into multiple rows
countpattern	Creates new column based on the times of appearance of the given pattern from the designated column
nest	Designates multiple columns and makes an arrangement or map using their values to create a new column
unnest	Extracts a specific element from an arrangement or map to create a new column

3.2.5 Create a data snapshot

Click the 'snapshot' button in the Wrangled dataset edit screen to save a snapshot.

A. Data snapshot type: File

Select 'File' as the data snapshot type to save the created snapshot in the server as a file.

The screenshot shows a 'Create a snapshot' dialog box. At the top, it says 'Create a snapshot'. Below that, there are four numbered steps:

- 1 Snapshot type:** A radio button group where 'File' is selected (indicated by a blue outline) and 'HIVE' is unselected (indicated by a grey outline).
- 2 File format:** A dropdown menu showing 'CSV'.
- 3 Compression:** A dropdown menu showing 'NONE'.
- 4 Full Data ETL engine:** A dropdown menu showing 'Embedded Engine'.

At the bottom of the dialog box are two buttons: 'Cancel' (in a light grey box) and 'Done' (in a dark grey box).

- 1 Data snapshot type:** Select the type of database where the data snapshot is saved.
- 2 File format:** Select the file format in which to save the data snapshot.
- 3 Compression:** Select the compression type to save the data snapshot.
- 4 Full Data ETL engine:** Select the engine to be used for data ETL (extraction, transformation and loading).
 - Embedded engine: Suitable for detecting small data.
 - Spark: Suitable for detecting large data.

B. Data snapshot type: HIVE

Select 'HIVE' as the data snapshot type to save the created snapshot in metatron's internal Hive as a database.

Create a snapshot

1 Snapshot type File HIVE

2 DB name

3 Table name

4 File format

5 Compression

6 Overwrite method

7 Partition keys

8 Full Data ETL engine

9 Profile job result when creating a snapshot

Cancel Done

- 1 **Data snapshot type:** Select the type of database where the data snapshot is saved.
- 2 **DB name:** Select the database where the data snapshot is saved.
- 3 **Table name:** Select the table name where the data snapshot is saved.
- 4 **File format:** Select the file format in which to save the data snapshot. Select the file format in which the recorded value is saved in Hive.
- 5 **Compression:** Select the compression type to save the data snapshot.
- 6 **Overwrite method:** Select how to save tables with the same name.
 - Overwrite: Overwrites the snapshot in the existing table.
 - Append: Add snapshot data to the existing table.
- 7 **Partition keys:** Designate columns to use as partition keys:

8 Full Data ETL engine: Select the engine to be used for data ETL (extraction, transformation and loading).

- Embedded engine: Suitable for detecting small data.
- Spark: Suitable for detecting large data.

9 Choose columns to use as partition keys: Select whether to use a partition key function.

3.3 Data snapshot

Data snapshot is a function to apply the preparation rules defined in a Wrangled dataset to the ingested data, and save its result. The saved data can be used as a data source for analysis and visualization. In this way, the original data passes data preparation and is saved as an optimal form for analysis.

3.3.1 Data snapshot management home

This screen displays the name and creation time of data snapshots, the name of a created dataflow and Wrangled dataset, success/failure, elapsed time, etc.

The screenshot shows a user interface for managing data snapshots. At the top, there is a status filter with radio buttons for All, Success, and Fail. Below the filter is a search bar labeled "Search by name of snapshots". To the right of the search bar, it says "There are 27 lists". The main area is a table with the following columns: Snapshot, Dataflow | Dataset, Status, Elapsed time, and Created. The table contains five rows of data:

Snapshot	Dataflow Dataset	Status	Elapsed time	Created
winemag-v2 [W] 20180424_062250	winemag-v2 dataflow winemag-v2 [W]	Success	00:00:20.00	2018-04-24 15:22 by ad...
winemag-v2 [W] 20180423_053613	winemag-v2 dataflow winemag-v2 [W]	Success	00:00:20.00	2018-04-23 14:36 by pol...
winemag-v2 [W] 20180423_052714	winemag-v2 dataflow winemag-v2 [W]	Success	00:00:17.00	2018-04-23 14:27 by pol...
HONG_TEST [W] 20180417_053135	TEST HONG_TEST [W]	Success	00:00:01.00	2018-04-17 14:31 by ad...
hong_test [W] 20180417_010504	hong hong_test [W]	Success	00:00:01.00	2018-04-17 10:05 by ad...

- ① **Status:** Search the data snapshot by its success/failure status.
 - All: Displays all data snapshots regardless of the success/failure of data snapping.
 - Success: Displays snapshots where data snapping is successful.
 - Failure: Displays snapshots where data snapping has failed.
- ② **Search by name:** Search a created data snapshot by its name.
- ③ **Number of data:** Displays the number of data snapshots searched from the current list.
- ④ **Data snapshot list:** Displays data snapshots that meet the established criteria. Click a data source to see its details. (See Paragraph 3.1.2.)
- ⑤ **Delete:** Hover the mouse over the data snapshot to display the trash icon. Click the icon to delete the data snapshot.

3.3.2 Data snapshot details

Click a snapshot listed in the data snapshot management home to display information about the data snapshot such as the Hive table name, number of lines and capacity, and to check the actual contents of data as follows.

The screenshot shows the METATRON DISCOVERY interface with a data snapshot details page for 'winemag-v2 [W]'. The page is divided into several sections:

- Summary Bar:** Shows Valid (100%), Mismatched (0%), and Missing (0%) rates.
- Data Preview Table:** Displays a sample of data from the 'sale' table, including columns: ab_no, ab_country, and ab_description.
- Detailed Metadata:**
 - Database:** winemag-v2
 - Table:** sale [W]
 - Summary:** 129,975 rows, 14 Columns
 - Elapsed time:** 00:00:20.00
 - Created:** 2018-04-24 15:22
 - Dataflow:** winemag-v2 dataflow
 - Dataset:** winemag-v2 [W]
 - Origin imported dataset:** winemag-v2

- 1 **Name:** Dataset used by the snapshot.
- 2 **Success rate:** Rate of records saved in the Hive table successfully.
 - Valid: Rate of records saved in the Hive table successfully because all column formats match.
 - Mismatched: Rate of records failed to save in the Hive table because column formats do not match.
 - Missing: Rate of missing records
- 3 **Snapshot data history:** Displays the history of data snapped in the snapshot.
- 4 **Database:** Name of the Hive database where the snapshot is saved.
- 5 **Table:** Name of the Hive table where the snapshot is saved.
- 6 **Summary:** Number of rows (records) and columns of data snapped in the snapshot.
- 7 **Elapsed time:** Elapsed time to snap the data.
- 8 **Created on:** Time when the snapshot was created.
- 9 **Dataflow:** Displays the dataflow used to create the snapshot. Click to search/edit the dataflow.
- 10 **Dataset:** Displays the Wrangled dataset used to create the snapshot. Click to search/edit the dataset.
- 11 **Origin imported dataset:** Displays the Imported dataset (original data) of the snapshot.

4. Notebook Management

metatron Discovery provides functions to utilize external analysis tools such as Jupyter and Zeppelin through a module called a notebook. To utilize these functions, it is necessary to link the server where the external analysis tool is installed. This server is called the 'notebook server'.

4.1 Notebook server

In this menu, a user can register a new notebook server or search/edit registered notebook servers.

4.1.1 Search notebook server list

Used to view a list of available notebook servers. A user can add a notebook server, and edit or delete its connection information.

The screenshot shows the 'Notebook Management' interface with the 'Notebook Server' tab selected. At the top, there is a 'Type' dropdown set to 'ALL' (1) and a search bar (2). To the right, there are buttons for 'Add a server' (4) and 'Delete selections' (5), and a message indicating 'There are 7 lists' (3). The main area displays a table of registered servers (6) with columns for Server name, Type, Host, Port, Creator (ID), and Last updated. The servers listed are: jwc_jupyter, jwc_zeppelin, zeppelin-default, jupyter-default, test N -ttt', dstest zeppelin -zeppelin, and dstest-jupyter.

Server	Type	Host	Port	Creator (ID)	Last updated
jwc_jupyter	jupyter	metatron-poc-a04	8889	관리자 (admin)	2018-04-23 by 관리자
jwc_zeppelin	zeppelin	metatron-poc-a04	8080	관리자 (admin)	2018-04-20 by 관리자
zeppelin-default	zeppelin	zeppelin.mclouud.sktelecom.com	80	관리자 (admin)	2017-11-23 by 관리자
jupyter-default	jupyter	jupyter.mclouud.sktelecom.com	80	관리자 (admin)	2017-11-22 by 관리자
test N -ttt'	zeppelin	test DS host	9000	관리자 (admin)	2017-11-21 by 관리자
dstest zeppelin -zeppelin	zeppelin	52.231.201.148	8080	관리자 (admin)	2017-11-21 by 관리자
dstest-jupyter	jupyter	52.231.201.148	8888	관리자 (admin)	2017-11-21 by 관리자

- ① **Type:** Search a registered notebook server by the linked external analysis tool (Jupyter/Zeppelin).
 - All: Displays all notebook servers regardless of the external analysis tool.
 - Jupyter: Displays notebook servers linked to Jupyter.
 - Zeppelin: Displays notebook servers linked to Zeppelin.
- ② **Search by server name:** Search a registered notebook server by its name.
- ③ **Number of data:** Displays the number of data sources searched from the current list.
- ④ **Add a server:** Click to create a new notebook server.
- ⑤ **Delete selections:** Deletes notebook servers selected in the left check boxes from the notebook server list.
- ⑥ **Notebook server list:** Displays notebook servers that meet the established criteria. Click a notebook server to edit its setting.

4.1.2 Register a new notebook server

Click the 'Add a server' button in the notebook management screen to display a pop-up window to register a notebook server as follows.

Add a notebook server

1 Type jupyter

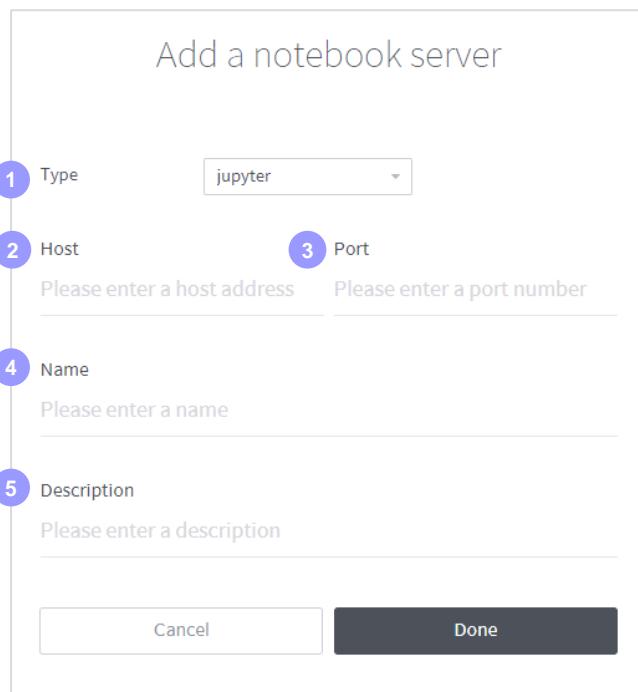
2 Host Please enter a host address

3 Port Please enter a port number

4 Name Please enter a name

5 Description Please enter a description

Cancel Done



- 1 **Type:** Select the external analysis tool installed in the notebook server to be registered.
- 2 **Host:** Enter a host value of the notebook server to be registered.
- 3 **Port:** Enter the port number of the notebook server to be registered.
- 4 **Name:** Enter the name of the notebook server to be registered (mandatory).
- 5 **Description:** Enter the description of the notebook server to be registered.

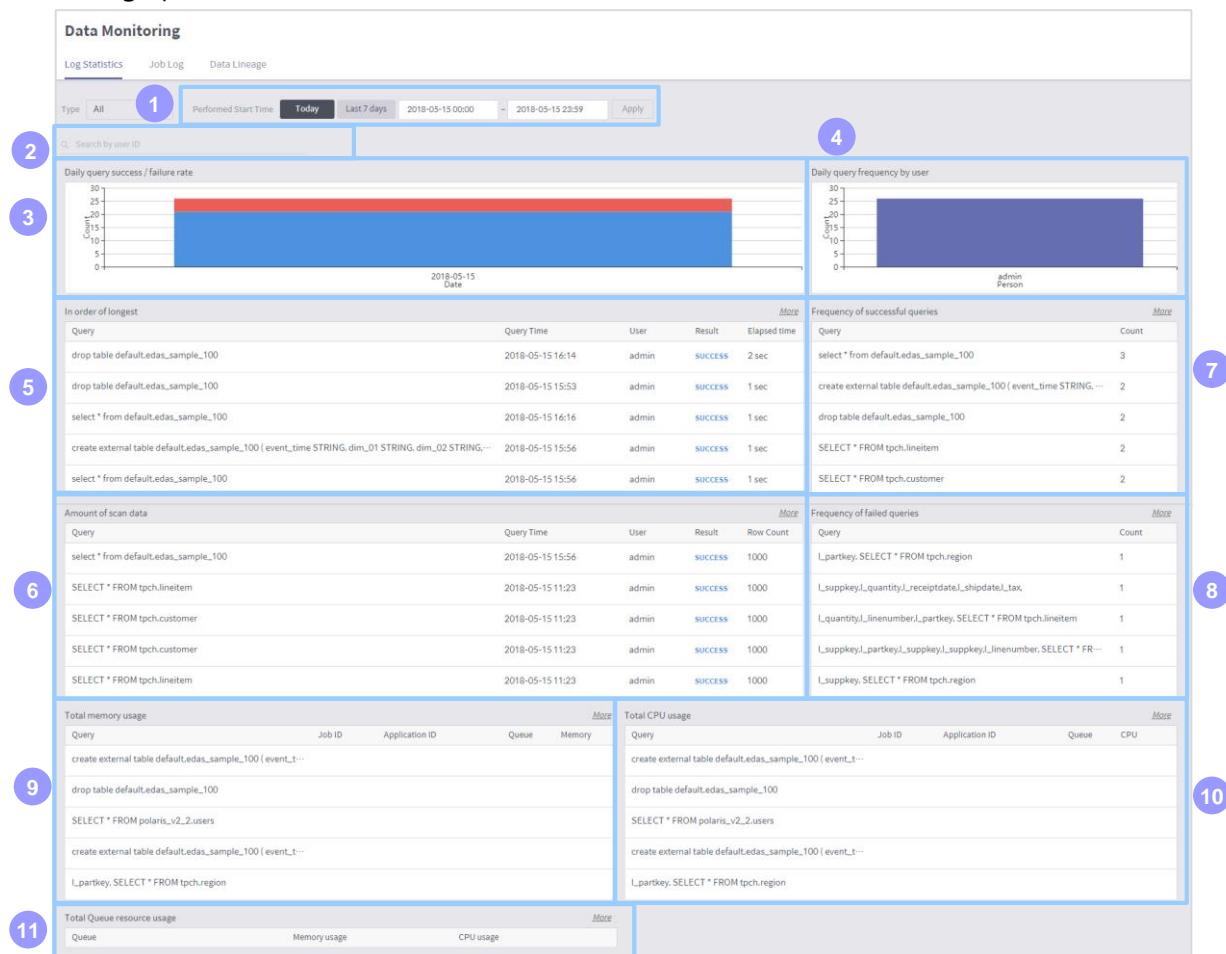
5. Data Monitoring Management

Data monitoring is a function to monitor various data logs generated in metatron's Staging DB (internal Hive DB) and workbench (external DB).

5.1 Log analysis

5.1.1 Log analysis home screen

This screen displays various statistics related to the performance of queries in metatron Discovery. Click a query from the search results to move to the View detail screen of that query. (See Paragraph 5.2.2.)



- 1 Performance Start Time:** Enter the period of the data log to be included in the query analysis.
- 2 Search User ID:** Search by ID of the user who executed the query.
- 3 Query success/failure rate:** Displays the success/failure rate of queries performed in metatron.
- 4 Query frequency by user:** Graph indicating the frequency of each user who performed the query. Click a bar to view the Job Log performed by the user.
- 5 In order of longest:** Displays the performed queries in order of longest elapsed time.

- 6 **Amount of scan data:** Displays the performed queries in order of highest amount of scanned data.
- 7 **Frequency of successful queries:** Displays the performed queries in order of highest frequency of success.
- 8 **Frequency of failed queries:** Displays the performed queries in order of highest frequency of failure.
- 9 **Total memory usage:** Displays the performed queries in order of highest total memory usage.
- 10 **Total CPU usage:** Displays the performed queries in order of highest total CPU usage.
- 11 **Resource usage by queue:** Displays the amount of resources consumed in each YARN queue of the Hadoop environment.

5.2 Job log

Used to search the history of all queries performed in metatron.

5.2.1 Job log home

In this screen, a user can search the query history by applying custom search criteria.

The screenshot shows the 'Data Monitoring' interface with the 'Job Log' tab selected. At the top, there are three main search filters: 1) Status (radio buttons for All, Success, Fail), 2) Limited Elapsed time (dropdowns for All, Over 10 ms, Over 30 ms, Over 60 ms, and a date range input), and 3) Performance Start Time (dropdowns for All, Today, Last 7 days, and date range inputs). Below these is a search bar 4) 'Search by job and application'. To the right of the search bar is a message 5) 'There are 1308 lists'. The main area displays a table of query logs with columns: Status, Job name, Job ID, Application ID, Queue, User (ID), Started time, and Elapsed time. The table contains seven entries, all labeled 'SUCCESS' under the 'Status' column, showing various SQL statements and their execution details.

Status	Job name	Job ID	Application ID	Queue	User (ID)	Started time	Elapsed time
SUCCESS	select * from polaris_v2_2.users where user_name = 'ad...				admin	2018-05-15 16:59	1ms
SUCCESS	SELECT * FROM polaris_v2_2.users				admin	2018-05-15 16:57	2ms
SUCCESS	SELECT * FROM polaris_v2.users				admin	2018-05-15 16:57	1ms
SUCCESS	select * from default.edas_sample_100				admin	2018-05-15 16:16	1 sec
SUCCESS	create external table default.edas_sample_100 (event_...				admin	2018-05-15 16:16	1 sec
SUCCESS	drop table default.edas_sample_100				admin	2018-05-15 16:14	2 sec
SUCCESS	select * from default.edas_sample_100				admin	2018-05-15 15:56	1 sec

- 1 **Status:** Search queries performed by success/failure status.
- 2 **Limited Elapsed time:** Search queries with long elapsed time. User can select the standard time.
- 3 **Performance Start Time:** The standard time for searching queries. This time is based on the start time of each query's performance.
- 4 **Search by Job or application:** Search remaining queries with the current history by query statement or Application ID.
- 5 **Number of data:** Displays the number of queries searched from the current list.
- 6 **Job list:** Displays queries that meet the established sort criteria. Click a data source to see its details. (See Paragraph 5.2.2.)

5.2.2 Job log details

Click a query listed in the job log home to view various information and the history of that query.

The screenshot shows the 'Job log details' page for a query. At the top, it displays the query: '(SELECT * FROM polaris_v2.users)'. To the right, it says 'Recently performed on 2018-05-15 16:57 by admin'. Below this, there are sections for 'Data Information' and 'Connection' (MySQL), and a 'Query History' section. A numbered callout (1-8) points to various parts of the interface.

Data Information

1	Status	SUCCESS
2	Job name	SELECT * FROM polaris_v2.users
3	Started time	2018-05-15 16:57
4	Elapsed time	1ms
5	User	admin

Connection

6	Connection	Type	MYSQL
		Host	metatron-poc-a01
		Port	3306
		JDBC URL	jdbc:mysql://metatron-poc-a01:3306/

Query History

Query date	User	Elapsed time	Result	Action
2018-05-15 16:59	관리자	1 ms	SUCCESS	Detail >
2018-05-15 16:57	관리자	2 ms	SUCCESS	Detail >
2018-05-15 16:57	관리자	1 ms	SUCCESS	Detail >
2018-05-15 13:49	관리자	3 ms	SUCCESS	Detail >
2018-05-11 14:12	관리자	55 ms	SUCCESS	Detail >

Plan **Run Plan Query**

- 1 **Status:** Displays the success/failure status of the query.
- 2 **Job name:** Performed query statement.
- 3 **Start time:** Time when the query was started.
- 4 **Elapsed time:** Time taken to perform the query.
- 5 **User:** ID of the user who performed the query.
- 6 **Connection:** For a query performed in workbench, displays information of the relevant data connection.
- 7 **Query History:** For a query performed in workbench, the history of the latest 5 queries performed in the database and their results. Click Detail to display the query statement in a new window.
- 8 **Plan:** Perform the query performance plan.

5.3 Data lineage

In data lineage, a user can analyze a Hive log performed from metatron's internal Hive database to check the flow of data ETL (extraction, transformation and loading). Workflow is performed through query statement, so the data lineage tracking function is based on the query log.

5.3.1 Data lineage management home screen

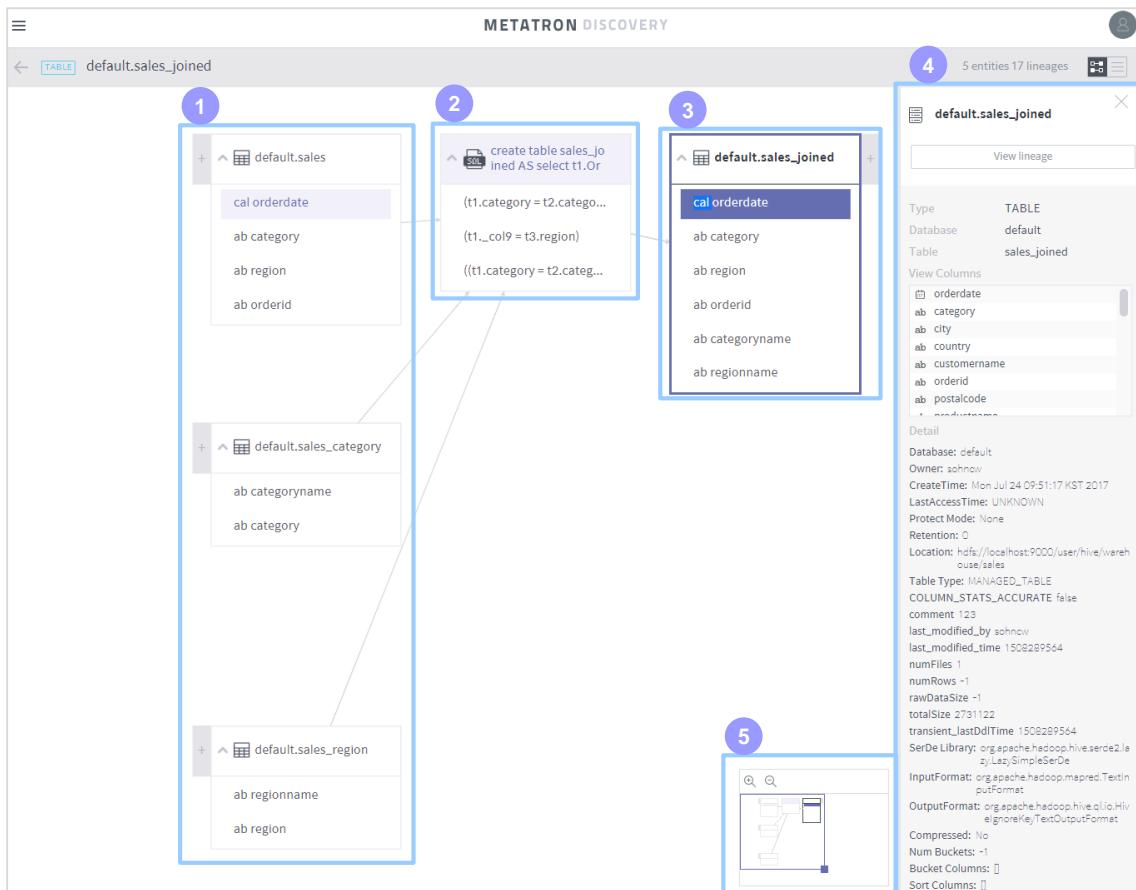
The screenshot shows the METATRON DISCOVERY interface with the Data Monitoring section selected. The Data Lineage tab is active. A search bar labeled "Search lineage by keyword" is highlighted with a blue box. To its right, a radio button group for selecting entity types is also highlighted. The results list, labeled with a blue circle containing the number 3, displays a table with 7 entries:

Database	Table
default	sales
default	sales_category
default	sales_joined
default	sales_joined3
default	sales_region
default	test
default	test333

- ① **Search:** Input a keyword to search data. The keyword is determined by the entity type selected on the right.
- ② **Entity type:** Select entity type to track workflow.
 - Table: Displays tables related to the performed workflow.
 - Column: Displays columns related to the performed workflow.
 - SQL: Displays the SQL query statement related to the performed workflow.
 - Workflow: Displays performed workflows.
- ③ **Entity list:** Displays entities that meet the established sort criteria. Click a data source to see its details. (See Paragraph 5.3.2.)

5.3.2 Data lineage details

Click an entity listed in the data lineage management home to view information of the workflow linked to that entity.



- ① **Original data table box:** Displays the name of the table containing the original data used to perform each query in the workflow. Click the button to search columns stated in the query statement. Select the box to search information on the relevant table, and select the column to color the linked query and columns of other tables.
- ② **SQL query list box:** Displays the SQL query statement performed for the workflow. Click the button to search the contents of the clause that defines the relation between the original data column and result data column. Select box to search information on the relevant SQL, and select column to color the linked query and linked columns.
- ③ **Result data table box:** Displays the name of the table containing the result data of workflow queries. Click the button to search the columns stated in the query statement. Select the box to search information on the relevant table, and select the column to color the linked query and columns of other tables.
- ④ **View details:** Click a box in workflow to display the details of the table or SQL on the right of the screen. Select the table box to search the meta data of the relevant table, and select the SQL box to search information on the query statement.
- ⑤ **Mini-map:** Used to zoom in/zoom out/move the screen.



metatron 2.0

Part 3

Using Workspace



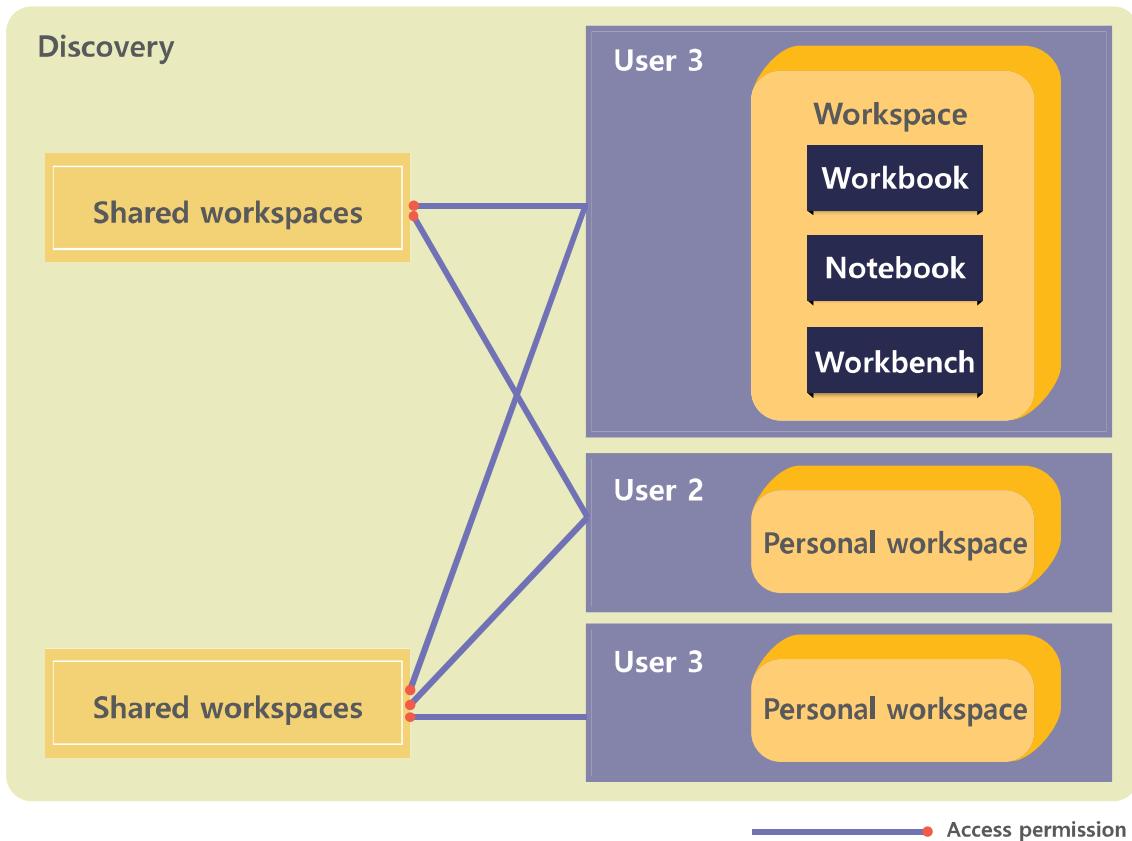
Copyright ©  SK telecom

The content of this document must not be copied,
distributed or used in part or in whole without the prior approval of SK Telecom.

Contents

1. Overview of Workspace	3
2. Workspace Management Home.....	4
2.1 Composition of the workspace management home screen	4
2.2 Folder item.....	6
2.3 Entity item	6
2.4 Copy/Move/Delete folder and entity	7
3. Shared Workspace List.....	8
4. Create a Shared Workspace.....	9
5. Set Access Permission for a Shared Workspace	10
5.1 Permission schema	10
5.1.1 Search permission schema	10
5.1.2 Change permission schema setting	11
5.2 Set shared member & group	12

1. Overview of Workspace



Workspace is a space to store metatron Discovery analysis modules such as workbook, notebook and workbench. Workspace is divided into a personal workspace and a shared space.

- **Personal workspace:** A private workspace assigned to each Discovery member. Only the member can access his/her personal workspace.
- **Shared workspace:** A public workspace for multiple users. This space is used to share the process and results of analysis with other users. An owner or administrator of a shared workspace can grant various levels of access to Discovery members.

This part is divided as follows.

Chapter 2 Workspace management home: Describes the composition and UI of the workspace home screen.

Chapter 3 Shared workspace list: 'Shared workspace list' is a page that lists a user's accessible shared workspaces. The user can filter and search for a desired work list.

Chapter 4 Create a shared workspace: Describes how to create a new shared workspace.

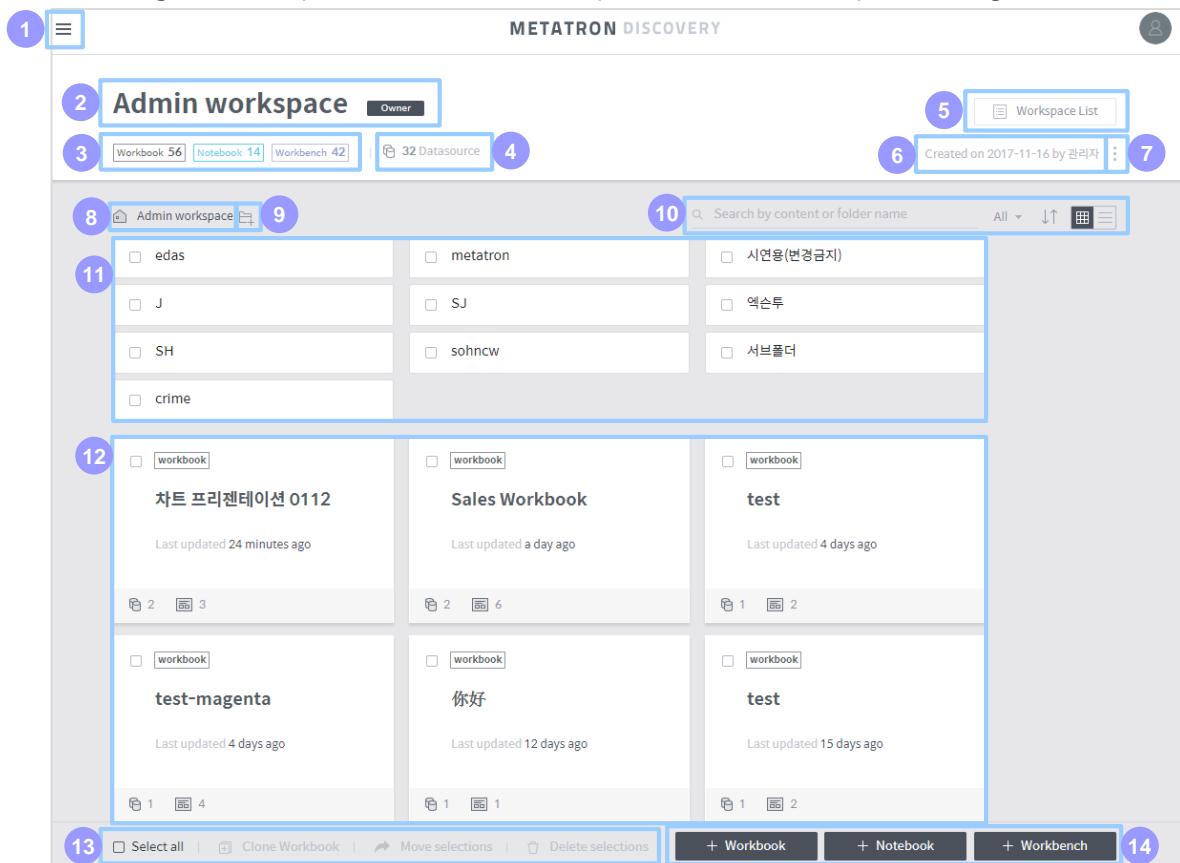
Chapter 5 Set access permission for a shared workspace: To collaborate in a shared workspace, each user must have various roles and permissions. This menu is used to set the permission level of each role and assign these permission levels to each user.

2. Workspace Management Home

In this screen, a user can perform the management functions in metatron Discovery's analysis modules (workbook, notebook and workbench).

2.1 Composition of the workspace management home screen

The following is a description of the overall composition of the workspace management home.



- 1 **Main menu button:** Click this button to open a panel to access another workspace.
- 2 **Workspace information:** Displays the name and description of the workspace. If the logged-in user is the owner of the workspace, an 'Owner' icon will be displayed next to the name of the workspace.
- 3 **Status of registered entities:** Displays the number of entities registered in the workspace by entity type.
- 4 **Data source:** Displays the number of data sources used in the workspace. Click this area to show a list of these data sources.
- 5 **Workspace List:** Click this button to show a list of shared workspaces. (See Chapter 3 for more information.)
- 6 **Creation information:** Displays the creation date and creator of the workspace.

- 7 More:** Displays the creation date and creator of the workspace.
- Edit the name and description: Edit the name and description of the workspace.
 - Set shared member & group: Set the user and group who can access the workspace. (See Paragraph 5.2 for more information.)
 - Set notebook server: Set the access information for the external analysis tool used in the notebook module.
 - Set permission schema: Set the access permission of each user role for the workspace. (See Paragraph 5.1 for more information.)
 - Change owner: Change the owner of the workspace.
 - Delete workspace: Delete the workspace.
- 8 Workspace path:** Displays the current location in the workspace. Click on an upper folder listed in the path to move to that folder.
- 9 Create a folder:** Click to create a new folder in the current location.
- 10 Sort/align entity list:**
- Search: Search for an entity or folder in the workspace by name.
 - Entity type combobox: Search for only a specific entity type in workbook, notebook and workbench.
 - Sort: Sort folders or entities by name or update time.
 - View format: Select grid view or list view as a format to list entities in the workspace.
- 11 Folder list:** Displays folders in the current location that meet search criteria. Click a folder to move to that folder. (For more information on each folder, see Paragraph 2.2.)
- 12 Entity list:** Displays entities in the current location that meet search or sort criteria. Click an entity to move to the home screen of that entity. (For more information on each entity, see Paragraph 2.3.)
- 13 Select/copy/move/delete entity:** Select all entities, or copy, move or delete an entity. (See Paragraph 2.4 for more information.)
- 14 Create an entity:** Buttons used to create a specific type of entity in the workspace. (See Chapter 2 of Part 4, Chapter 3 of Part 5, and Chapter 2 of Part 6, respectively, for more information on how to create.)

2.2 Folder item

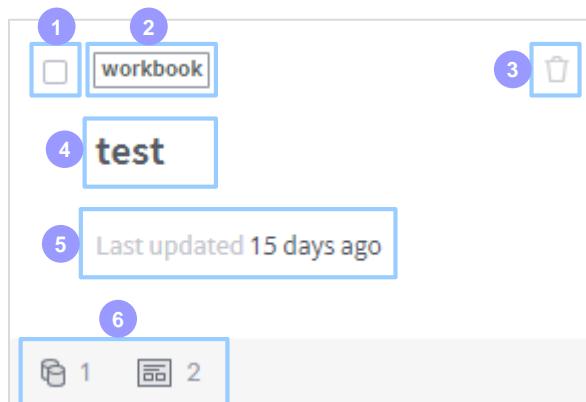
Each folder item is displayed as follows.



- ① **Check box:** Used to select the folder. The user can clone, move or delete the selected folder.
- ② **Name:** Name of the folder.
- ③ **Edit:** Click to edit the name of the folder. To display this button, you must hover the mouse over the folder item.
- ④ **Delete:** Click to delete the folder. To display this button, you must hover the mouse over the folder item.

2.3 Entity item

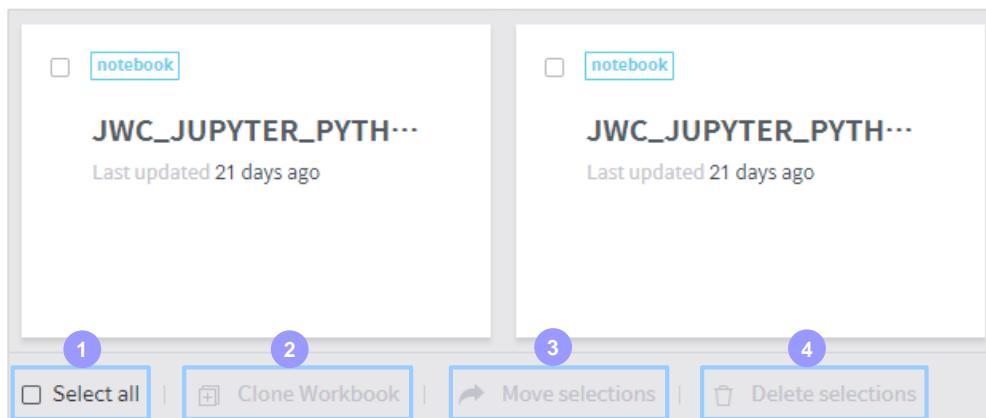
Each entity item is displayed as follows.



- ① **Check box:** Used to select the entity. The user can clone, move or delete the selected entity.
- ② **Entity type:** Displays the type of entity (workbook/notebook/workbench).
- ③ **Delete:** Click to delete the entity. To display this button, you must hover the mouse over the entity item.
- ④ **Name:** Name of the entity.
- ⑤ **Update time:** Displays the time of the latest entity update.
- ⑥ **Number of data sources/dashboards:** This is an exclusive area for the workbook entity.
 - The number next to the  icon is the number of data sources linked to the workbook.
 - The number next to the  icon is the number of dashboards registered to the workbook.

2.4 Copy/Move/Delete folder and entity

A user can copy, move or delete folders and entities in the workspace. Select a folder or entity to be copied, moved or deleted. Operation buttons will be activated in the lower-left corner of the workspace home screen.



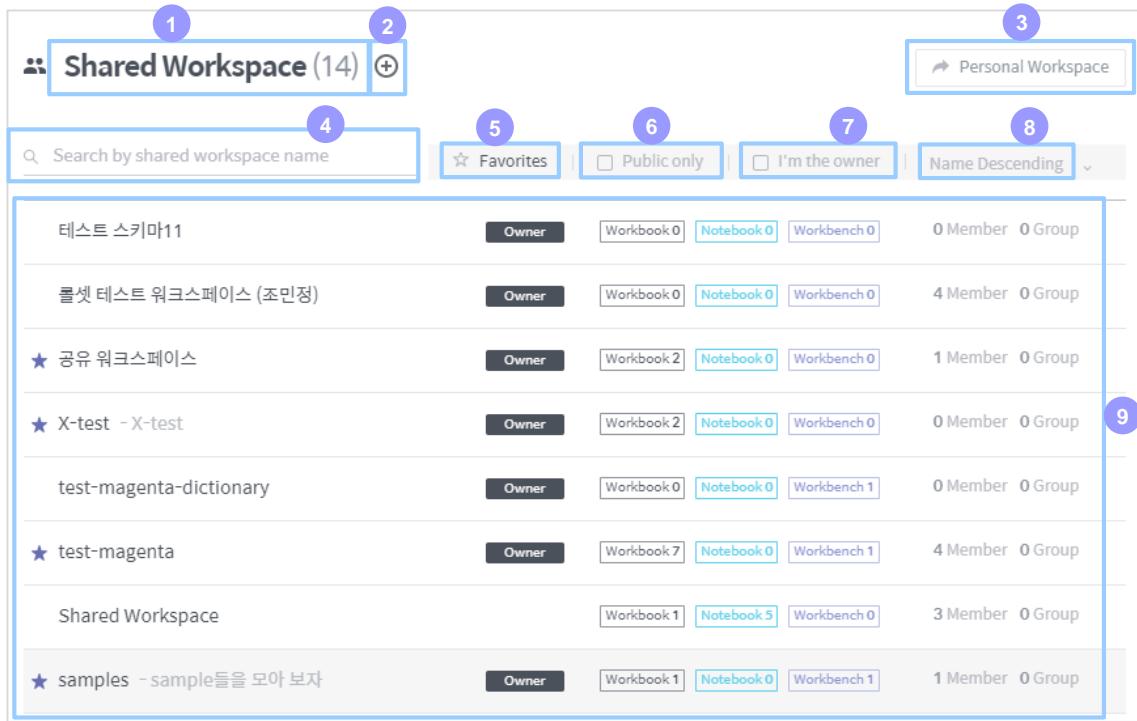
- 1 **Select all:** Select all items in the current folder list and entity list.
- 2 **Clone Workbook:** This is an exclusive function for the workbook. Click this button to clone the selected workbook.
- 3 **Move selections:** Move the selected folder or entity. In the case of workbook, a user can move it to another workspace. For other items, the user can move it to another folder in the same workspace. However, it is impossible to move it when the workbook is selected with another entity.
- 4 **Delete:** Delete the selected folder or entity.

3. Shared Workspace List

The shared workspace list screen is used to view the list of all shared workspaces accessible by the logged-in user and to move to a specific workspace. This screen can be accessed via two methods.

- Click the  button at the top-left of the Discovery screen to open the main panel, and click 'Workspace list >'.
- Click 'Workspace List' at the top-right of the workspace home screen.

The shared workspace list screen is composed as follows.



The screenshot shows a list of shared workspaces. Each workspace entry includes the workspace name, owner status, and counts for Workbooks, Notebooks, and Workbenches, along with member and group counts. A vertical callout number 9 points to the bottom right corner of the list area.

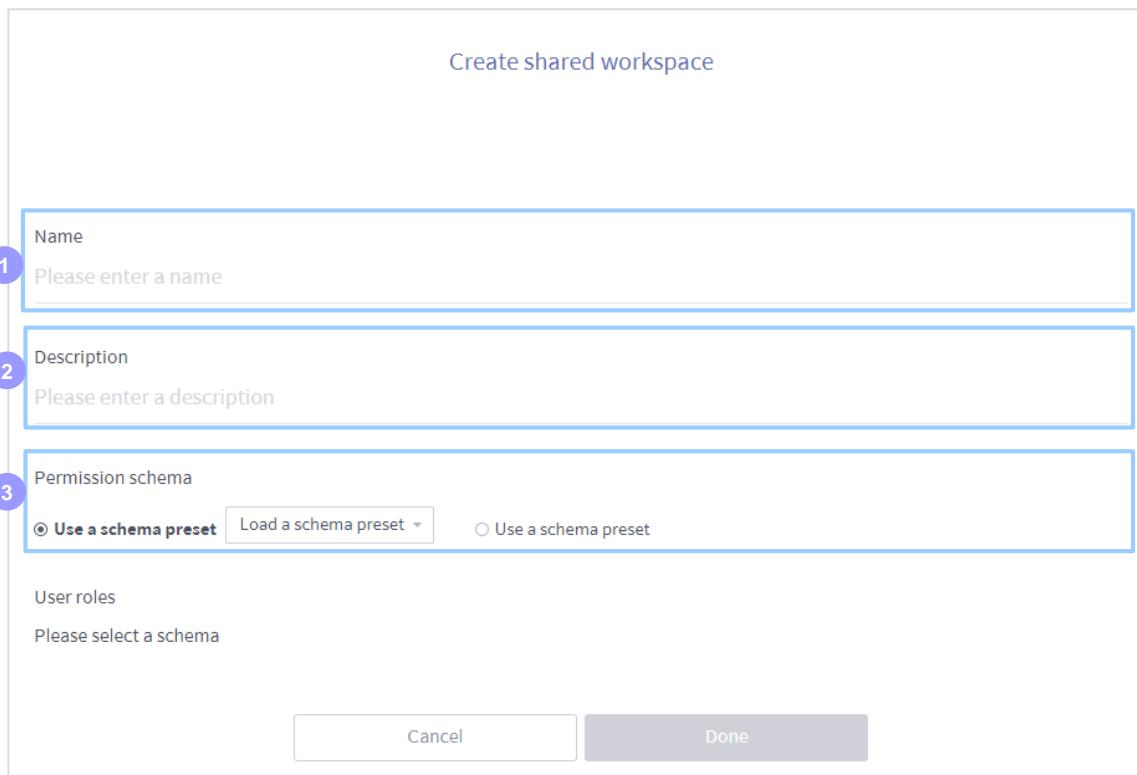
Workspace Name	Owner	Workbooks	Notebooks	Workbenches	Members	Groups
테스트 스키마11	Owner	Workbook 0	Notebook 0	Workbench 0	0 Member	0 Group
콜셋 테스트 워크스페이스 (조민정)	Owner	Workbook 0	Notebook 0	Workbench 0	4 Member	0 Group
★ 공유 워크스페이스	Owner	Workbook 2	Notebook 0	Workbench 0	1 Member	0 Group
★ X-test - X-test	Owner	Workbook 2	Notebook 0	Workbench 0	0 Member	0 Group
test-magenta-dictionary	Owner	Workbook 0	Notebook 0	Workbench 1	0 Member	0 Group
★ test-magenta	Owner	Workbook 7	Notebook 0	Workbench 1	4 Member	0 Group
Shared Workspace		Workbook 1	Notebook 5	Workbench 0	3 Member	0 Group
★ samples - sample들을 모아 보자	Owner	Workbook 1	Notebook 0	Workbench 1	1 Member	0 Group

- ① **Number of shared workspaces:** Displays the number of shared workspaces in the list.
- ② **Add a shared workspace:** Click this button to move to the screen to add a shared workspace. (See Chapter 4 for more information on how to add.)
- ③ **Move to a personal workspace:** Click this button to move to the personal workspace of the logged-in user.
- ④ **Search:** Search a shared workspace by its name.
- ⑤ **Favorites:** Sort only workspaces designated as favorites.
- ⑥ **Anonymous only:** Sort only workspaces set as public.
- ⑦ **I'm the manager:** Displays a list of workspaces where the logged-in user is the administrator.
- ⑧ **Name ascending/descending:** Sort names of a shared workspace in ascending/descending order.
- ⑨ **Workspace list:** Displays workspaces that meet the established sort criteria. Click a workspace to move to that workspace.

4. Create a Shared Workspace

A new shared workspace is created as follows.

1. Click the  button in the shared workspace list to open the screen to create a new shared workspace.
2. Refer to the descriptions below and fill in the entries.



The screenshot shows the 'Create shared workspace' dialog box. At the top, it says 'Create shared workspace'. Below that are three input fields numbered 1, 2, and 3. Field 1 is for 'Name' with placeholder text 'Please enter a name'. Field 2 is for 'Description' with placeholder text 'Please enter a description'. Field 3 is for 'Permission schema' with two options: ' Use a schema preset' and ' Load a schema preset'. Below these fields is a section for 'User roles' with placeholder text 'Please select a schema'. At the bottom are 'Cancel' and 'Done' buttons.

- 1 **Name:** Enter a name for the shared workspace.
- 2 **Description:** Enter a description of the shared workspace.
- 3 **Workspace permission schema:** Set the permission schema for each role of the shared workspace.
 - Use a schema preset: Import the permission schema defined by the administrator.
 - Use a custom schema: Define a new permission schema. (For more information on how to define a new permission schema, see Paragraph 5.1.1.)
3. Click 'Done' to finish creating a workspace.

5. Set Access Permission for a Shared Workspace

Setting the access permission for a shared workspace is composed of the following two steps.

- Set the access permission for each user role (See Paragraph 5.1 'Set permission schema'.)
- Grant a suitable user role for each user or user group (See Paragraph 5.2 'Set shared member & group'.)

5.1 Permission schema

5.1.1 Search permission schema

Click the  icon at the top-right of the shared workspace home screen and click 'Set permission schema' to view the defined permission schema as follows.

Set permission schema												
User roles of 테스트 워크스페이스 (조인정)												
User role	Default role	Workbook			Notebook			Workbench			Workspace	
		View	Create	Edit any	View	Create	Edit any	View	Create	Edit any	Create folders	Set config.
Manager	<input type="radio"/>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor	<input checked="" type="radio"/>	✓	✓	-	✓	✓	-	✓	✓	-	✓	-
Watcher	<input type="radio"/>	✓	-	-	✓	-	-	✓	-	-	-	-

In the above example, Manager, Editor and Watcher are defined as user roles. As shown in this example, 'permission schema' is a group of user roles where an individual's access permission is defined to each role.

Attributes of each column for each user role are as follows.

Default role

For the new user or user group, the user role designated as a default role is granted.

Permission for each type of workbook/notebook/workbench entity

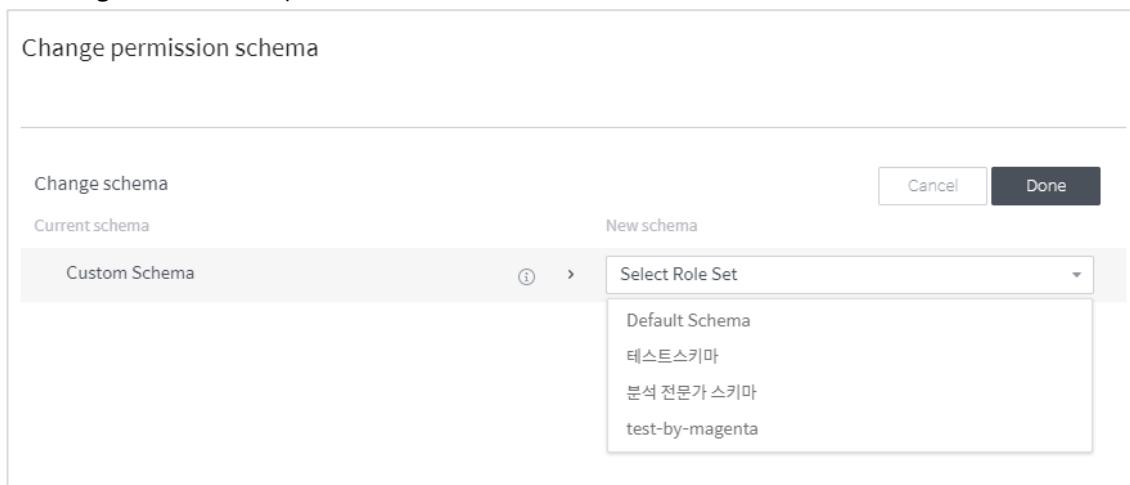
- View: A user can view data by accessing the entity of the relevant type.
- Create: A user can create, edit or delete entities of the relevant type.
- Edit any: A user can edit or delete entities of the relevant type created by another user.

Workspace permission

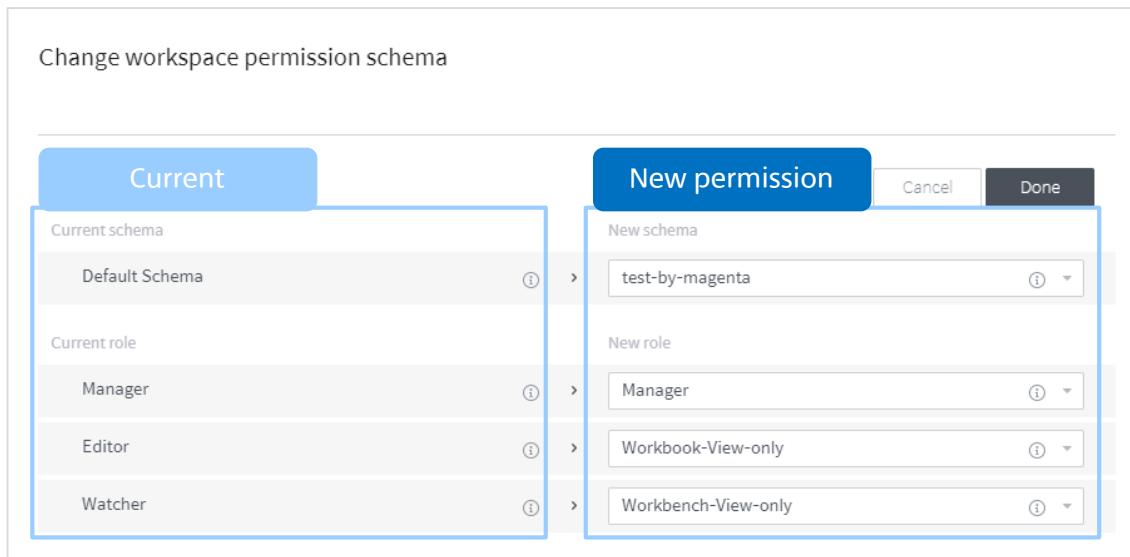
- Create folders: A user can create, edit or delete folders in the workspace.
- Set config.: A user can edit the name and description of the workspace, and change the workspace permission schema.

5.1.2 Change permission schema setting

Click the 'Change schema' button in the permission schema search screen to display the screen to change the defined permission schema as follows.



Click 'Select Role Set' combobox on the right to display schemata defined by the administrator. Use 'Custom RoleSet' at the bottom of the list to set a new user role. Select one to display the following screen. (If you select 'Custom RoleSet', you must define the permission for each user role first. Click the button to move to the permission setting screen, and set the permission by user role referencing the descriptions in Paragraph 5.1.1.)

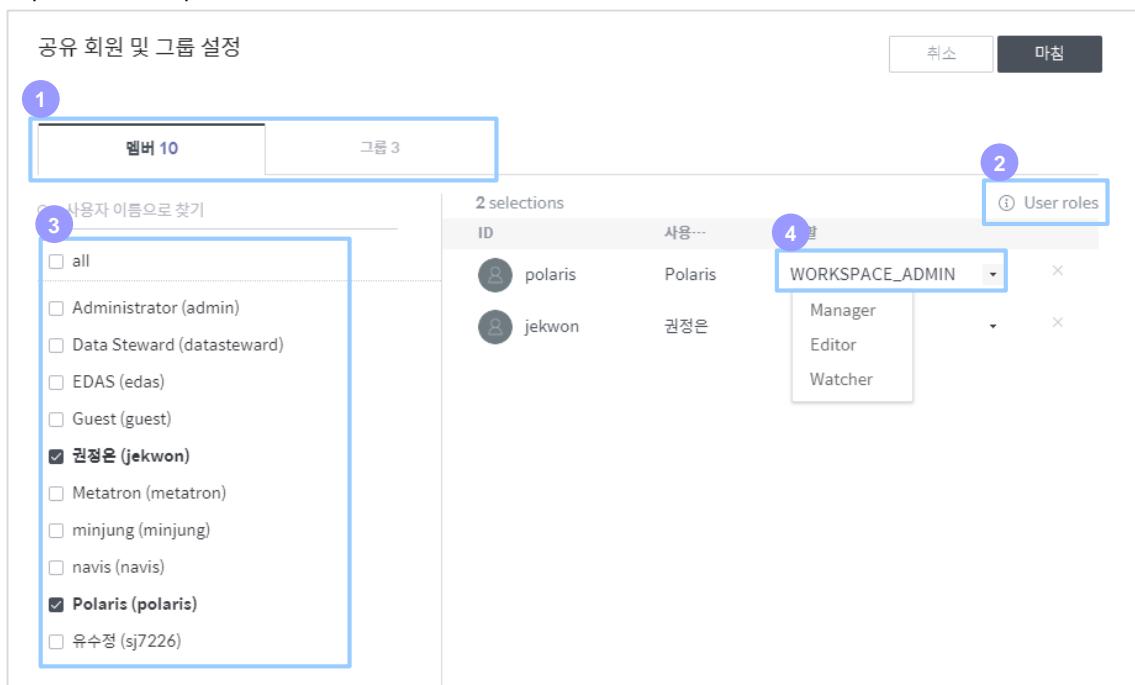


In this example, each user role of the current permission schema is substituted with the user role defined in the new permission schema. Hover the mouse over the icon next to the name of each user role to display the permission assigned to the user role.

Click 'Done' to finish setting the permission schema.

5.2 Set shared member & group

Click the icon at the top-right of the shared workspace home screen, and click 'Set shared member & group' to display the screen to set a shared member and group as follows. In this example, each user role defined in the permission schema is assigned to each user or user group. Assign the user role referencing the following description, and click 'Done' to finish setting workspace access permission.



1 Select unit of user role assignment

- Member tab: Assign a user role to each user.
- Group tab: Assign a user role to each user group. (A user group can be granted administrator permission.)

2 User roles: Click to display the information for the permission schema (the definition of permission by user role) as a pop-up window.

3 Member/group list: Lists users (groups in the case of the group tab) registered in Discovery. Click a user (group) in the list to add it to the role assignment area on the right. Click an added user (group) to remove it from the area on the right.

4 Assign user role: Click this combobox to display user roles defined in the active permission schema. Select a role to be assigned to the user (group).



metatron 2.0

Part 4

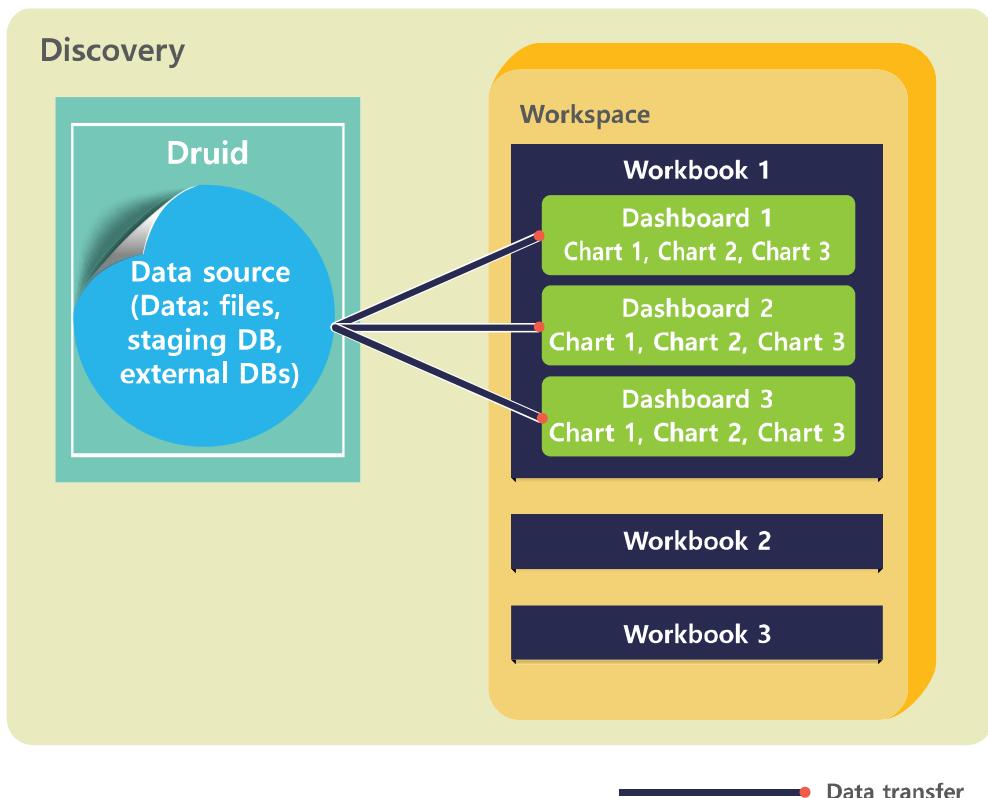
Using Workbook



Contents

1. Workbook Outline	3
2. Create a Workbook	4
3. Create a Dashboard.....	6
4. Manage a Workbook	11
4.1 Dashboard list.....	12
4.2 Dashboard detail view	14
4.2.1 Dashboard - Basic screen	14
4.2.2 Dashboard - data source information dialog box.....	17
4.2.3 Dashboard - presentation view screen.....	19
4.2.4 Dashboard - Edit screen.....	20
5. Create/Manage Chart.....	28
5.1 Overview of the chart home screen	28
5.2 Data column list	29
5.2.1 Composition of data column list.....	29
5.2.2 Add a custom column.....	30
5.2.3 Dimension and measure.....	31
5.3 Pivoting.....	32
5.3.1 What is 'pivoting'?	32
5.3.2 The concept of column/row/cross shelves	33
5.4 Chart type	35
5.5 Chart filter	39
5.5.1 Filters included automatically.....	39
5.5.2 Chart filter panel	40
5.5.3 Chart filter dialog box	41
5.6 Chart style setting	46
5.6.1 Chart style setting menu	46
5.6.2 'Common Setting' items by chart type.....	49

1. Workbook Outline



Workbook is the visual data analysis module based on the metatron Discovery engine, Druid. The main features are as follows.

- It provides fast and flexible data analysis using a time series-based multi-dimensional data source.
- Each dashboard contains visualization widgets for various charts and texts, so users can utilize it as a report for presentations.
- Frequently used algorithms such as clustering, prediction line, or trend line can be implemented as GUI.

This part is divided as follows.

Chapter 2 Create a Workbook: 'Workbook' functions as an independent report. This chapter describes how to create a new workbook.

Chapter 3 Create a Dashboard: Dashboard means each slide of workbook. This chapter describes how to create a new dashboard in workbook, and to link that dashboard to a data source.

Chapter 4 Manage a Workbook: This chapter describes how to sort a list of dashboards that compose a workbook and to configure various widgets in each dashboard.

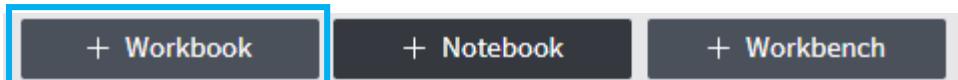
Chapter 5 Create/manage a Chart: A chart is another widget that makes up the dashboard, but it has many unique concepts and UI menus, so it is necessary to describe only the chart in this part.

2. Create a Workbook

In metatron Discovery, 'workbook' functions as an independent data analysis report. Once a workbook is created, a user can store many 'dashboards' in the workbook and present them in the proper order.

A workbook is created as follows.

1. Click the '+ Workbook' button at the bottom of the workspace. The screen used to create a workbook will be displayed.



2. Enter a name (mandatory) and description of the workbook to be created and click 'Done'.

With the 'Continue to create a dashboard of a new workbook' checkbox selected, the creation of a workbook proceeds directly to 'Create a Dashboard' page. This page is required because a workbook cannot work without dashboards in it.

The screenshot shows a dialog box titled 'Create a Workbook'. It contains two input fields: 'Name' with placeholder text 'Please enter a name' and 'Description' with placeholder text 'Please enter a description'. Below these fields is a checked checkbox labeled 'Continue to create a dashboard of a new workbook'. At the bottom are two buttons: 'Cancel' and 'Done'.

Create a Workbook

Name
Please enter a name

Description
Please enter a description

Continue to create a dashboard of a new workbook

Cancel Done

3. For more information on how to create a dashboard, see Chapter 3.

The screenshot shows a 'Create a Dashboard' interface. At the top, there's a title 'Create a Dashboard' with a 'Cancel' button. Below it is a message 'Please select a datasource'. A search bar labeled 'Search by datasource name' is followed by a table listing data sources. The table has columns: No., Datasource, Type, Used in, and Updated. A filter 'Only view open data' is checked, and a dropdown 'Type' is set to 'All'. The data sources listed are:

No.	Datasource	Type	Used in	Updated
32	salesqa [Open data]	Collection type	All Workspaces	2018-05-15
31	StateCrimeRates [Open data]	Collection type	All Workspaces	2018-05-15
30	null_test [Open data]	Collection type	All Workspaces	2018-05-15
29	sales - sales data (2011-2014) [Open data]	Collection type	All Workspaces	2018-05-15
28	sale_lookup_sample [Open data]	Collection type	All Workspaces	2018-05-15
27	Excel-test - Excel-test [Open data]	Collection type	All Workspaces	2018-05-15
26	test-hive - test-hive	Collection type	1 Workspaces	2018-05-15
25	고트코운기과 [Open data]	Collection type	All Workspaces	2018-05-15

At the bottom, there are 'More' and 'Next' buttons.

4. A user can check the new workbook in the workspace screen as follows. Click that workbook to see a screen where you can use the screen.

The screenshot shows the 'METATRON DISCOVERY' workspace screen for 'Admin workspace'. It displays a grid of workbooks. The top navigation includes 'Owner', 'Workspace List', and a creation date 'Created on 2017-11-16 by 권리자'. The workspace contains 57 workbooks, 14 notebooks, and 42 workbenches. The workbooks are organized into several categories:

- Folders:** Admin workspace, edas, J, SH, crime.
- Workbooks:**
 - workbook • New: test (Last updated a minute ago)
 - workbook: 차트 프리젠테이션 0112 (Last updated 43 minutes ago)
 - workbook: Sales Workbook (Last updated a day ago)
 - workbook: test (Last updated 4 days ago)
 - workbook: test-magenta (Last updated 4 days ago)
 - workbook: 你好 (Last updated 12 days ago)
- Notes:** 0, 2, 3, 6.
- Workbenches:** 1, 2, 4, 1.
- Others:** 시연용(변경금지), 엑셀, 서브폴더.

At the bottom, there are buttons for 'Select all', 'Clone Workbook', 'Move selections', 'Delete selections', and '+ Workbook', '+ Notebook', '+ Workbench'.

3. Create a Dashboard

A dashboard is contained in a workbook, and provides functions to analyze and visualize a specific data source based on the customer's needs. Therefore, an important step to create a dashboard is linking a data source.

A dashboard is created as follows.

1. From the list of data sources disclosed in the workspace, select a data source to be linked to the dashboard. Only one root data source can be selected, and the user can select another data source to be joined to the subsequent screen.

The screenshot shows the 'Create a Dashboard' interface. At the top, there is a search bar labeled 'Search by datasource name' (1) and a checkbox 'Only view open data' (2). A dropdown menu 'Type' is set to 'All' (3). Below these are two tabs: 'Open data' (4) and 'Collection type' (5). The main area displays a list of data sources:

No.	Datasource	Type	Used in	Updated
32	salesqa [Open data]	Collecti...	All Workspaces	2018-05-15
31	StateCrimeRates [Open data]	Collecti...	All Workspaces	2018-05-15
30	null_test [Open data]	Collecti...	All Workspaces	2018-05-15
29	sales - sales data (2011~2014) [Open data]	Collecti...	All Workspaces	2018-05-15 ✓
28	sale_lookup_sample [Open data]	Collecti...	All Workspaces	2018-05-15
27	Excel-test - Excel-test [Open data]	Collecti...	All Workspaces	2018-05-15
26	test-hive - test-hive	Collecti...	1 Workspaces	2018-05-15
25	고트고우기과 [Open data]	Collecti...	All Workspaces	2018-05-15

Below the list is a 'More' button. To the right, a detailed view of the selected data source 'sales' is shown (6), listing its properties and dimensions/measures. At the bottom are 'Cancel' and 'Next' buttons.

- 1 **Search by data source name:** Search a data source allowed in the relevant workspace by its name.
- 2 **View only open data:** Display only a data source designated as an 'open data source'.
- 3 **Type:** Display connection type or collection type data sources.
- 4 **Data source list:** Displays data sources that meet the established criteria.
- 5 **Data source information:** Displays an overview of the data source selected in the list.

2. Refer to the description below and add other data sources to be joined to the top data source selected above. Click 'Next' if you get the result you want.

Basic screen

The screenshot shows the 'Create a Dashboard' interface. At the top, there's a header 'Create a Dashboard' with three small circular icons. Below it is a 'Data source tree' section labeled '1'. This section shows a connection between two data sources: 'sales' (represented by a bar chart icon) and 'sale_lookup_sample' (represented by a table icon). There are three icons for managing the join relationship: a pencil icon for edit, a checkmark icon for edit, and a trash bin icon for delete. To the right of the tree, there are buttons '+ Add a datasource to join' and '+ Add a datasource to join'. Below the tree is a large table labeled 'Result data' labeled '2'. The table has 31 columns and 1000 rows. The columns include sales.OrderDate, sales.Category, sales.City, sales.Country, sales.CustomerName, sales.OrderID, sales.PostalCode, and sales.ProductName. The table contains several rows of data, such as Darren Powers from Houston, United States, and Mick Brown from Philadelphia, United States. At the bottom of the table are 'Previous' and 'Next' navigation buttons.

sales.OrderDate	sales.Category	sales.City	sales.Country	sales.CustomerName	sales.OrderID	sales.PostalCode	sales.ProductName
2011-01-04 00:00:00	Office Supplies	Houston	United States	Darren Powers	CA-2011-103800	77095	Message Book Wirebo...
2011-01-05 00:00:00	Office Supplies	Naperville	United States	Phillina Ober	CA-2011-112326	60540	Avery 508
2011-01-05 00:00:00	Office Supplies	Naperville	United States	Phillina Ober	CA-2011-112326	60540	GBC Standard Plastic ...
2011-01-05 00:00:00	Office Supplies	Naperville	United States	Phillina Ober	CA-2011-112326	60540	SAFCO Boltless Steel S...
2011-01-06 00:00:00	Office Supplies	Philadelphia	United States	Mick Brown	CA-2011-141817	19143	Avery Hi-Liter EverBol...
2011-01-07 00:00:00	Furniture	Henderson	United States	Maria Etezadi	CA-2011-167199	42420	Global Deluxe High-Ba...
2011-01-07 00:00:00	Office Supplies	Athens	United States	Jack OBriant	CA-2011-166054	30605	Dixon Prang Watercolo...
2011-01-07 00:00:00	Office Supplies	Henderson	United States	Maria Etezadi	CA-2011-167199	42420	Alliance Super-Size Ba...

1 **Data source tree:** Displays join relationships between data sources as a tree structure.

- Edit the join relationship.
- Edit the join relationship. (Same function of the icon)
- Delete the join relationship.
- Click '+ Add a datasource to join' to display a dialog box to join a new data source. (For more information on this dialog box, see the description on the next page.)

2 **Result data:** Displays the results table of joining a data source.

Data join dialog box

The screenshot shows the 'Join' dialog box with the following interface elements:

- Master datasource:** Displays the 'sales' table with columns: OrderDate, Category, City, Country. Rows include: 2011-01-04 00:00:00, Office Supplies, Houston, United States; 2011-01-05 00:00:00, Office Supplies, Naperville, United States.
- Datasource to join:** Displays the 'safe_lookup_sample' table with columns: City, user_id, current_datetime. Rows include: Henders... test03 2018-01-24 00:00:00; Naperville test01 2018-01-24 00:00:00.
- Join key configuration:** A grid where you can map columns from both datasets. Step 3 shows 'Column' selected for both rows.
- Join type:** Options include Inner, Left, Right, and Full outer. Step 4 highlights the 'Left' option.
- Preview results:** Shows the joined data with 31 columns and 1000 rows. Step 5 highlights the preview area.

- 1 Master data source:** Displays information on the master data source of the new data source to be joined.
- 2 Data source to be joined:** Select a data source to be joined to the master data source.
- 3 Add a join key:** 'Join key' is a key to define the join relationship between the master data source and the data source to be joined. Select a column to be joined from each data source, and click this button to add a new join key. In this case, the data type defined in the column of each data source must be identical.
- 4 Join type:** Select how to join and transform a data source. To help you understand, each join type is explained using the following tables as an example.

Master data source

Product name (join key)	Price
A	\$22.11
B	\$9.23
C	\$8.99
D	\$10.10

Data source to be joined

Product name (join key)	Sales
B	100
D	200
E	50

- Inner: Joins records with a join key present in both the master data source and the data source and includes them in the resulting table. (Intersection of two data sources)

Product name (join key)	Price	Sales
B	\$9.23	100
D	\$10.10	200

- Left: Imports data from the right data source (data source to be joined) based on the data value inside the join key column of the left data source (master data source to join), and shows it in the results table. Among the records in the right data source, records whose join key value are not present in the left data source are discarded.

Product name (join key)	Price	Sales
A	\$22.11	null
B	\$9.23	100
C	\$8.99	null
D	\$10.10	200

- Right: Imports data from the left data source (master data source to join) based on the data value inside the join key column of the right data source (data source to be joined), and shows it in the results table. Among the records in the left data source, records whose join key value are not present in the right data source are discarded.

Product name (join key)	Price	Sales
B	\$9.23	100
D	\$10.10	200
E	null	50

- Full Outer: Joins by importing all data of both data sources based on the data value inside the join key column and includes them in the resulting table. (Union of two data sources)

Product name (join key)	Price	Sales
A	\$22.11	null
B	\$9.23	100
C	\$8.99	null
D	\$10.10	200
E	null	50

- 5 **Preview results:** Displays the result value of data source joining.

3. Confirm information on the data source imported to create a dashboard, enter a name and description, and click 'Done' to create a new dashboard.

Create a Dashboard

Please complete dashboard creation

Workbook	test
Datasource	sales

Name
mydashboard

Description
Please enter a description

Previous Done

4. The new dashboard will be added to the home screen of the workbook. Click this to display the screen of the relevant dashboard.

METATRON DISCOVERY

← Workbook test : mydashboard Please enter dashboard description. Updated on 2019-05-15 20:23 by 관리자 Presentation view Edit dashboard

Dashboard(1) | Comment(0) |

Search

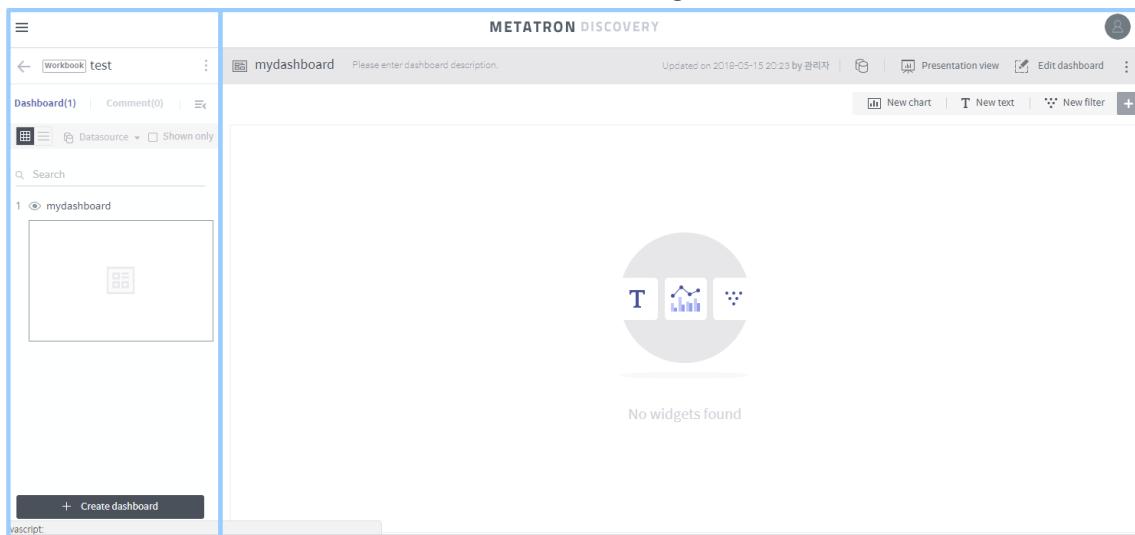
1 mydashboard

No widgets found

+ Create dashboard

4. Manage a Workbook

Select a workbook from the list on the workspace screen to display the home screen of the workbook. On the workbook home screen, the user can view, create, and edit the dashboard. In addition, the user can display various dashboards in slide show mode via presentation mode. The home screen of the workbook is divided into the following two areas.



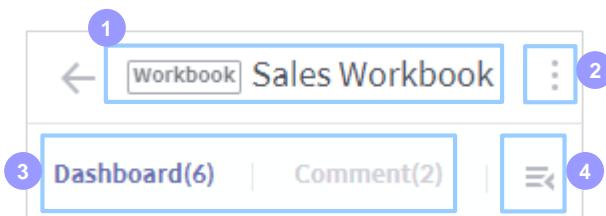
- **Dashboard list:** Used to configure the workbook and manage dashboard lists. (See Paragraph 4.1.)
- **Dashboard detail view** Displays detailed information on the dashboard selected in dashboard list area. In this area, a user can analyze, visualize and sort data through chart, text and filter widgets. (See Paragraph 4.2.)

4.1 Dashboard list

The dashboard list area can be switched into 'Dashboard Mode' or 'Comment Mode'. Each mode provides the following function.

- Dashboard: The user can add a new dashboard or list the registered dashboards.
- Comment: The user can add a new comment or list the registered dashboards. Comments can be written and viewed by all users who have access to the workbook.

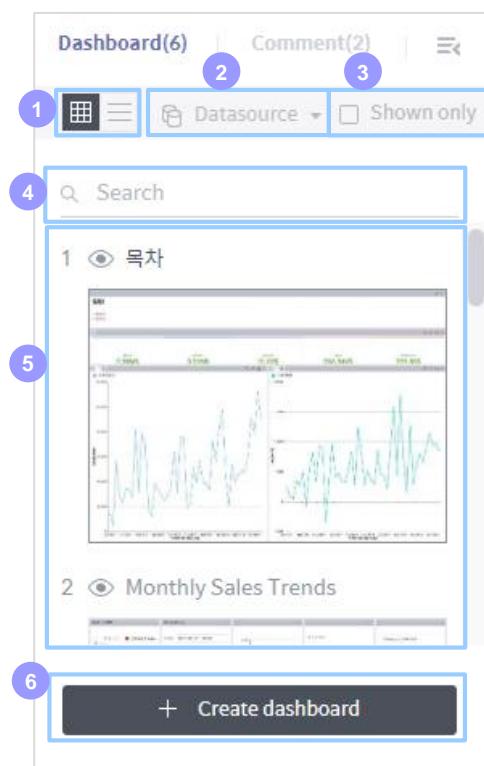
Dashboard/comment joint area



- ① **Name:** Name of the workbook.
- ② **More:** Click to edit the name/description of the workbook or delete the workbook. Also, it is possible to check the time of the latest update and creation of the workbook.
- ③ **Select dashboard or comment mode:** Select dashboard or comment mode as the mode to be displayed. The number in parentheses next to each selection indicates the number of dashboards/comments registered in the workbook.
- ④ **Fold:** Click this button to collapse or expand the dashboard list.

Dashboard mode

- ① **View format:** Select format to display a dashboard list. The user can select Thumbnail view or List view.
- ② **Data source:** Used to check data sources used in the active workbook. Check or uncheck the check box of each data source to search only dashboards using a specific data source.
- ③ **Shown only:** Search only dashboards set as Shown.
- ④ **Search Formula:** Search a dashboard by its name.
- ⑤ **Dashboard list:** Displays the dashboard list that meets search criteria. Hover the mouse over a dashboard to copy or delete that dashboard.



- 6 Create dashboard:** Click this button to display the screen to create a new dashboard. (See Chapter 3 for more information on how to create it.)

Comment mode

- 1 Comment list:** Displays comments registered in the workbook in order of latest registration date.
- 2 Add a Comment:** Create a new comment.
Press Enter to register the entered comment.
Press Shift + Enter to insert the line break.

나 2018.05.15 20:25
고객에 관련된 설명은 이 정도면 충분합니다.

나 2018.05.15 20:25
매출 측면에서 보다 상세한 설명이 필요해보입니다.

2 Share your thought on workbook
(Enter key | Shift + Enter for new line)

4.2 Dashboard detail view

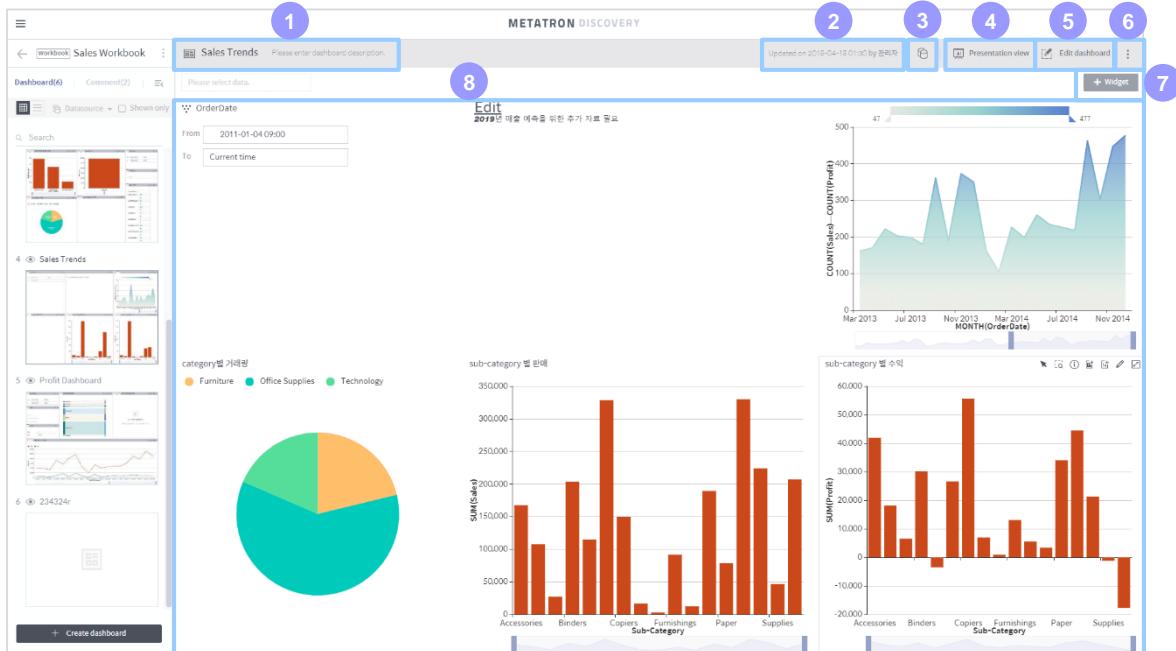
The dashboard detail view area displays detailed information on the dashboard selected in dashboard list area. In this area, a user can analyze, visualize and sort data through chart, text and filter widgets.

4.2.1 Dashboard - Basic screen

In the basic screen of the dashboard, a user can view charts registered in the dashboard and move to various screens to configure the dashboard.

A. Overall composition of the dashboard basic screen

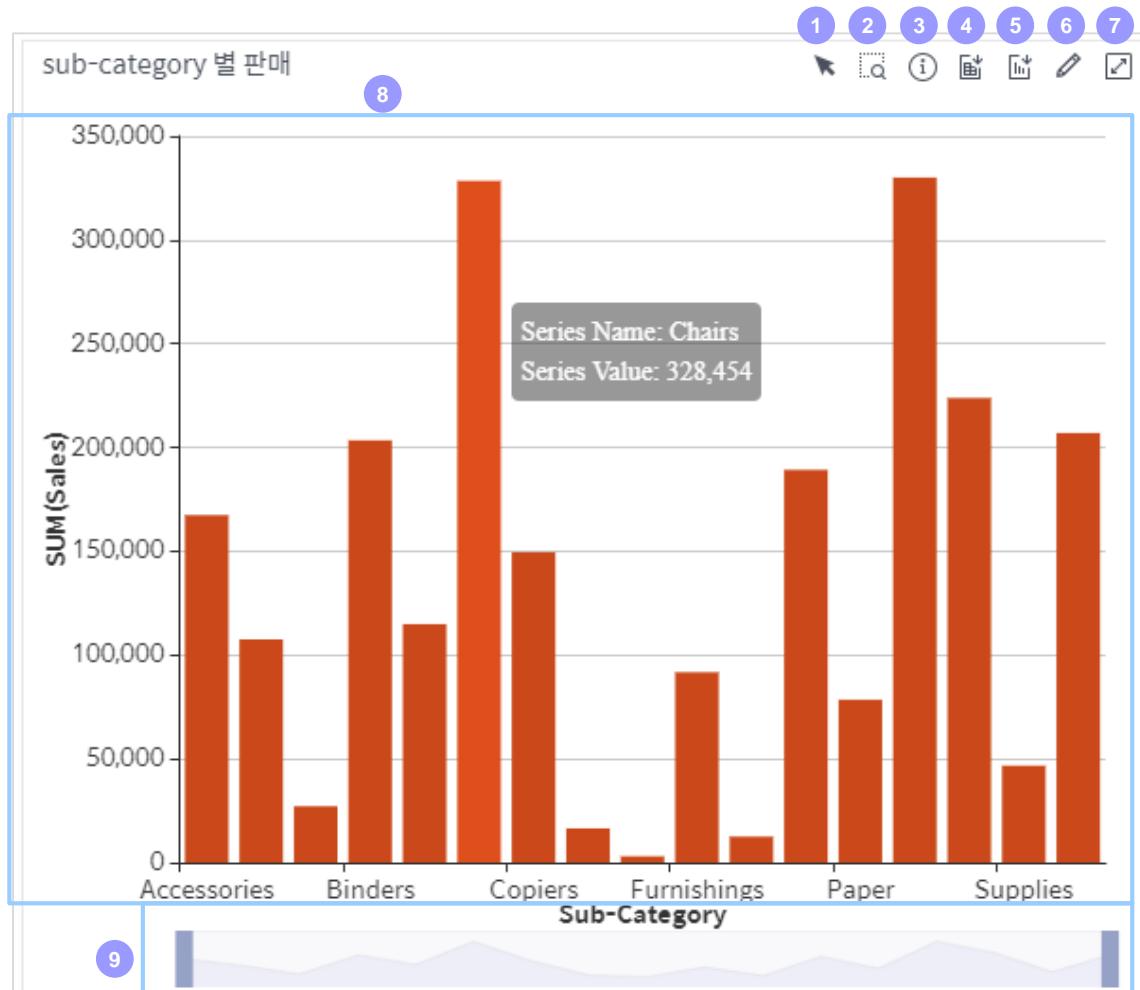
The following is a description of the overall composition of the basic screen of the dashboard.



- 1 **Name and description:** Name and description of the dashboard. Hover the mouse over to edit.
- 2 **Update information:** Displays the user name and time of the latest dashboard update.
- 3 **Data source:** Click to see information on the data source used in the dashboard as well as the related statistics/schema view. (For detailed information, see Paragraph 2.1.3 A)
- 4 **Presentation view:** Used to view the workbook with the appropriate UI for presentation. Click it to display a slide corresponding to the active dashboard. Also, the user can move to another dashboard. (See Paragraph 4.2.3.)
- 5 **Edit dashboard:** Click the button to display a screen to edit widgets on the dashboard. (See Paragraph 4.2.4.)
- 6 **More:** Click it to clone or delete the dashboard. Also, the user can check information on the recent modification and creation of the dashboard.
- 7 **Add a widget:** The menu to add a widget in the dashboard. Currently chart, text and filter widgets are available. (See Paragraph 4.2.4 D)
- 8 **Widget layout area:** Displays widgets configured to be shown in the layout screen of the dashboard.

B. Chart widget box in the dashboard basic screen

The following is a description of the chart widget box in the widget area. Hover the mouse over the box to display the widget's settings icons at the top right of the screen.



- ① **Select a data area:** Select how to choose a data item in a chart graph with the mouse cursor. Select a specific data item(s) to filter and display all charts in the dashboard based on the relevant dimension category.
- ② **Zoom in/out for chart:** Used to zoom in/out of the chart screen. Press the button to reset.
- ③ **Chart info:** Displays pivoted data information when creating the chart.
- ④ **Save data table:** Save the chart's data information as a local file.
- ⑤ **Save chart image:** Save an image of the chart as a jpg file.
- ⑥ **Edit:** Click to display a dialog box to edit the widget.
- ⑦ **Expand to full screen:** Expands the chart to the entire dashboard detail view area.
- ⑧ **Chart graph area:** Area to display the relevant chart graph. Select a specific data item(s) to filter and display all charts in the dashboard based on the relevant dimension category.
- ⑨ **Chart mini-map:** Displays a value distribution map of the chart by data category. Apply a filter to reduce the scope of the mini-map.

C. Text widget box in the dashboard basic screen

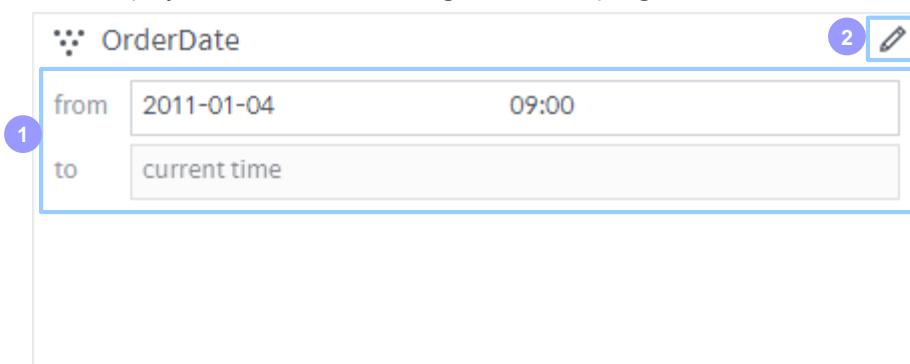
The following is a description of the text widget box in the widget area. Hover the mouse over the box to display icons to edit the widget at the top right of the screen.



- ① **Text area:** Area to display the written text.
- ② **Edit:** Click to display a dialog box to edit the widget.

D. Filter the widget box in the dashboard basic screen

The following is a description of the filter widget box in the widget area. Hover the mouse over the box to display icons to edit the widget at the top right of the screen.



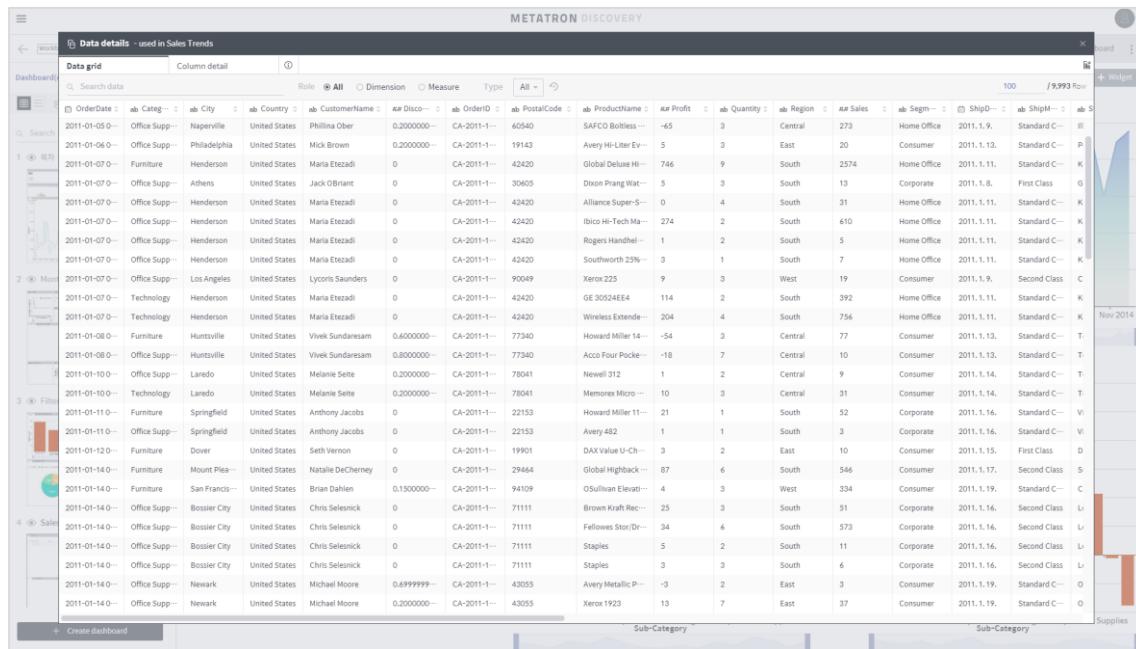
- ① **Filter area:** Area to display the configured filter. A user can edit the scope of the filter.
- ② **Edit:** Click to display a dialog box to edit the details of the widget.

4.2.2 Dashboard - data source information dialog box

Click the  button in the dashboard basic screen to display a dialog box displaying information of the data source used in the dashboard. This dialog box is composed of 3 tabs (Data grid, Column detail, Dashboard data information).

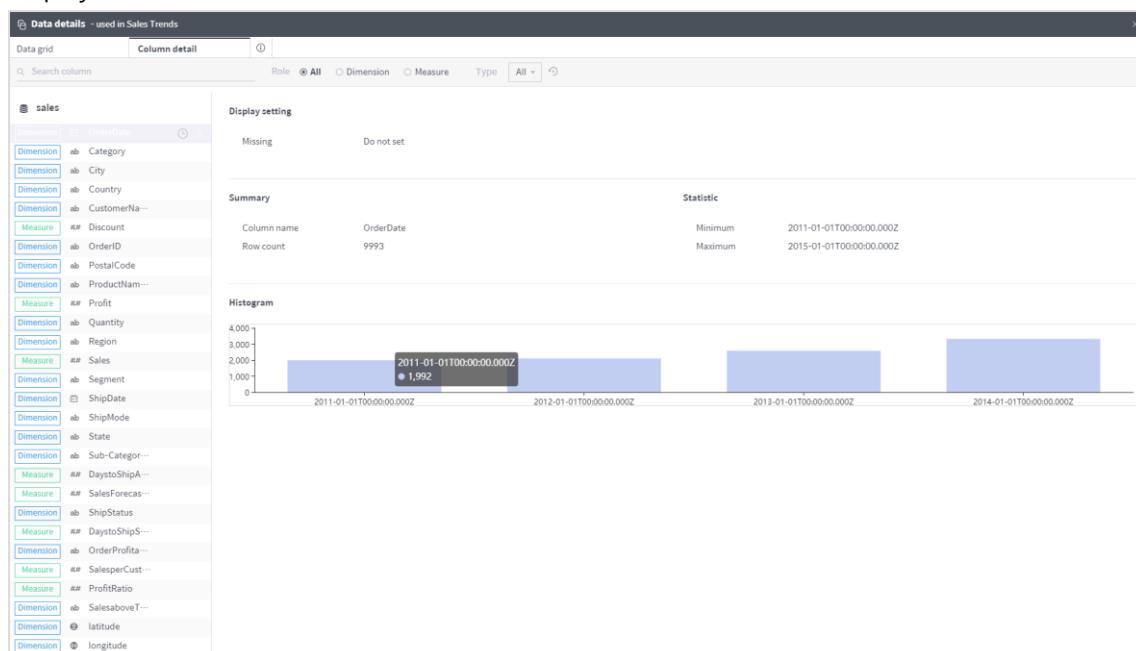
Data grid tab

Displays all record values of the data source.



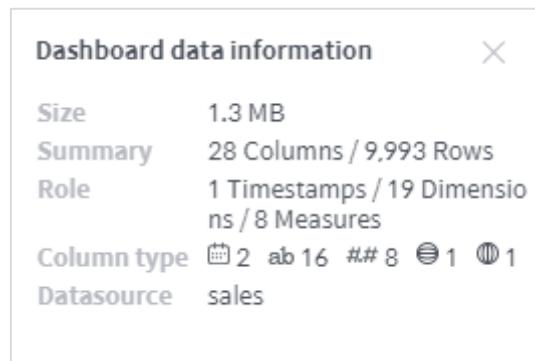
Column detail tab

Displays detailed information in each column of the data source.



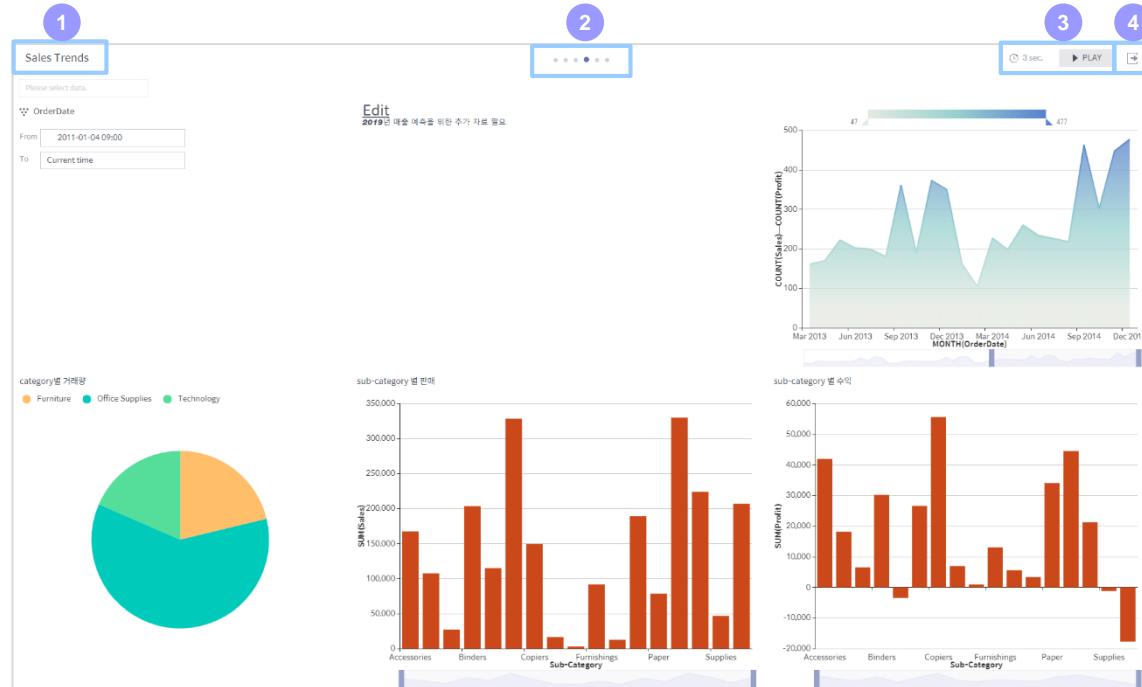
Dashboard data information tab

Displays an overview of the data source.



4.2.3 Dashboard - presentation view screen

Click 'Presentation View' in the dashboard basic screen to view workbook dashboards with the appropriate UI for presentation. With this UI, a user can easily report and share data analysis results.

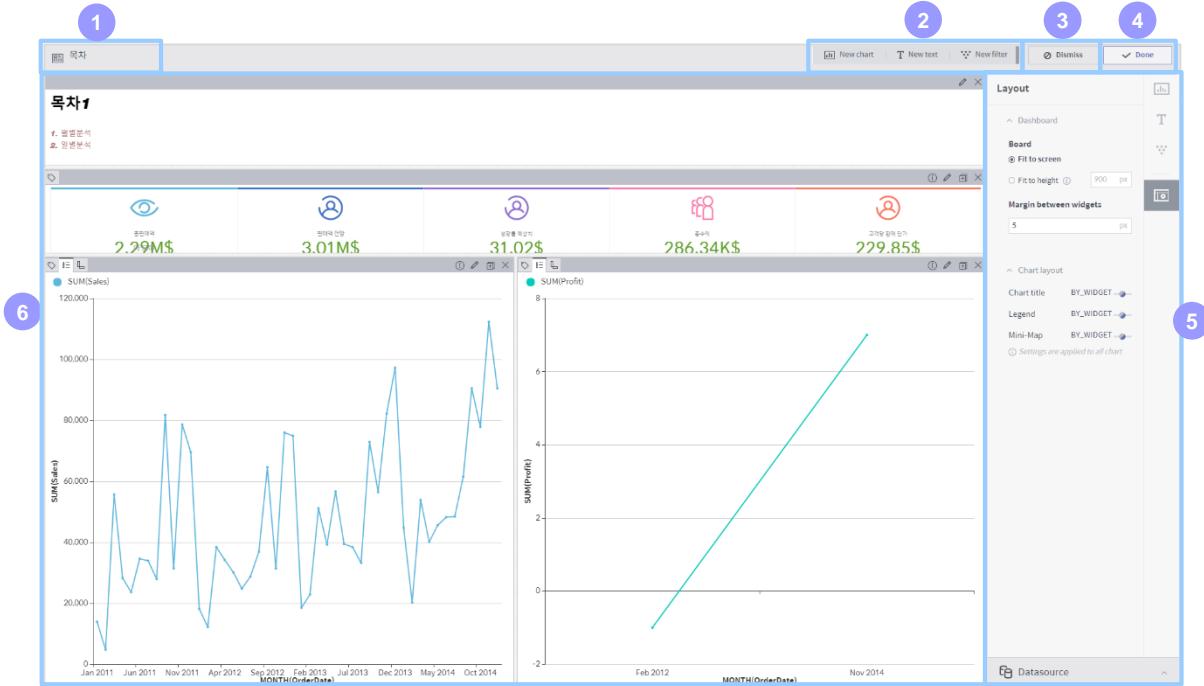


- ① **Name:** Name of the current dashboard.
- ② **Slide navigation:** Each circle represents each dashboard in the workbook. For example, if the user clicks the 4th circle, the 4th dashboard slide will be displayed and that circle will be highlighted.
- ③ **Auto slide show setting:** Select the time interval and click PLAY to start the auto slide show. Slides will be changed in the selected time interval.
- ④ **Exit:** Close the presentation view and return to the workbook/dashboard basic screen.

4.2.4 Dashboard - Edit screen

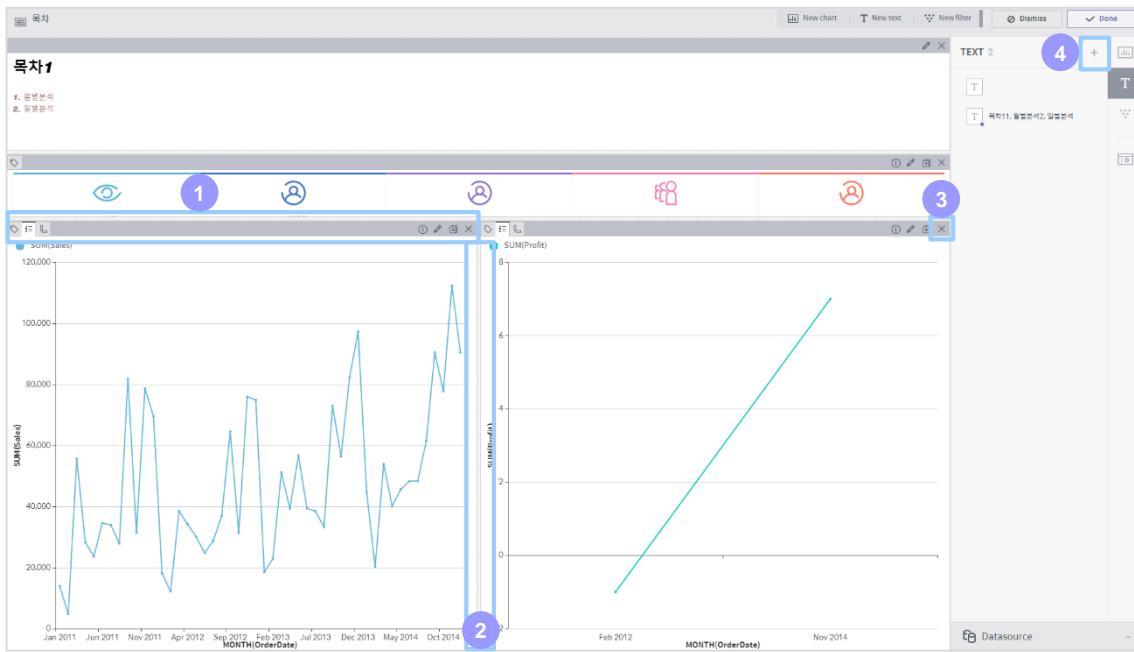
Click 'Edit Dashboard' in the dashboard basic screen to move to a screen to edit the composition of the dashboard. A user can add a widget, edit the dashboard, set the hierarchy and change the layout.

A. Overall composition of the dashboard edit screen



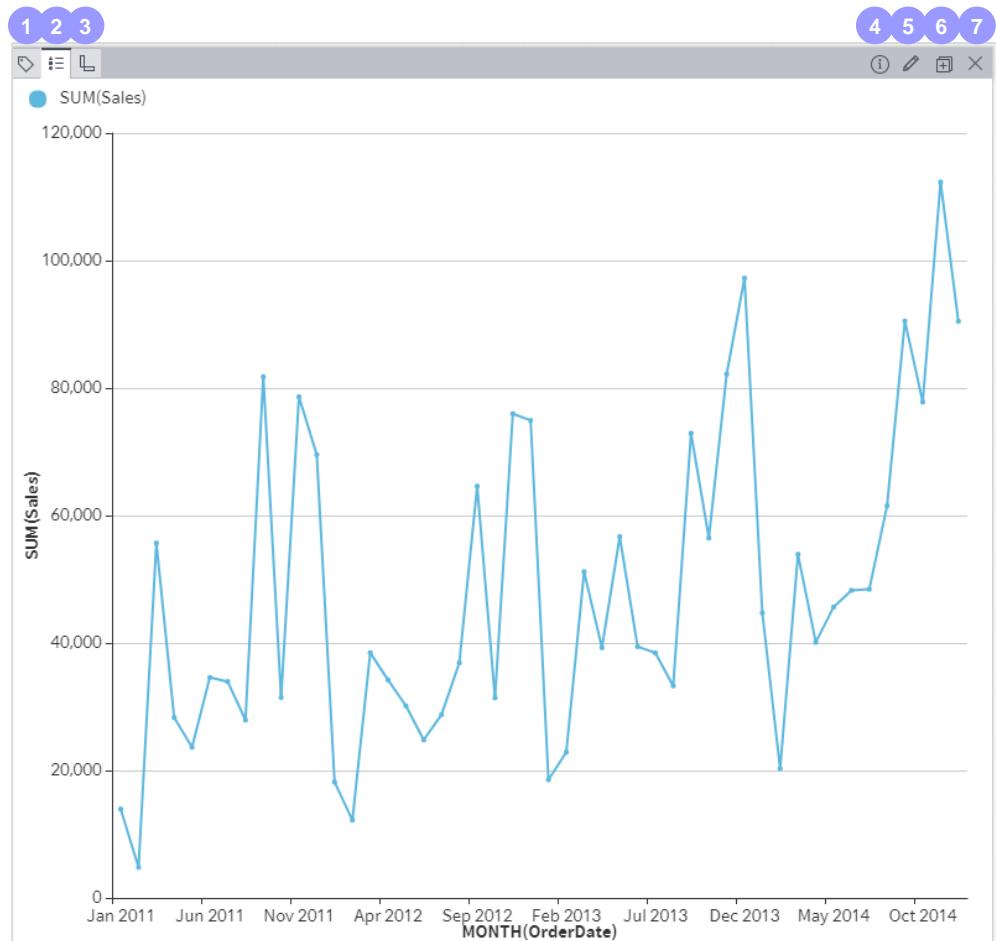
- 1 **Name:** Displays the name of the dashboard.
- 2 **Add widget menu:** The menu to add a widget such as chart, text and filter widgets.
- 3 **Dismiss:** Click to exit the current screen without saving changes.
- 4 **Done:** Click to save changes and exit the current screen.
- 5 **Widget layout area:** Edits the arrangement, display and attributes of widgets configured to be shown in the layout of the dashboard. (See Paragraph B and C for more information.)
- 6 **Panel area:** Used to add/edit/delete various widgets and edit settings related to how to display the dashboard. (See Paragraph D for more information.)

B. Widget arrangement setting



- 1 Change widget location:** Drag the title of a widget to change the location of the widget.
- 2 Adjust widget width:** Move the distance between widgets to adjust their width.
- 3 Add a widget to the screen:** Drag a widget from the widget list in the right panel to the left widget layout area. The widget will be added to the layout area.
- 4 Delete a widget from the screen:** Click the **X** button in a widget shown in the widget layout area. That widget will be deleted from the layout area.

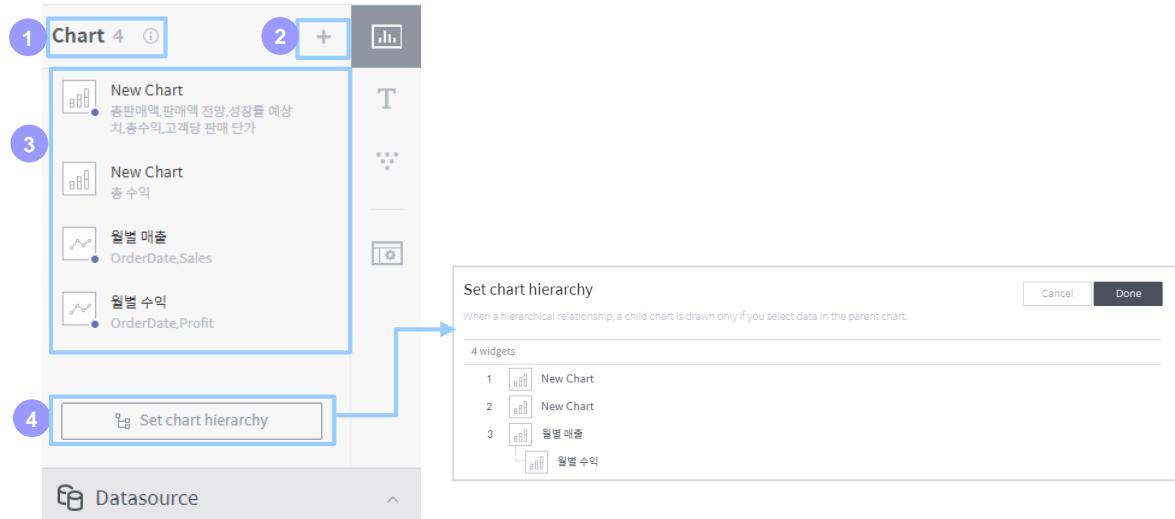
C. Individual widget edit area



D. Panel area in dashboard edit screen

Chart widget panel

In the chart widget panel, a user can add/edit/delete a chart in the dashboard.



- ① **Number of chart widgets:** Displays the number of chart widgets registered in the active dashboard.
- ② **Add a chart widget:** Used to create a new chart widget in the dashboard. (See Chapter 5 for more information on how to add it.)
- ③ **Chart widget list:** Lists chart widgets registered in the active dashboard. Hover the mouse over a widget to edit or delete it. The relevant icon will be displayed. Drag a widget to the widget layout area to display the widget in the layout area.
- ④ **Set chart hierarchy:** Used to set parent/child relationships of charts in the dashboard. Select a data item from the parent chart. The child chart will be filtered by that item. To set the hierarchy, drag a chart to be set as 'child' under its parent chart. Once the chart hierarchy is set, the user can check the modified structure in the chart menu.

Text widget panel

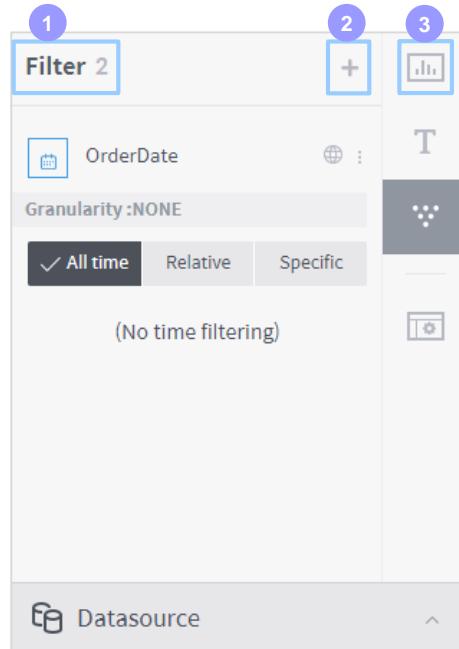
In the text widget panel, a user can add/edit/delete a text widget in the dashboard.



- ① **Number of text widgets:** Displays the number of text widgets registered in the active dashboard.
- ② **Add a text widget:** Used to create a new text widget in the dashboard.
- ③ **Text widget list:** Lists text widgets registered in the active dashboard. Hover the mouse over a widget to edit or delete it. The relevant icon will be displayed. Drag a widget to the widget layout area to display the widget in the layout area.

Filter widget panel

In the filter widget panel, a user can add/edit/delete a filter widget in the dashboard.

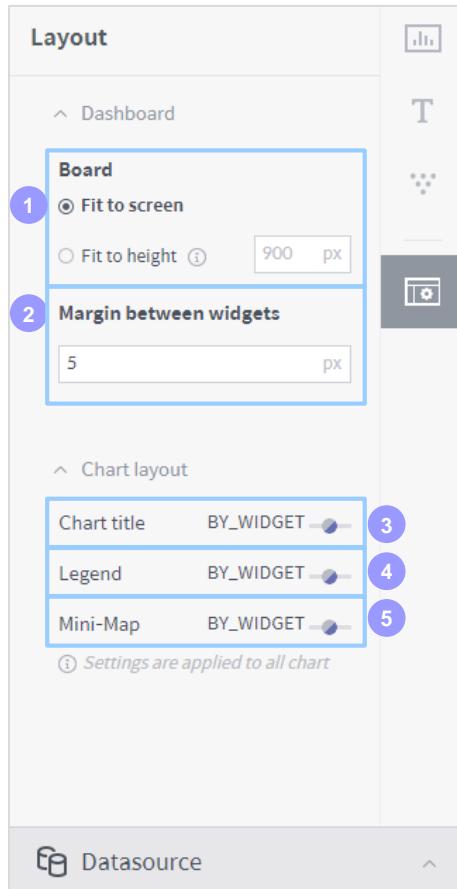


- ① **Number of filter widgets:** Displays the number of filter widgets registered in the active dashboard.
- ② **Add a filter widget:** Used to create a new filter widget in the dashboard.
- ③ **Filter widget list:** Lists filter widgets registered in the active dashboard. Hover the mouse over a widget to edit or delete it. The relevant icon will be displayed. Drag a widget to the widget layout area to display the widget in the layout area.

This panel has the same functions and structure of the chart filter panel in the chart home. For more detailed information, see Paragraph 5.5.2. Any filter created here is applied with the 'global' attribute, which makes the filter applied to all charts.

Layout panel

In the layout panel, a user can adjust some of the settings on how to arrange widgets and display an individual widget in the widget layout area.



1 Set board height

- Fix to screen: Match the height of the board to the screen.
- Fix to height: Set the height of the board as a specific pixel value.

2 Margin between widgets:

Used to set the margin between widgets shown in the widget layout area.

3 Chart title:

Used to set the title display of all chart and filter widgets in the widget layout area.

4 Legend:

Used to set the legend display of all chart widgets in the widget layout area.

5 Mini-map:

Used to set the mini-map display of all chart widgets in the widget layout area.

Data source panel

In the data source panel, a user can view/edit information of the linked data source, as well as conveniently add/delete a column filter.

The screenshot shows the 'Datasource' panel with two main sections: 'Dimension' and 'Measure'. The 'Dimension' section contains several columns: OrderDate (selected), Category, City, Country, CustomerName, OrderID, PostalCode, ProductName, and Quantity. The 'Measure' section contains four measures: Discount, Profit (selected), Sales (selected), and DaystoShipActual.

Dimension
OrderDate
Category
City
Country
CustomerName
OrderID
PostalCode
ProductName
Quantity

Measure
Discount
Profit
Sales
DaystoShipActual

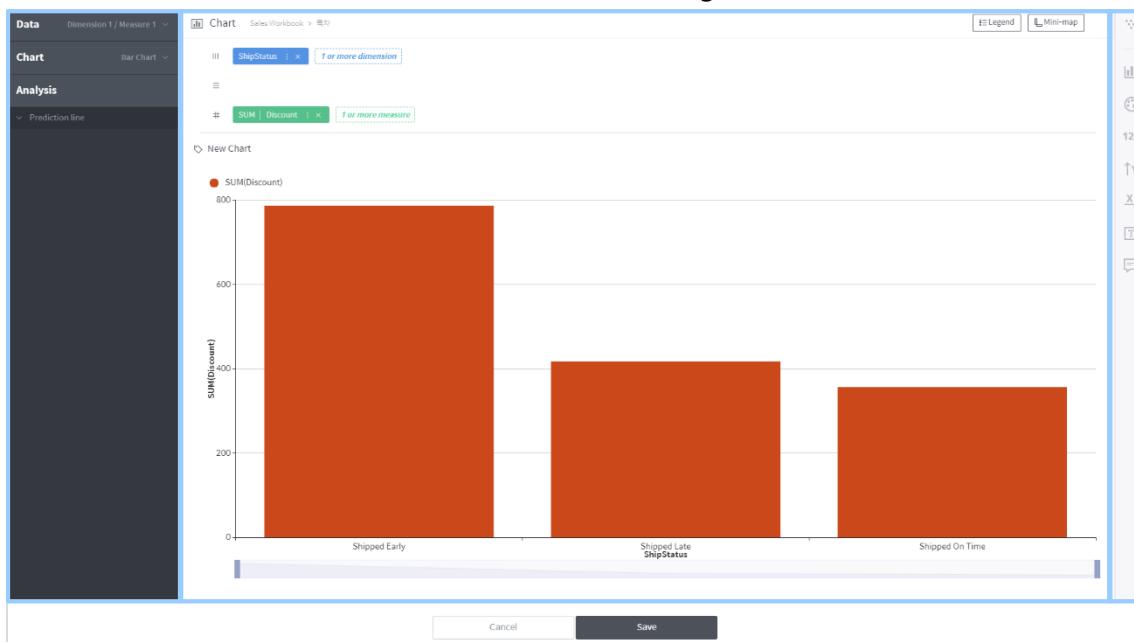
This panel has the same functions and structure of the data column list in the chart home. For more detailed information, see Paragraph 5.2. However, remember that the set/canceled filter in this panel is a dashboard filter, and the set/canceled filter in the chart home is a chart filter.

5. Create/Manage Chart

Each dashboard in the workbook is fundamentally composed of various charts that visualize the analyzed data. This section describes some of the concepts that you need to know to create a chart for data analysis, as well as for the UI to configure charts in Discovery.

5.1 Overview of the chart home screen

The home screen of the chart is divided into the following three areas.

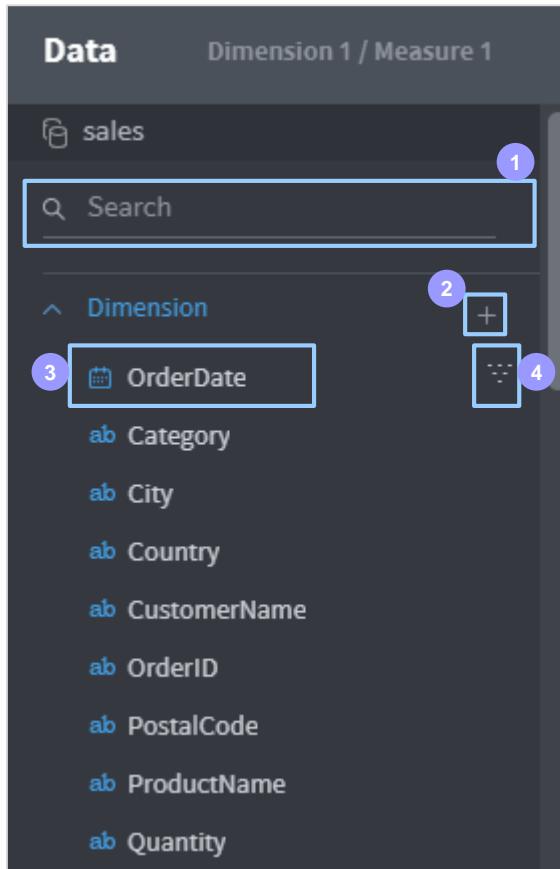


- ① **Column/Chart selection area:** The UI is set up in order of actions required to create a chart. A user can pivot a chart by selecting Data (data column list), and visualize data by selecting Chart (chart type list). In addition, the user can load analysis conditions to a chart using Analytics.
- ② **Visualization area:** This area is composed of the pivotable shelf area and the visualization area where the actual chart is drawn. Once data and chart are selected in the column/chart selection area, the chart is drawn in this area.
- ③ **Option area:** Used to customize the appearance and display of a chart. The option area is composed of filter, palette, axis, numeric expression and chart expression.

5.2 Data column list

5.2.1 Composition of data column list

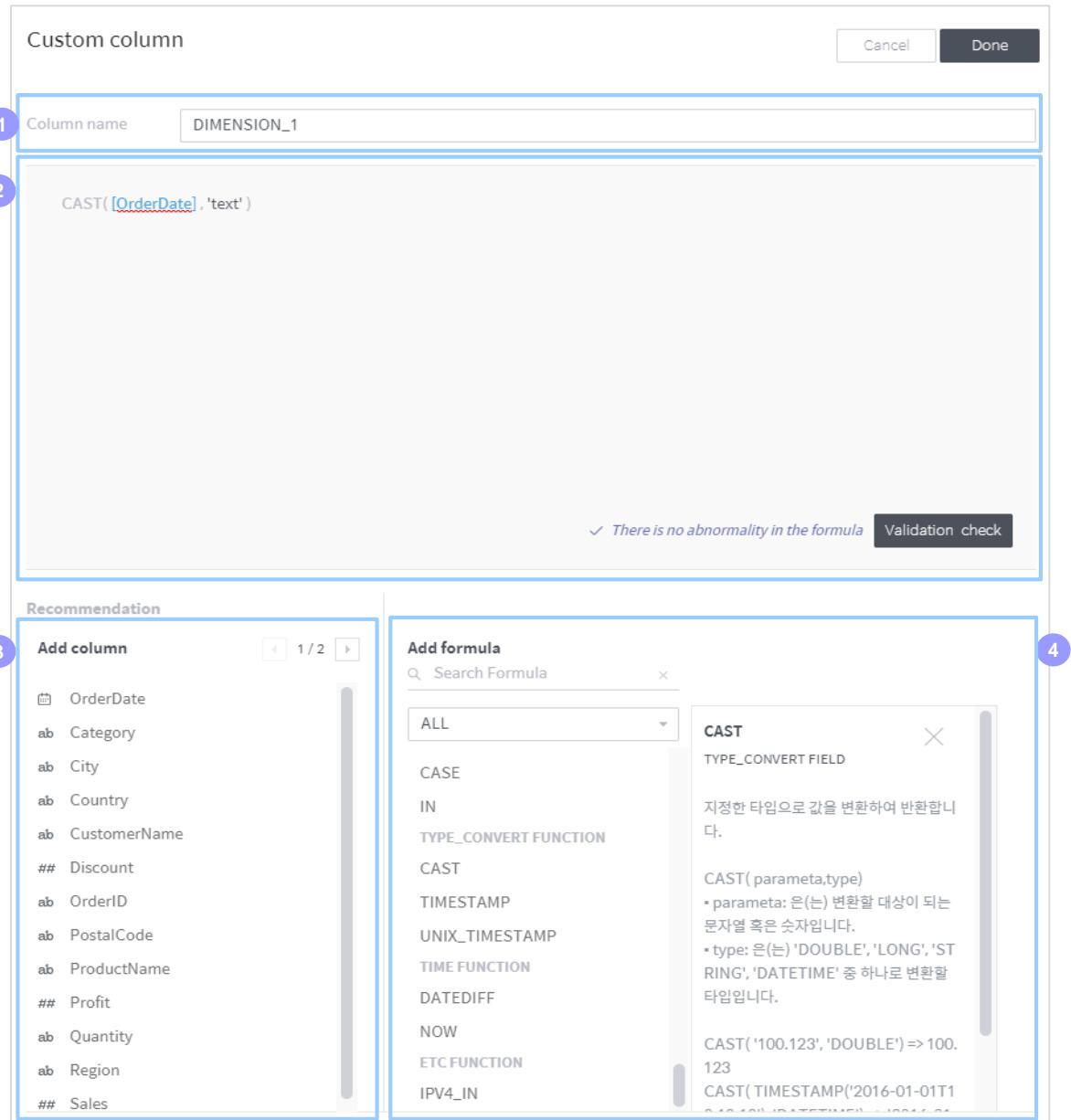
In the data column list, a user can view/edit linked data source information, as well as conveniently add/delete a column filter.



- ① **Search by column name:** Search for a column in the data source by its name.
- ② **Add custom column:** Click to open the dialog box to create a new column by combining/processing columns in the data source. The added custom column can be used in the entire area of the dashboard.
- ③ **Set/cancel filter:** Hover the mouse over a column to display this button. Click to set the column as a chart filter, and click again to cancel the chart filter. In columns set as a filter, the icon is displayed regardless of mouseover.
- ④ **More:** Used to check additional information on the column and set an alias.
 - Column detail: Used to check the overview and data values of the column.
 - Alias: Used to set a column alias. A formal column name can contain only alphanumeric characters and some special characters, and space is not allowed. Therefore, register a unique alias for more convenient analysis. The alias is applied to the entire area of the dashboard.
 - Value alias: An alias can be set to each data value in the column. The alias is applied to the entire area of the dashboard.

5.2.2 Add a custom column

Click the button in the data source column list to open the dialog box to add a custom column. A user can create a new column necessary to create a chart by applying various formulas to the existing columns in the data source.



Custom column

Column name: DIMENSION_1

Code area: CAST([OrderDate], 'text')

Validation check: ✓ There is no abnormality in the formula

Recommendation

Add column: OrderDate, Category, City, Country, CustomerName, Discount, OrderID, PostalCode, ProductName, Profit, Quantity, Region, Sales

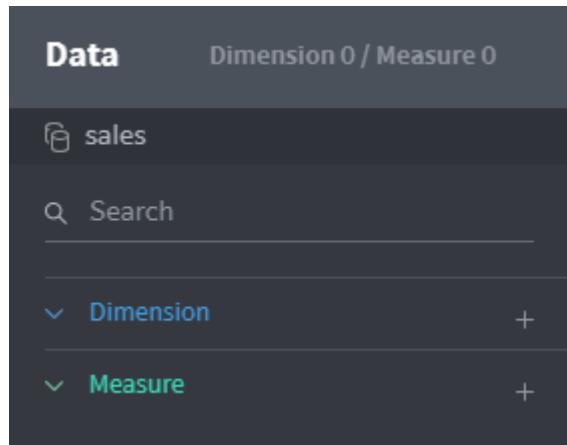
Add formula: ALL, CASE, IN, TYPE_CONVERT FUNCTION, CAST, TIMESTAMP, UNIX_TIMESTAMP, TIME FUNCTION, DATEDIFF, NOW, ETC FUNCTION, IPV4_IN

CAST
TYPE_CONVERT FIELD
지정한 타입으로 값을 변환하여 반환합니다.
CAST(parameta,type)
• parameta: 익(는) 변환할 대상이 되는 문자열 혹은 숫자입니다.
• type: 익(는) 'DOUBLE', 'LONG', 'STRING', 'DATETIME' 중 하나로 변환할 타입입니다.
CAST('100.123', 'DOUBLE') => 100.123
CAST(TIMESTAMP('2016-01-01T1'))

- 1 **Column name:** Place to write the name of a custom column.
- 2 **Code area:** Place to write the code to create a custom column. Click an item from the following column and formula list. The item will be typed in this area automatically.
- 3 **Add column:** A list of the existing columns in the data source. Click a column in the list. The column will be typed in the code area automatically.
- 4 **Add formula:** A list of formulas supported by metatron. Click a formula in the list. The formula will be typed in the code area automatically, and the typing cursor will be automatically moved to the parameter input field. For more information on purpose, instructions for use and example of each formula, see the help text on the right of screen.

5.2.3 Dimension and measure

Columns of a data source linked to the dashboard are divided into a dimension column and a measure column as follows. To fully utilize the chart function of Discovery, you must clearly understand the concepts of dimension and measure.



The screenshot shows the 'Data' panel of the Metatron interface. At the top, it says 'Dimension 0 / Measure 0'. Below that is a section for 'sales' with a file icon. A search bar labeled 'Search' is present. Two expandable sections are shown: 'Dimension' (with a plus sign) and 'Measure' (with a plus sign). The 'Measure' section is currently expanded, showing its contents.

Dimension column

A categorical data column with the following characteristics.

- A categorical (not aggregated) data field (e.g.: Category, Region, Organization, etc.)
- Criteria to display measure.

Measure column

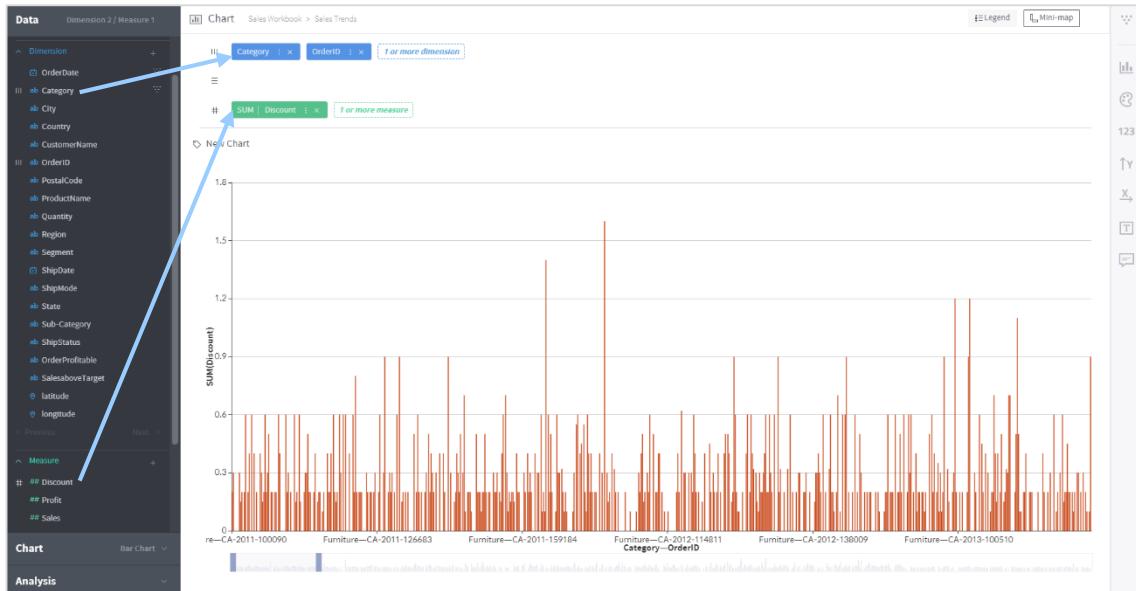
A quantitative data field with the following characteristics.

- A field containing aggregated or quantitative information (e.g.: Sales, etc.)
- Data displayed in the chart based on the criteria provided by dimension

5.3 Pivoting

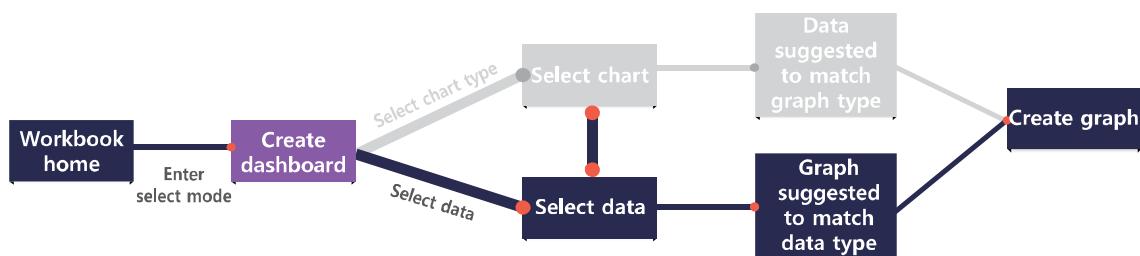
5.3.1 What is 'pivoting'?

'Pivoting' or 'shelfing' means a procedure to select a column in the column/chart selection area and place it on the column/row/cross shelf in the shelf area.



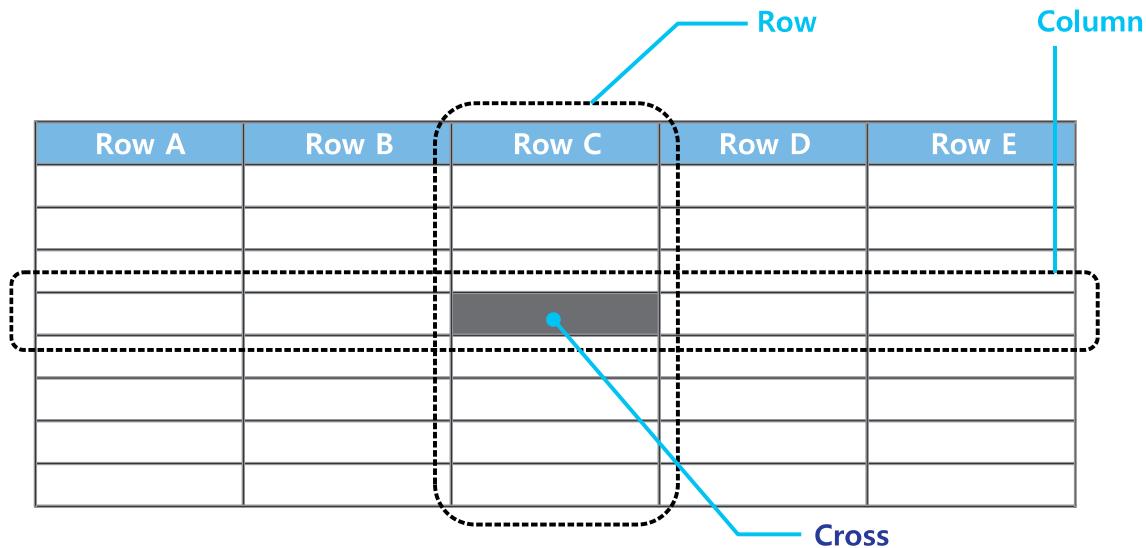
The above figure displays pivoting of two dimension columns in the column shelf, and pivoting of one measure column in the cross shelf. The chart displays the data of pivoting columns.

Mandatory/Recommended column types of each shelf are different by chart type. Select the chart type before pivoting to show the necessary column types for the shelf.

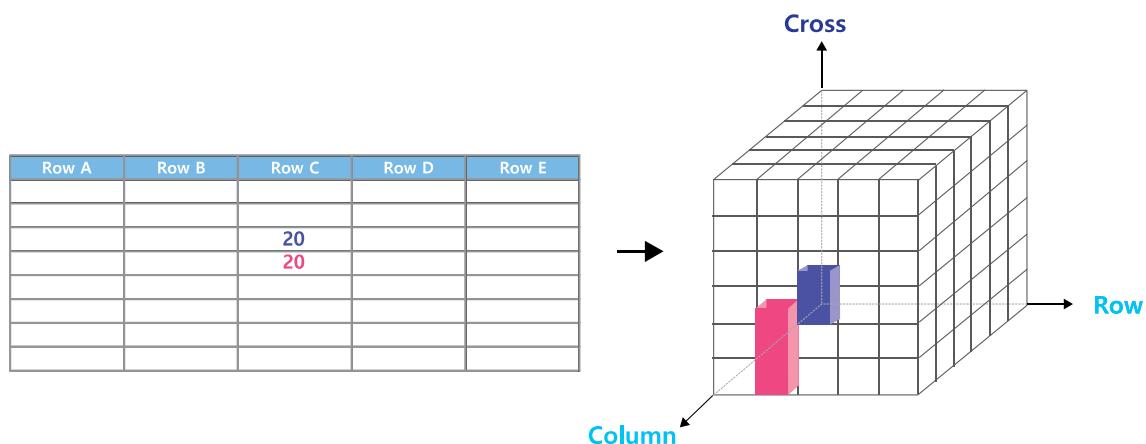


5.3.2 The concept of column/row/cross shelves

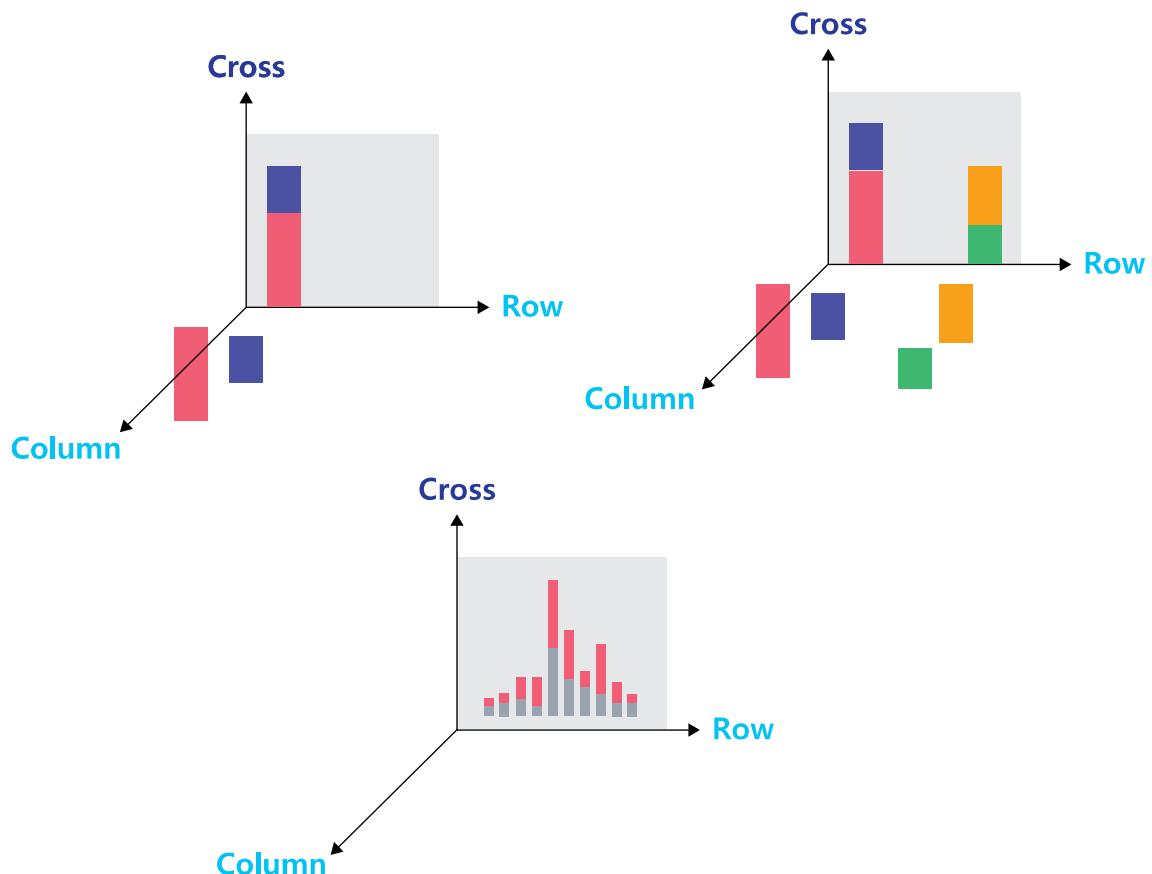
Think of the structure of Excel to understand the concept of column/row/cross shelves. As shown below, a column/row defines the block, and cross defines the value to be entered in the block.



metatron's column/row/cross does similar things. Excel displays column/row/cross in a two-dimensional grid, but metatron displays column/row/cross in a three-dimensional cube. As a tool of OLAP Data Discovery, metatron searches data in a three-dimensional space via OLAP Cube. (For more information on OLAP, see metatron Outline in Part 1.) The following chart is an axis figure of column/row/cross values expressed as a three-dimensional cube in metatron.

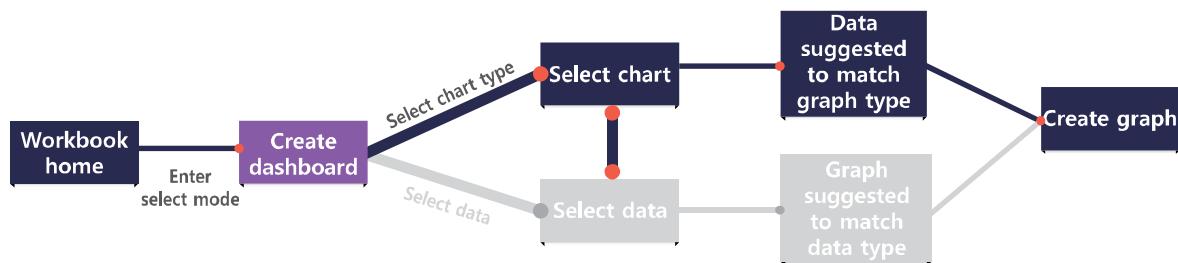


We can assume that if values of an Excel grid are displayed in a three-dimensional chart, multiple cross values will create various bars. However, metatron displays a chart as a two-dimensional cross section, so bars will be stacked based on columns and rows. As a result, metatron will display the two-dimensional chart in a similar way to the gray area in the bottom figure.



5.4 Chart type

metatron Discovery provides more than 20 chart types. If the user pivots a column before selecting a chart, suitable charts will be highlighted in purple.



The following table summarizes creation conditions and use attribute/type/example for each chart.

Chart Name/Icon	Creation Condition	Use Attribute	Use Type	Use Example
	Column: 1 or more dimensions & cross: 1 or more measures	Compares the value of each item.	Used to compare groups or view trends over time. Very effective when the trend is fluctuating.	Comparison of sales and profits by product
	Column or row: 1 or more dimensions & cross: 1 or more measures	Displays the cross data of each item as text.	Used to view measures that meet certain criteria. Useful to check detailed data and accurate values. Not good for visualization.	Detailed sales data by year
	Column: 1 or more dimensions & cross: 1 or more measures	Displays change of data over time.	Used to view trends over time. If the trend is not fluctuating, a line chart is more effective than a bar chart.	Monthly sales trend
	Column: 1 measure & row: 1 measure & cross: 1 or more dimensions	Displays the relation between various items.	Used to define the relation between two variables.	The relation between product sales and profit

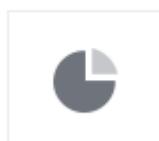
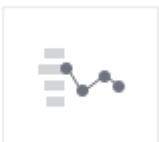
Chart Name/Icon	Creation Condition	Use Attribute	Use Type	Use Example
	Column or row: 1 or more dimensions & cross: 1 or more measures	Displays the cross data of each item as a color distribution.	Used to compare two variables intuitively based on color and size. Used to emphasize the visual elements of a table chart.	Sales of each product by region
	Cross: 1 or more dimensions, 1 or more measures	Ratio of each item over the whole	Used to compare components of something.	Comparison of a web browsers' market share
	Column: 1 dimension in time attribute & cross: 1 or more measures	Displays characteristics values for processing status.	Used to quickly deliver information on the current organizational performance.	Number of customers introduced in this year or organizational performance index
	Cross: 1 or more measures	Displays main indicators with trends.	Used to compare the distribution of each group or indicate the target of a specific value.	Comparison of the distribution of delay time by airplane type
	Column: 1 or more dimensions, row: 1 dimension, cross: 1 measure	Indicates an increase and decrease of value over time.	Used to emphasize an increase and decrease of value over time.	Monitoring change of the number of team members for a certain period, or stock
	Column: 1 dimension in time attribute & cross: 1 measure	Displays the sum when adding or subtracting a value.	Used to summarize and emphasize important words.	Summary of customer comments

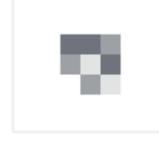
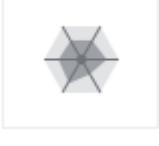
Chart Name/Icon	Creation Condition	Use Attribute	Use Type	Use Example
 Word Cloud	Cross: 1 or more dimensions, 1 measure	Displays the size of text proportional to its appearance frequency.	Used to emphasize various types of information.	Simultaneous monitoring of price and sales by product
 Combo Chart	Column: 1 or more dimensions & cross: 2 or more and less than 4 measures	Compares data by combining bar and line charts.	Used to visualize hierarchy data.	Monitoring of sales by product (major class- medium class- minor class)
 Treemap	Column: 1 dimension & row: 1 or more dimensions & cross: 1 measure	Displays hierarchy data as a group of overlapped quadrangles.	Used to intuitively compare various measurement targets.	Comparison of products evaluated by 5 factors of quality
 Radar Chart	Cross: 1 dimension, 1 or more measures	Displays various evaluation factors from the central point.	Used to view the flow of generated data.	Monitoring the flow of a project task
 Network Diagram	Subject shelf: 1 dimension & target shelf: 1 dimension & connecting shelf: 1 measure	Connection type diagram displaying factors with dependency	Used to monitor the quantitative flow of data.	Monitoring the energy flow in the factory
 Sankey Diagram	Column: 3 or more dimensions & cross: 1 measure	Displays proportion of flow by the width of the connection line.	Used to view data proportion.	Monitoring of profit by region

Chart Name/Icon	Creation Condition	Use Attribute	Use Type	Use Example
 Gauge Chart	Column: Row: 1 or more dimensions & cross: 1 measure	Visualizes performance for the established target.	Used to compare groups or view trends over time. Very effective when the trend is fluctuating.	Comparison of sales and profits by product

5.5 Chart filter

A chart filter limits the scope of each column's data to be shown in the chart. This part describes how to set and use chart filters and consists of the following.

- **5.5.1 Filters included automatically:** Describes basic filters that do not need to be added.
- **5.5.2 Chart filter panel:** Describes the chart filter panel shown on the right of the chart home. In this panel, a user can conveniently search and set registered filters.
- **5.5.3 Chart filter dialog box:** Describes the chart filter dialog box to open the chart filter panel. With this dialog box, a user can add or configure a chart filter.

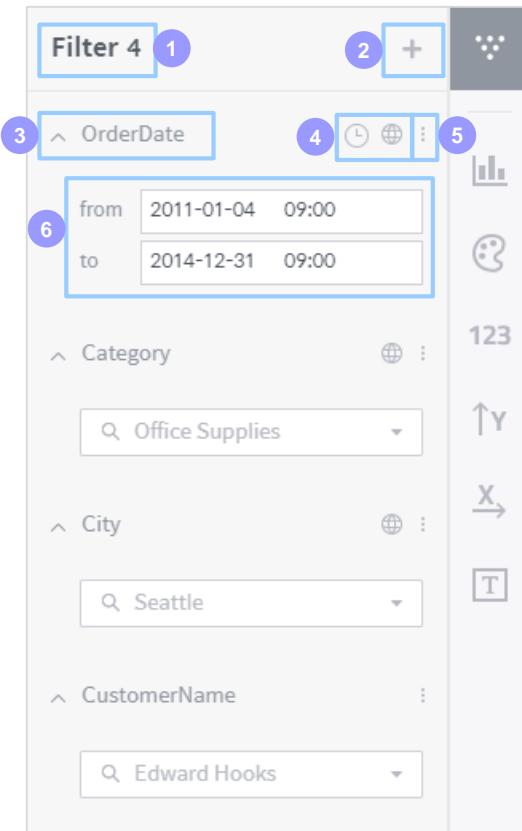
5.5.1 Filters included automatically

The following column filters are included automatically so it is not necessary to add a chart filter.

- **Timestamp column filter:** Due to the time-series characteristic of metatron Discovery, time condition filtering is required.
- **Recommended filter:** Column filters designated as a 'recommended filter' in the data source.
- **Dashboard filter with a 'global' attribute:** Filters applied to all charts registered in the dashboard.

5.5.2 Chart filter panel

Click the  icon at the top of the options area to display the chart filter panel as follows. In this panel, a user can search and edit the basic information of registered filters.



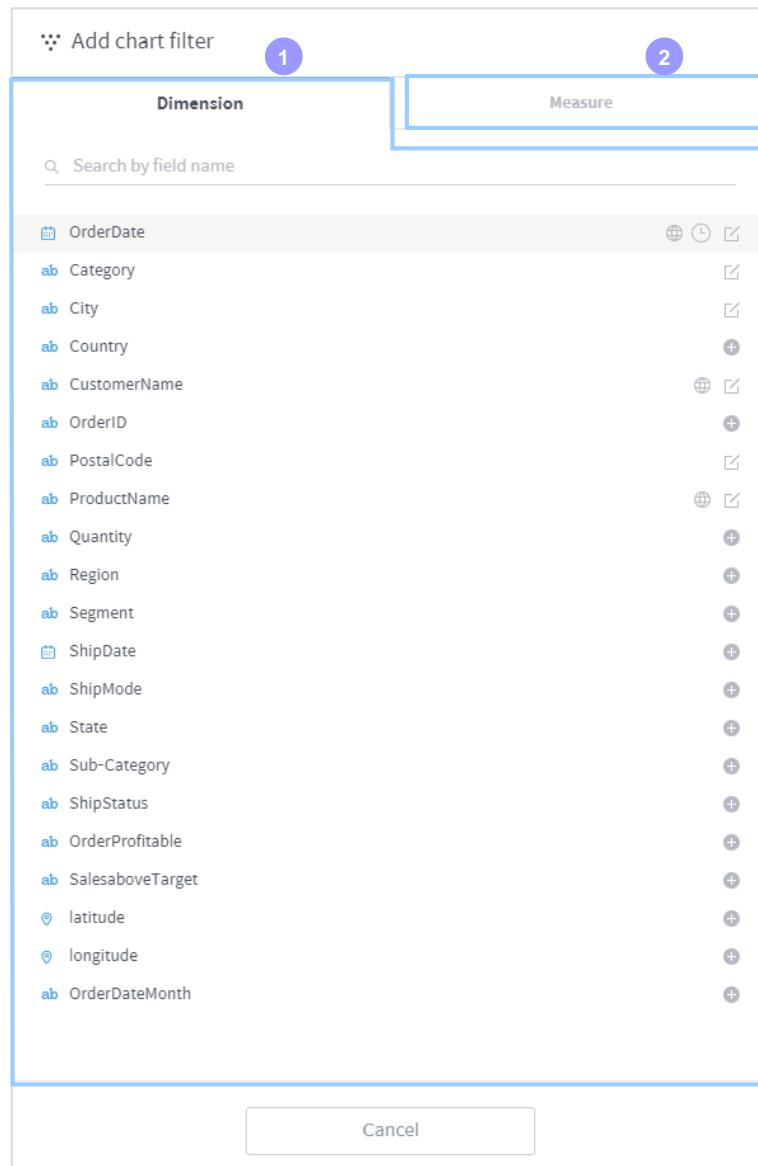
- ① **Filter number:** Displays the number of the currently registered filter.
- ② **Add/edit filter:** Click to add a new filter or display the dialog box to configure an existing filter.
- ③ **Filter name:** Displays the column name of the filter.
- ④ **Filter attribute:** Displays the basic attributes of the filter as an icon.
 - : Indicates that the column of the filter is a timestamp column.
 - : Indicates that the filter is a dashboard filter with a 'global' attribute.
- ⑤ **More:** Used to reset or configure the filter.
- ⑥ **Filtering scope:** Set the scope of data values to be shown in the chart.

5.5.3 Chart filter dialog box

Click the button in the top of the chart filter panel or click the button in each filter area to open the chart filter dialog box. With this dialog box, a user can add a new filter or configure an existing filter.

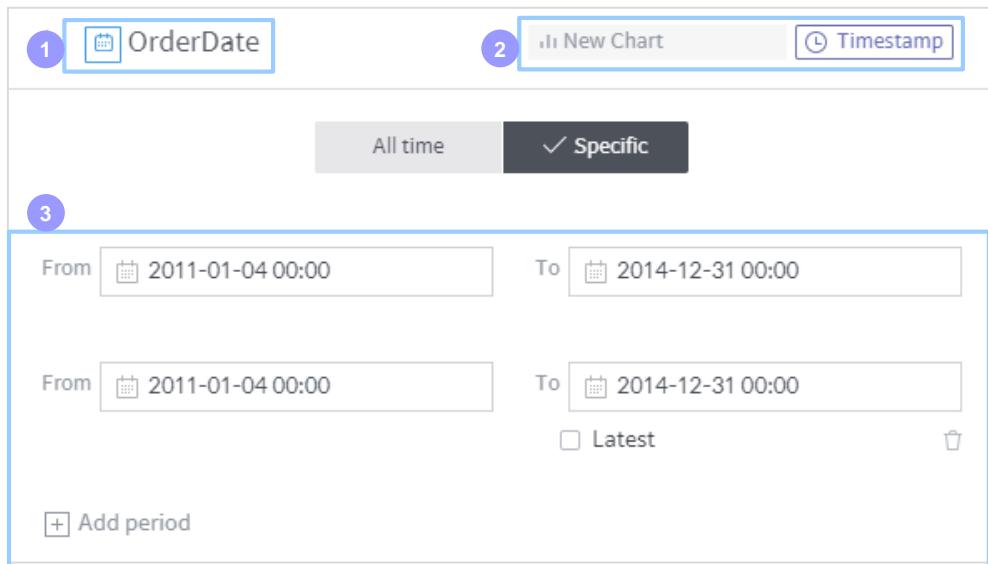
A. Composition of chart filter dialog box

The chart filter dialog box is divided into the Dimension and Measure sections as shown below:



- ① **Dimension:** You can select a dimension from the connected data source to make a filter. For how to configure a dimension filter, see Subsection C.
 - indicates a timestamp column, for which a timestamp filter can be configured. For how to configure a timestamp filter, see Subsection B.
- ② **Measure:** You can select a measure from the connected data source to make a filter. For how to configure a measure filter, see Subsection D.

B. Timestamp column filter setting



- 1 Name:** Displays the name of the column where the filter is applied.
- 2 Filter attribute:** Displays the basic attributes of the filter.
 - **Timestamp:** Indicates that the filter column is a timestamp column.
 - **Chart name:** Displays the name of the chart to which the chart applies.
- 3 Set default status:** Set the time range to be shown in the chart.
 - **All:** Displays the contents of columns during the entire period on the chart without applying time filtering.
 - **Period:** Displays the contents of columns during a specific period on the chart. Click 'Set to current time' to set the latest reference time as the present time.
 - **Add/delete time range:** Use **[+] Add period** and **[trash]** icons in the lower right to add or delete a time range.

C. Dimension column filter setting

The screenshot shows the Metatron Using Workbook interface for managing chart filters. On the left, a list of customer names is displayed with their respective counts. On the right, a detailed view of the filter settings is shown.

Filter Settings (Right Side):

- Wildcard:** values starting with "AFTER"
- Condition:** sum of values is equal to 10
- Limitation:** values ranked in the top 10 of (SUM)

Customer Name List (Left Side):

Customer Name	Count
Aaron Bergman	6
Aaron Hawkins	11
Aaron Smayling	10
Adam Bellavance	18
Adam Hart	20
Adam Shillingsburg	25
Adrian Barton	20
Adrian Hane	16
Adrian Shami	3
Aimee Bixby	12
Alan Barnes	14
Alan Dominguez	12
Alan Haines	8
Alan Hwang	13

- ① **Name:** Displays the name of the column where the filter is applied.
- ② **Chart name:** Displays the name of the chart to which the chart applies.
- ③ **Select range:** Select a range to be shown on the chart by filtering data categories included in the column of the selected filter.
 - Single item: Select one data category and display it on the chart.
 - Multiple items: Select multiple data categories and display them on the chart.
- ④ **Search by name:** Search data values in the selected filter by their name.
- ⑤ **Filtering:** Filters data categories to be displayed.
 - Wild card: Search data in the selected filter using a specific letter. For example, to view only data beginning with 'L', enter L in 'The first word' and click Apply. To view only data ending with '89', enter 89 in 'The last word' and click Apply. To view only data containing 'Cart', enter Cart in 'Contain' and click Apply.
 - Condition: View measures in the column by applying conditions. Select measures to display in the leftmost field. Data whose SUM/AVG/COUNT/MIN/MAX of the measure is equal/over/under/not less than/not more than the target value will be displayed.
 - Limitation: View the measure in the column by filtering with an upper or lower ranking of the established conditions. Select the measure to be filtered.

SUM/AVG/COUNT/MIN/MAX of the measure will be displayed after filtered by an upper/lower ranking adjusted by the number selected by the user.

- 6 **Sort:** Select how to sort data in the list.
 - Frequency ascending/descending Sorts data columns in ascending/descending order by frequency.
 - Alphanumeric ascending/descending Sorts data columns in ascending/descending order by name.
- 7  icon: Used to indicate specific items from data categories in the list. Activate the only icon next to the item to be displayed, and click the  icon at the top to display only the selected item.
- 8 **Defined value:** Used to add a data category without a column as a filter condition. This function is required to create a filter in advance for a data category that may be added later.

D. Measure column filter setting



- ① **Name:** Displays the name of the column where the filter is applied.
- ② **Chart name:** Displays the name of the chart to which the chart applies.
- ③ **Select range:** Select the minimum and maximum data values to be shown on the chart, from the column of the selected filter.

5.6 Chart style setting

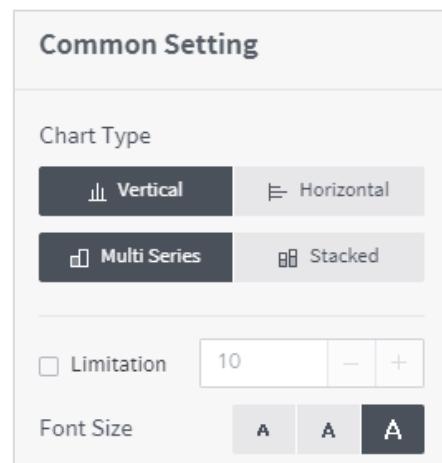
After data pivoting, an option menu to set the chart style will be displayed on the right. The composition of the menu is different by chart type. 5.6.1 part describes the setting items applied to all chart types, and 5.6.2 part describes the 'Common Setting' items of each chart type.

5.6.1 Chart style setting menu

This part describes how to set each item of the chart style setting menu. Please note that among the following items, some items may be unavailable in a specific chart type.

Common setting

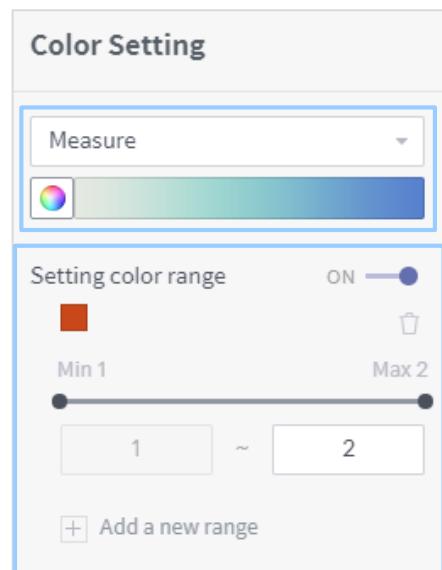
Defines the shape of the chart. Items of a common setting are different by chart type. See *** for more detailed information on the common setting of each chart type.



Color setting

Defines various colors in the chart.

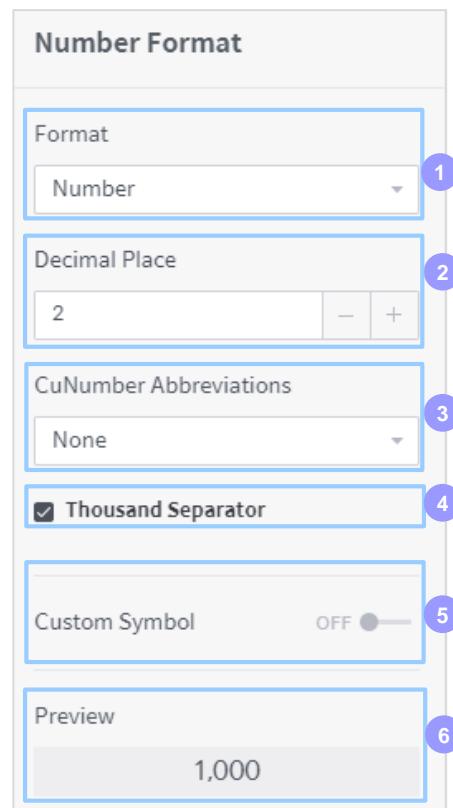
- ① **Graph color setting:** Set criteria to classify the color of items to display data in a graph, and select the color theme.
 - Series: Classifies color based on the type of measure.
 - Dimension: Classifies color based on the type of dimension.
 - Measure: Classifies color based on the size of the measure.
- ② **Text color:** Select the text color of the dimension.
- ③ **Area color:** Select the background color of the table.
- ④ **Color range setting:** This item is displayed when 'Measure' is selected as criteria to classify the data indication color. Set 'ON' to change the color by the range of the measure. The color range can be subdivided at will from the lowest section. To add a new section, adjust the maximum value of the final section first and click 'Add a new range'.



Number format

Defines how to display data values to be shown as text in the chart graph. To use this function, enable the label display function from the data label setting menu.

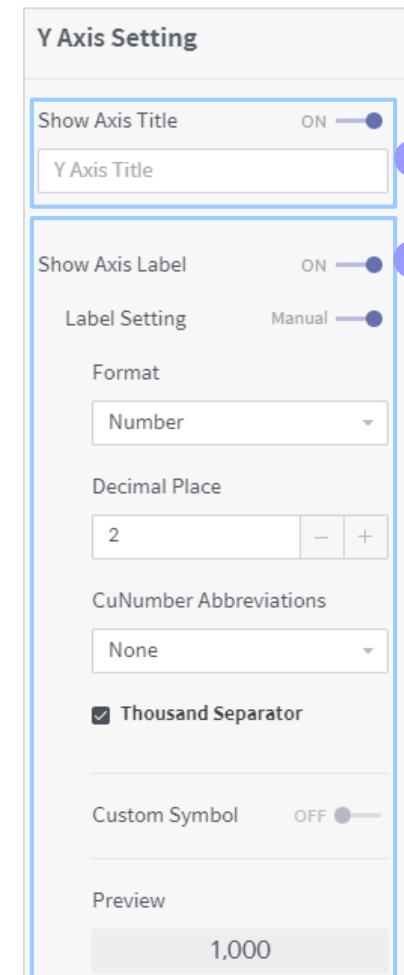
- ① **Display format:** Select the display format of a data value among number, currency, percent and exponent.
- ② **Decimal place setting:** Select the decimal places for a data value.
- ③ **Number abbreviation setting:** Set K (thousands), M (millions) or B (billions) as an abbreviation for a large data value. Select 'Automatic Adjustment' to set the most proper unit automatically based on digits of data values.
- ④ **Use thousands separator:** Select whether to display data values using a thousands separator.
- ⑤ **Customer symbol setting:** Insert custom text before/after data values.
- ⑥ **Preview:** Displays an actual result of the defined number format.



Y-axis setting (based on vertical chart type)

Defines how to display the Y-axis of the chart. Change the chart type as 'Horizontal Type' from Common Setting to switch between X-axis and Y-axis settings.

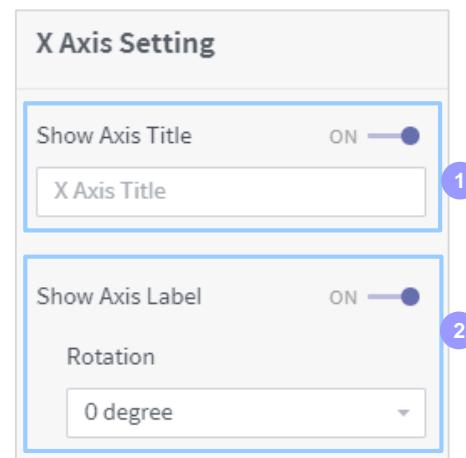
- ① **Enter axis title:** Used to enter the title of the Y-axis of the chart. Do not use this function to hide the title of the Y-axis.
- ② **Show label:** Select whether to display the data label on the Y-axis of the chart. Do not use this function to hide the data label of the Y-axis.
 - **Label setting:** Set the number format displayed on the data label of the Y-axis. Set 'Auto' to apply settings of 'Number Format', or 'Manual' to set a format for the data label of the Y-axis.



X-axis setting (based on vertical chart type)

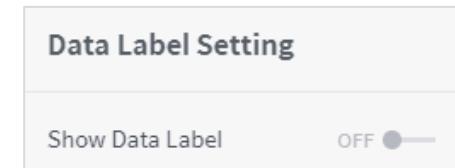
Defines how to display the X-axis of the chart. Change the chart type as 'Horizontal Type' from Common Setting to switch between X-axis and Y-axis settings.

- ① **Enter axis title:** Used to enter the title of the X-axis of the chart. Do not use this function to hide the title of the X-axis.
- ② **Show label:** Select whether to display the data label on the X-axis of the chart. Do not use this function to hide the data label of the X-axis.
 - Label rotation: Select the angle of the data label of the X-axis as 0/45/90 degrees.



Data label setting

Select whether to display the data value on the graph of the chart.



5.6.2 'Common Setting' items by chart type

This part describes how to style the 6 most popular charts (bar chart, table, line chart, scatter chart, heatmap and pie chart).

A. Bar chart

Displays data values in each category of the dimension column as a bar.



1 Chart type: Defines the shape of the chart.

- Vertical Type: Displays data values as a vertical bar based on the vertical dimension axis.
- Horizontal Type: Displays data values as a horizontal bar based on the horizontal dimension axis.
- Parallel Type: If 2 or more measures are selected, bars representing each measure are displayed in parallel.
- Overlapped Type: If 2 or more measures are selected, all measures are displayed in one bar.

2 Limitation: Set the number of columns to be shown in the chart.

B. Table

Creates a table block based on the categories of dimension columns pivoted to column/row shelves. The corresponding measures will be displayed in the cross area as text.

Segment		
Consumer	Corporate	Home Office
155	88	24
45	19	6

1 Chart type: Defines the shape of the chart.

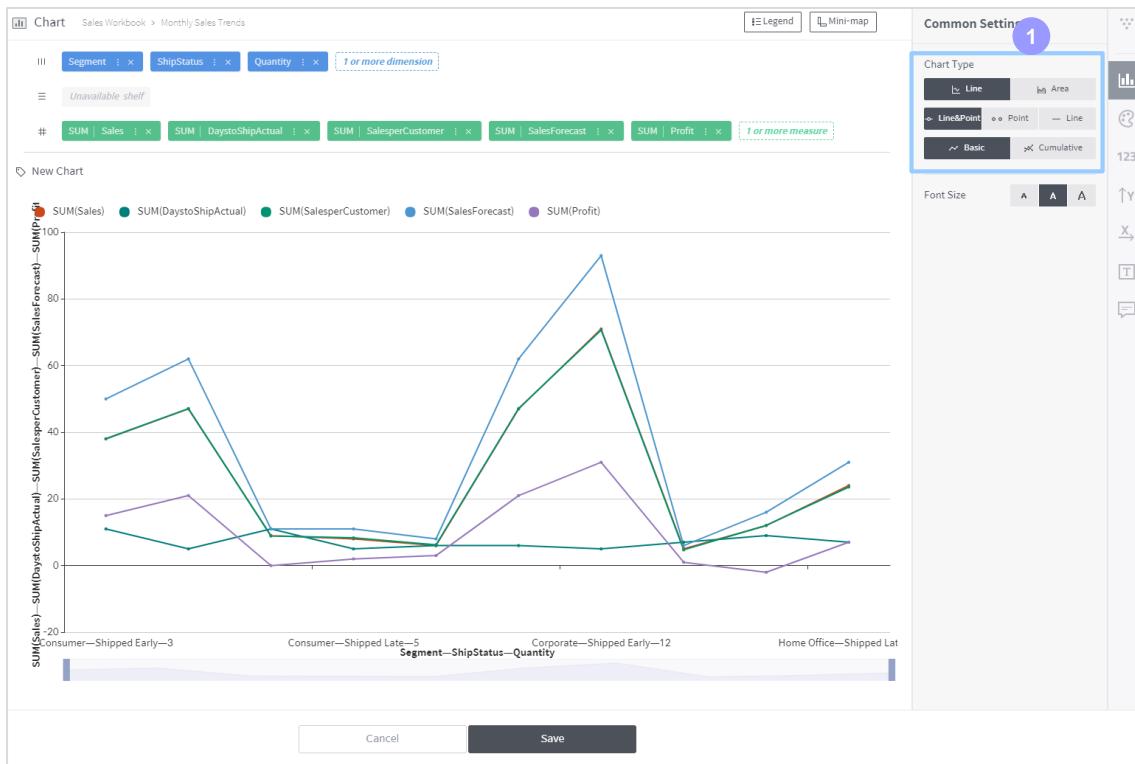
- Pivot data: Aggregate measures with the same dimension category for classification in one cell (SUM, MIN, MAX, etc.).
- Original data: Displays all original measures based on a specific dimension column without aggregation.
- Vertical view: Displays measures vertically in the table. This function is not available when using the original data type.
- Horizontal view: Displays the table horizontally when using pivot data type. Displays measures horizontally in the table.

2 Show Head Column: Set text alignment in the head column as horizontal or vertical.

When using the original data, the head column is mandatory. When using pivot data type, the head column is optional.

C. Line chart

Displays data values in each category of the dimension column as dots. Dots of adjacent category items are connected with each other. Used to view the trend.

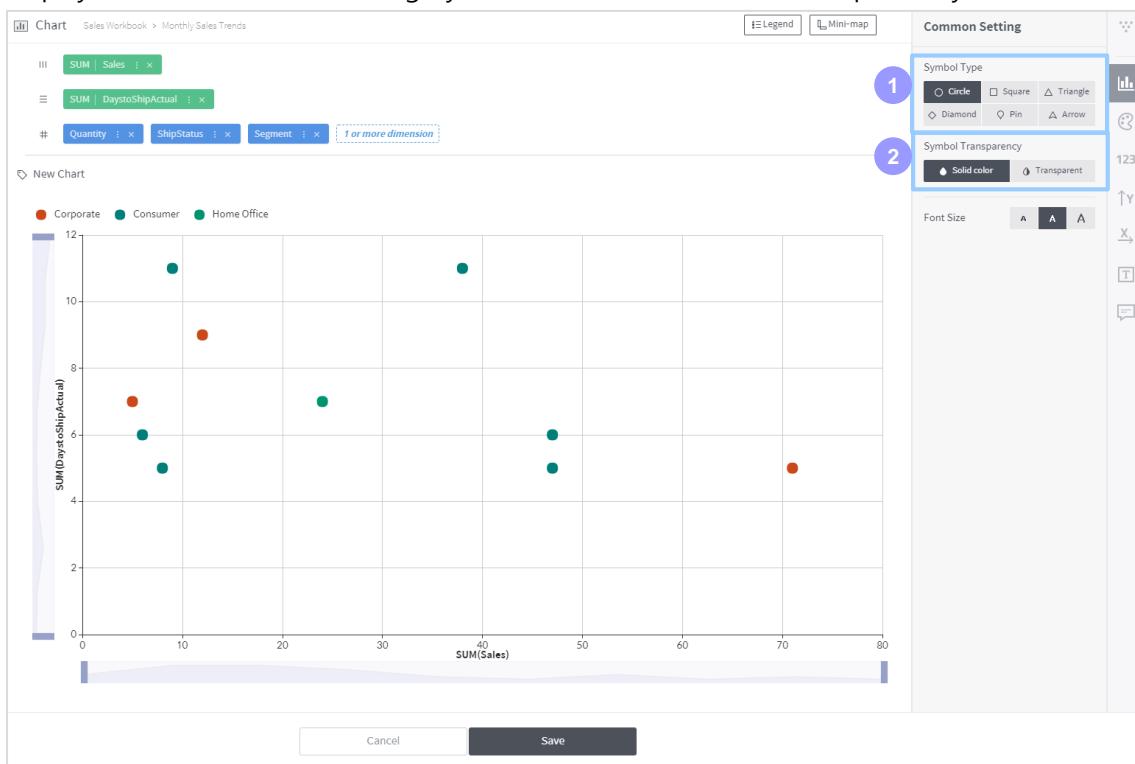


① Chart type: Defines the shape of the chart.

- Line type: Displays the chart by connecting lines from the measure.
- Area type: Displays the chart by applying color to the area created by lines.
- Line & Point: Displays both dots based on the measure and lines connecting the dots.
- Point: Displays only dots based on the measure.
- Line: Displays only connection of lines.
- Basic type: Displays measures on the chart.
- Accumulation type: Displays the accumulation of measures on the chart.

D. Scatter chart

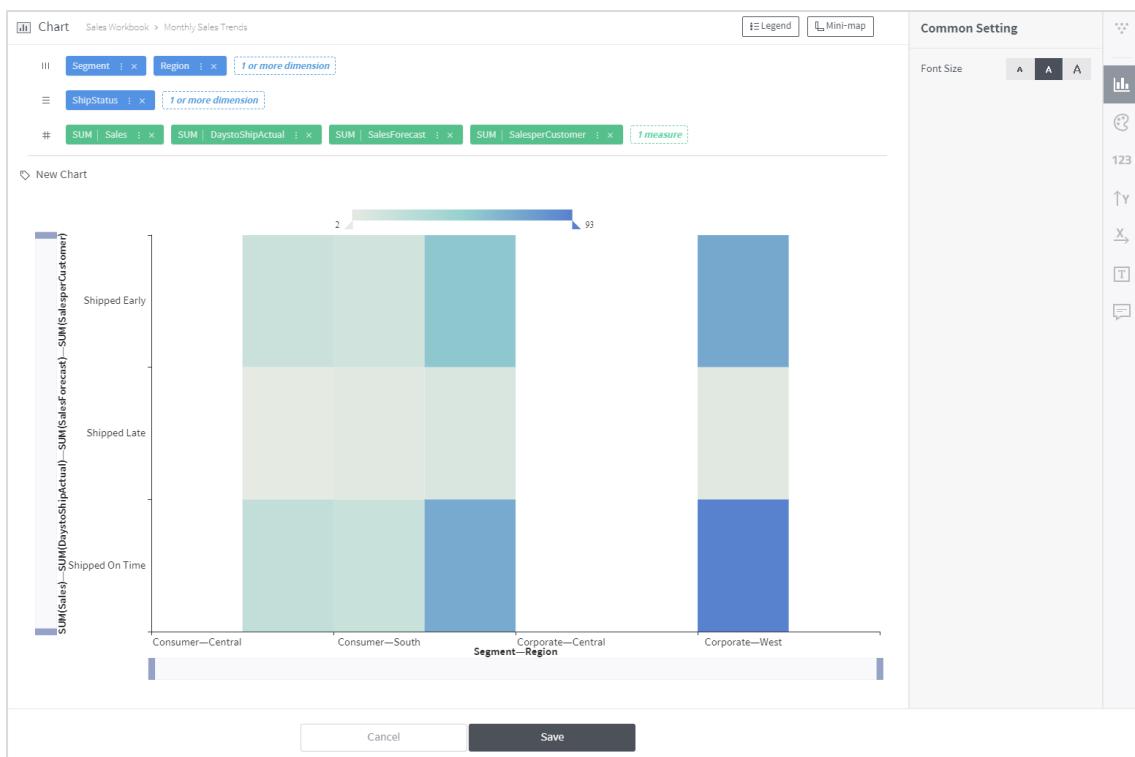
Displays data values in each category of the dimension column as a specific symbol.



- 1 **Symbol type:** Set the shape of the symbol to be shown on the chart.
- 2 **Symbol transparency:** Set the transparency of the symbol to be shown on the chart.
Select solid color or transparent.

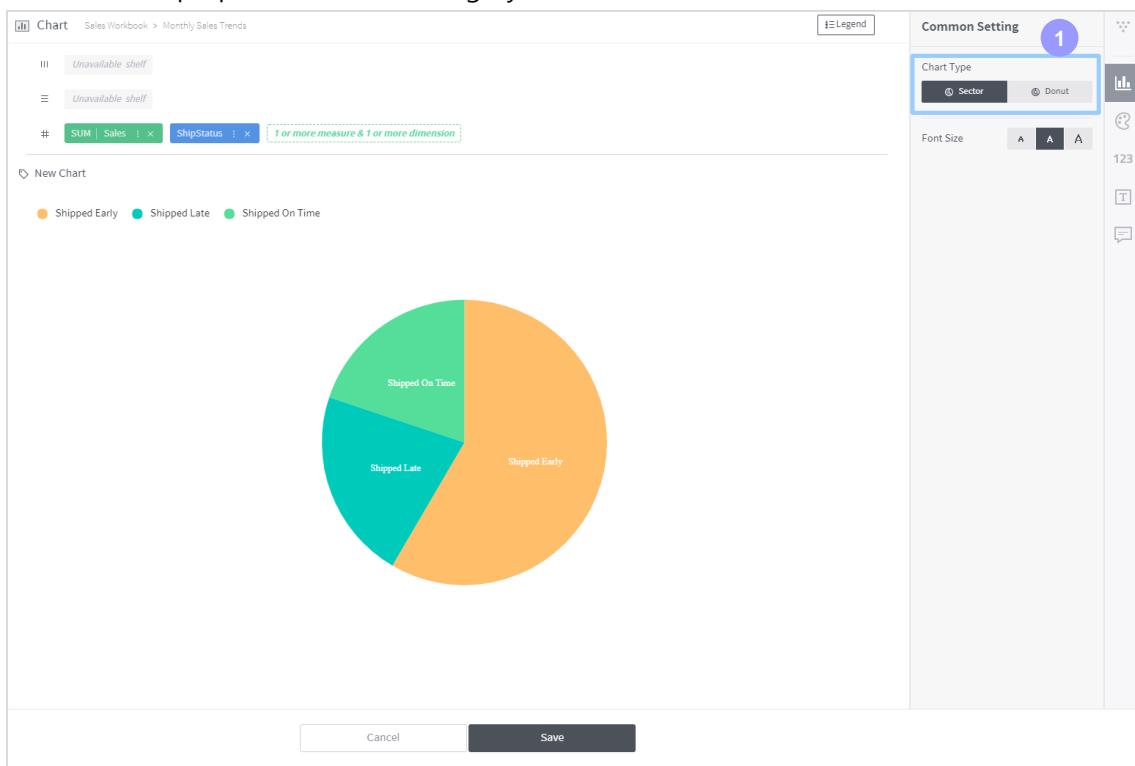
E. Heatmap

Displays each data value of the measure column on the cross shelf as a color. For a larger data value, a stronger color will be applied. Heatmap does not provide any common setting items.



F. Pie chart

Visualizes the proportion of each category item of the dimension column.



① **Chart type:** Defines the shape of the chart.

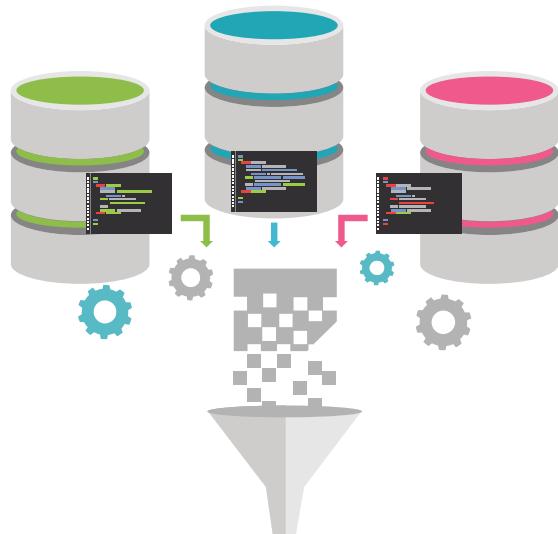
- Fan type: Displays the chart as a pie.
- Donut type: Displays the chart as a donut.



metatron 2.0

Part 5

Using Notebook



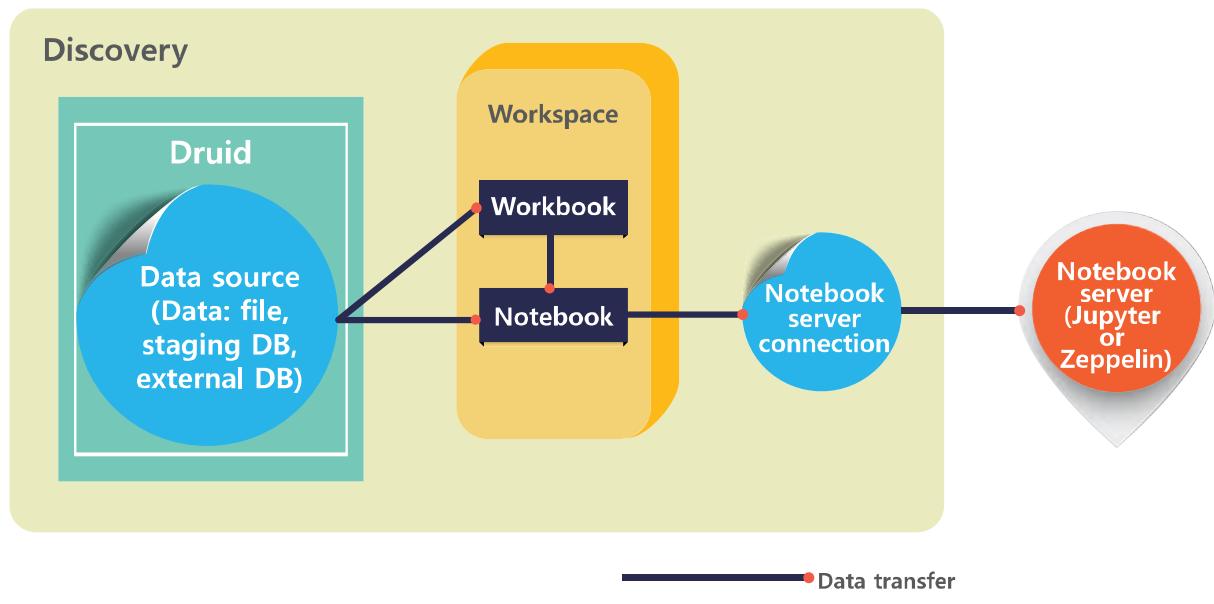
Copyright ©  SK telecom

The content of this document must not be copied,
distributed or used in part or in whole without the prior approval of SK Telecom.

Contents

1. Notebook Outline	3
2. Initial Notebook Server Setting	4
3. Create a Notebook	5
4. Use Functions of a Notebook.....	11
4.1 Detailed notebook search	11
4.2 Notebook coding.....	12
4.3 Register a notebook API.....	13

1. Notebook Outline



metatron Discovery Workbench provides an environment for data analysis **based on machine learning**. The main features are as follows.

- Data to be analyzed can be loaded from the data source saved in the Druid engine, or extracted from a dashboard or chart inserted in a workbook.
- It supports the following external analysis tools and languages.

External analysis tools	Available languages
Jupyter	R
	Python
Zeppelin	Spark

Using these tools, analysts can analyze/sort/forecast data as they wish and share their analysis process with others.

This part is divided as follows.

Chapter 2 Initial Notebook Server Setting: To link Jupyter or Zeppelin mentioned above, a server to host these tools must be registered in metatron Discovery. This chapter describes how to register a Jupyter or Zeppelin server.

Chapter 3 Create a Notebook: Each document that stores machine learning-based analysis code is called a 'Notebook'. This chapter describes how to create a Notebook, link data to be analyzed, and decide on the tool and language for analysis.

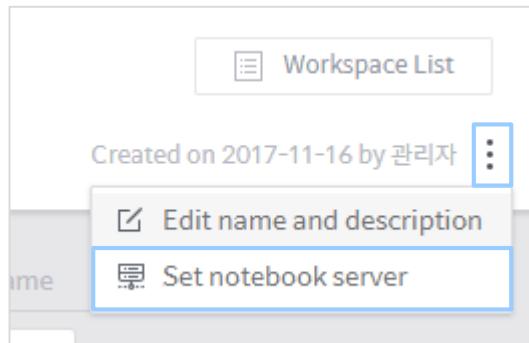
Chapter 4 Utilize a Notebook: This chapter describes how to process coding from a newly created notebook and execute/share results.

2. Initial Notebook Server Setting

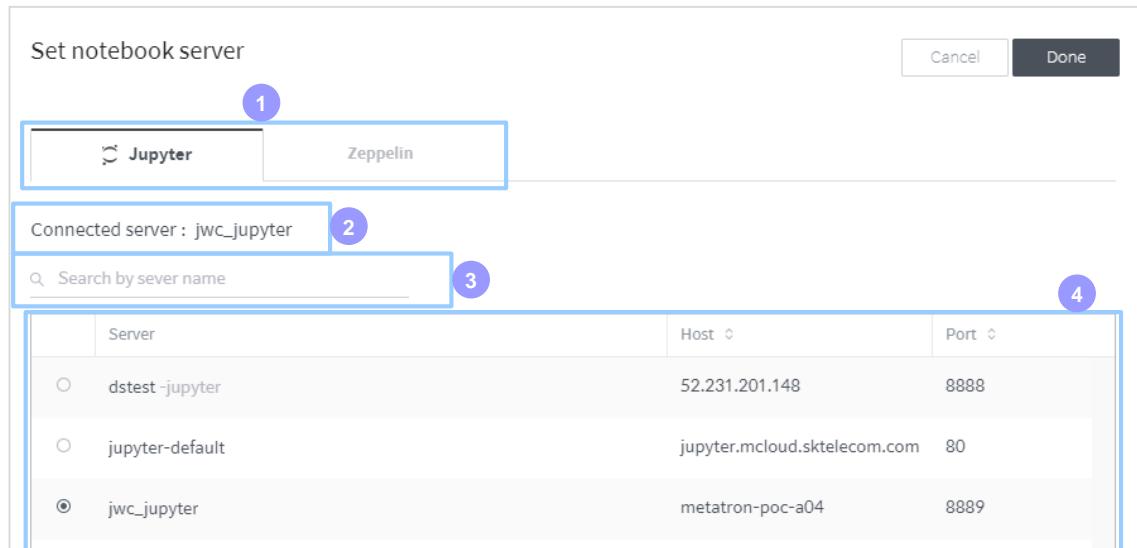
To analyze the data in the workspace using Notebook, a user must initialize the notebook server. To initialize the Notebook server, the Notebook server must be registered. For more information, see '4. Notebook Management' of 'Part 3 Management'.

The procedure for initializing the notebook server is as follows.

1. Click the  button in the top-right corner of the workspace and select 'Set notebook server'.



2. Once the Set notebook server screen is displayed, refer to the description below of each item and select the notebook server that you want. Then click 'Done'.



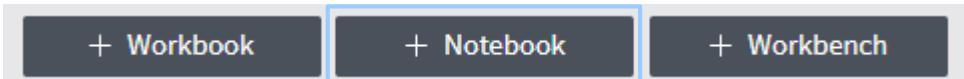
- 1 **Select server type:** Among Jupyter or Zeppelin servers registered by the user, click the notebook server to be linked to the workspace.
- 2 **Connected server:** The name of the currently selected notebook server in the server list.
- 3 **Search by server name:** Search a registered notebook server by its name.
- 4 **Server list:** Displays notebook servers that meet the established criteria.

3. Create a Notebook

Once the notebook server is initialized, a user can create a notebook. To create a notebook, the user must set 'data to be analyzed' and 'tool and language for analysis'.

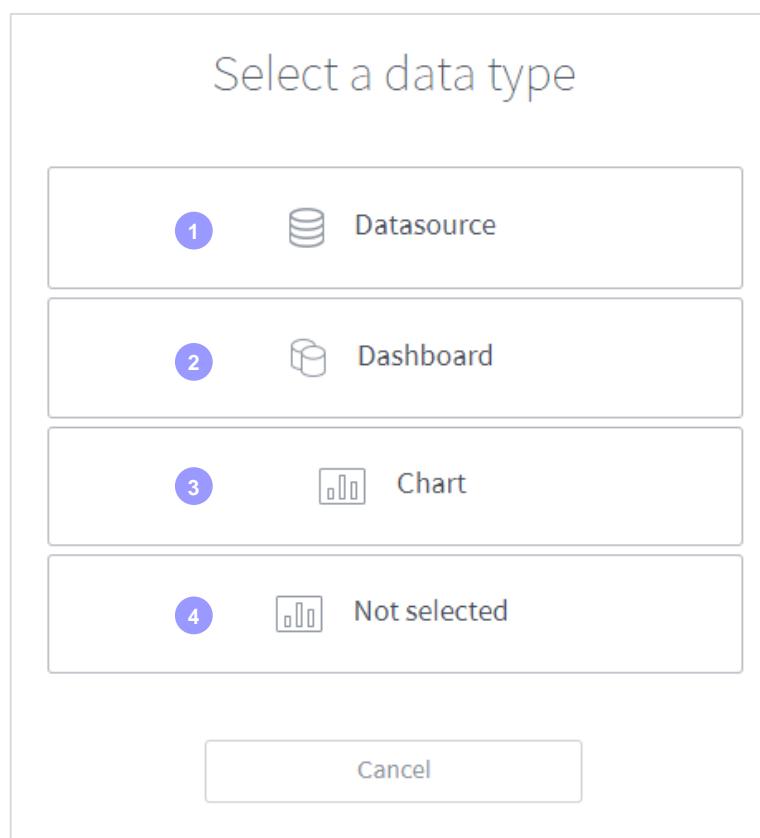
The notebook is created as follows.

1. Click the '+ Notebook' button at the bottom of the workspace. The screen to select the type of data required to create a notebook will be displayed.



+ Workbook + Notebook + Workbench

2. In the Notebook creation screen, select the type of data to be analyzed in the Notebook.



- 1 **Data source:** Load data from the data source allowed in the workspace and analyze it.
- 2 **Dashboard:** Load data from a dashboard in the workbook saved in the relevant workspace and perform an analysis.
- 3 **Chart:** Load data from a chart saved in the relevant workspace and perform an analysis.
- 4 **Not selected:** Select this option if you want Zeppelin to analyze in SPARK.

3. Select the data source, dashboard or chart to be analyzed, and click 'Next'.

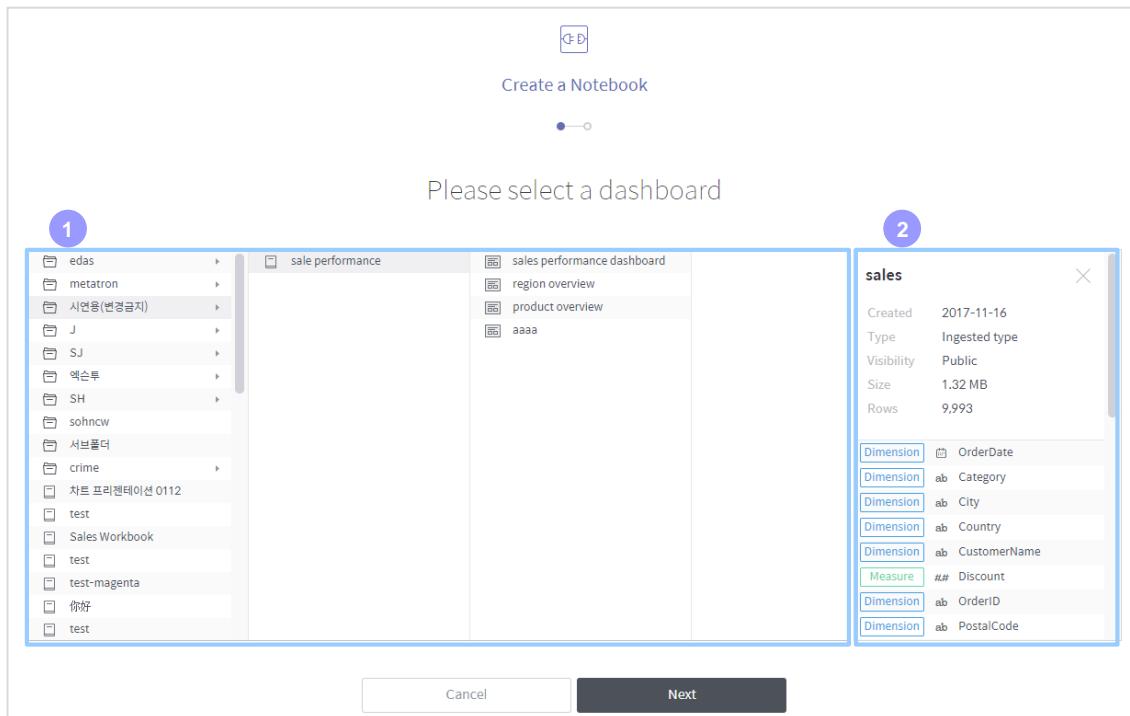
If 'data source' is selected as a data type

The screenshot shows the 'Create a Notebook' interface with the following components:

- Search Bar:** A search input field labeled 'Search by datasource name' (1).
- Filter Options:** Buttons for 'Only view open data' (2) and 'Type' dropdown (3) set to 'All'.
- Data Source List:** A table (4) listing data sources with columns: No., Datasource, Type, Used in, and Updated. The table contains 10 entries, with the last one, 'edas_sample_100_01', highlighted.
- Selected Data Source View:** A detailed view on the right for 'edas_sample_100_01' (5), showing:
 - Created: 2018-05-15
 - Type: Ingested type
 - Visibility: Public
 - Size: 31.40 MB
 - Rows: 3,700,900
 - Dimensions: edas_sample_100.event_time, edas_sample_100.dim_01, edas_sample_100.dim_02, edas_sample_100.dim_03, edas_sample_100.dim_04, edas_sample_100.dim_05, edas_sample_100.dim_06, edas_sample_100.dim_07
 - Measures: edas_sample_100.me_01
- Buttons:** 'Cancel' and 'Next' at the bottom.

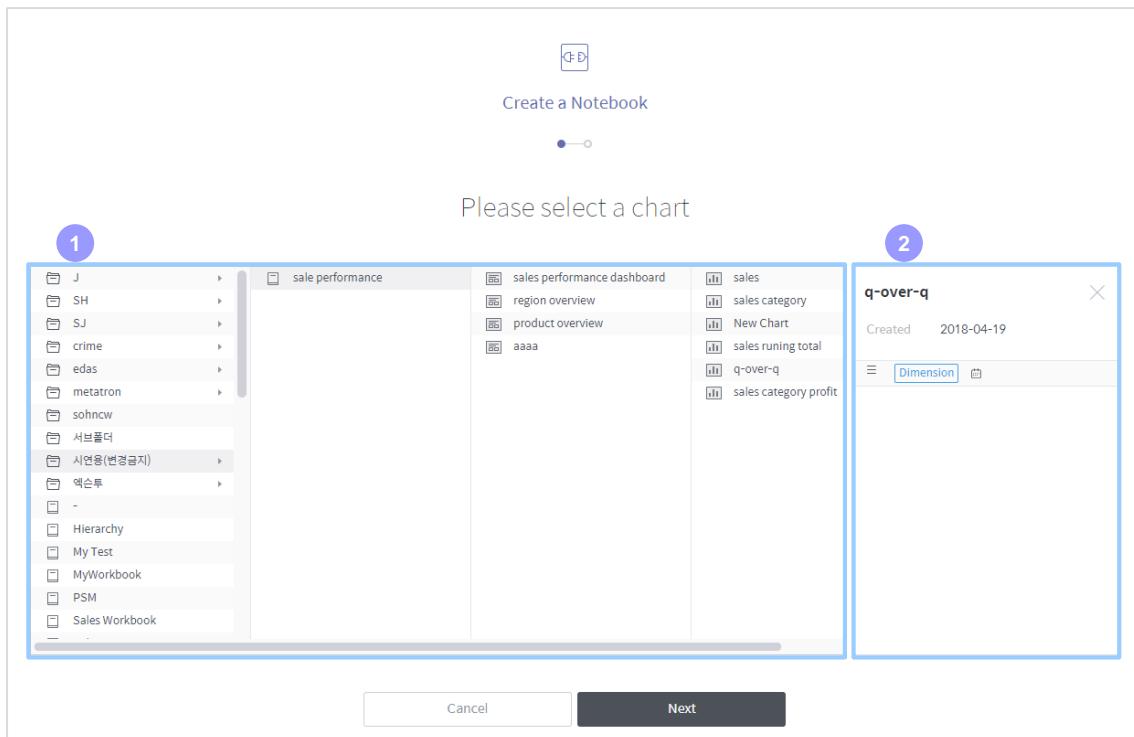
- 1 **Search by data source name:** Search a data source allowed in the relevant workspace by its name.
- 2 **View only open data:** Display only a data source designated as an 'open data source'.
- 3 **Type:** Display connection type or collection type data sources.
- 4 **Data source list:** Displays data sources that meet the established criteria.
- 5 **Data source information:** Displays an overview of the data source selected in the list.

If 'dashboard' is selected as a data type



- 1 Dashboard search area:** Select a workbook in the workspace to search dashboards within the workbook. Then select the dashboard containing data to be analyzed.
- 2 Dashboard information area:** Displays an overview of the dashboard selected in the search area.

If 'chart' is selected as a data type



- 1 **Chart search area:** Select a workbook in the workspace, and select a dashboard to search its charts. Then select the chart containing data to be analyzed.
- 2 **Chart information area:** Displays an overview of the chart selected in the search area.

4. Enter the information of the external analysis tool and language to be linked to the notebook, and enter the name and description of the notebook. Then click 'Done'.

The screenshot shows the 'Create a Notebook' interface. At the top, there's a 'Create a Notebook' button and a progress bar. Below that, a message says 'Please complete notebook creation'. The main area is divided into five numbered steps:

- 1: Chart (A preview of a chart titled 'sale performance > sales performance dashboard > q-over-q').
- 2: Server type (A dropdown menu set to 'zeppelin').
- 3: Develop language (A dropdown menu set to 'SPARK').
- 4: Name (A text input field placeholder 'Please enter a name').
- 5: Description (A text input field placeholder 'Please enter a description').

At the bottom, there are 'Previous' and 'Done' buttons.

- 1 **Data source:** The data source selected in the previous step.
- 2 **Server type:** Select Jupyter or Zeppelin as a server type. However, to do this, the relevant server type must be set in the initialized notebook server.
- 3 **Development language:** Select the development language for analysis. If you select Jupyter, you can analyze data with R or PYTHON, and Zeppelin analyzes data with SPARK.
- 4 **Name:** Enter the name of the notebook.
- 5 **Description:** Enter the description of the notebook.

5. Once a notebook is created, a user can see the new notebook in the workspace as follows.
Click to bring up a screen where you can use the notebook.

The screenshot shows the 'Admin workspace' section of the Metatron Discovery interface. At the top, there are tabs for 'Workbook 55', 'Notebook 14' (which is selected), and 'Workbench 42'. Below the tabs, it says '33 Datasource' and 'Created on 2017-11-16 by 관리자'. A search bar and filter options ('All') are also present. The main area displays a grid of 9 items, each with a checkbox, type, name, and last update time:

Type	Name	Last updated
notebook	test	a few seconds ago
workbook	edas-1	5 hours ago
workbench	workbench	a day ago
workbench	new	a day ago
workbook	test-magenta	a day ago
notebook	dfads	a day ago
notebook	edas	(no update time)
notebook	J	(no update time)
notebook	SH	(no update time)
notebook	crime	(no update time)
notebook	metatron	(no update time)
notebook	SJ	(no update time)
notebook	sohncw	(no update time)
notebook	시연용(변경금지)	(no update time)
notebook	엑스투	(no update time)
notebook	서브폴더	(no update time)

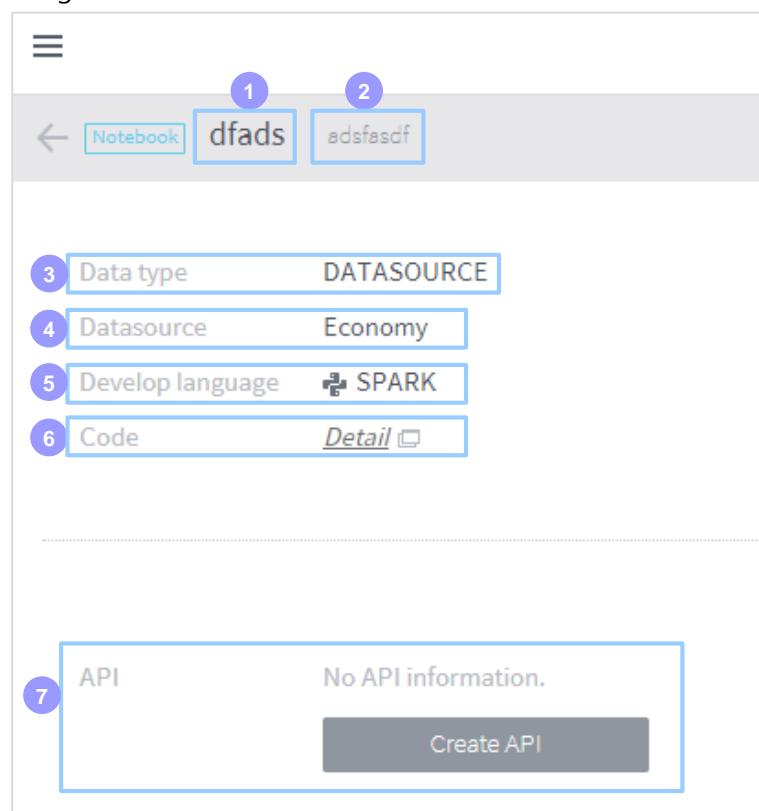
At the bottom, there are buttons for 'Select all', 'Clone Workbook', 'Move selections', 'Delete selections', and three buttons for creating new items: '+ Workbook', '+ Notebook', and '+ Workbench'.

4. Use Functions of a Notebook

Prepare the script in the newly created notebook that will be checked by exporting the result. Using the REST API, share the analysis codes that have been developed and the execution results with other users and systems.

4.1 Detailed notebook search

Select a notebook to be analyzed in the workspace. The detailed search screen will be displayed as follows. A user can search data type, data source name, development language and code entered when creating the notebook.



- ① **Name:** Name of the notebook.
- ② **Description:** Description of the notebook.
- ③ **Data type:** Data type used in the notebook (data source/dashboard/chart).
- ④ **Data source:** Name of data source/dashboard/chart imported from the notebook.
- ⑤ **Development language:** Development language used in the notebook.
- ⑥ **Code:** When clicked, the web-based notebook from the external tool (Jupyter / Zeppelin) that the user applies to the notebook will launch in a new window. In that notebook, a user can use the programming language to code. (See Paragraph 4.2.)
- ⑦ **API:** API Information for running script in the notebook. With this function, a user can run script on a regular basis. Click 'Create API' to display a screen to create API. (See Paragraph 4.3.)

4.2 Notebook coding

Click 'Details' in the code menu of the detailed notebook search screen. A new window for coding the notebook will appear. At the top of this window, a code to load the dataset is inserted. Execute the relevant cell to load a dataset in JSON into the dataset object.

The screenshot shows a notebook editor window with two sections:

- 1. load dataset**

```
In [ ]: library(RMetis)
```

```
In [ ]: dataset <- datasources.get(client.url('metatron.mcloud.sktelecom.com', 80), 'ds-37')
```
- 2. analyze**

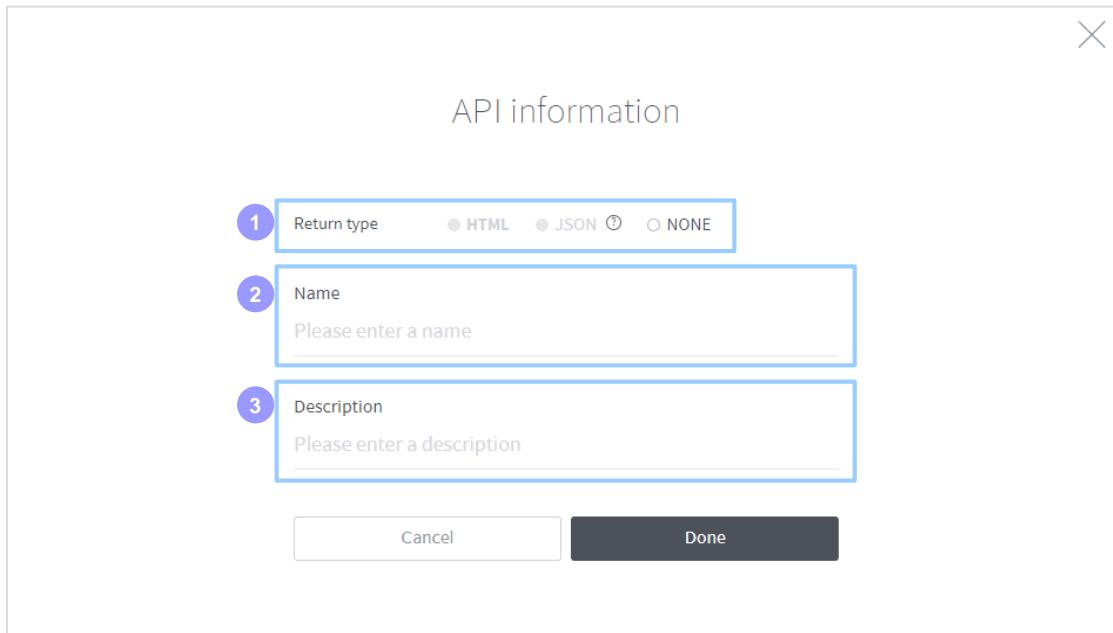
```
In [ ]:
```

The above screen is displayed when 'R' is selected as the development language. The same dataset loading cell is inserted in 'Python'. If you click "Run" in the middle of coding, you can check the coding result within the notebook module itself. After coding is complete, click "Save."

metatron Discovery does not provide a detailed description because it only implements interoperability with these open source external analysis tools. For more information on notebook coding languages and interfaces, see the documents on the relevant analysis tool and language.

4.3 Register a notebook API

Once a notebook is created, a user can return the result by calling the REST API. If the API is registered, a URL to search the relevant result will be generated, so analysts can share analysis logics with others.



1 Return type: Select the type to use to return the results of notebook.

- HTML: The screen displays all of the execution results of the notebook script in HTML.
- JSON: Returns a JSON object in a custom format created in a notebook script. In this case, the `response.write(...)` function provided by metatron Discovery will be used. The following is an example code for using the `response.write` function.
 - R-based notebook: `response.write(list(coefficient = 2, intercept = 0))`
 - Python-based notebook: `response.write({'coefficient' : 2.5, 'intercept' : 0})`
- None: Runs the notebook script but does not provide a return value.

2 Name: Name of the API to be registered.

3 Description: Description of the API to be registered.

Once a user has entered all the API information and clicked 'Done', the API creation will be completed, and the user can see the following REST API URL. Click 'Result' to search the URL execution result value in a pop-up window.

Name	RESTful API
Description	
URL	http://metatron.mcloud.sktelecom.com/api/notebooks/rest/354599d4-444a-43ab-b966-fdadffd12e7e
Return type	HTML
API result	Result
Edit API Delete API	



metatron 2.0

Part 6

Using Workbench



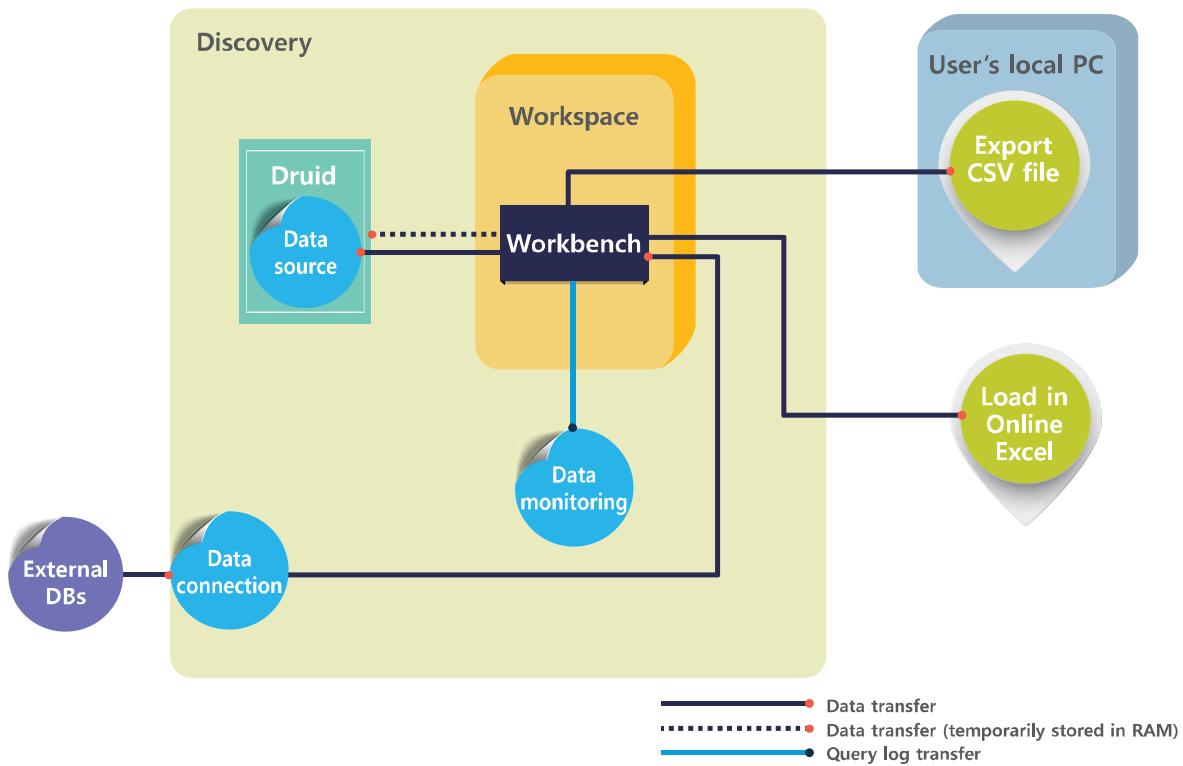
Copyright ©  SK telecom

The content of this document must not be copied,
distributed or used in part or in whole without the prior approval of SK Telecom.

Contents

1. Workbench Outline	3
2. Create a Workbench.....	4
3. Utilize a Workbench.....	6
3.1 Basic information area.....	6
3.2 Schema and table area.....	7
3.3 Query editor area.....	8
3.4 Query result area	9
3.5 Extra tools area	10
3.5.1 Global Variable Editing.....	10
3.5.2 Search Query History List.....	11
3.5.3 Workbench navigation	12

1. Workbench Outline



metatron Discovery Workbench provides an environment for data pre-processing and analysis based on SQL. The main functions are as follows.

- A user can search various external databases simultaneously using the work space.
- A user can search/select linked tables and columns easily, and open a view with more detailed information
- Query edit tools are embedded and query results can be checked in real time and used in various ways.
 - Query results can be downloaded as a local file or exported to an online Excel.
 - Query results can be immediately visualized to help the analyst see how the data tables look.
 - Query results can be stored as data sources so that they can be used on Workbook or Notebook for analysis.

This part is divided as follows.

Chapter 2 Create a Workbench: Each document storing SQL-based analysis query is called a 'Workbench'. This chapter describes how to create a Workbench and link a data connection to be analyzed.

Chapter 3 Utilize a Workbench: This chapter describes how to create/execute a query from the created notebook and visualize its results.

2. Create a Workbench

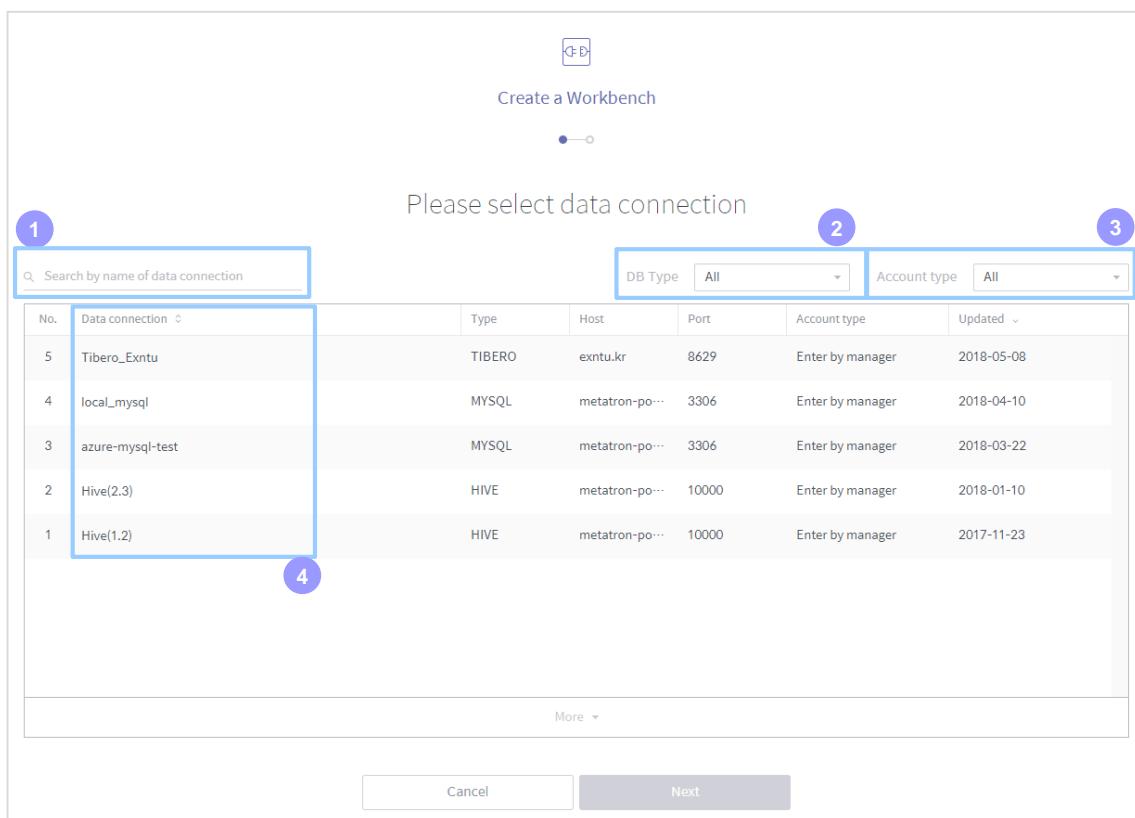
To use a workbench in the relevant workspace, a data connection for the Workbench must be established. For more information, please see '2.2 Data Connection' of 'Part 2 Management'.

A Workbench is created as follows.

1. Click the '+ Workbench' button at the bottom of the workspace. A screen to link a data connection for data analysis will be displayed.

+ Workbook + Notebook + Workbench

2. Select the data connection for the workbench to be linked and used by the user, and click 'Next'.



- 1 **Search by data connection name:** Search a data connection allowed in the relevant workspace by its name.
- 2 **DB Type:** See a data connection by database type (Oracle/MySQL/Hive/Presto/Tibero). Select 'All' to see data connections regardless of database type.
- 3 **Account Type:** See a data connection by the established account type (Enter by manager/Use user account/Input on connection). Select 'All' to see data connections regardless of account type.
- 4 **Data Connection:** Displays data connections that meet the established criteria.

3. Confirm the information of the selected data connection and enter a name and a description to create a workbench.

Create a Workbench

Complete workbench creation

Name	Hive(2.3)
DB Type	HIVE
Host	metatron-poc-h03
Port	10000

Name
New

Description
Please enter a description

Previous Done

4. Once the workbench is created, a user can see a list of new and existing Workbenches in the workspace screen as follows. Click the list to open a screen where you can use the functions of Workbench.

METATRON DISCOVERY

Admin workspace Owner

Workbook 53 Notebook 14 Workbench 42 | 32 Datasource

Created on 2017-11-16 by 관리자

Admin workspace

Search by content or folder name

All ▾

edas	metatron	시연용(번경금지)
J	SJ	엑스투
SH	sohnkw	서브폴더
crime		

new

Last updated a few seconds ago

workbook

test-magenta

Last updated an hour ago

notebook

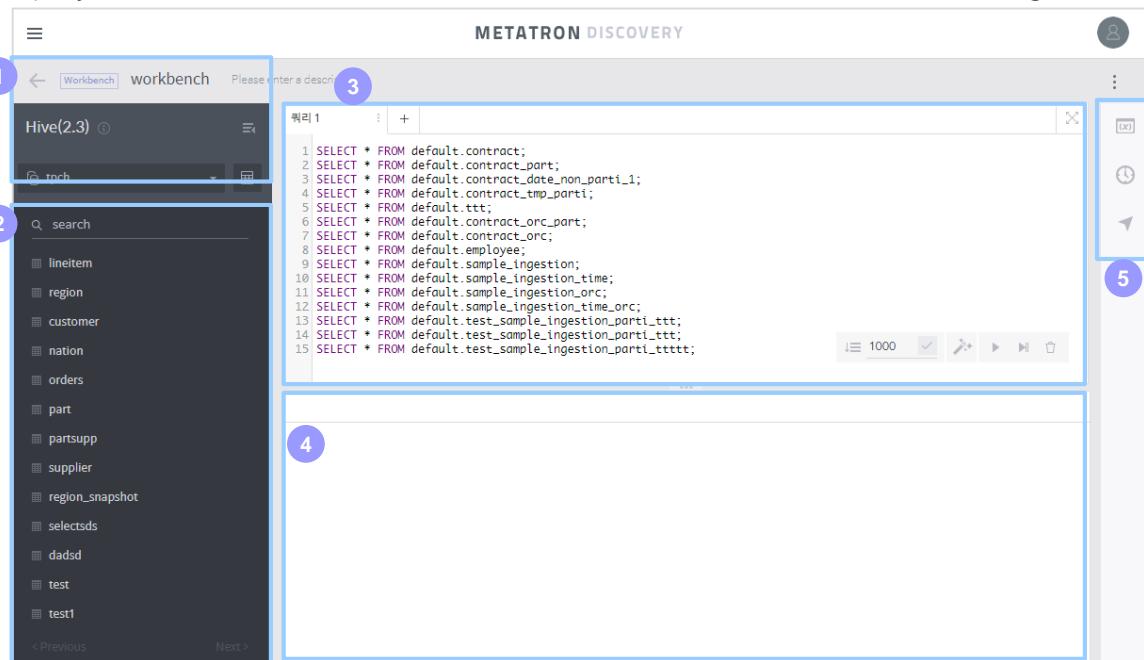
dfads

Last updated an hour ago

Select all | Clone Workbook | Move selections | Delete selections | + Workbook | + Notebook | + Workbench

3. Utilize a Workbench

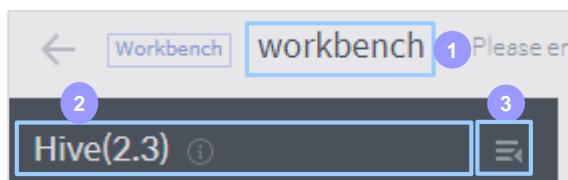
In the workbench, a user can edit and manage an SQL database easily, as well as visualize and save the query results in various forms. The workbench screen is divided into the 5 following areas.



- ① Basic Information Area (see Paragraph 3.1)
- ② Schema and Table Area (see Paragraph 3.2)
- ③ Query Editor Area (see Paragraph 3.3)
- ④ Query Result Area (Paragraph 3.4)
- ⑤ Extra Tools Area (Paragraph 3.5)

3.1 Basic information area

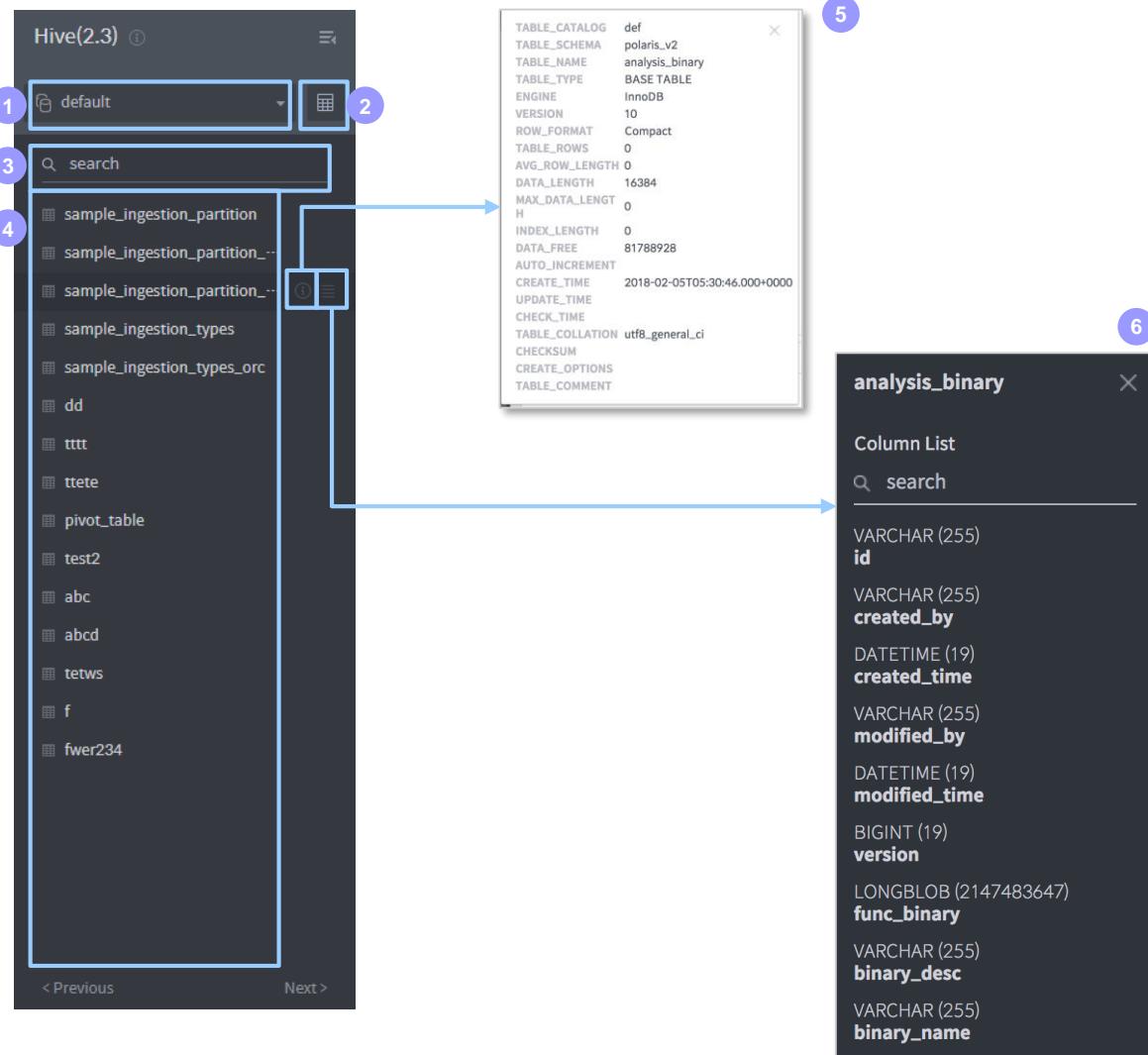
This area displays the information of the active workbench.



- ① **Name:** Name of the Workbench. Click it to change the workbench's name.
- ② **Data Connection:** Name of the data connection linked with the workbench. Click the ⓘ icon to see more information.
- ③ **=:** UI button to collapse or expand the panel

3.2 Schema and table area

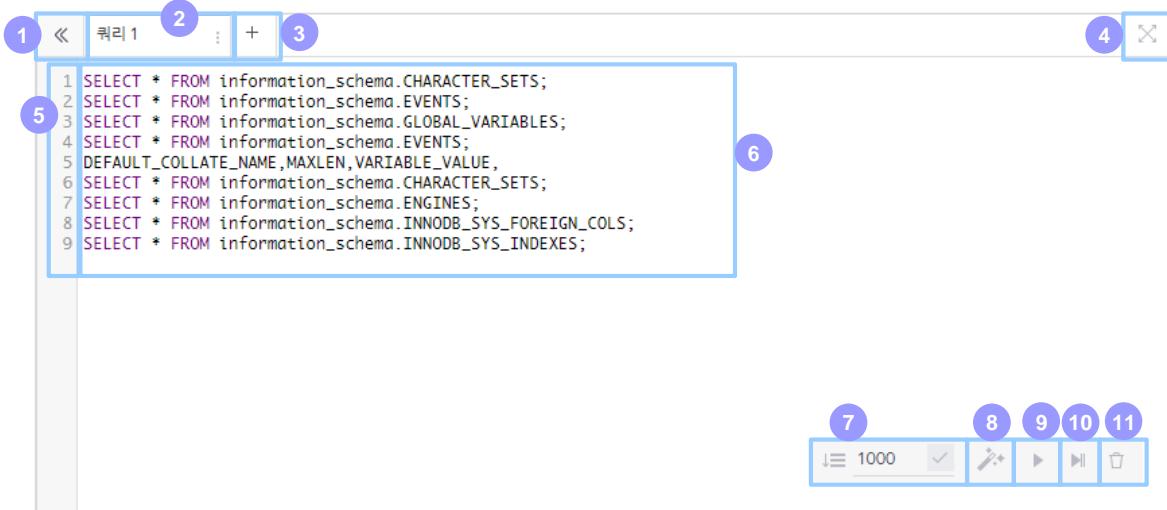
UI function to input a specific database, table and column in the query editor conveniently.



- ① **Database name:** Displays the name of the selected database. By default, the first table of a data connection registered in the workbench will be selected.
- ② **Database list:** Used to change the selected database. Click to search all databases included in the data connection. Select one database to replace the currently selected database.
- ③ **Schema information:** Displays the table list of the selected database, and information of all the columns and records in each table.
- ④ **search:** Search tables registered in the selected database by their name.
- ⑤ **Table name:** Select the table that contains the required data. Then the SELECT * FROM query for the table will be entered in the query editor on the right side automatically.
- ⑥ **Table information:** Displays basic table information.
- ⑦ **Column list:** Displays the name of all columns in the table and the data type of each column. Click a column name to enter it in the query editor automatically.

3.3 Query editor area

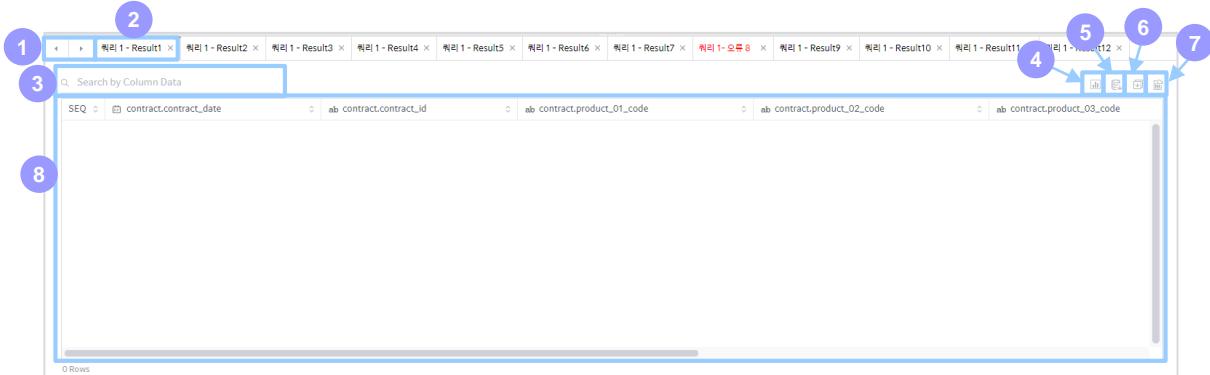
Editor screen to create and execute a query.



- 1 «: Displays previously created tabs if there are too many tabs. If there are not enough tabs, the user can see a list of current tabs. Click a tab to move to it.
- 2 **Tab:** To manage queries, a user can divide them into multiple tabs and execute or save a query. Click the :: button to edit a tab title or delete a tab.
- 3 +: Click this button to add a new tab.
- 4 ✖: Click this button to minimize or maximize the query editor area to full screen.
- 5 **Query row:** Displays the row number of query code.
- 6 **Editor screen:** Write the query statement in this area. A user can execute a multiple or single query. Insert ';' at the end of every query statement to divide queries to be executed. An auto-complete function is provided.
- 7 **Max. number of rows to be searched:** Set how many lines of query results will be displayed.
- 8 **SQL BEAUTIFIER:** Click this button to align the created query statements by query syntax standards. Select a query to be aligned and click the button to sort it.
- 9 **ALL SQL EXECUTE:** Execute all queries in the query statement. (Shortcut: Ctrl + Enter)
- 10 **SELECTED SQL EXECUTE:** Execute a specific query in the query statement where the mouse is located, or execute queries in the selected area by dragging the mouse. (Shortcut: Command + Enter)
- 11 **CLEAR SQL:** Clear all query statements.

3.4 Query result area

Once a query is executed, the result will be displayed in the query results tab. Every query result will be accumulated continuously, but the user can select and delete a specific results tab. The query result is displayed as a text grid, and Chart Preview, Save as Datasource, Export CSV file, and Online Excel functions are provided.



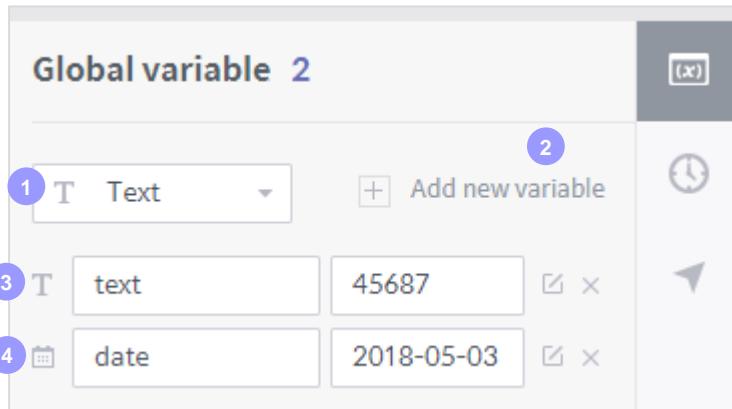
- ① : Displays previously created tabs if there are too many tabs for a screen. If there are not too many tabs, the user can see a list of current tabs. Click a tab to move to it.
- ② **Tab:** Name of the results tab. Click the button to delete a tab.
- ③ **Search by Column Data:** Search columns and values in the results.
- ④ **Chart Preview:** Draw a chart based on the query result. The chart is only for visualization; it is not reflected in the actual workspace. (For more information on how to control, see "Chapter 5 Create/Manage a Chart" in "Part 4 Workbook".)
- ⑤ **Save as Datasource:** Save the query result as a data source in the workspace. A pop-up to create a data source will be displayed, and procedures such as selecting a data connection and table will be replaced with the results of the workbench. As a result, the schema definition and ingestion cycle will proceed immediately. (For more detailed procedures, see Paragraph 2.1.3 of 'Part 3 Management').
- ⑥ **Export CSV file:** Download the query result as a local file (CSV).
- ⑦ **Online Excel:** Click the button to display the results in Online Excel. Online Excel is an online Excel application implemented in HTML5 named Spread Sheet. A browser containing the online Excel application will be opened in a new window, and not in the active workbench.
- ⑧ **Data History:** The data history output by executing a query. A user can copy the output data to the clipboard.

3.5 Extra tools area

The extra tools area is composed of useful tools for Workbench.

3.5.1 Global Variable Editing

If a certain syntax is used repeatedly and you have to execute queries by changing only its contents, set the syntax as a 'global variable' for convenient use.



- 1 **Variable type:** For a variable type, calendar and text are provided.
- 2 **Add new variable:** Select the variable type you want and click the 'Add new variable' button. The relevant global variable will be added in the query editor area.
- 3 **Name:** Enter the name of the variable.
- 4 **Variable value:** In calendar, select the date. In text, select the value.

3.5.2 Search Query History List

Search the history and results of an executed query or load that query.

The screenshot shows a 'Query History' list with the following columns: No., Time, Action, Resp., and Result. The 'Action' column displays various SQL queries. The 'Result' column indicates the status of each query. Row 30 is marked as 'Fail' with a red circle icon. A blue box highlights this row. Numbered callouts point to specific elements:

- 1: A search bar at the top left.
- 2: A delete all button in the top right corner.
- 3: The header of the main table.
- 4: A '+ More' button at the bottom left.

No.	Time	Action	Resp.	Result
37	05-16 16:08	SELECT * FROM default.contract	1 sec	Success
36	05-16 16:08	SELECT * FROM default.contract_ part	14 sec	Success
35	05-16 16:08	SELECT * FROM default.contract_ date_non_parti_1	1 sec	Success
34	05-16 16:08	SELECT * FROM default.contract_t mp_parti	1 sec	Success
33	05-16 16:09	SELECT * FROM default.ttt	1 sec	Success
32	05-16 16:09	SELECT * FROM default.contract_ orc_part	1 sec	Success
31	05-16 16:09	SELECT * FROM default.contract_ orc	1 sec	Success
30	05-16 16:09	SELECT * FROM default.employee	1 sec	Fail
29	05-16 16:09	SELECT * FROM default.sample_in gestion	1 sec	Success
28	05-16 16:09	SELECT * FROM default.sample_in gestion_time	1 sec	Success
27	05-16 16:09	SELECT * FROM default.sample_in gestion_orc	1 sec	Success
26	05-16 16:09	SELECT * FROM default.sample_in gestion_time_orc	1 sec	Success
25	05-16 16:09	SELECT * FROM default.test_samp le_ingestion_parti_ttt	1 sec	Success
24	05-16 16:09	SELECT * FROM default.test_samp le_ingestion_parti_ttt	1 sec	Success
23	05-16 16:09	SELECT * FROM default.test_samp le_ingestion_parti_ttttt	1 sec	Success

- 1 **Search query:** Search queries in the query history.
- 2 **Delete all:** Delete all query data in the query history.
- 3 **Query list:** Display information of queries executed after the last deletion of the query history in order of execution. Click a query in the list to copy the relevant query statement to the query editor.
- 4 **+ More:** Click this button to see more query data.

3.5.3 Workbench navigation

Used to move to another workbench. Click the target workbench to move.

Workbench Navigation (15)		
No.	Workbench name	Updated
42	workbench	2018-05-16
41	new	2018-05-16
40	metatron_metadata	2018-05-15
39	stage_03	2018-05-15
38	Tibero	2018-05-08
37	로컬 확인용	2018-04-30
36	aaa	2018-04-30
35	gg	2018-04-26
34	test-Magenta	2018-04-20
33	teddypark	2018-04-19
32	test_workspace	2018-04-16
31	경제지표	2018-04-09
30	111	2018-04-05
29	ddd	2018-04-05
28	test	2018-04-04

- 1 **Search a workbench:** Search workbenches saved in the workspace.
- 2 **Workbench list:** Display all workbenches saved in the workspace. Click a workbench in the list to move to that workbench.