SK telecom

metatron

Introduction of Metatron Discovery

SKT Metatron
Development Team

Contents


Intelligence Platform for Smart Factories
Big data- and AI-based technology development and commercialization
metatron
SK Telecom's Industrial Intelligence Platform
Machine Vision
Appliance
discovery
Anomaly
HMP

Metatron Discovery

Overview

metatron discovery

Big data- and AI-based data discovery and analytics

Performance Efficiency

Business Insights

Customer Understanding

Machine learning, big data, and visualization technologies are combined so that even non-experts can obtain highly valuable insights from the data easily.

Metatron Discovery – Overview

“Big data- and AI-based data discovery and analytics”

Big Data
(real time, batched)

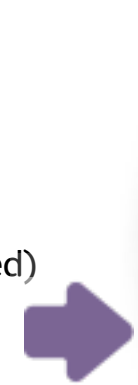
IoT Data

User Data

Application Data

Operation Data

...



metatron discovery

Managements

Tools

Engines

Anomaly detection

Cause analysis

Monitoring

Alarms

Time-series optimization
ML/DL analysis modeling
Real-time statistics

Real-time indicator management

Service fault analysis

Fraud detection

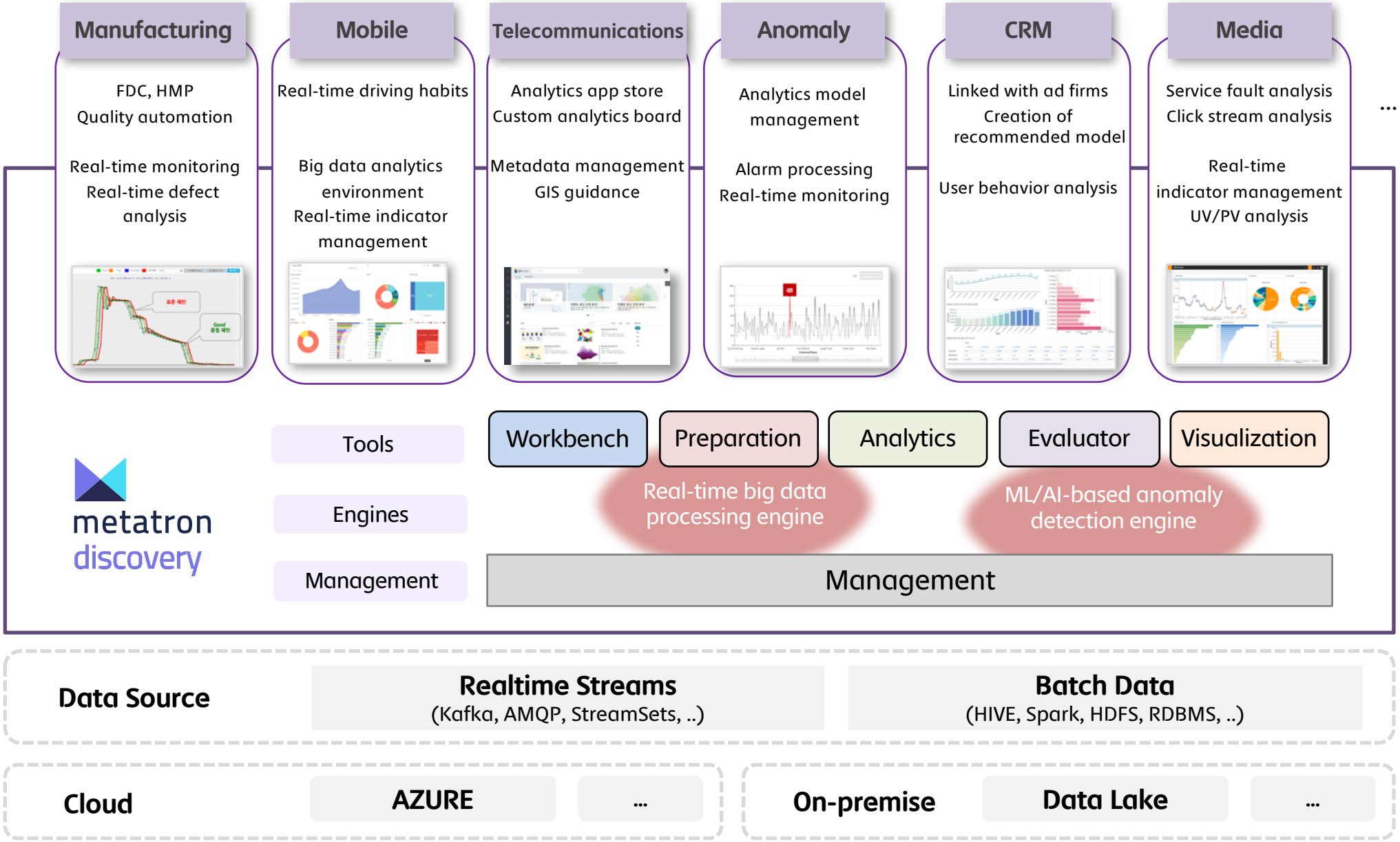
User behavior analysis (CRM)

APM
(Asset Performance Monitoring)

Click stream analysis

...

Metatron Discovery – Use Cases



Data Source

Realtime Streams
(Kafka, AMQP, StreamSets, ..)

Batch Data
(HIVE, Spark, HDFS, RDBMS, ..)

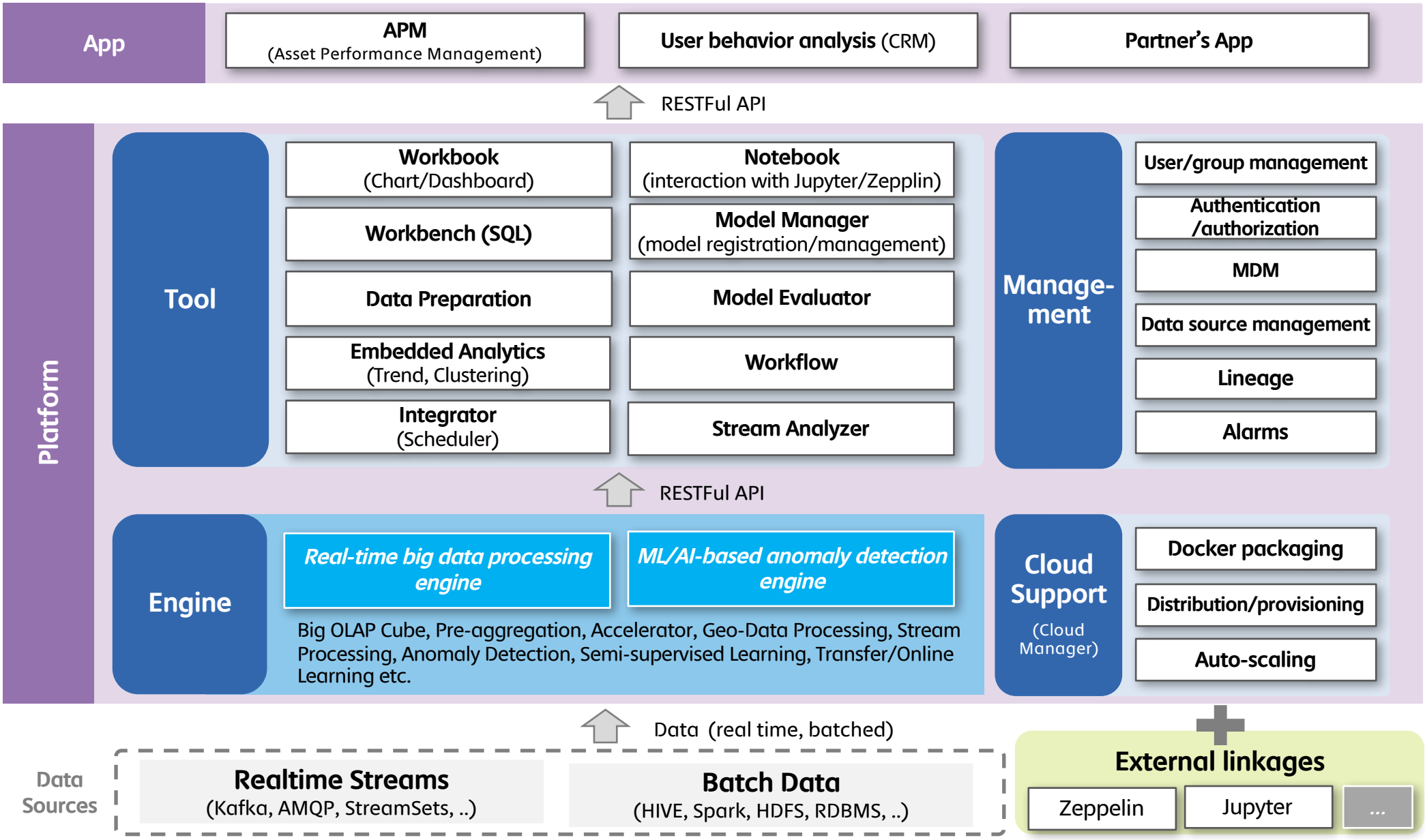
Cloud

AZURE ...

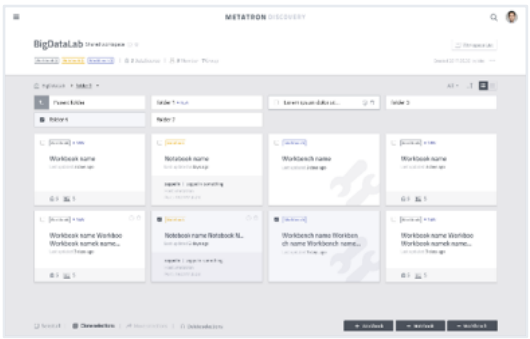
On-premise

Data Lake ...

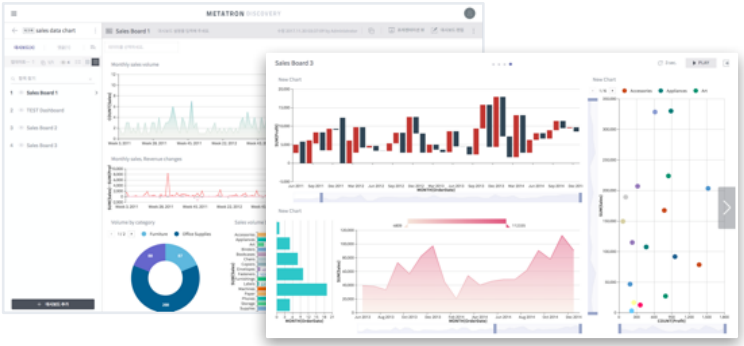
Metatron Discovery – Architecture



Metatron Discovery – Tools & Management



Workspace

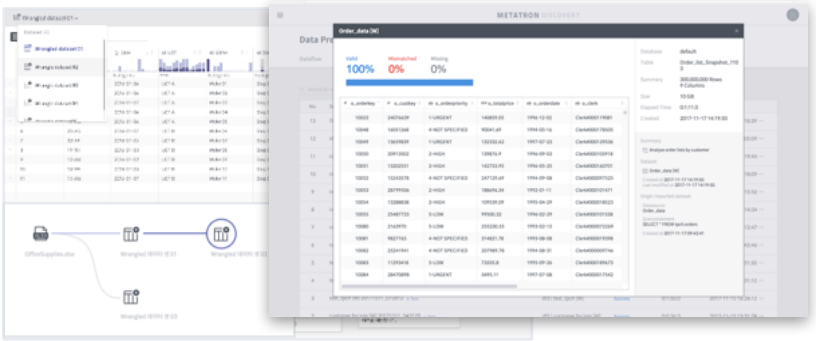


Notebook - DashBoard/Chart

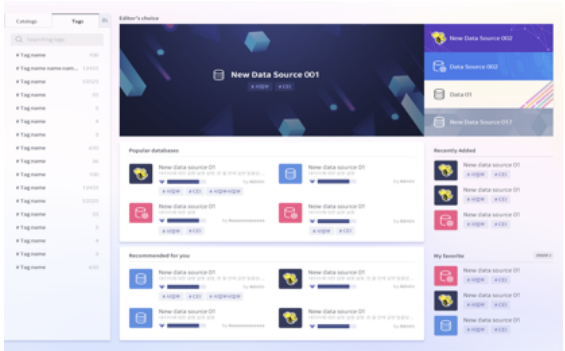
Workbench – SQL Explorer

Embedded Analytics

Data Lineage



Data Prep. – Data Transformer

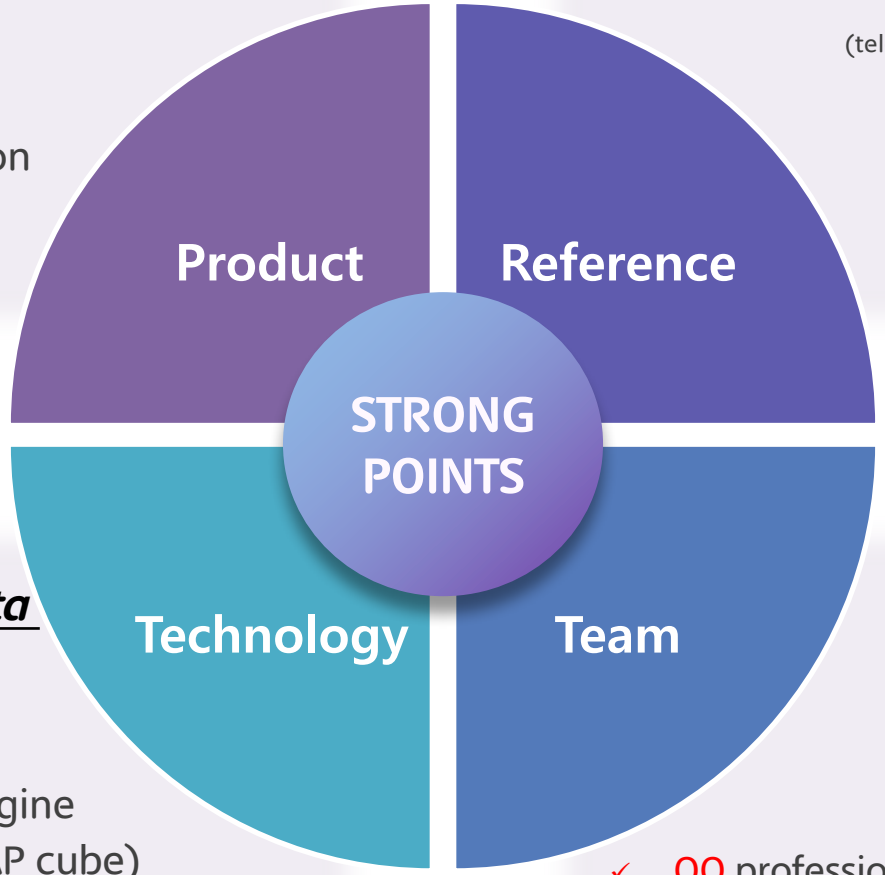


MDM

Metatron Discovery – Features and Strengths

Unified solution for data discovery

- ✓ Easy preparation
- ✓ Fast & intuitive visualization
- ✓ Powerful management



Tested with various use cases

- ✓ Applied to 0 fields
  (telecommunications, finance, manufacturing, etc.)
- ✓ Applied to 00 sites
  (site A, site B, site C, etc.)
- ✓ Capable of processing up to 0000 queries/day

Interactive time-series data processing and analytics optimization

- ✓ Sub-second processing engine
- ✓ Big data handing (big OLAP cube)
- ✓ Automated ML/DL support

World-class experts in the fields of big data, analytics and IT

- ✓ 0 PMC members
- ✓ 00 professionals with 10+ years of experience
- ✓ 0 experts with PhDs

Metatron Discovery – Roadmap

Until Now

V1.0 (~'16.4Q)	V2.0 (~'17.4Q)	V3.0 (~'18.4Q)
Core Functions	**End-to-end Functions**	**Enhanced Functions**
• Chart • Dashboard • User/Group Management • Data Source Management • Workbench	• UX 2.0 • Embedded Analytics (Trend, Prediction, Clustering) • Analytic Support (Zeppelin, Jupyter) • Data Management (Lineage & Job Log) • Data Preparation	• Data Management (MDM) • Vertical-specific Functions • Realtime Functions • Map Analysis • Data Collection Agent • Model Evaluator • Docker Packaging • Anomaly Detection

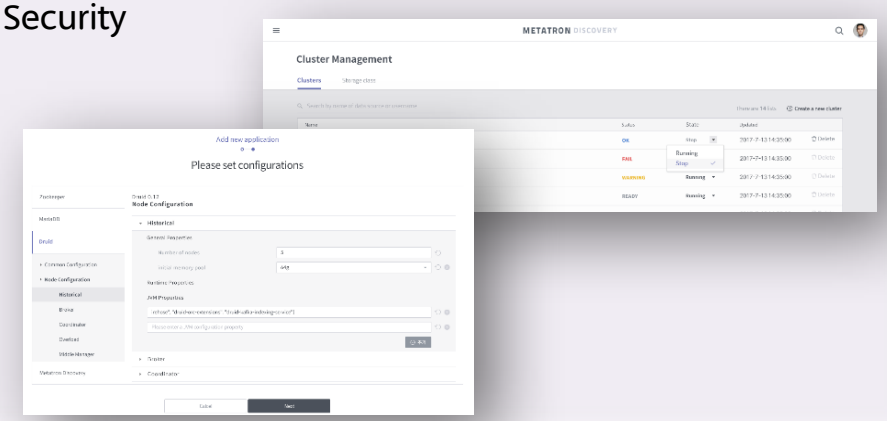
Next Step — *'19 ~ '21*

Cloud PaaS/SaaS	Enhanced Analytics (ML/AI)	Enhanced Dev. Support
▪ Metering/Charging ▪ High Availability ▪ Security	▪ Automated ML/DL Support ▪ Advanced Analytics ▪ Predictive/Prescriptive	▪ Enhance APIs ▪ Pluggable Architecture ▪ Community

Metatron Discovery – Next Step

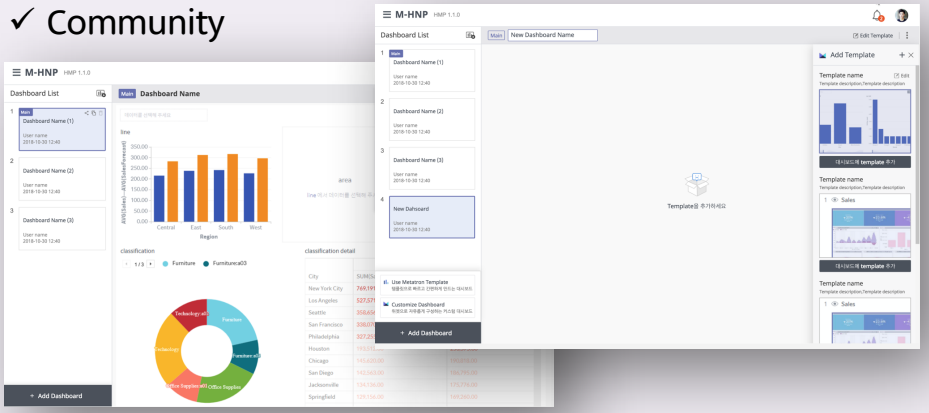
Cloud PaaS/SaaS

- ✓ Metering/Charging
- ✓ High Availability
- ✓ Security



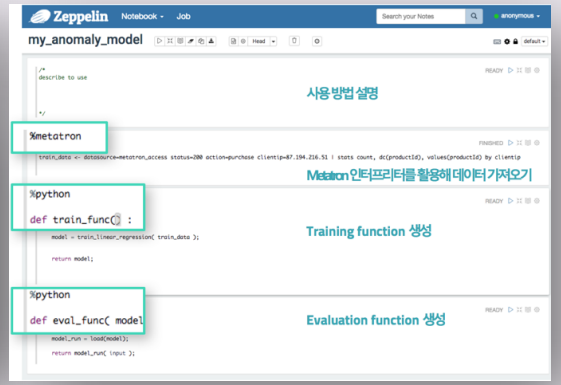
Enhanced Development Support

- ✓ Enhance APIs
- ✓ Pluggable Architecture
- ✓ Community



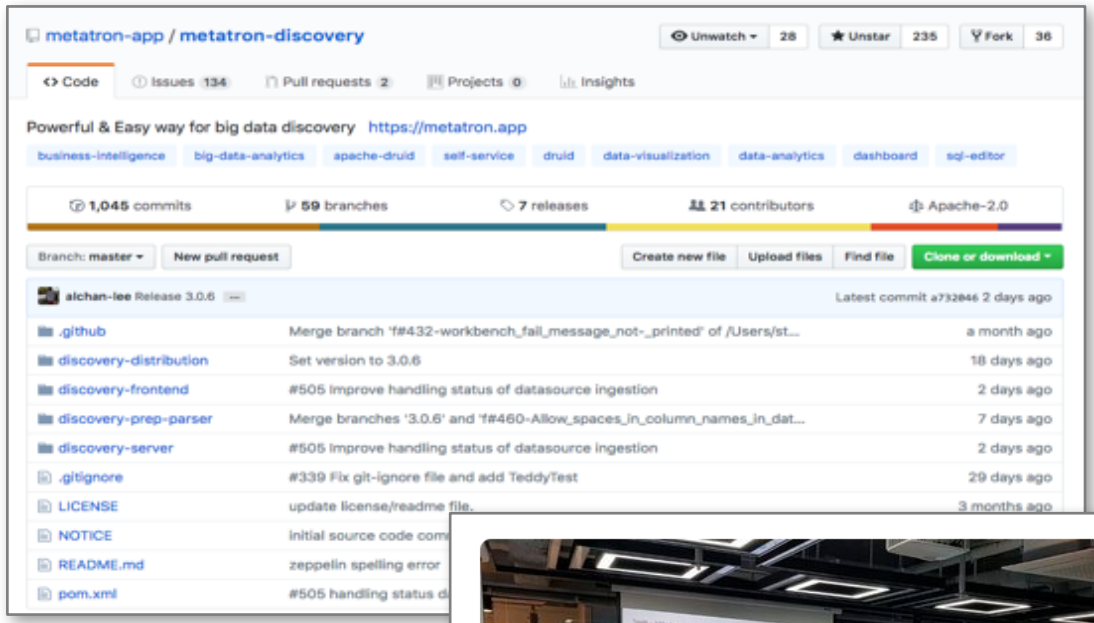
Enhanced Analytics (ML/AI)

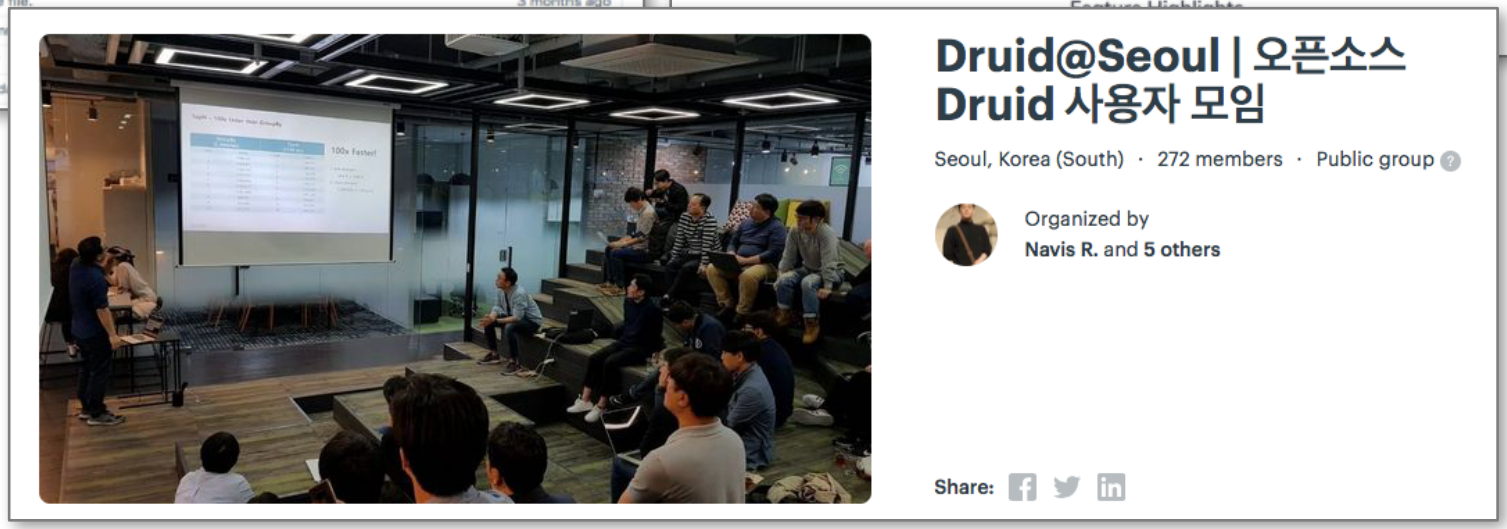
- ✓ Automated ML/DL Support
- ✓ Advanced Analytics
- ✓ Predictive/Prescriptive
- ✓ Co-working with Zeppelin



Metatron Discovery – Open Source & Activities

- Website: https://metatron.app/
- Source code: https://github.com/metatron-app/
- Board meetings held offline once a month to share information on version releases, discuss plans for future releases, etc.
- Druid meetups held offline once a quarter, where developers and users share about Druid technology and application experiences, etc.
  https://www.meetup.com/Druid-Seoul/





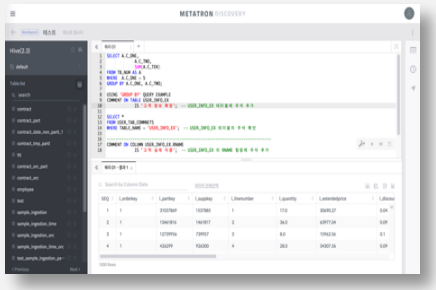
Metatron Discovery

Use Cases

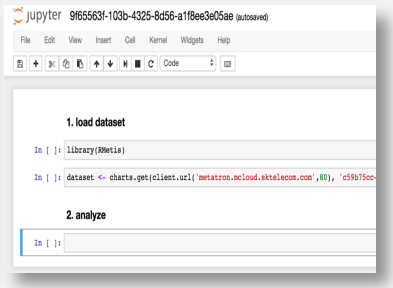
Use Case – Building an N/W Data Analytics Portal (Telecommunications)

metatron discovery

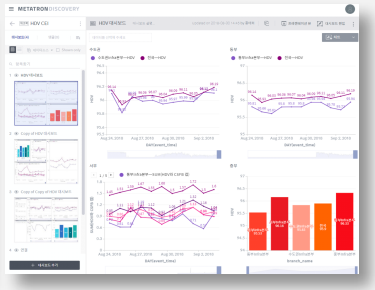
SQL



Analytics



Visualization



Tools

Workbench | **Notebook** | **Workbook, ...**

Management

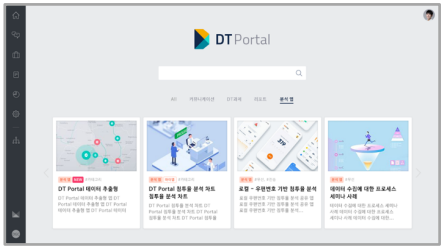
MDM, ...

Real-time big data processing and analytics engine

DW/Mart Data

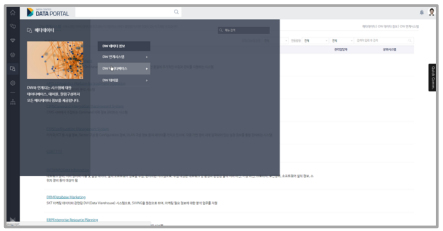
Hadoop DW Cluster
(Data warehouse with integrated infrastructure center)

Portal



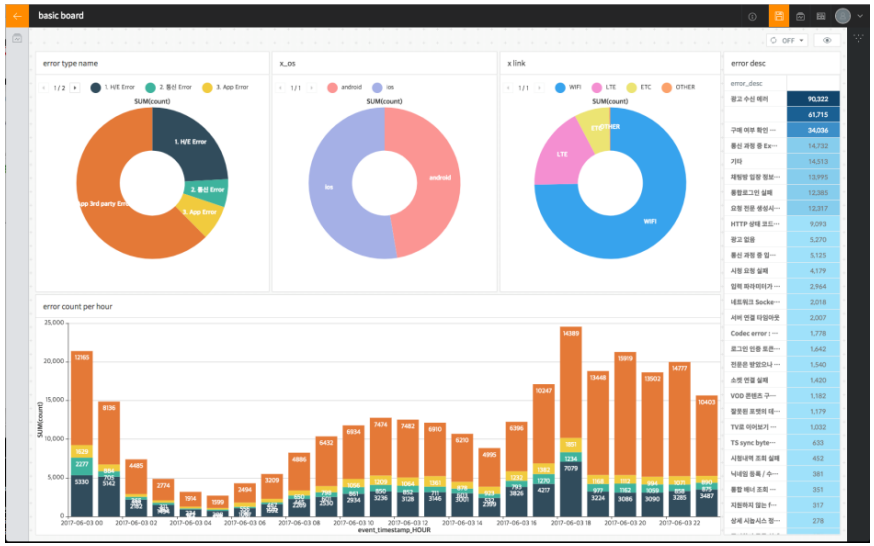
Analytics app place

Metadata search

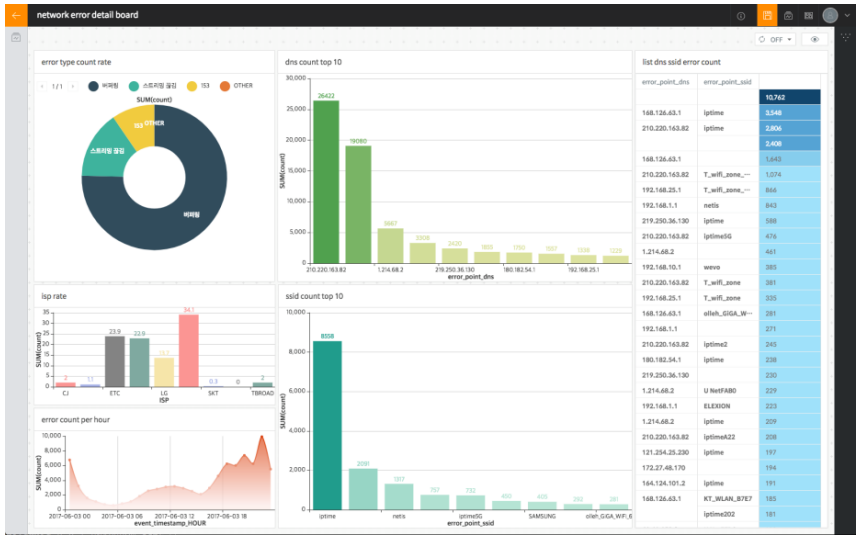


Use Case – Log Analysis

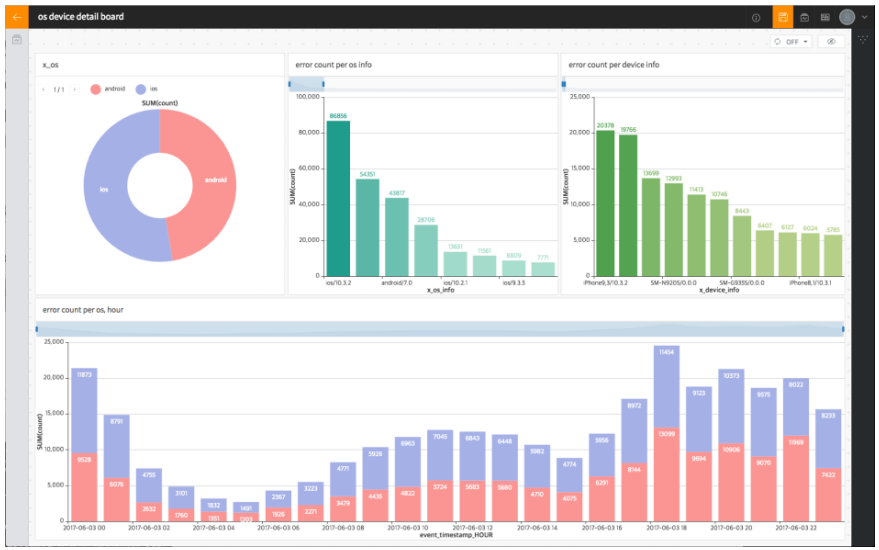
- Error monitoring



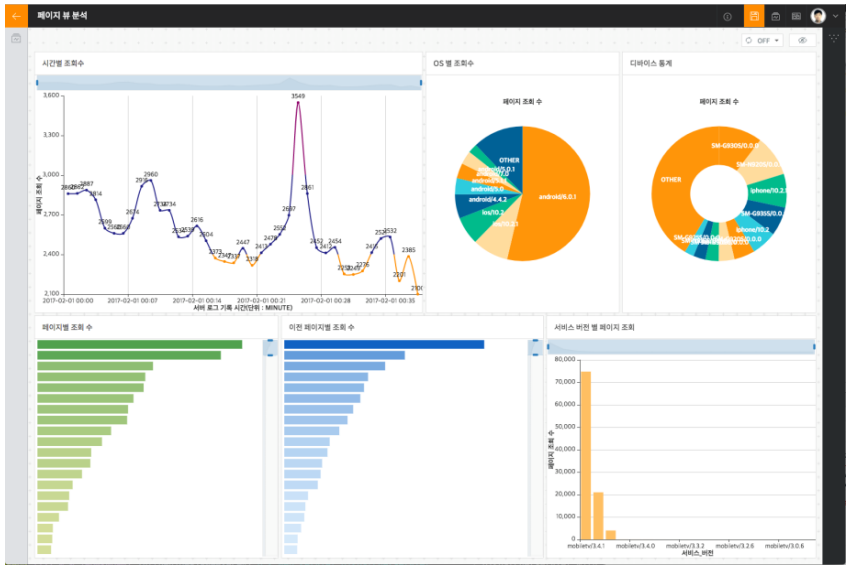
- Network error detail board



- OS device detail board



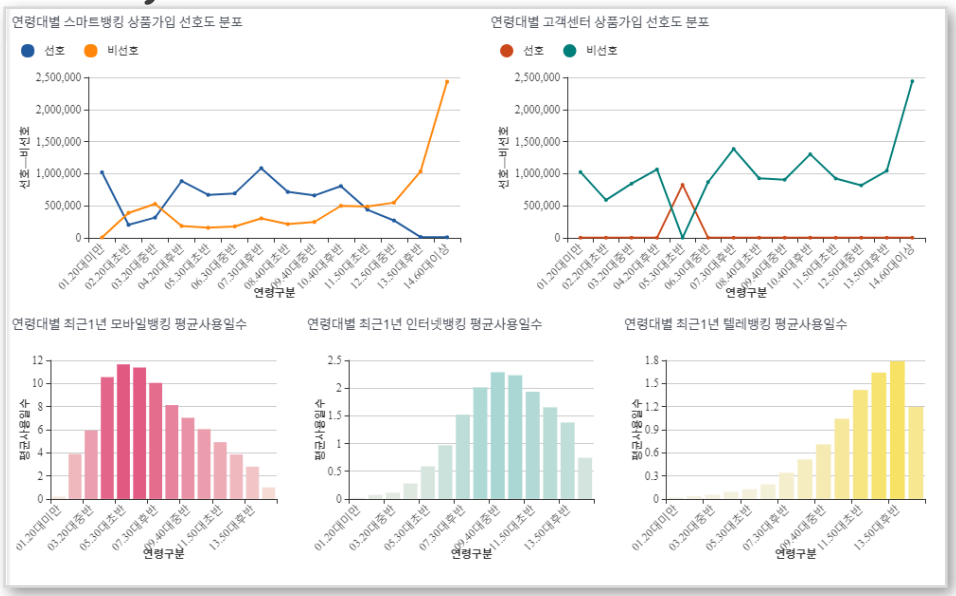
- Page view analysis



Use Case – Finance

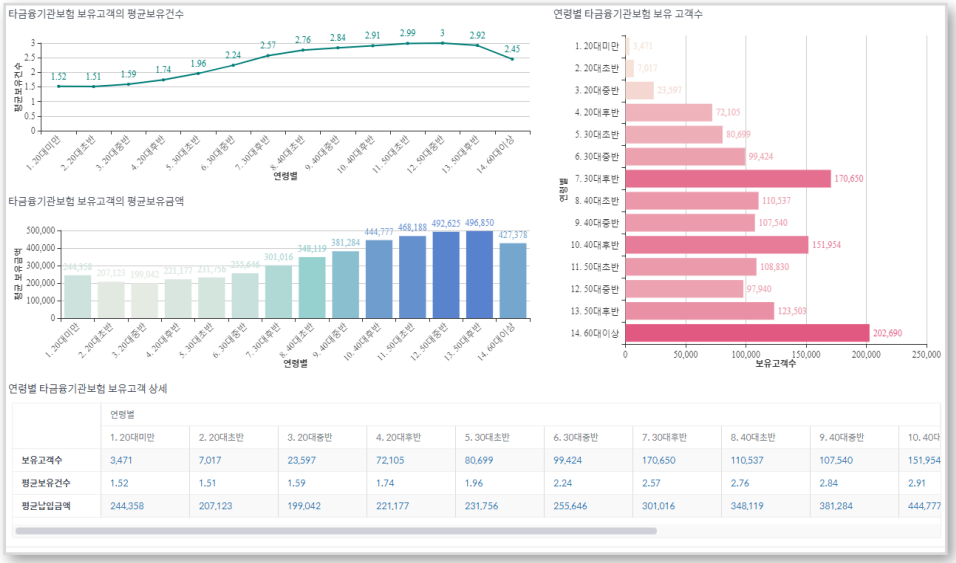
Analysis 1

Analysis 2



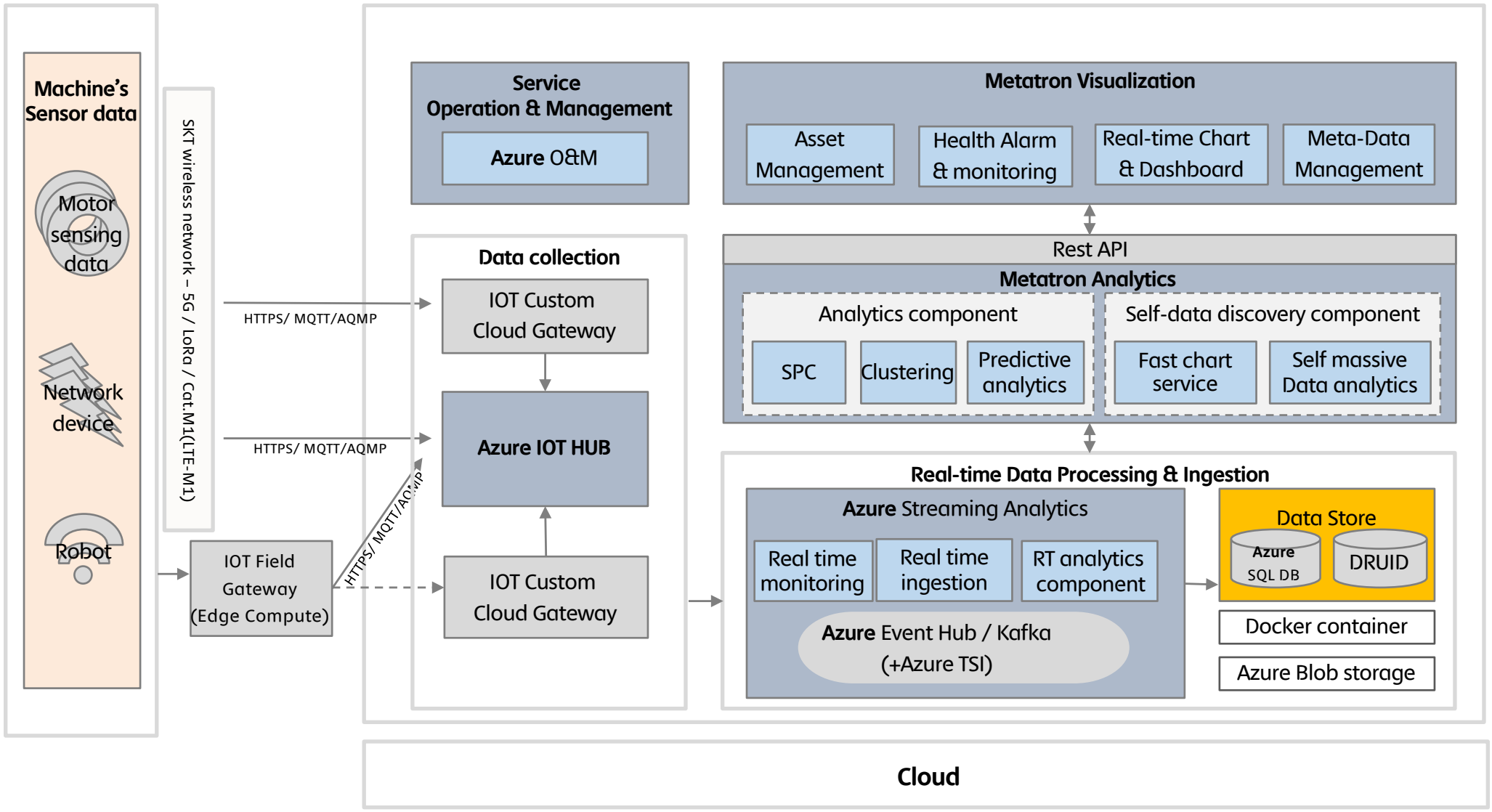
Analysis 3

Analysis 4



Use Case – Manufacturing

Machine Health Monitoring & Predictive Maintenance Analysis Service

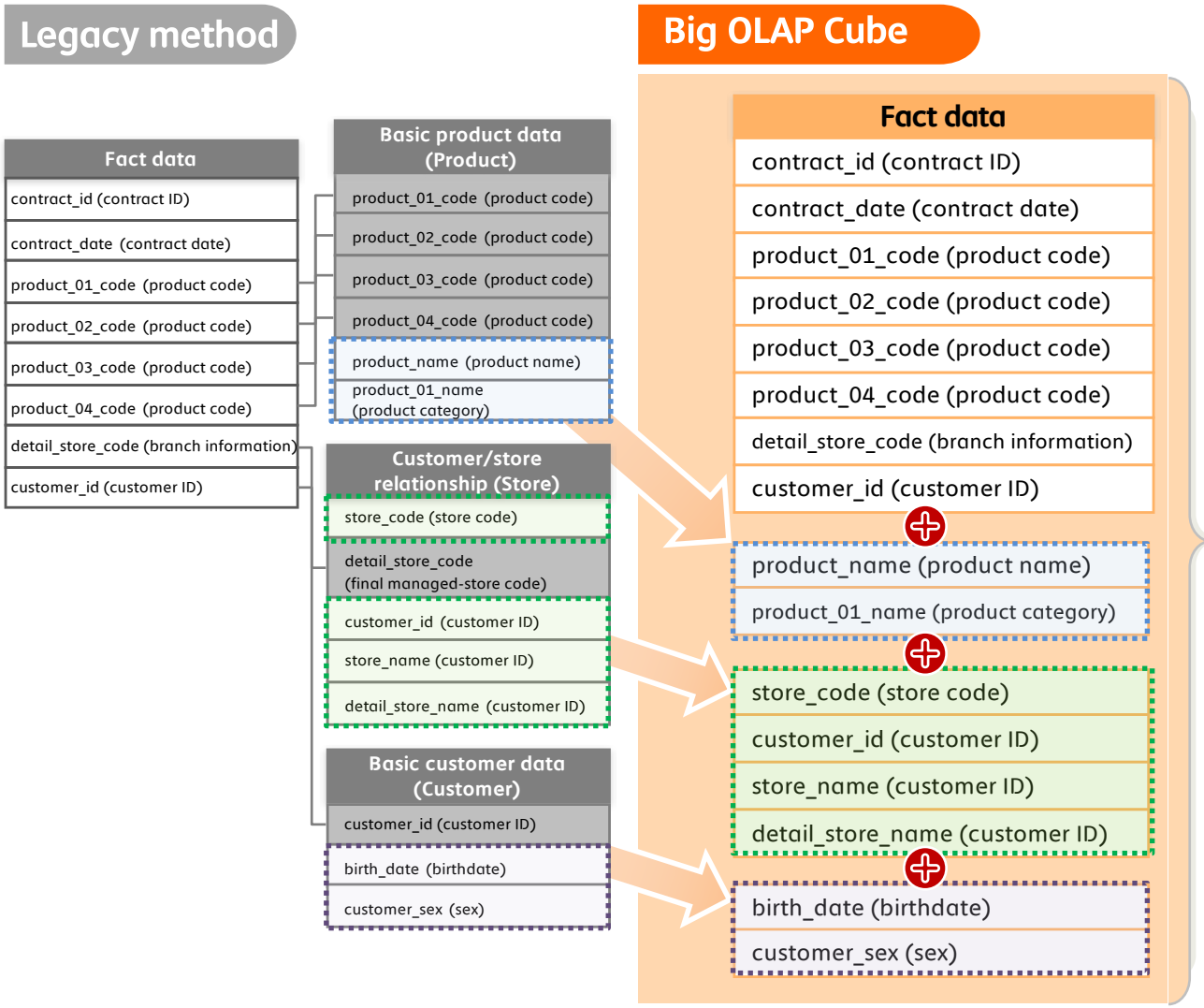


Metatron Discovery

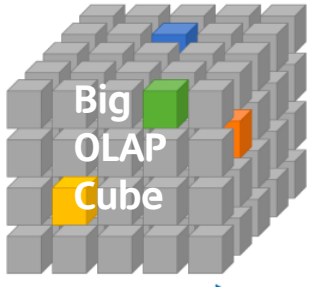
Technology

Major Features - Big OLAP Cube

❖ Creation of one big OLAP cube (data mart) by combining various dimension data based on a large amount of fact data



druid



Joined to create a big OLAP cube

Advantage 1 ETL expense minimization

- Reduce ETL expenses for data mart creation by minimizing the number of data marts
- Minimize the effects of structural changes by reducing the volume of mart data
- Capable of meeting diverse demands by saving all fact data

Advantage 2 Fast speed and free schema changes

- A distributed architecture enables storing large amounts of data and ensures fast data processing
- With a dynamic schema approach, schema changes do not require schema redefining.

Advantage 3 Real-time processing

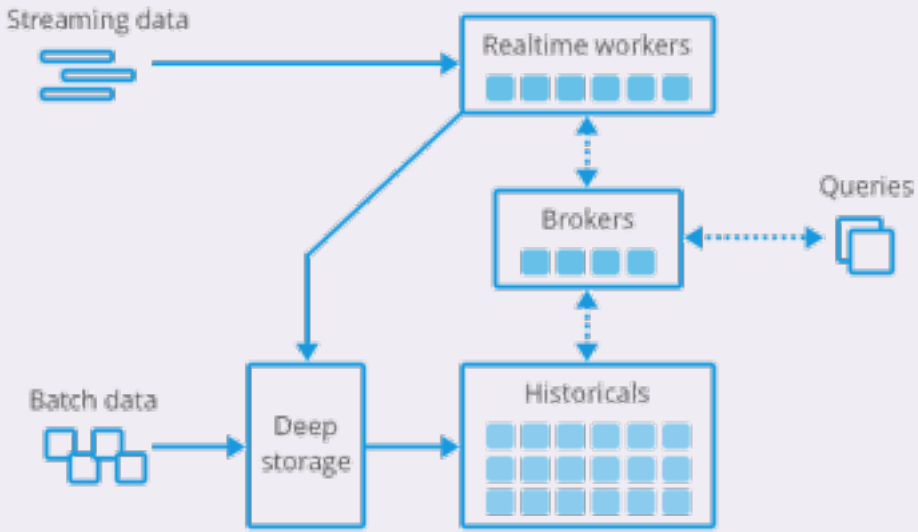
- Can process data at the record level in real time as tables are stored with no data loss

Major Features - Sub-Second Processing Engine (Optimized Druid for Metatron)

- Druid is a data processing engine optimized for time series data that was developed under the leadership of ImplyData and MetaMarkets starting in 2012.
- SKT developed core functions required for Metatron, and now has branched off as an independent project.

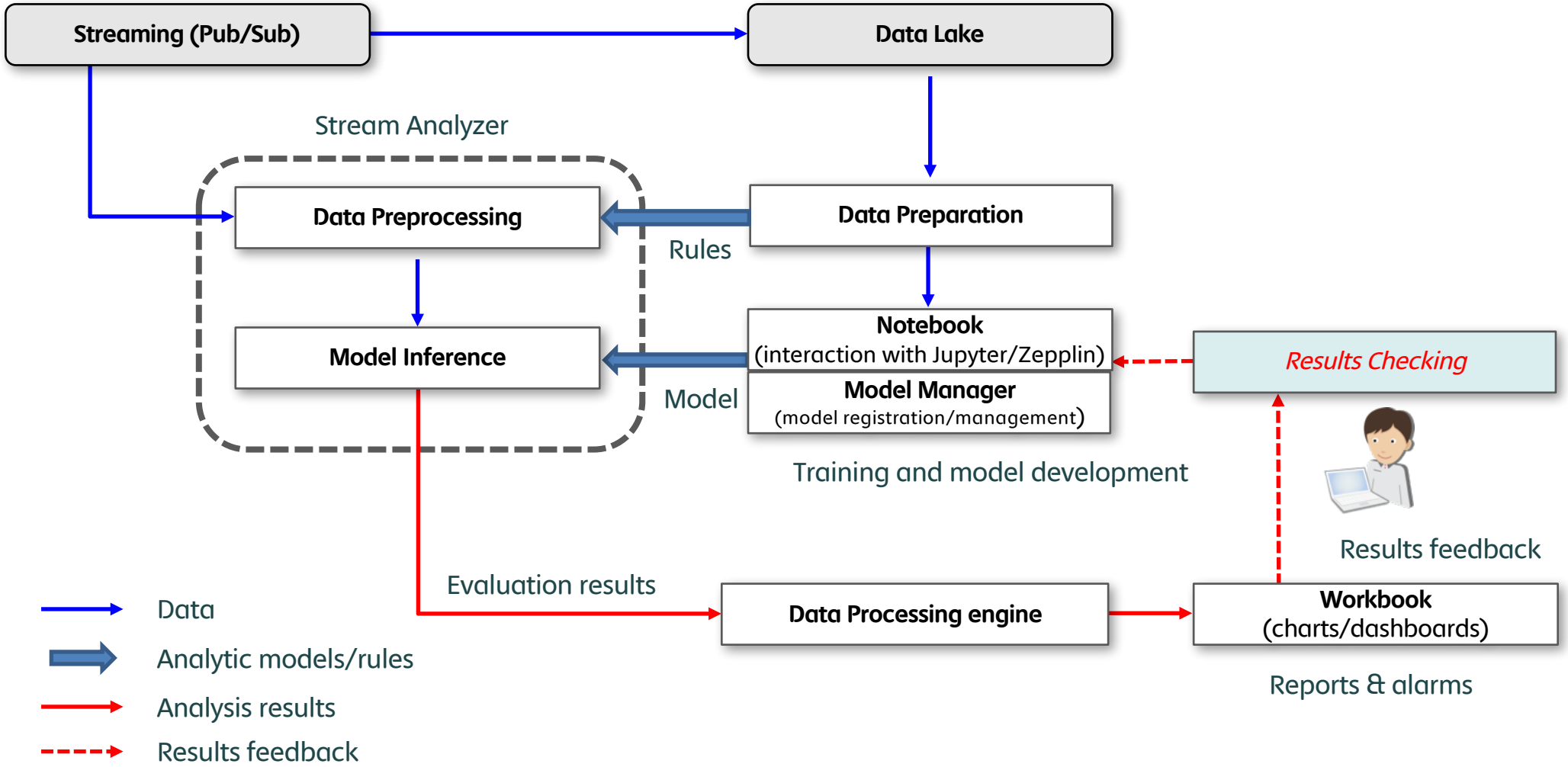
- **Development in open ecosystem**
  - Hosts Druid user meetups in S. Korea
  - Contributes opinions and codes (300+) to Druid on Github and Google Groups
- **Not just simple use of open-source code, but in-house development of core functions, performance improvements, etc, (under a separate branch) including:**
  - Joins between data sources
  - Data search (Lucene index, spatial index)
  - Extended query set (sketch queries)
  - Query statistics
  - Window functions
  - Virtual column map type
  - Multi-valued metric
  - ...

- Supports both real-time and batch processing
- Over time, data is transferred to memory, local storage, and then deep storage, by which the engine is capable of even terabytes of big data.
- Each function module (query processing, storage, and indexing) can be maintained on separate servers, which can be scaled out to accommodate increasing usage.



Major Features – Automated ML/DL Support

- Automation of the entire process from ML/DL-based analytics model development to commercial application
- Reduced development schedules and lower costs by eliminating development processes for fitting a newly developed model into the system
- Employment of incremental learning, etc. to build systems that gradually improve over cycles

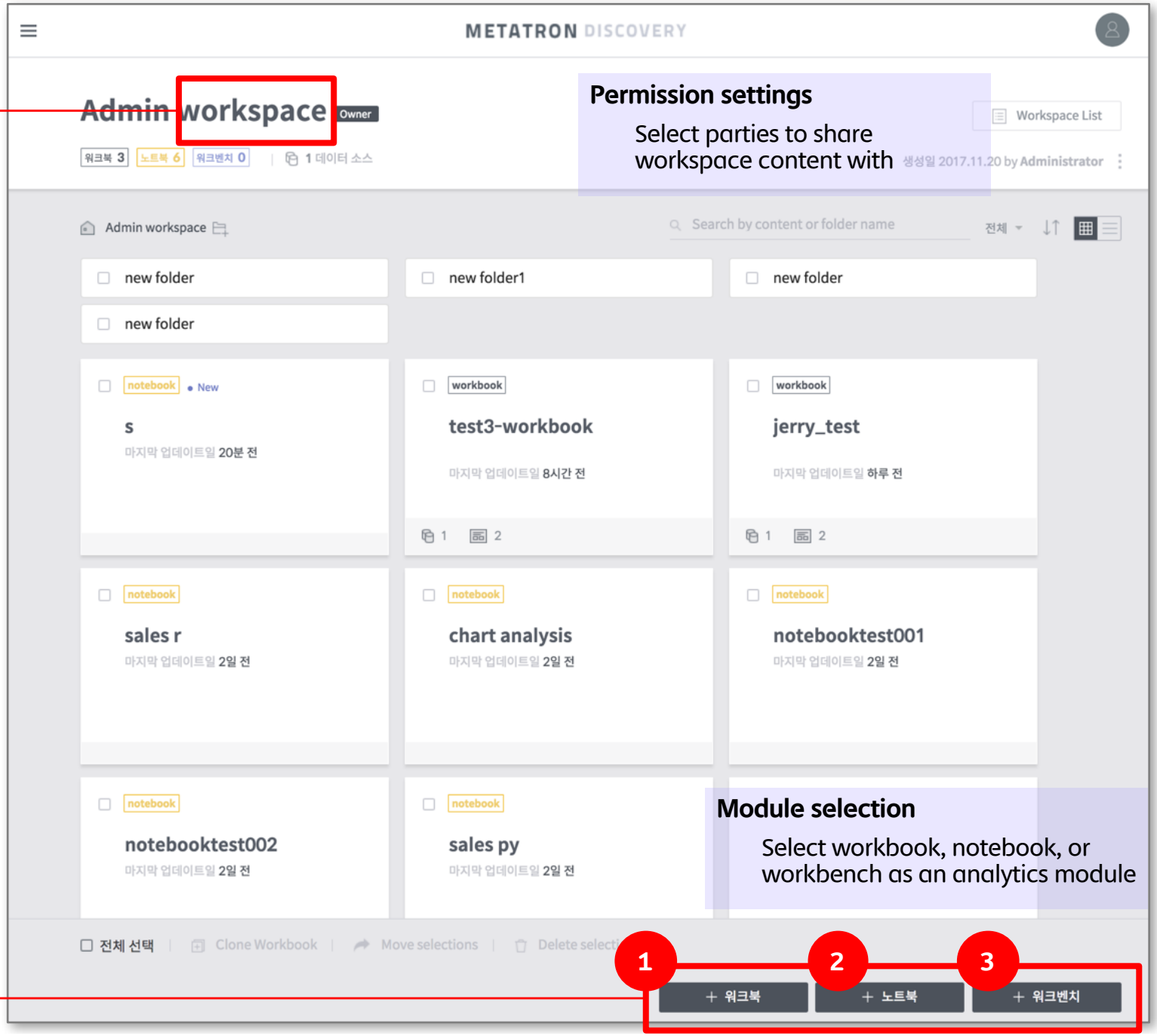


Metatron Discovery Functions

https://metatron.app/index.php/documents/

The Analytics Environment – Workspace

Workspace:
Analytics space
(personal, shared)



1. Workbook:
   Visualization-based analytics

2. Notebook:
   ML-based advanced analytics

3. Workbench:
   SQL-based analytics

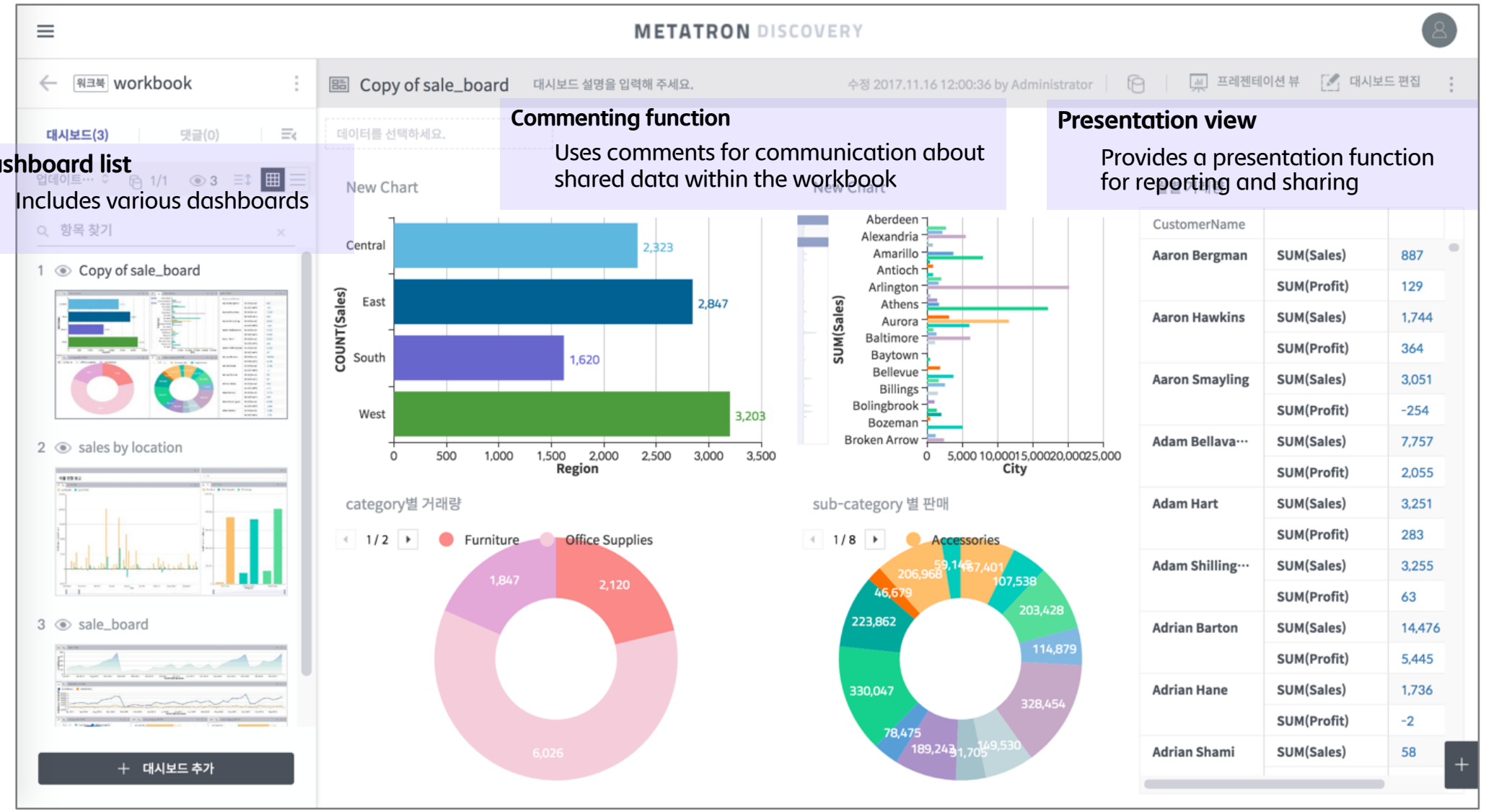
1 Workbook – Visualization-Based Analytics

■ The PowerPoint-type environment provides a familiar and attractive user experience.

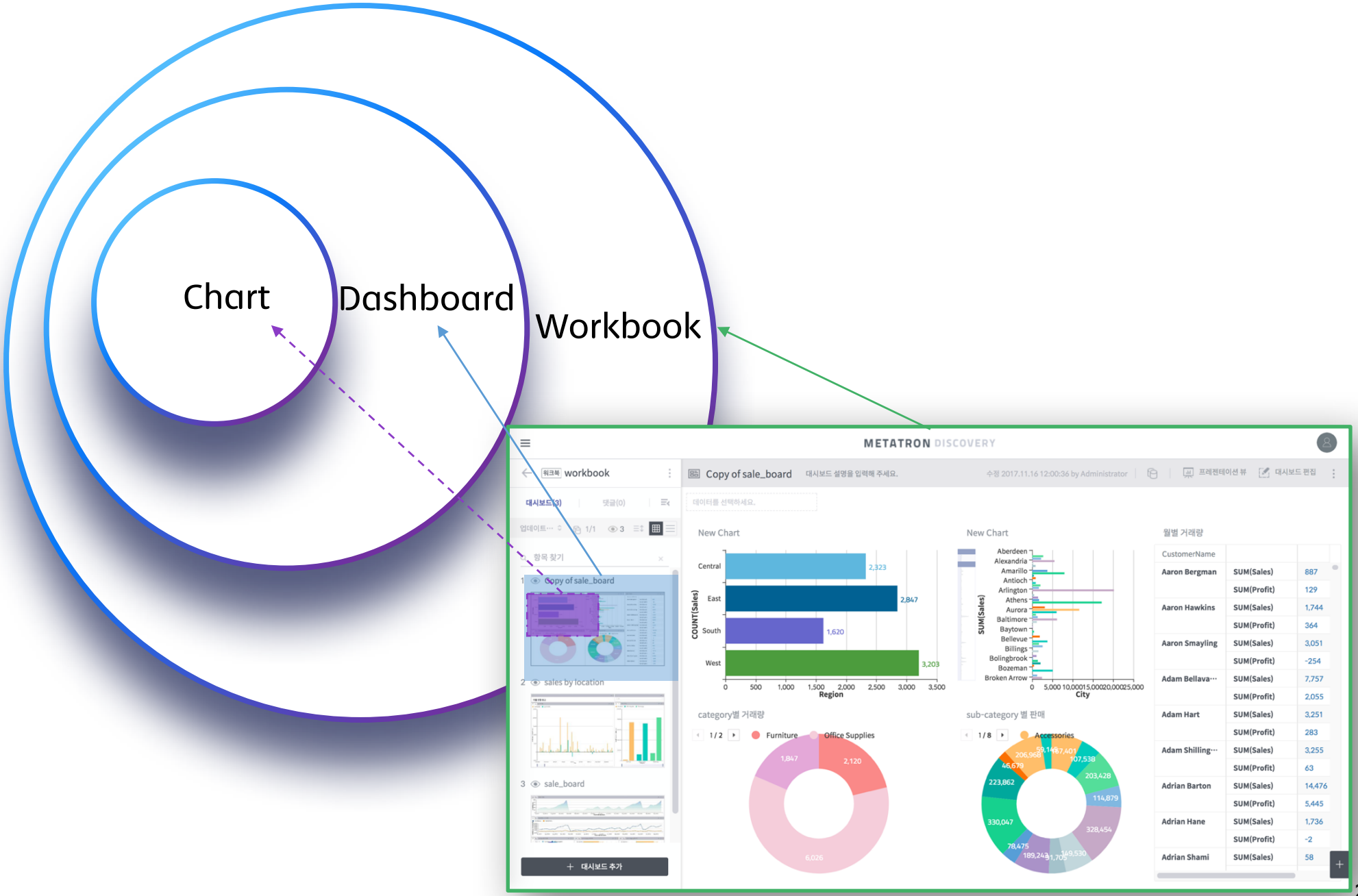
Dashboard list
Includes various dashboards

Commenting function
Uses comments for communication about shared data within the workbook

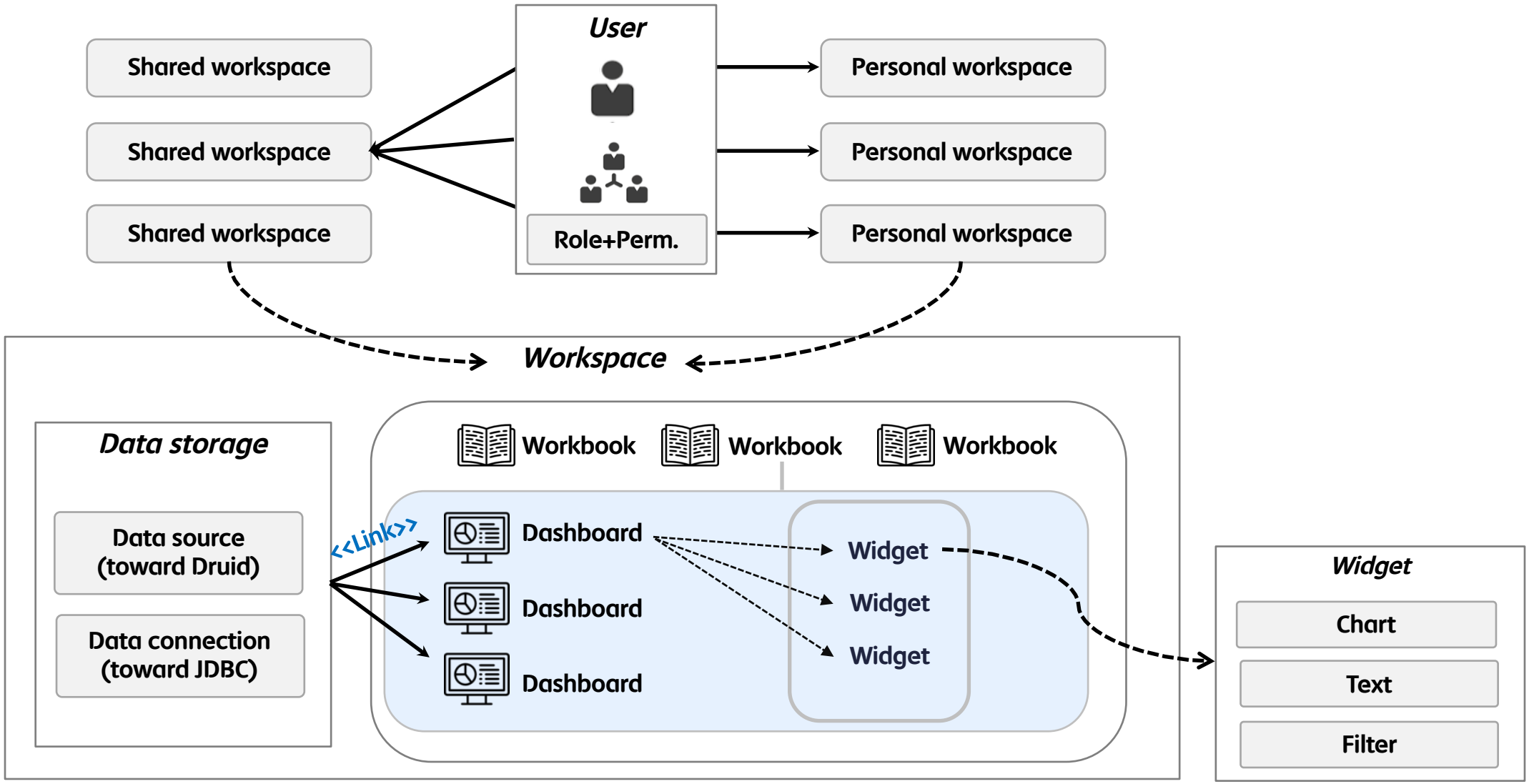
Presentation view
Provides a presentation function for reporting and sharing



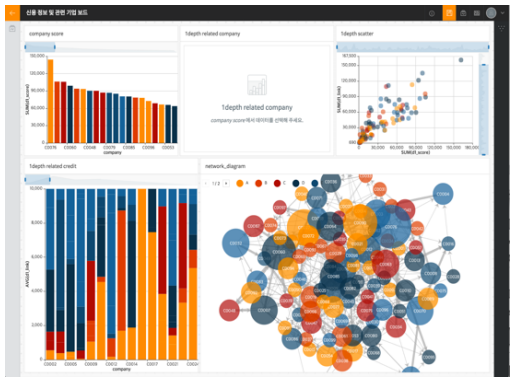
(Reference) Workbook hierarchy



Workbook structure



Workbook Operation (Data Visualization)



⑥①

⑤⑦

Polaris
(Data Discovery Agent)
③

②(ingested)

Big Data Cluster
Druid
Broker Nodes
Historical Nodes
Real-Time Nodes
Deep Storage (HDFS)

④

②'(temporary-linked)

②'

Machine Learning Cluster
Spark Cluster

External DB Cluster
Hive, Presto, etc

① **User command**

② **Data collection from big data clusters (for ingested type)**
or external databases (for temporary-linked type)

③ **Visual processing**

④ **Machine learning algorithm if necessary**
(embedded analytics)

⑤ **User responses**

⑥ **Additional commands with previously collected data**

⑦ **User Response**

Server Node

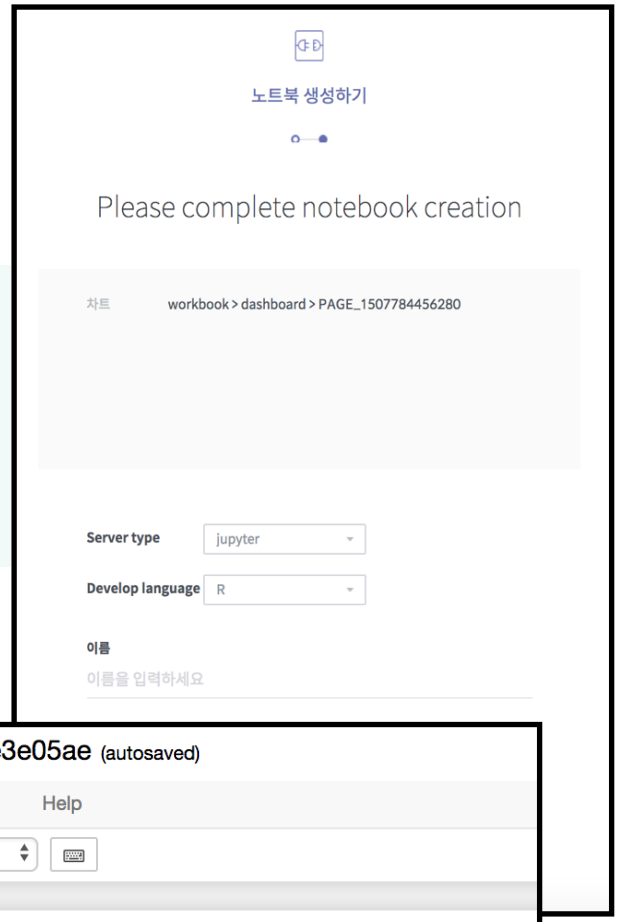
2 Notebook – ML-Based Advanced Analytics

■ Selection of analytics target

- Choose data source, dashboard, or chart
- Select data for analysis

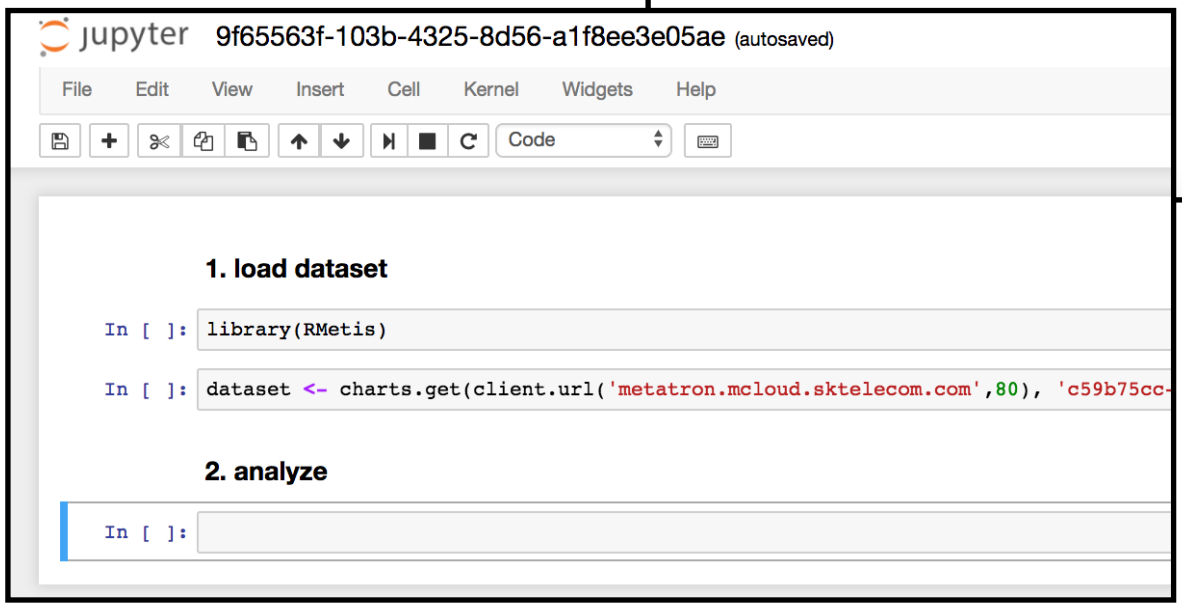
■ Notebook details entry and analysis

- Select server type
- Select development language (Jupyter: R/PYTHON, Zeppelin: Spark)
- Enter notebook name and description



■ Jupyter notebook pop-up

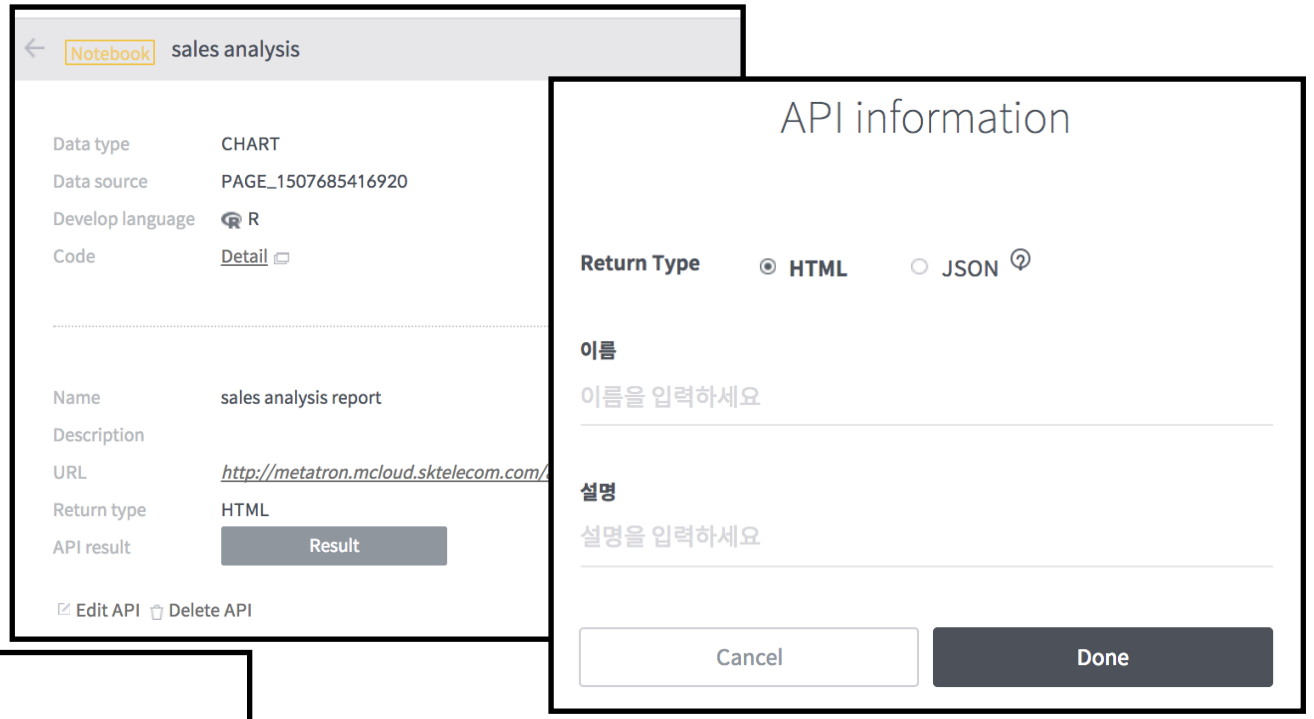
- Write code to load the target dataset to analyze with the development language
- Saved by analyst after writing R code

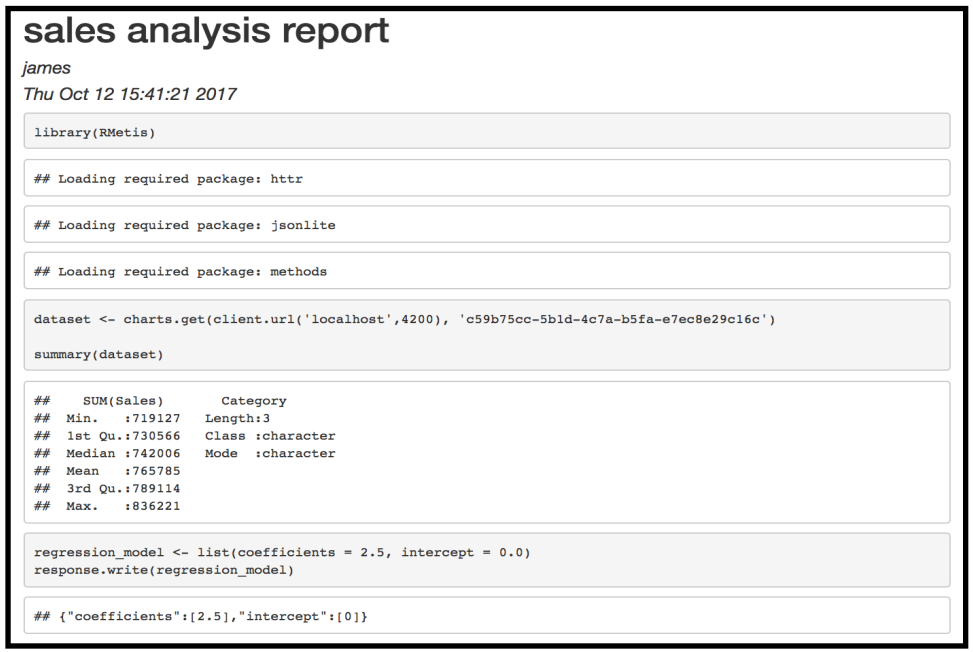


Notebook – ML-Based Advanced Analytics

■ Notebook API creation

- Select return type (html, json)
- notebook code
  response.write(_user_object_)





sales analysis report

james

Thu Oct 12 15:41:21 2017

```
library(RMetis)
```

```
## Loading required package: httr
```

```
## Loading required package: jsonlite
```

```
## Loading required package: methods
```

```
dataset <- charts.get(client.url('localhost',4200), 'c59b75cc-5b1d-4c7a-b5fa-e7ec8e29c16c')

summary(dataset)
```

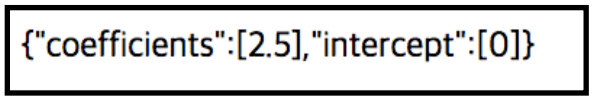
```
##    SUM(Sales)       Category
##  Min.   :719127   Length:3
##  1st Qu.:730566   Class :character
##  Median :742006   Mode  :character
##  Mean   :765785
##  3rd Qu.:789114
##  Max.   :836221
```

```
regression_model <- list(coefficients = 2.5, intercept = 0.0)
response.write(regression_model)
```

```
## {"coefficients":[2.5],"intercept":[0]}
```

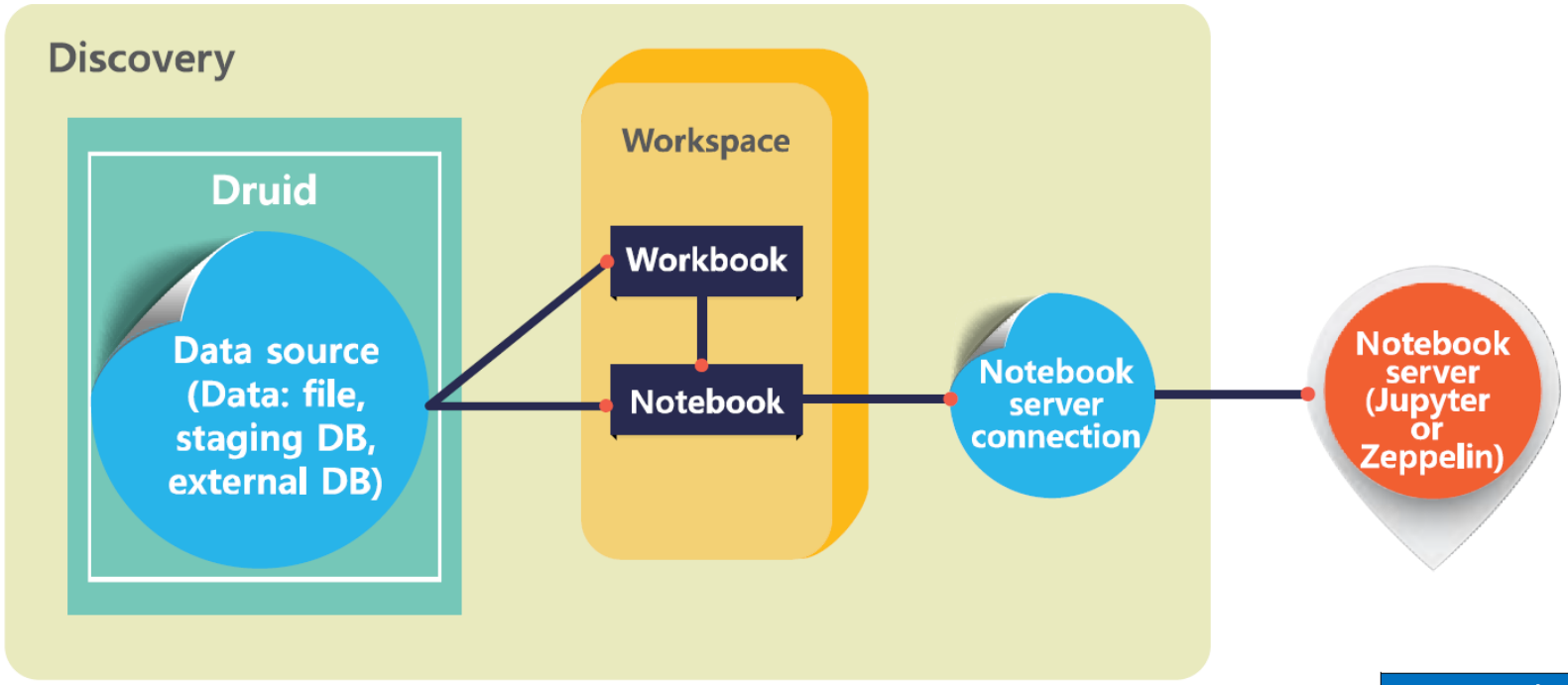
■ RESTful service provided through generated URL

- Click to view results



{"coefficients":[2.5],"intercept":[0]}

Notebook Structure

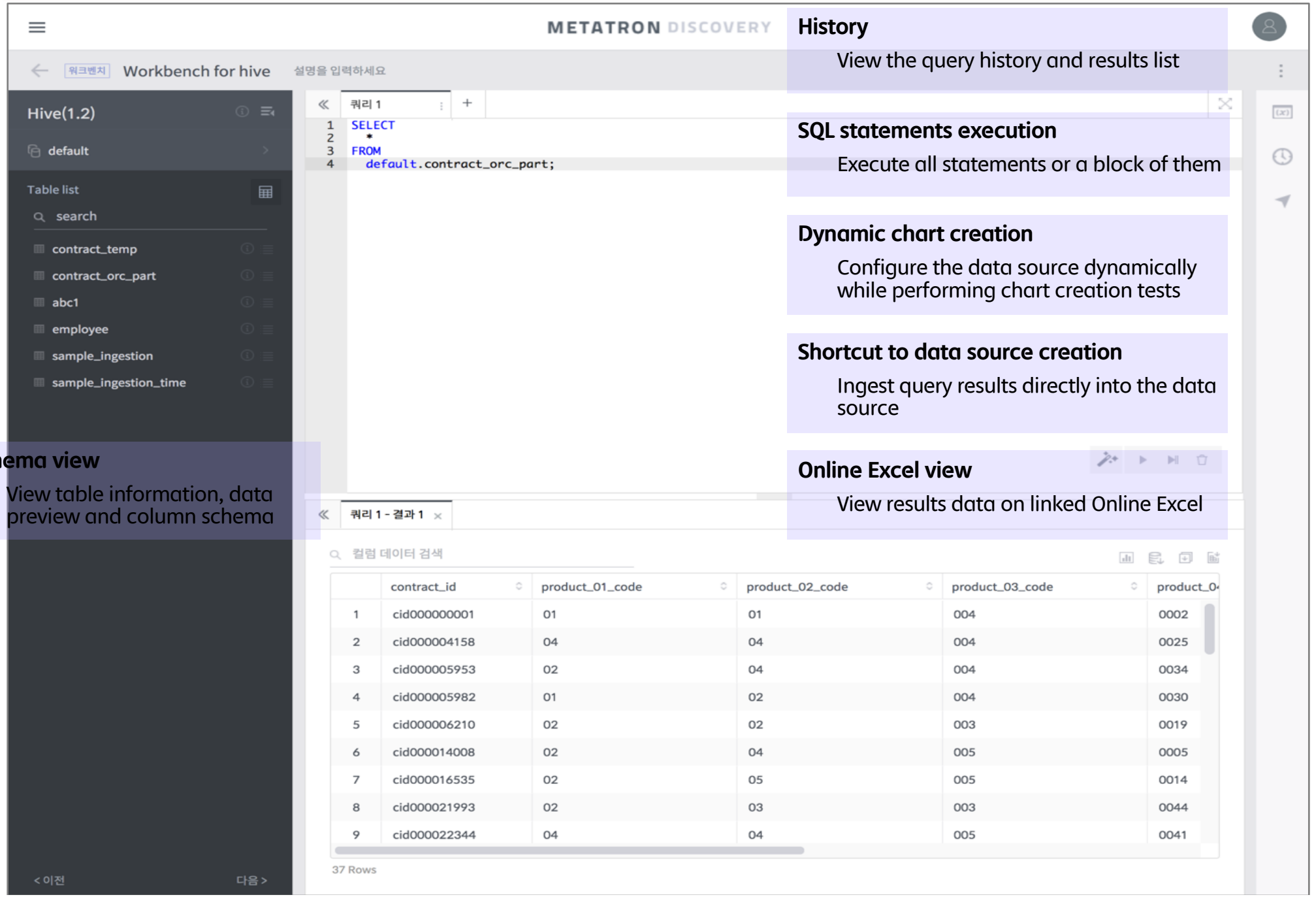


[Notebook structure]

External analysis tools	Available languages
Jupyter	R
	Python
Zeppelin	Spark

[External analytics tools and available languages]

3 Workbench – SQL-Based Analytics



History

View the query history and results list

SQL statements execution

Execute all statements or a block of them

Dynamic chart creation

Configure the data source dynamically while performing chart creation tests

Shortcut to data source creation

Ingest query results directly into the data source

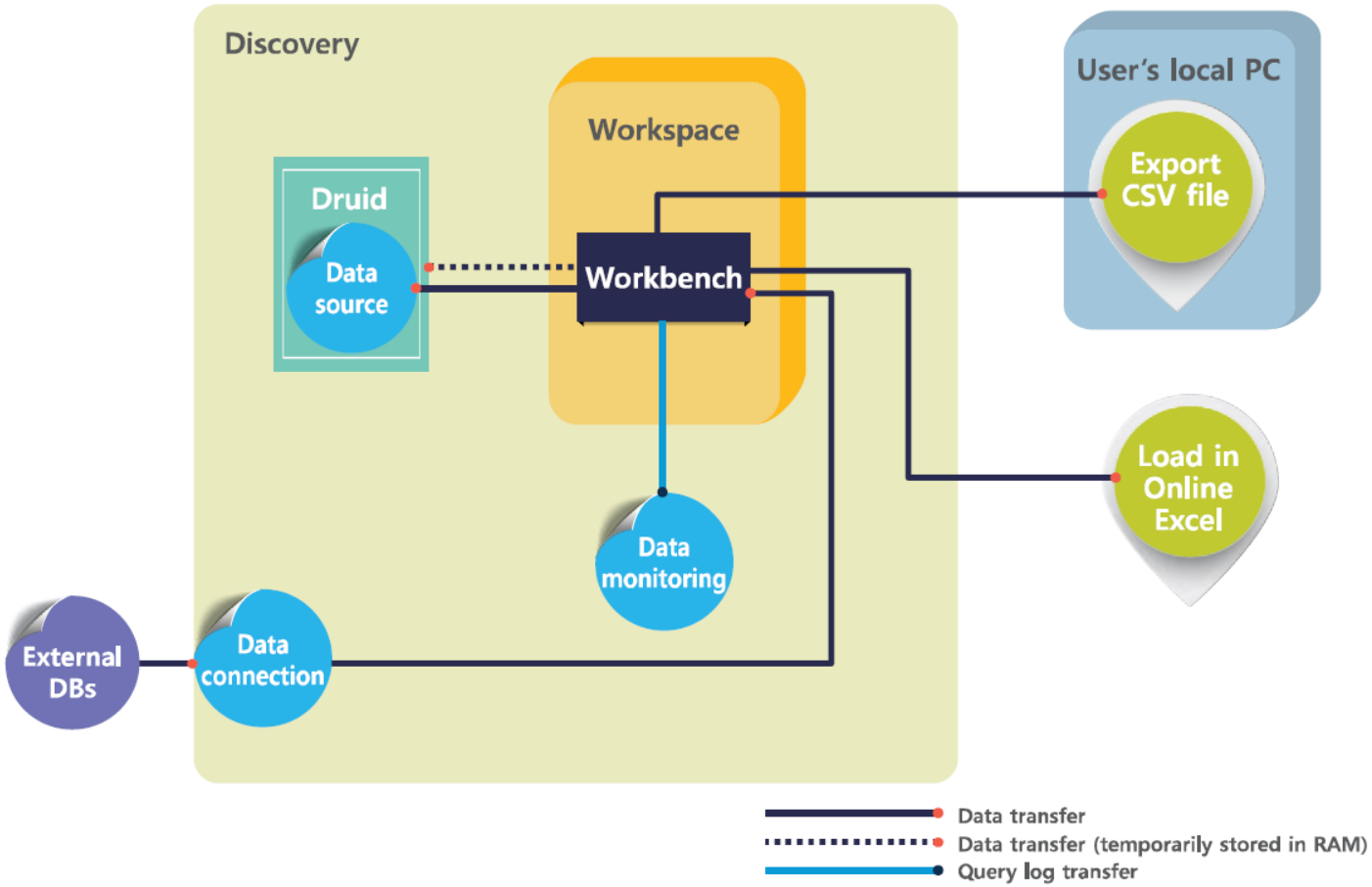
Online Excel view

View results data on linked Online Excel

Schema view

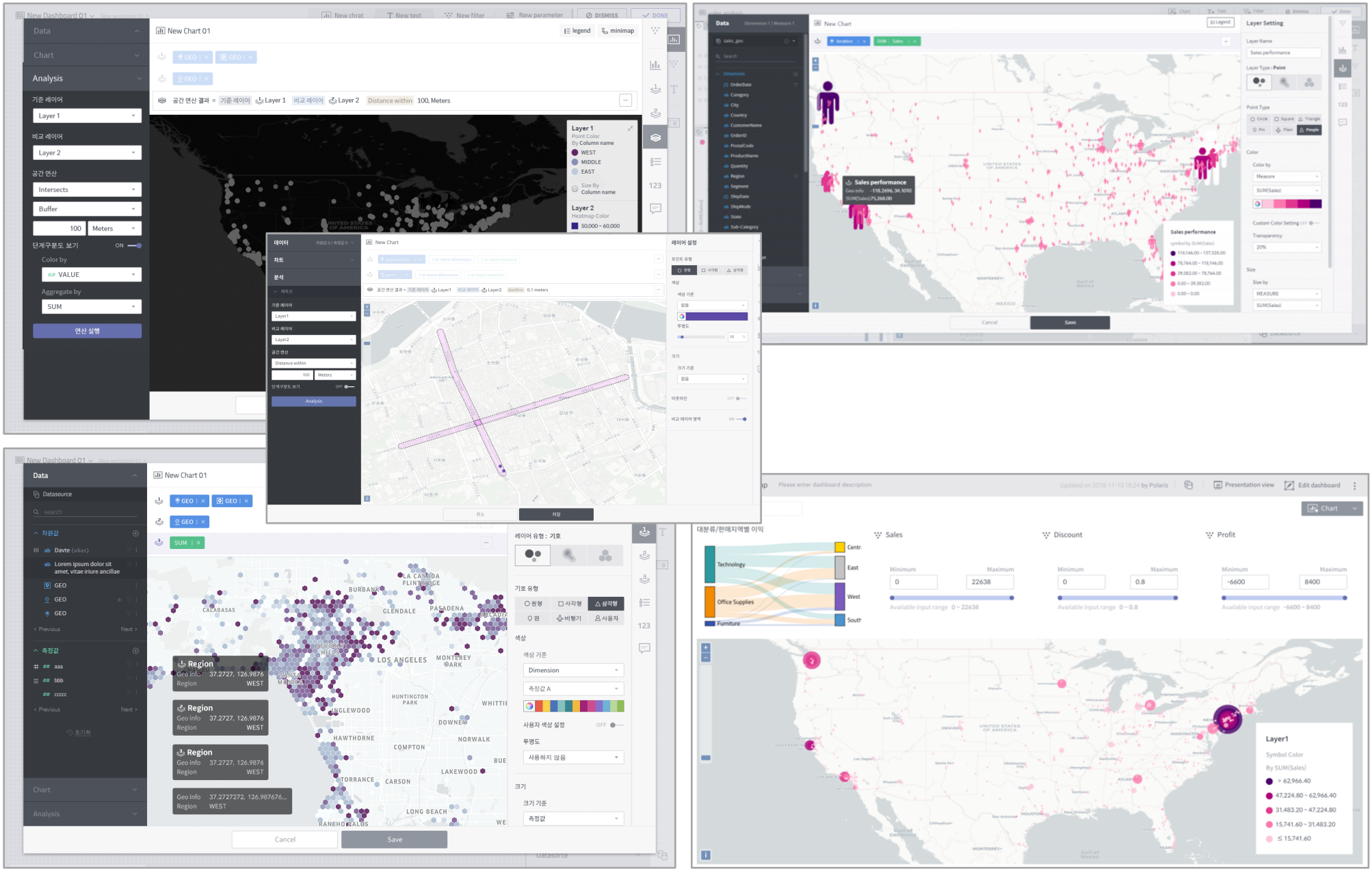
View table information, data preview and column schema

Workbench Structure and Features



- Various external databases (Oracle, MySQL, Hive, Presto, Tibero) on different servers can be loaded in one space.
- Easy and free exploration on connected database schemas facilitates views and selection of tables and columns.
- A built-in query editor enables easy querying, adding, deleting and editing of data.
- Query results are shown in real time and downloadable as local files.
- Data outputs are visualized with a variety of charts by importing them into the Druid engine.
- Query logs are accessible in the data monitoring menu.

Map-based Analytics



Data Preparation

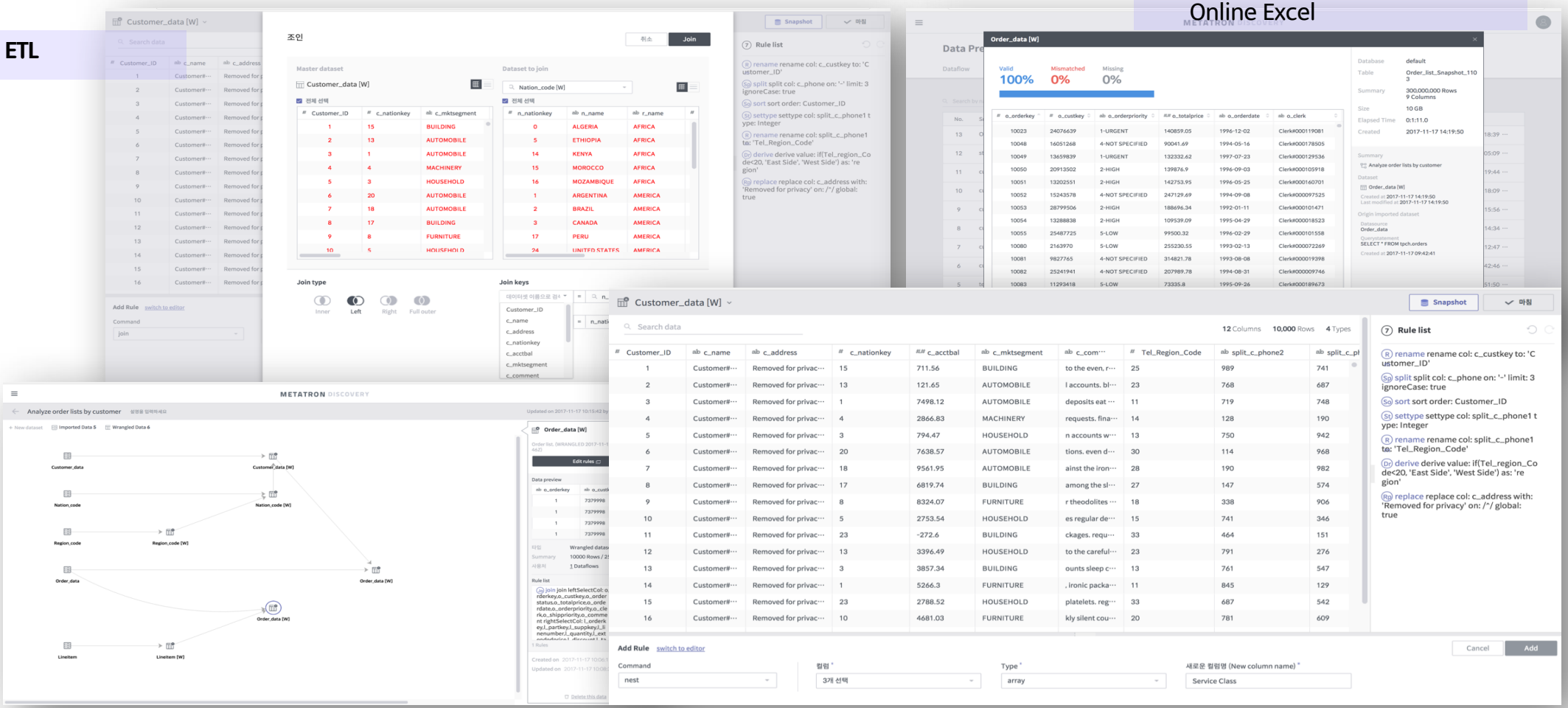
- **ETL tasks for data visualization and analytics that used to be boring and repetitive can now be performed easily by anyone**
  - A consistent abstraction level ensured for different data sources (RDB, Hive, log files, etc.)
  - Back-and-forth, step-by-step editing by looking at and comparing pre- and post-processed sets of sample data

ETL functionality for visualizing and analyzing data

Creation of data snapshots

View results data on linked Online Excel

ETL

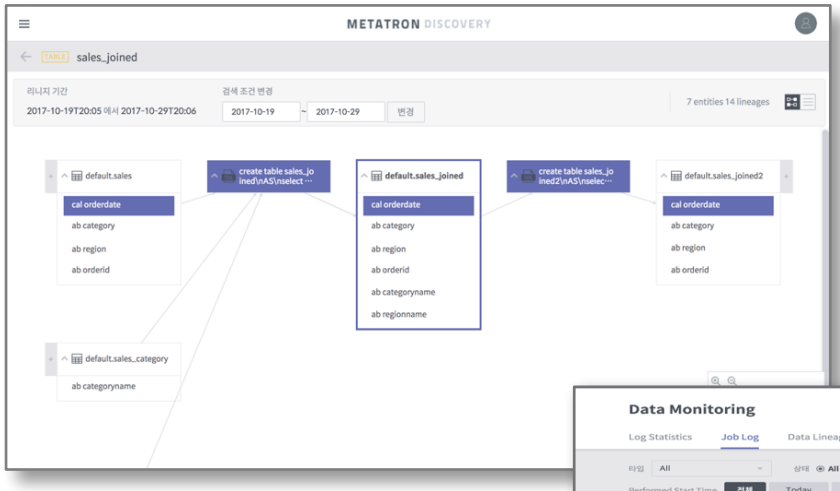


Creation of data flows

Application of rules

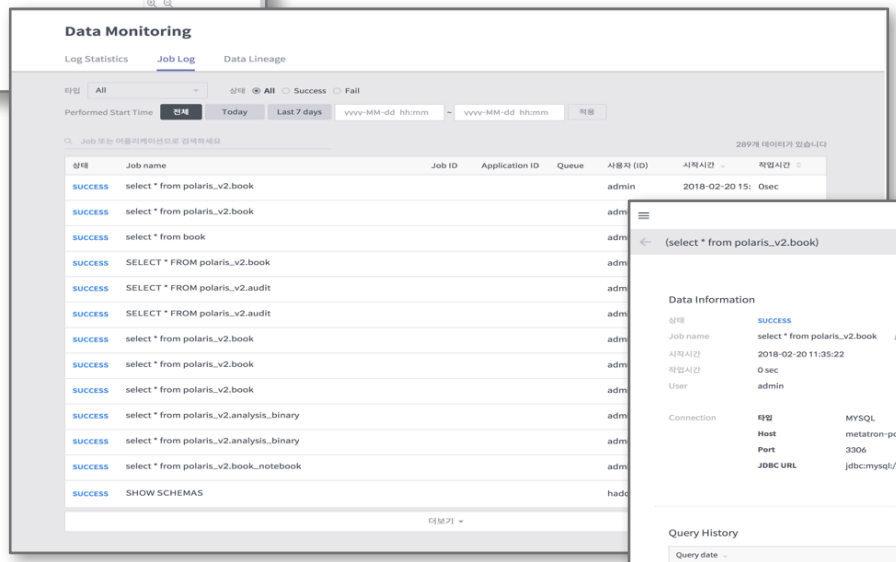
Data Management (Lineage & Job Log)

- **Data Lineage: Manages relationships between pre- and post-processed data using query logs**
- **Job Log: Manages the history of queries issued in workbenches and various other statistical information**



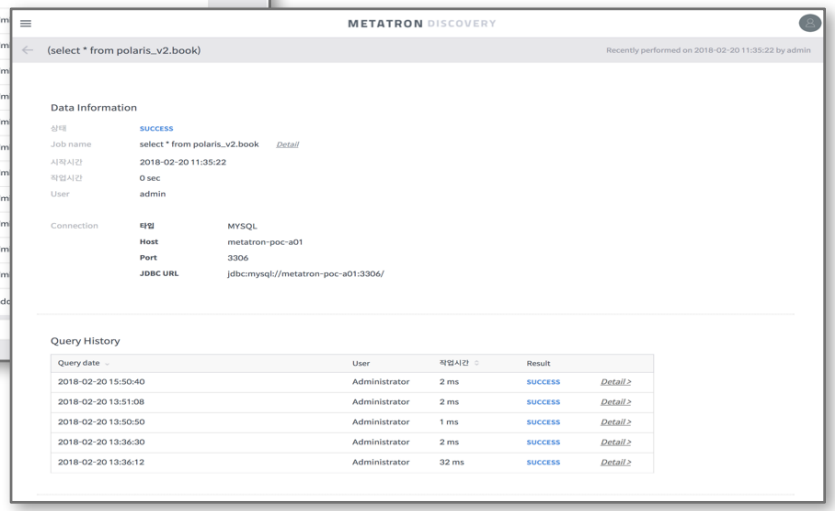
Column history using graphing

An MDM view of the connections of each data query with tables and columns



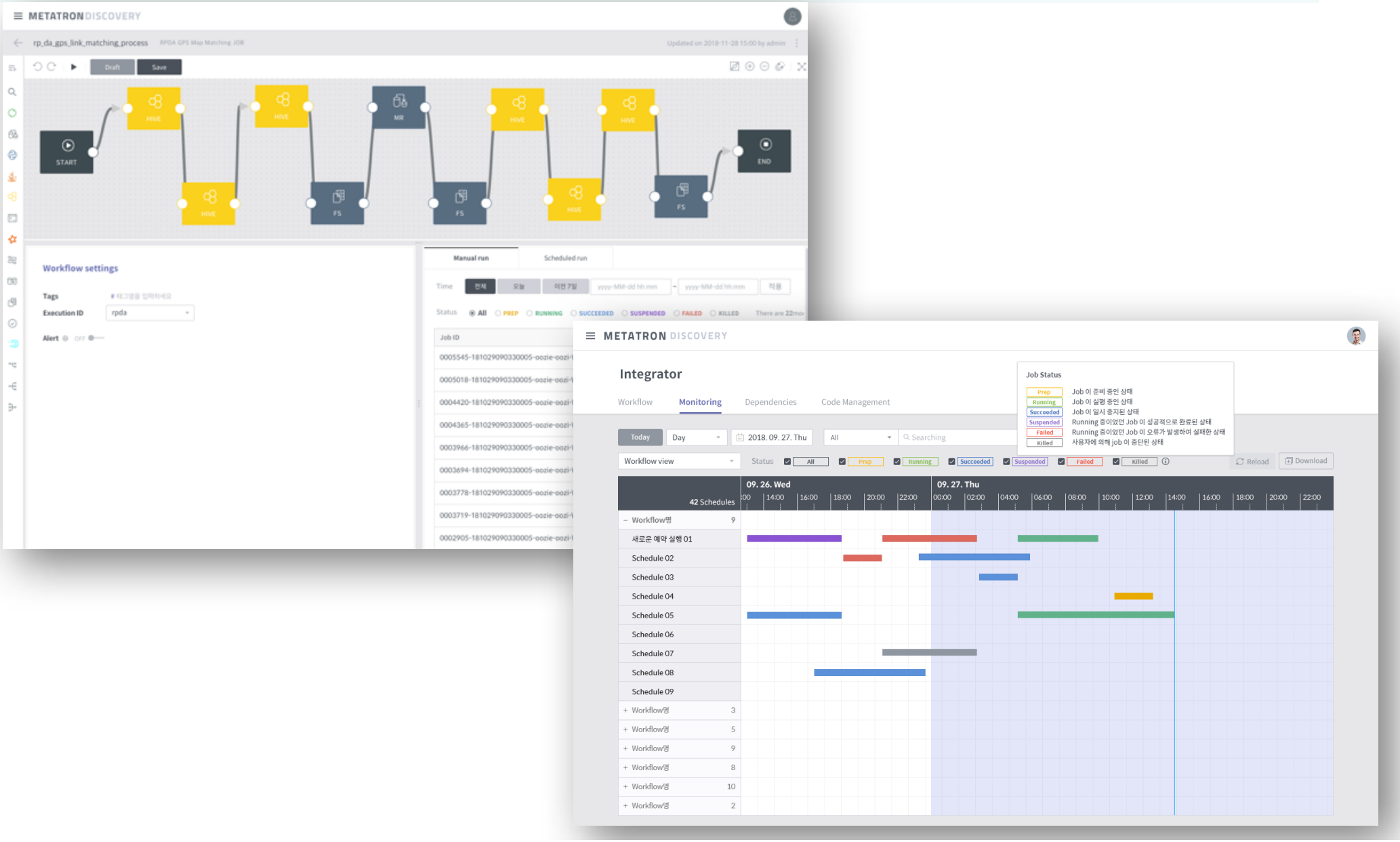
View and search of query history

Detail view of individual queries



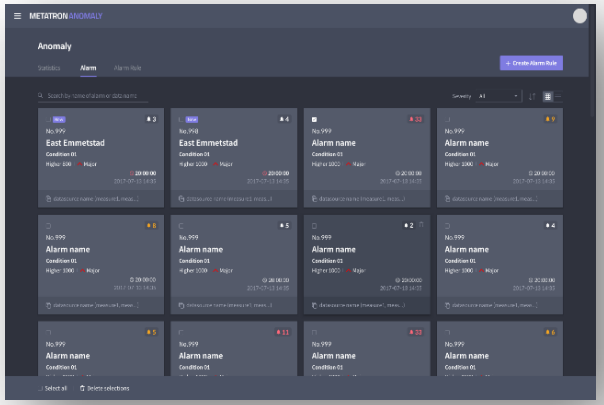
Integrator

■ **Provides a GUI for registering and managing periodic jobs and viewing their processing results**

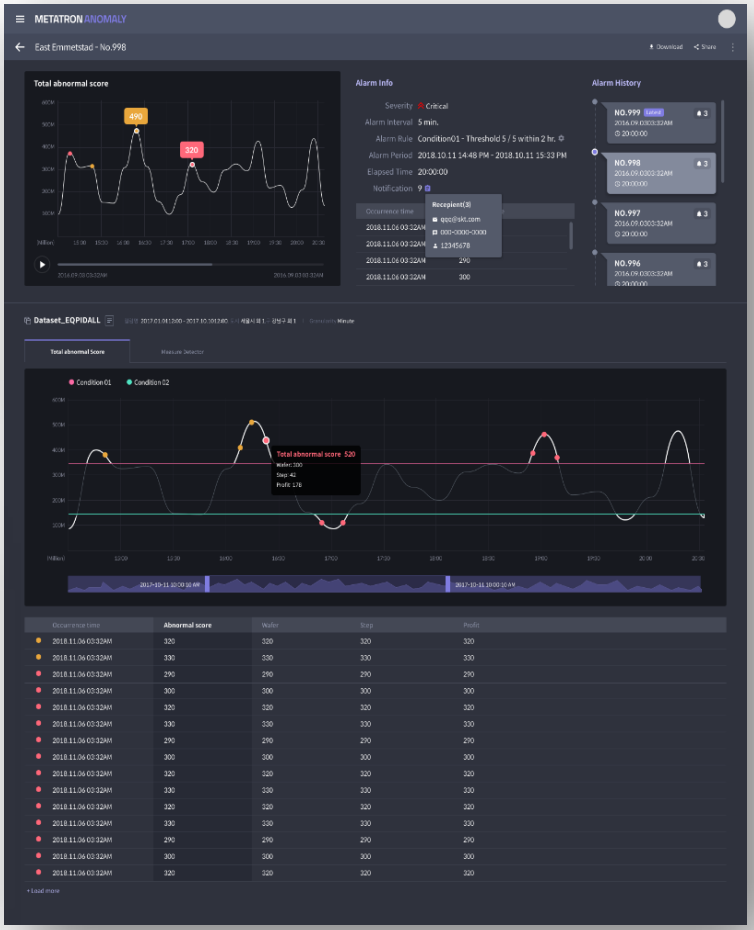


Anomaly Detection

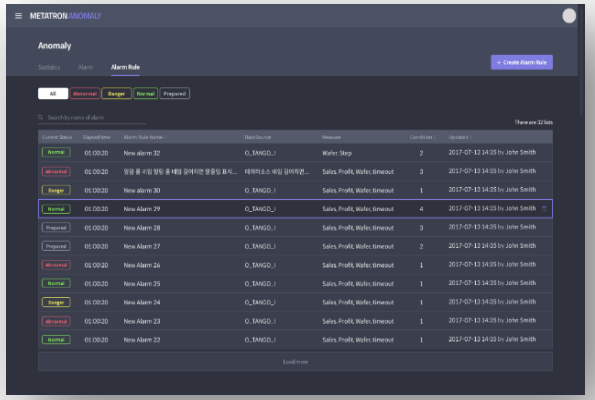
- **Detects anomalous situations using an ML prediction model and generates alarms on them for immediate actions**



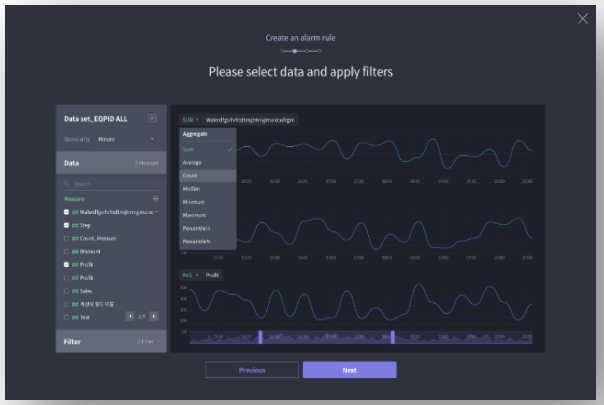
Alarm list



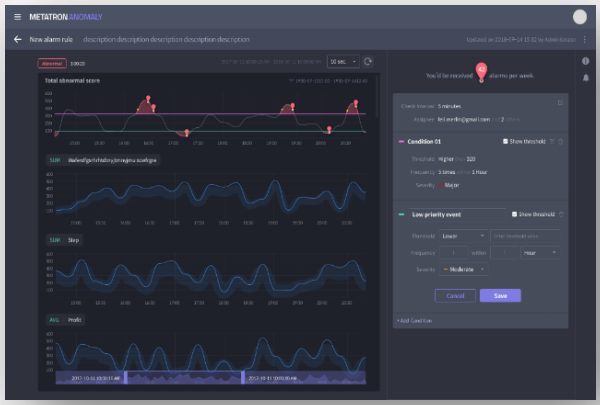
Alarm details & report



Alarm rule creation



Alarm rule creation



Alarm rule details

END

metatron@sk.com

SK telecom