


Username
Enter Username

Password
Enter password

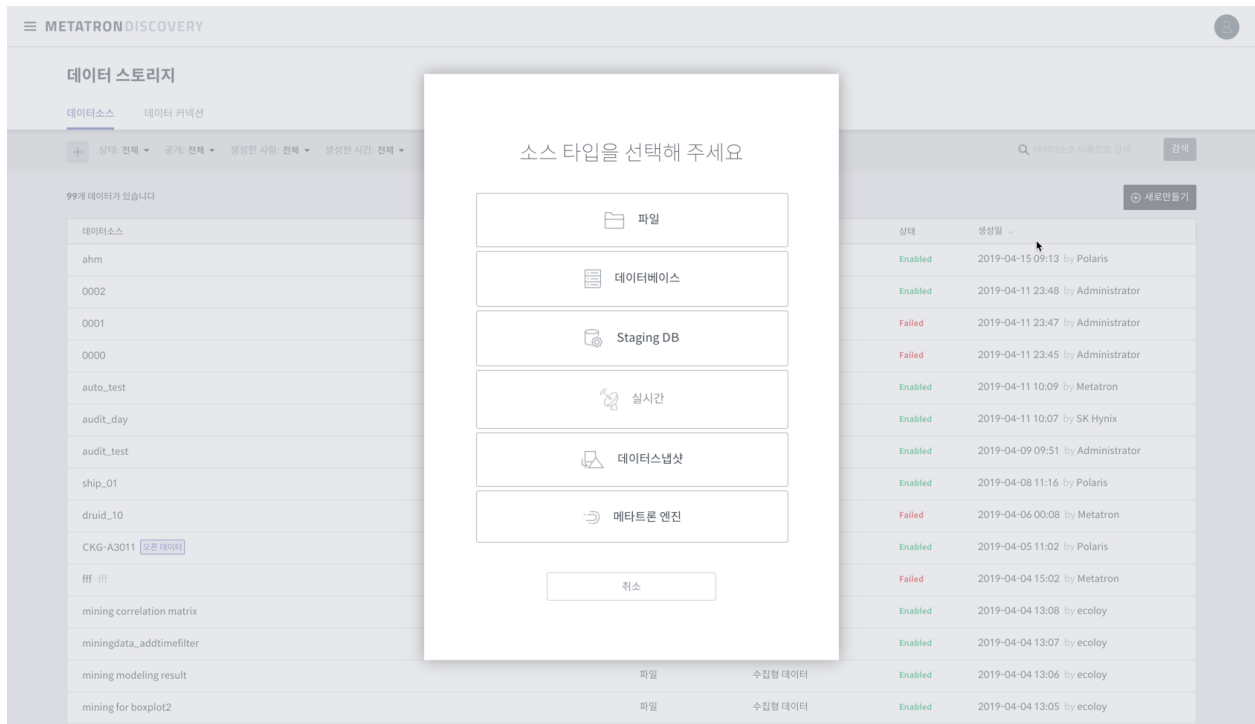
☐ Remember me

LOGIN

[Forgot password?](#) | [Sign up](#)

[Site Terms of Use](#)
Copyright © SK Telecom Co., Ltd. All rights reserved.





데이터소스 생성하기 (파일)

데이터를 선택해 주세요

sales.csv

불러오기 또는 파일을 여거나 끌어다 놓으세요

35690 byte 31 열 100 / 100 행 1 타입							
OrderDate	Category	City	Country	CustomerName	Discount	OrderID	Pos
2011-01-04 00:...	Office Supplies	Houston	United States	Darren Powers	0.2	CA-2011-103...	770
2011-01-05 00:...	Office Supplies	Naperville	United States	Phillina Ober	0.2	CA-2011-112...	605
2011-01-05 00:...	Office Supplies	Naperville	United States	Phillina Ober	0.8	CA-2011-112...	605
2011-01-05 00:...	Office Supplies	Naperville	United States	Phillina Ober	0.2	CA-2011-112...	605
2011-01-06 00:...	Office Supplies	Philadel...	United States	Mick Brown	0.2	CA-2011-141...	191
2011-01-07 00:...	Furniture	Henderson	United States	Maria Etezadi	0	CA-2011-167...	424
2011-01-07 00:...	Office Supplies	Athens	United States	Jack OBriant	0	CA-2011-106...	306
2011-01-07 00:...	Office Supplies	Henderson	United States	Maria Etezadi	0	CA-2011-167...	424
2011-01-07 00:...	Office Supplies	Henderson	United States	Maria Etezadi	0	CA-2011-167...	424
2011-01-07 00:...	Office Supplies	Henderson	United States	Maria Etezadi	0	CA-2011-167...	424
2011-01-07 00:...	Office Supplies	Henderson	United States	Maria Etezadi	0	CA-2011-167...	424
2011-01-07 00:...	Office Supplies	Los Angeles	United States	Lycoris Saunders	0	CA-2011-130...	900
2011-01-07 00:...	Office Supplies	Manderson	United States	Maria Etezadi	0	CA-2011-167...	424

컬럼 구분자

라인 구분자

첫번째 행을 헤더로 사용합니다. (선택하지 않은 경우 새 행이 생성되고 헤더로 사용됨)

취소

다음

데이터소스 생성하기 (파일)

스키마 구성



필름 이름으로 검색해 주세요

☐	컬럼
☐	자원값abOrderDate
☐	자원값abCategory
☐	자원값abCity
☐	자원값abCountry
☐	자원값abCustomerName
☐	자원값abDiscount
☐	자원값abOrderID
☐	자원값abPostalCode
☐	자원값abProductName
☐	자원값abProfit
☐	자원값abQuantity
☐	자원값abRegion
☐	자원값abSales

역할 ☒ 전체 ☐ 차원값 ☐ 측정값 타입 전체

필름 추가

OrderDate		
데이터	50 행	설정
2011-01-04 00:00:00		역할 ☒ 차원값 ☐ 측정값 타입 ab 문자 빈 값 설정 ☐ 대체 ☐ 버림

시간 타입 또는 현재 시간 중 하나는 시간값으로 지정되어야 합니다

☒ 현재 시간 ☐ 시간 타입 필름 선택된 타임타입 필름이 없습니다

이전

다음

데이터소스 생성하기 (파일)

스키마 구성



필름 이름으로 검색해 주세요

☐	컬럼	
☐	필름값	OrderDate
☐	필름값	Category
☐	필름값	City
☐	필름값	Country
☐	필름값	CustomerName
☐	속성값	Discount
☐	필름값	OrderID
☐	필름값	PostalCode
☐	필름값	ProductName
☐	속성값	Profit
☐	속성값	Quantity
☐	필름값	Region
☐	속성값	Sales

역할 ☒ 전체 ☐ 차원값 ☐ 속성값 타입 전체

OrderDate

방법

데이터 latitude longitude

필름 이름

2011-01-04 00:00:00

2011-01-05 00:00:00

2011-01-05 00:00:00

2011-01-05 00:00:00

2011-01-06 00:00:00

2011-01-07 00:00:00

2011-01-07 00:00:00

2011-01-07 00:00:00

2011-01-07 00:00:00

2011-01-07 00:00:00

2011-01-07 00:00:00

2011-01-07 00:00:00

타입

타입스탬프는 변경이 불가능합니다

시간 표현 ☒ ☐ Unix time 사용

yyyy-MM-dd HH:mm:ss

타임존

+09:00 서울/대한민국/아시아

시간 타입 또는 현재 시간 중 하나는 시간값으로 지정되어야 합니다

☐ 현재 시간 ☒ 시간 타입 필름 OrderDate

이전

다음

데이터소스 생성하기 (파일)

수집 설정을 완료해 주세요



타임스탬프 설정

쿼리 단위

초

세그먼트 단위

시간

데이터 범위

2010-12-30 20 ~ 2011-01-25 04 609 세그먼트 단위 수

간격의 설정값은 타임스탬프 컬럼의 데이터값 범위와 같거나 커야하며, 이 구간에 의한 세그먼트 갯수가 10,000개가 넘을 수 없습니다.

종업

☒ 설정할 ☐ 설정안함

고급 설정

이전

다음

데이터소스 생성하기 (파일)
데이터소스 생성을 완료해 주세요

파일 sales.csv
타임스탬프 설정
• 세그먼트 단위 : 시간
• 세그먼트 범위 : 2010-12-30 20 ~ 2011-01-25 04
종류 설정함

이름
판매현황

설명
2010-2011 판매현황 데이터입니다.

이전

마침

METATRONDISCOVERY

← 판매현황 데이터

수정 2019-04-15 15:32 | Polaris

정보

데이터

설정 상세

모니터링

데이터 정보

설명
2010-2011년도 판매현황 자료입니다.

메타데이터로 이동

적재 타입

수집형 데이터

상태

ENABLED



타임스탬프 설정

위리 단위 SECOND
세그먼트 단위 HOUR
데이터 범위 2010-12-30 20 ~ 2011-01-25 04

히스토그램



할당

☐ 모든 워크스페이스에 데이터소스를 사용하도록 허용

소유

1 워크스페이스

METATRONDISCOVERY

← 판매현황 데이터

정보

데이터

데이터 정보

설명2010-2011 판매현황 데이터

적재 타입수집형 데이터

상태ENABLED

사이즈189.58 KB

인티벌2011-01-03T15:00:00Z

타임스탬프 설정쿼리 단위

세그먼트 단위

데이터 범위

히스토그램

할당

☐ 모든 워크스페이스에 할당

☒ 1 워크스페이스에 할당

할당

닫기

개인 워크스페이스 (1/59)

공통 워크스페이스 (0/18)

워크스페이스 또는 사용자 검색

워크스페이스

소유자 (사용자이름)

☐ 한정현 Workspace	한정현 (kazikai)
☐ 조민정 Workspace	조민정 (heesoo)
☐ 장지영 Workspace	장지영 (jjangdotcom)
☐ 이정윤 Workspace	이정윤 (arther720)
☐ 이정룡 Workspace	이정룡 (l1bfree)
☐ 윤준수 Workspace	윤준수 (yjscass)
☐ 유승호 Workspace	유승호 (shryu415)
☐ 송세리 Workspace	송세리 (srsong)
☐ 성승현 Workspace	성승현 (sshzozo)
☐ 박종호 Workspace	박종호 (tjajtaji)
☐ 문형권 Workspace	문형권 (sysmoon)
☐ 김상호 Workspace	김상호 (bboradoli)
☐ 김병길 Workspace	김병길 (jobofgod)
☐ tim-matatron.Matatron	tim-matatron (tim-matatron)

더보기

수정 2019-04-15 15:43 | Polaris

메타데이터로 이동



METATRONDISCOVERY

← 판매현황 데이터

수정 2019-04-15 15:54 | Polaris

정보

데이터

유통 상세

모니터링

데이터 정보

설명

2010-2011 판매현황 데이터

적재 타입

수집형 데이터

상태

ENABLED

사이즈

189.58 KB

인타입

2011-01-03T15:00:00.000Z

타입스펙트 설명

유통 단위

세그먼트 단위

데이터 범위

히스토그램

할당

☐ 모든 워크스페이스에 데이터소스를 사용하도록 허용

소문

1 워크스페이스

타데이터로 이동

'판매현황 데이터' 데이터소스를 전체 워크스페이스에 공개하시겠습니까?

취소

전체 공개

METATRONDISCOVERY

← 판매현황 데이터

수정 2019-04-15 15:54 | Polaris

정보

데이터

유통 상세

모니터링

데이터 검색

역할 ☒ 전체 ☐ 자원감 ☐ 측정감 타입

전체

63 행 다운로드

GeoPoint	OrderDate	Category	City	Country	CustomerName	Discount	OrderID	PostalCode	ProductName	Profit	Quantity	Region	Sales	Segment
GeoPoint	OrderDate	Category	City	Country	CustomerName	Discount	OrderID	PostalCode	ProductName	Profit	Quantity	Region	Sales	Segment
29.8941,-9...	2011-01-04 0...	Office Supp...	Houston	United States	Darren Powers	0.2	CA-2011-1...	77095	Message Book...	6	2	Central	16	Cor
41.7662,-8...	2011-01-05 0...	Office Supp...	Naperville	United States	Phillina Ober	0.2	CA-2011-1...	60540	Avery 508	4	3	Central	12	Hor
41.7662,-8...	2011-01-05 0...	Office Supp...	Naperville	United States	Phillina Ober	0.8	CA-2011-1...	60540	GBC Standard Pl...	-5	2	Central	4	Hor
41.7662,-8...	2011-01-05 0...	Office Supp...	Naperville	United States	Phillina Ober	0.2	CA-2011-1...	60540	SAFCO Bottless ...	-65	3	Central	273	Hor
39.9448,-7...	2011-01-06 0...	Office Supp...	Philadelphia	United States	Mick Brown	0.2	CA-2011-1...	19143	Avery Hi-Liter Ev...	5	3	East	20	Cor
37.8274,-8...	2011-01-07 0...	Furniture	Henderson	United States	Maria Etezadi	0	CA-2011-1...	42420	Global Deluxe Hi...	746	9	South	2574	Hor
33.9321,-8...	2011-01-07 0...	Office Supp...	Athens	United States	Jack OBriant	0	CA-2011-1...	30605	Dixon Prang Wat...	5	3	South	13	Cor
37.8274,-8...	2011-01-07 0...	Office Supp...	Henderson	United States	Maria Etezadi	0	CA-2011-1...	42420	Alliance Super-S...	0	4	South	31	Hor
37.8274,-8...	2011-01-07 0...	Office Supp...	Henderson	United States	Maria Etezadi	0	CA-2011-1...	42420	Ibico Hi-Tech Ma...	274	2	South	610	Hor
37.8274,-8...	2011-01-07 0...	Office Supp...	Henderson	United States	Maria Etezadi	0	CA-2011-1...	42420	Rogers Handhel...	1	2	South	5	Hor
37.8274,-8...	2011-01-07 0...	Office Supp...	Henderson	United States	Maria Etezadi	0	CA-2011-1...	42420	Southworth 25%-...	3	1	South	7	Hor
34.066,-11...	2011-01-07 0...	Office Supp...	Los Angeles	United States	Lycoris Saunders	0	CA-2011-1...	90049	Xerox 225	9	3	West	19	Cor
37.8274,-8...	2011-01-07 0...	Technology	Henderson	United States	Maria Etezadi	0	CA-2011-1...	42420	GE 30524EE4	114	2	South	392	Hor
37.8274,-8...	2011-01-07 0...	Technology	Henderson	United States	Maria Etezadi	0	CA-2011-1...	42420	Wireless Extende...	204	4	South	756	Hor
30.6448,-9...	2011-01-08 0...	Furniture	Huntsville	United States	Vivek Sundaresam	0.6	CA-2011-1...	77340	Howard Miller 14...	-54	3	Central	77	Cor
30.6448,-9...	2011-01-08 0...	Office Supp...	Huntsville	United States	Vivek Sundaresam	0.8	CA-2011-1...	77340	Acco Four Pocke...	-18	7	Central	10	Cor
27.5569,-9...	2011-01-10 0...	Office Supp...	Laredo	United States	Melanie Seite	0.2	CA-2011-1...	78041	Newell 312	1	2	Central	9	Cor
27.5569,-9...	2011-01-10 0...	Technology	Laredo	United States	Melanie Seite	0.2	CA-2011-1...	78041	Memorex Micro ...	10	3	Central	31	Cor
28.7689,-7...	2011-01-11 0...	Furniture	Cincinnati	United States	Anthony Lande	0	CA-2011-1...	73153	Howard Miller 11...	21	1	South	53	Cor

Metatron Workspace Cancel

Workspace List

Workbook 15 | Workbench 11 | 18 Datasource

Created on 2019-02-18 | Metatron

Metatron Workspace

Search for a content or folder

All

↑↓

☰

new folder New	new folder	metatron	
workbench New Druid data Last updated 12 hours ago druid connection	workbook New Sales & Profit Last updated 12 hours ago 2 1	workbook Bus Last updated 15 hours ago 2 2	workbench New test Last updated 1 day ago druid connection
workbook New auto_test Last updated 1 day ago 1 1	workbook SADFSF Last updated 3 days ago 1 1	workbench sdf Last updated 3 days ago druid connection	workbook test Last updated 3 days ago 2 2
workbook ferum Last updated 3 days ago 1 1	workbench ttt Last updated 4 days ago Presto-metatron-hadoop-01-8089	workbench ada123 Last updated 6 days ago MySQL-metatron-web-03-3306	workbook TestWithDruid Last updated 8 days ago 1 1

Select all

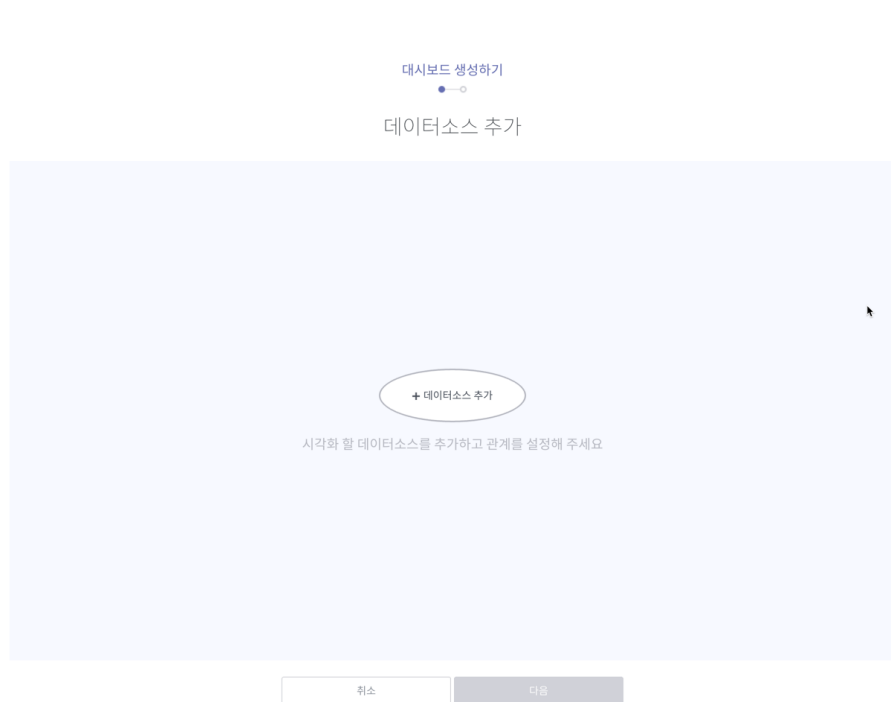
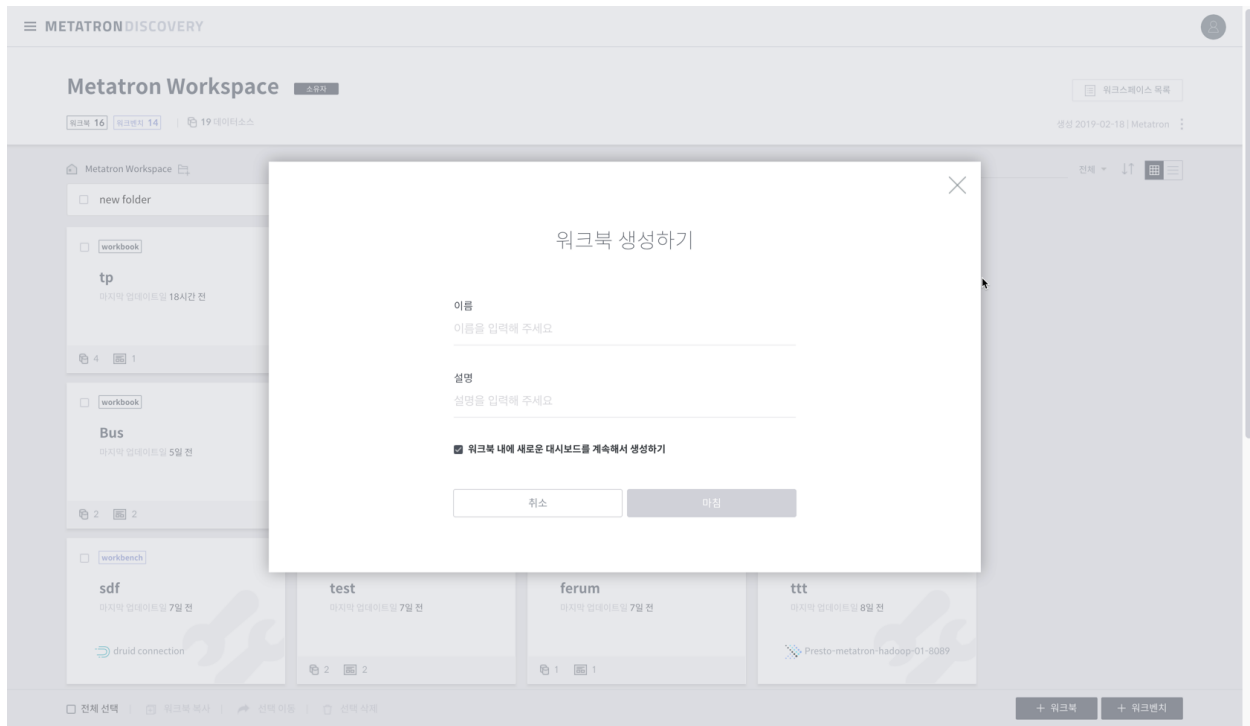
Clone Workbook

Move selections

Delete selections

+ Workbook

+ Workbench



데이터소스를 선택해 주세요

취소

마침

Q 데이터소스 이름 검색

오픈 데이터만 보기

타입

전체

No.	데이터소스 5	타입
☒	12 판매현황 데이터 - 2010-2011 판매---	수집형
☐	11 CKG-A3011	수집형
☐	10 xx	수집형
☐	9 zz	수집형
☐	8 geo	수집형
☐	7 sales - Sales data (2011-2014)	수집형
☐	6 Face Recognition Confusion Ma---	수집형
☐	5 Face Recognition Tuning Results	수집형
☐	4 Face Recognition Metrics	수집형
☐	3 cei_dong_test	연결형
☐	2 correlation matrix for mining	수집형
☐	1 iron mining data - for demo	수집형

더보기

판매현황 데이터

판매현황 데이터 이름

판매현황 데이터

설명

2010-2011 판매현황 데이터입니다.

타입

수집형

공개설정

공개

생성일

2019-04-15

사이즈

189.58 KB

Rows

63

좌원값

GeoPoint

좌원값

OrderDate

좌원값

Category

좌원값

City

좌원값

Country

좌원값

CustomerName

측정값

Discount

좌원값

OrderID

좌원값

PostalCode

좌원값

ProductName

측정값

Profit

측정값

Quantity

좌원값

Region

측정값

Sales

좌원값

Segment

좌원값

ShipDate

좌원값

ShipMode

좌원값

State

대시보드 생성하기

○ ●

대시보드 생성을 완성해 주세요

워크북

Sales & Profit

데이터소스

판매현황 데이터

이름

이름을 입력해 주세요

설명

설명을 입력해 주세요

이전

마침



1 sales performance

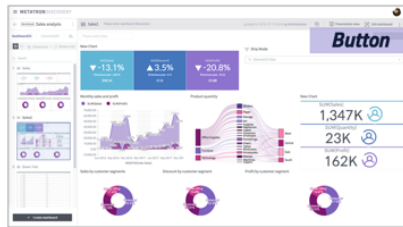


+ Create dashboard

Chart
Filter
Text



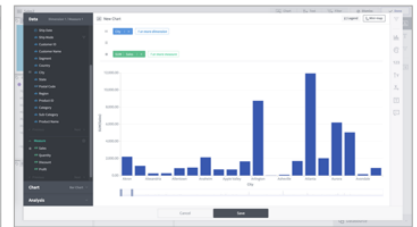
No widgets



1. [Dashboard] View



2. [Chart] Create Chart



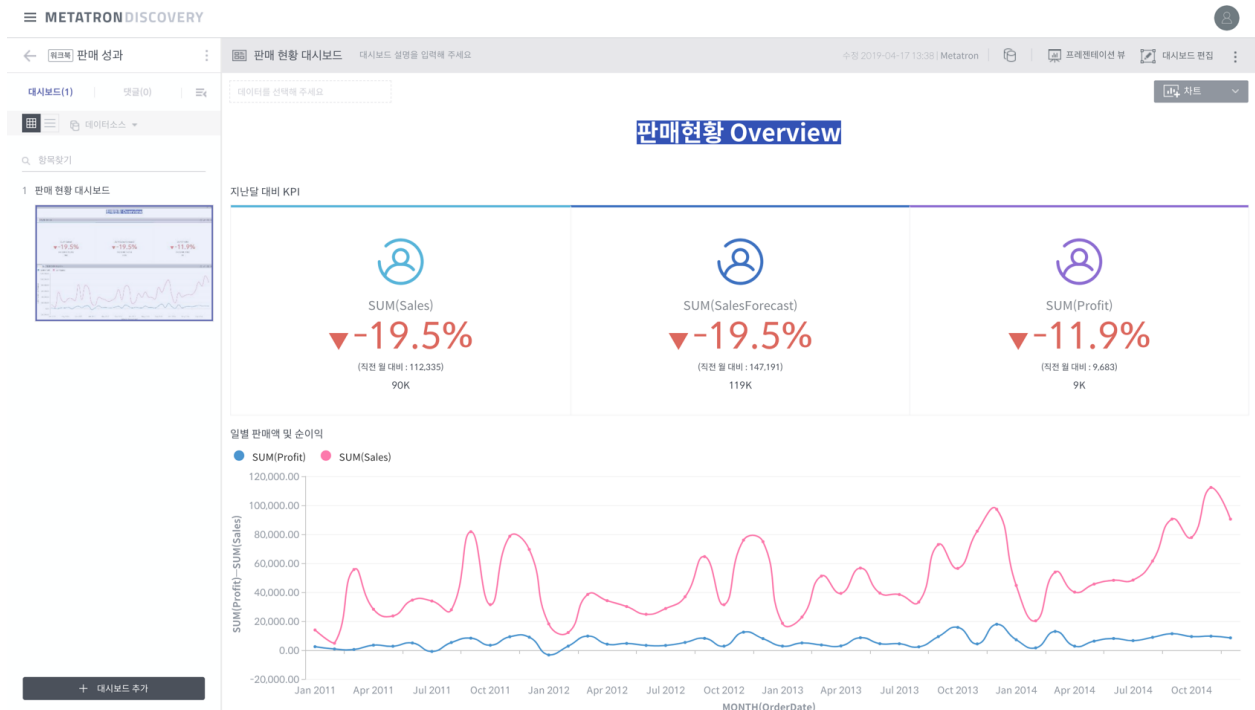
3. [Chart] Pivot Data

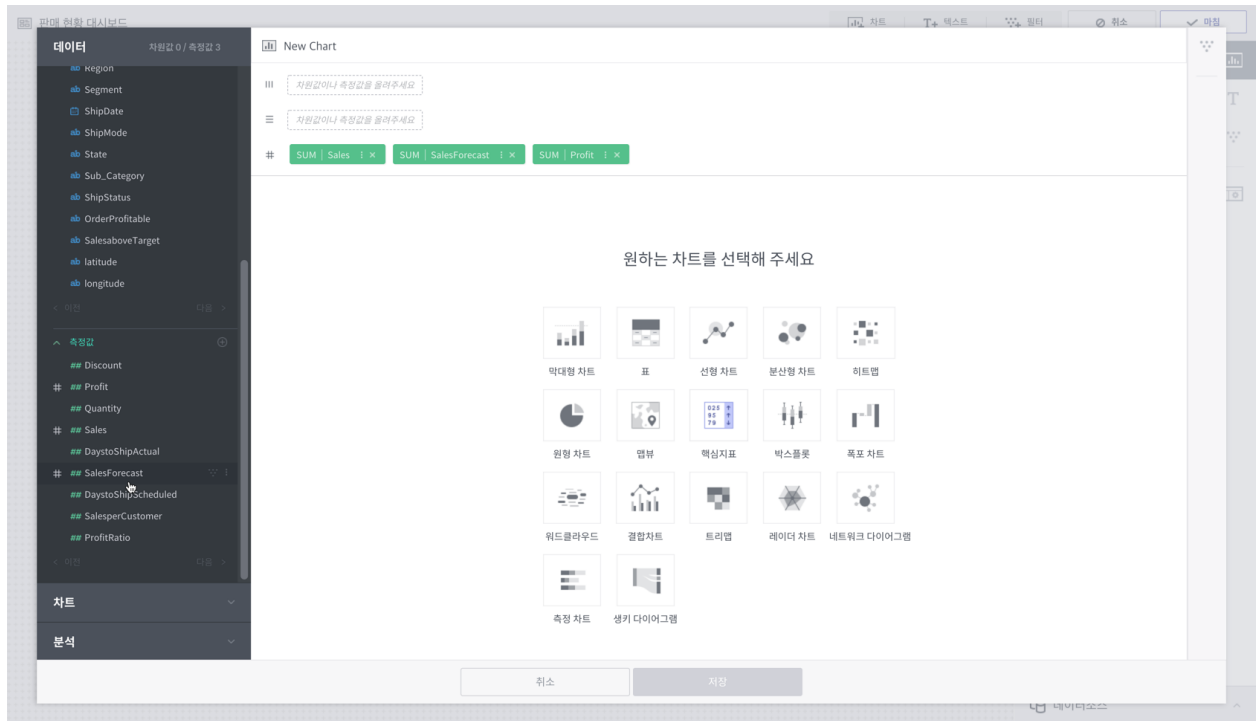
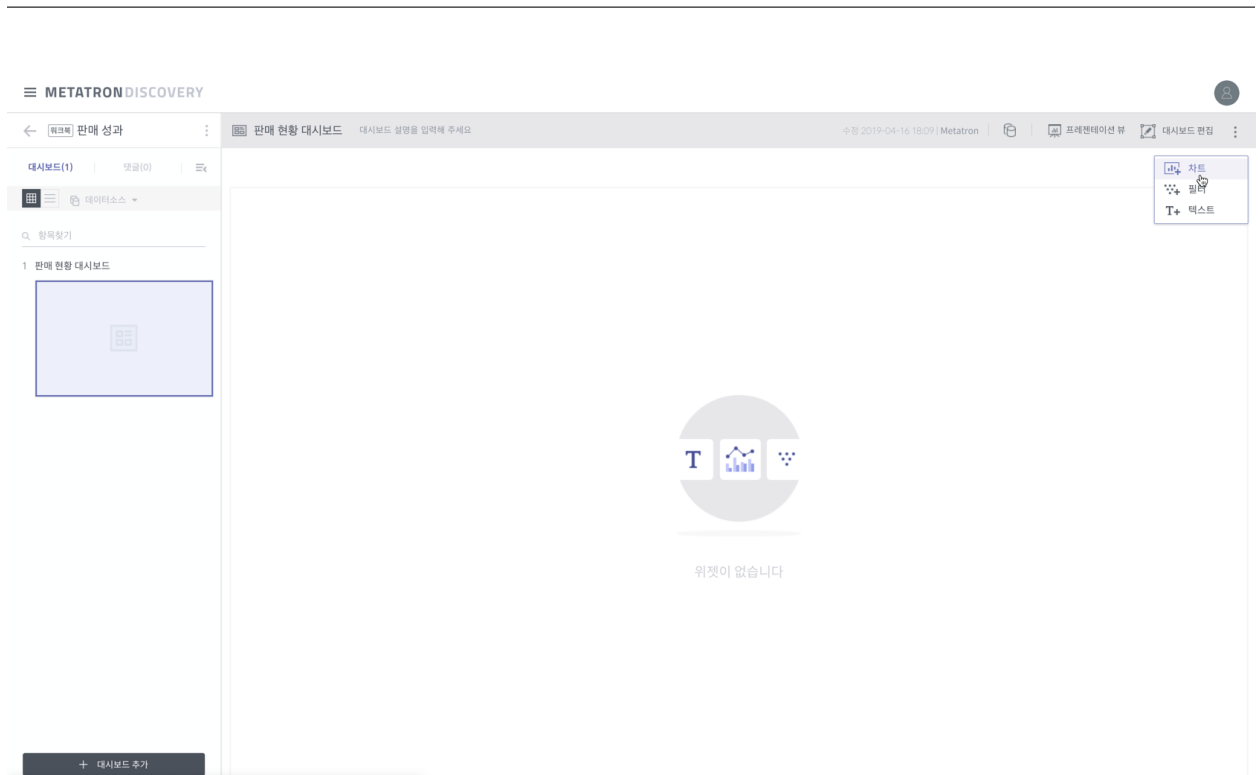


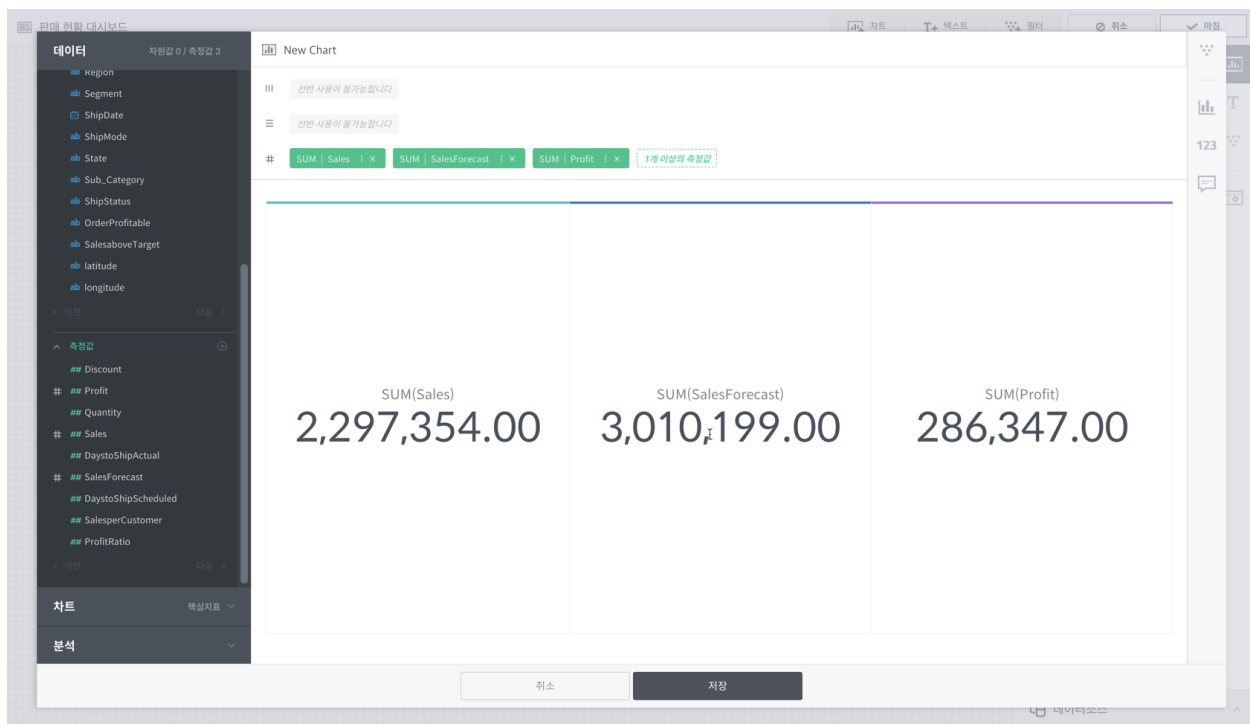
4. [Chart] Chart Decorate



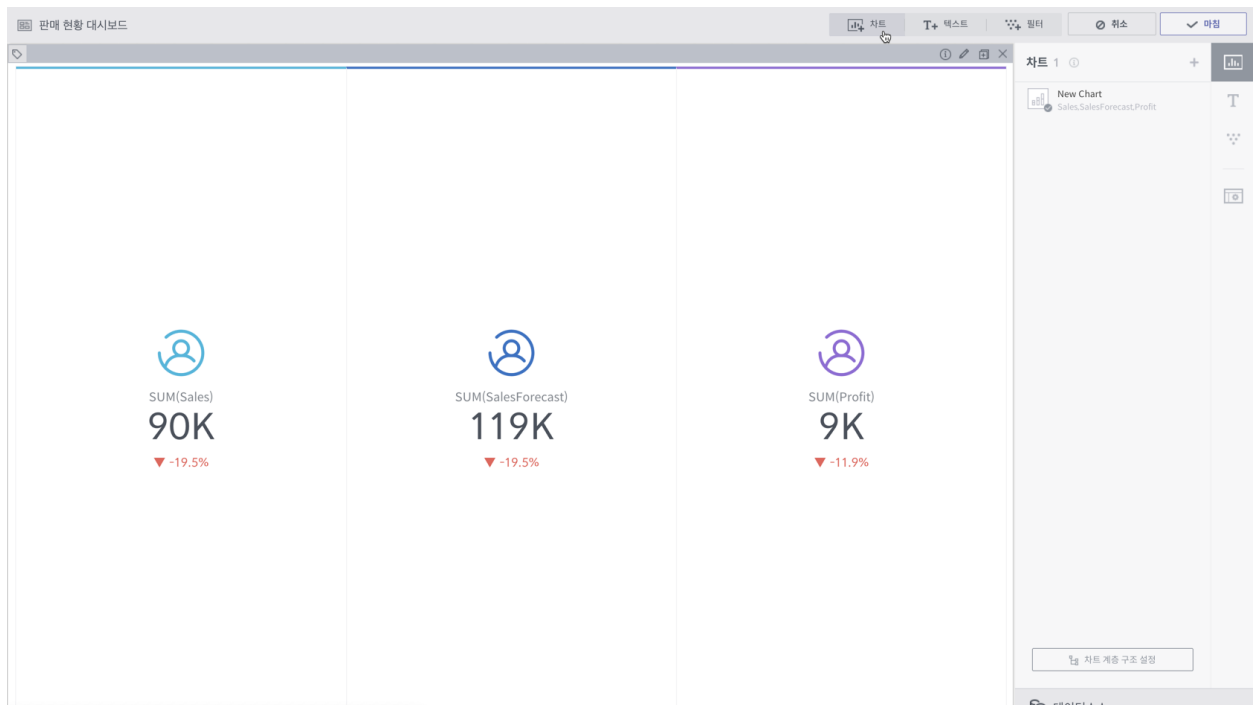
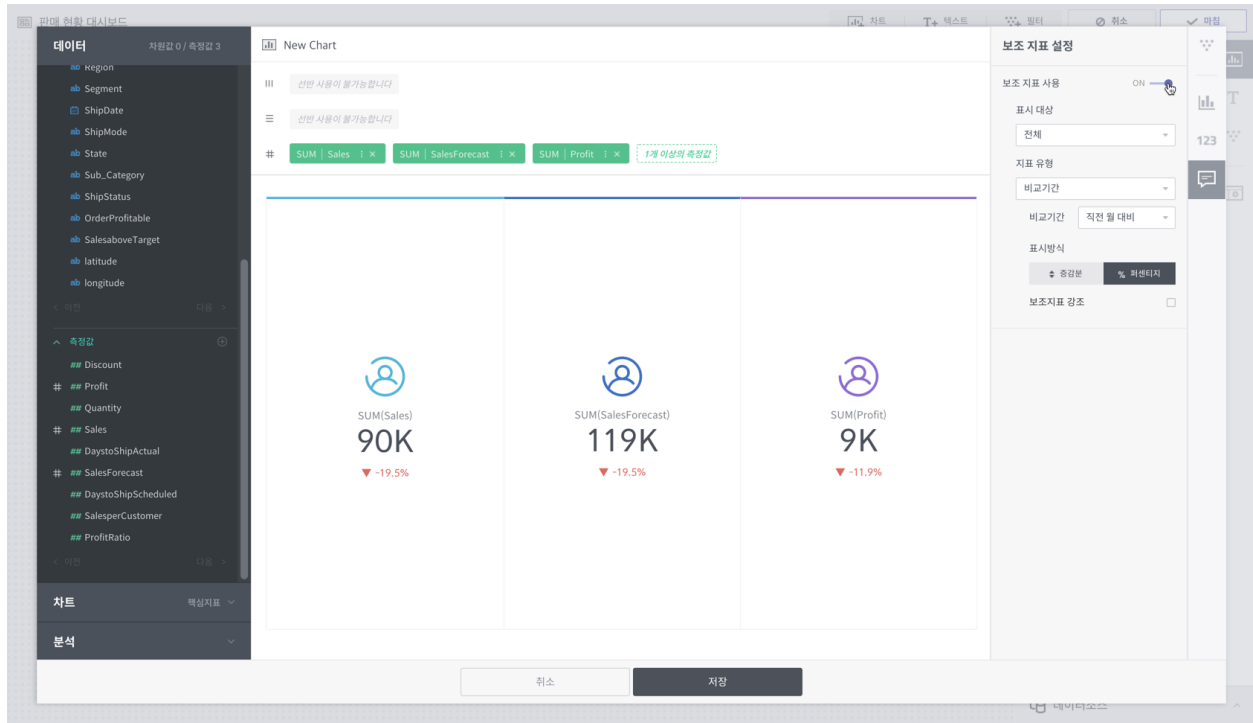
5. [Dashboard] Publish

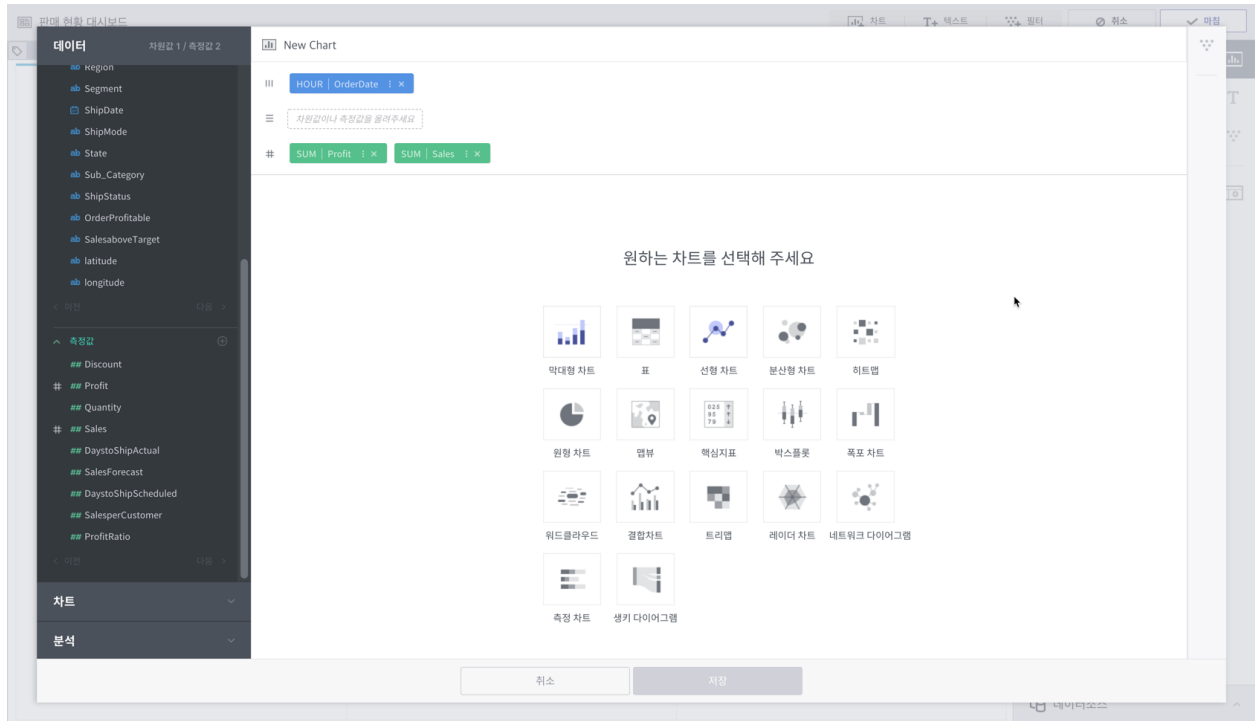


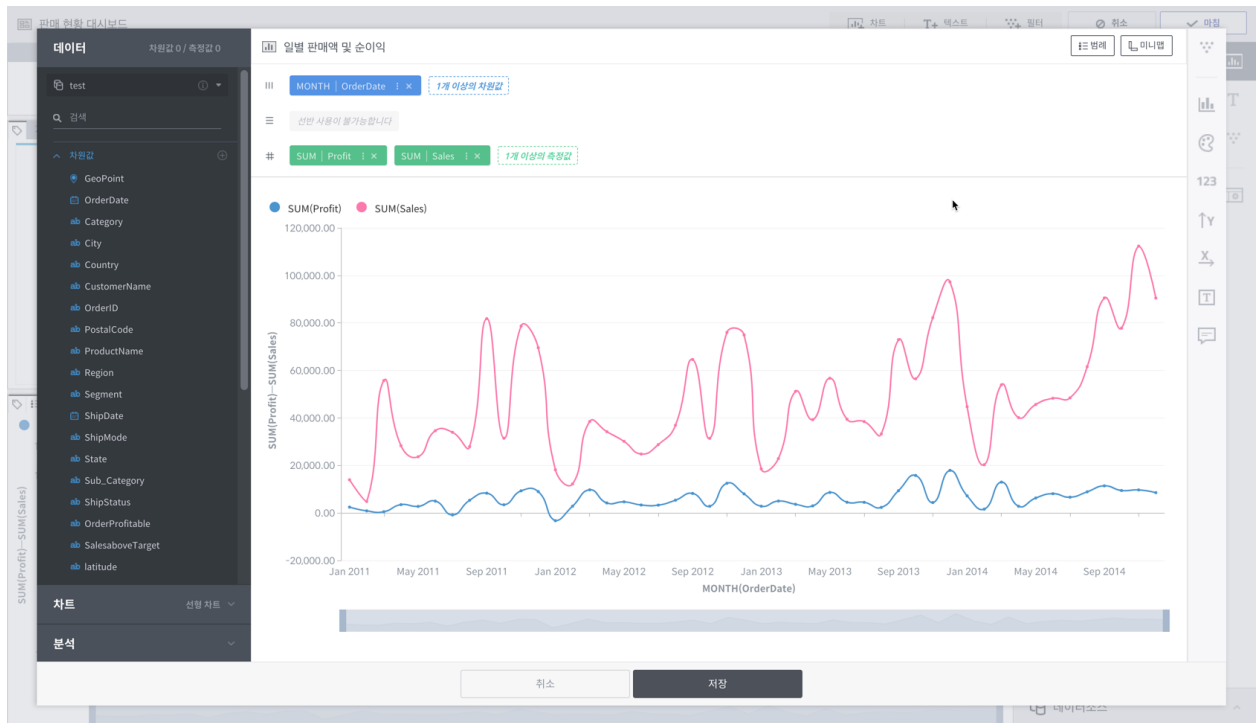
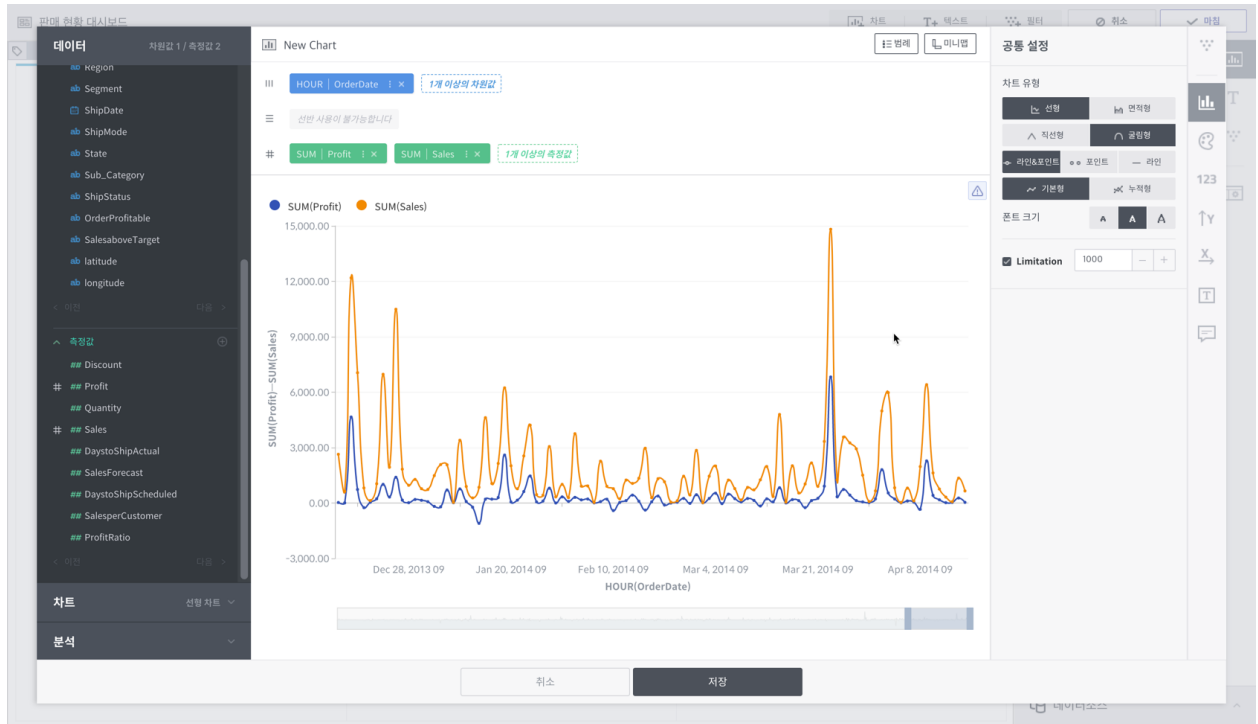


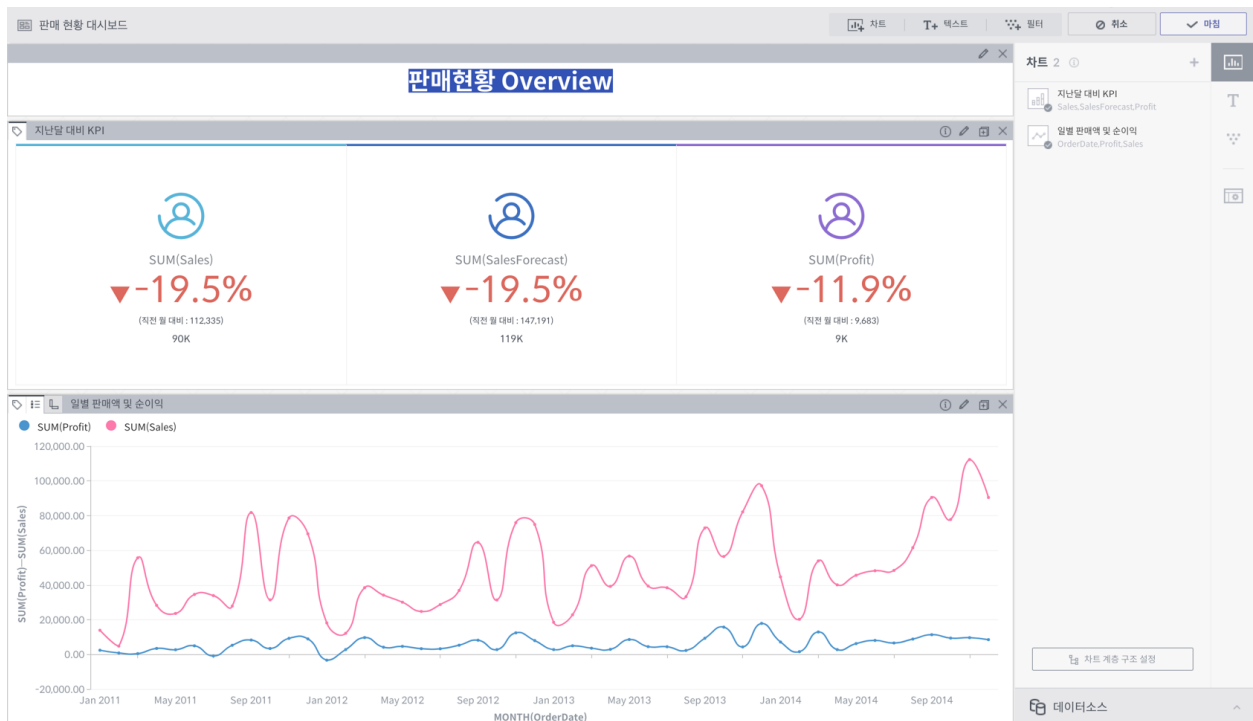
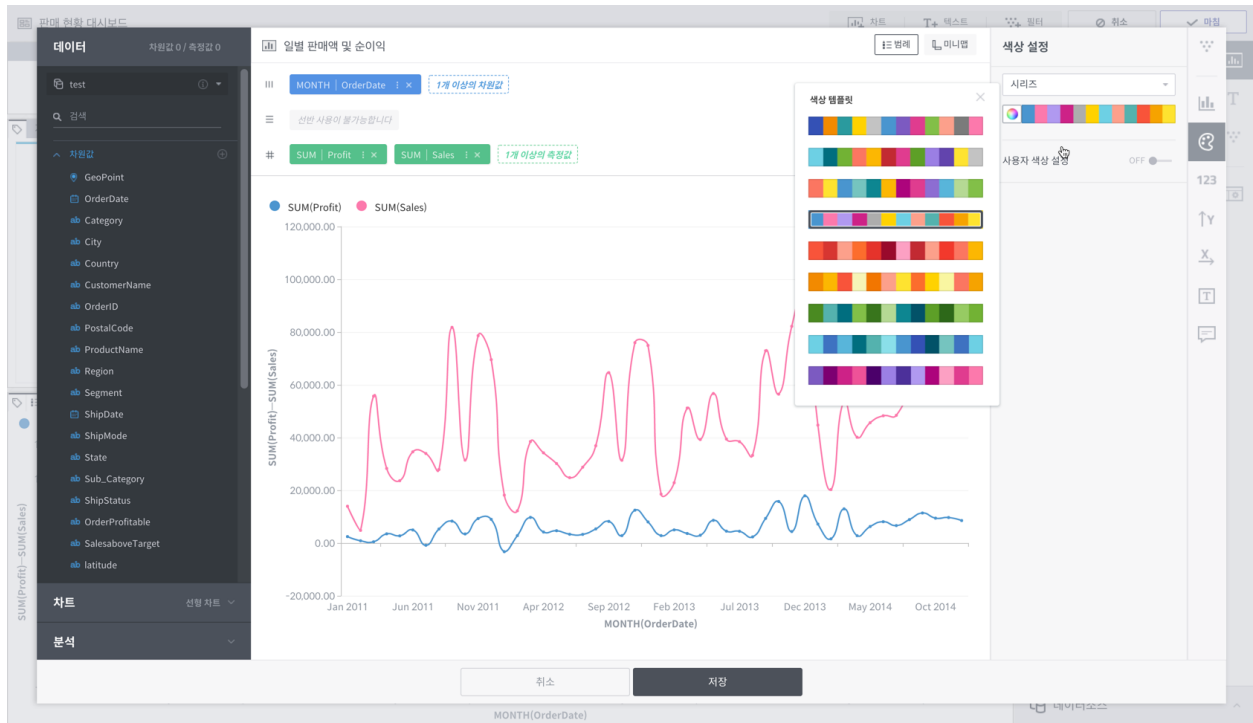


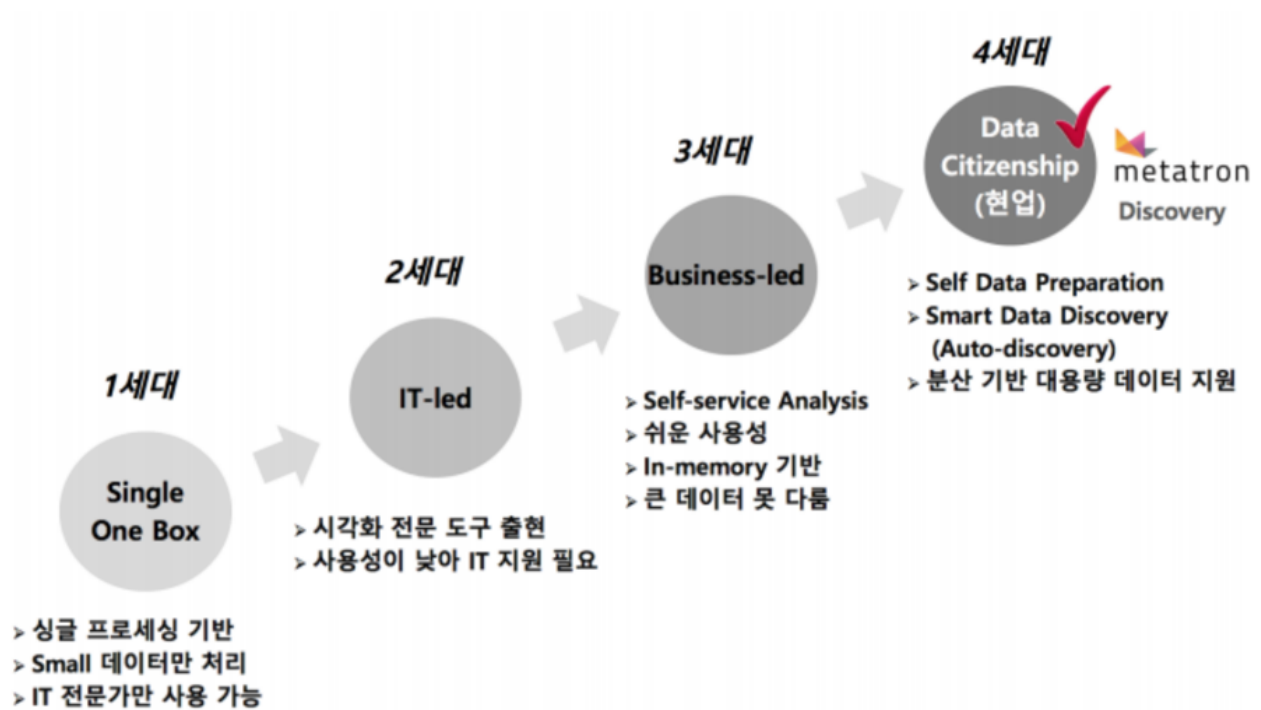


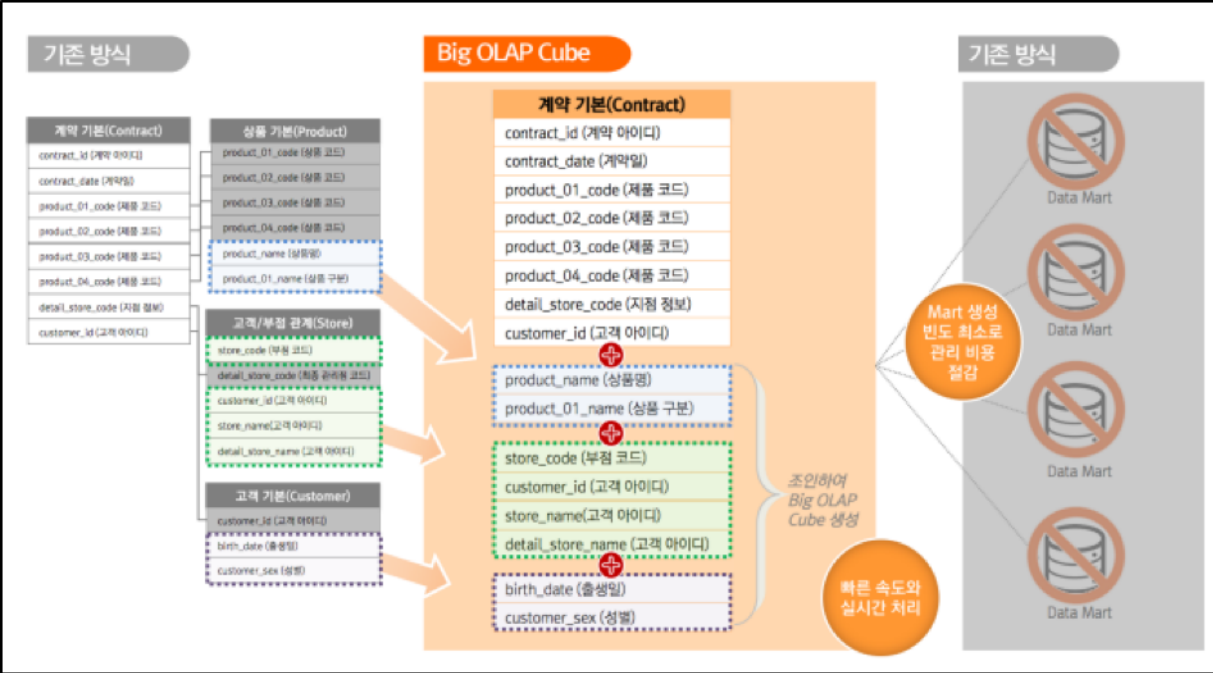






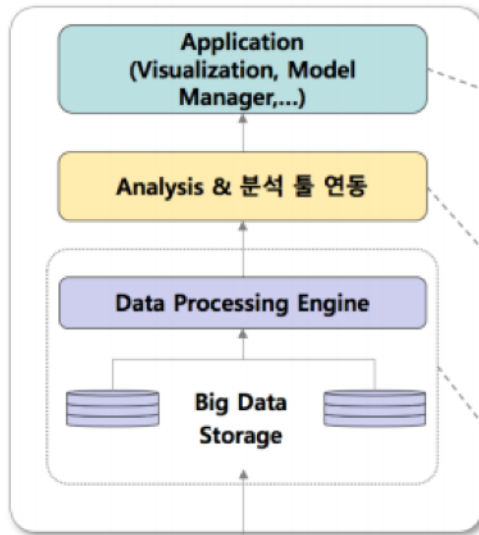






Single Solution for Data Discovery

【 Architecture 】



【 Features 】

App.

- > Massive Data의 빠른 시각화
- > 실시간 Dashboard를 통한 분석 결과 검증
- > Model Manager 통한 분석 모델 등록 및 실행

분석

- > Zepplin, Jupyter 연동으로 다양한 분석 가능
- > 산업별 특화 알고리즘 제공 (제조항 PCA, PLS 등)

저장 처리

- > Big OLAP Cube 생성 및 Dynamic 스키마 지원하여 ETL 비용 최소화
- > Time-series 데이터 Sub-second 처리
- > HDFS기반으로 전수 데이터의 저비용 저장 및 저장 속도 향상

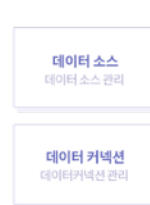
Master Data Source



Data Preparation & Ingestion



Data Storages (Big Mart)



Data Analytics & Visualization



데이터 모니터링



사용자 권한 및 계정



Worldcup match result

Please enter a description

Updated on 2018-07-11 14:56 by admin

+ New dataset

Imported Data 2

Wrangled Data 2

WorldCupMatches

WorldCups

WorldCups [W]

WorldCupMatches [W]

WorldCups [W]

Please enter description

Edit rules

Data preview

#	Year	Country	Win
1930	Uruguay	Uruguay	
1934	Italy	Italy	
1938	France	Italy	

TypeWrangled dataset

Summary21 Rows10 Columns

Used in1 Dataflows

Rule list

Convert row1 to header

Change 4 columns to Integer

Delete this data

Order_data [W]

Valid100%Mismatched0%Missing0%

#	o_orderkey	o_custkey	o_orderpriority	o_totalprice	o_orderdate	o_clerk
10023	24076639	1-URGENT	140859.05	1996-12-02	Clerk#000119081	
10048	16051268	4-NOT SPECIFIED	90041.69	1994-05-16	Clerk#000178505	
10049	13659839	1-URGENT	132332.62	1997-07-23	Clerk#000129536	
10050	20913502	2-HIGH	139876.9	1996-09-03	Clerk#000105918	
10051	13202551	2-HIGH	142753.95	1996-05-25	Clerk#000160701	
10052	15243578	4-NOT SPECIFIED	247129.69	1994-09-08	Clerk#000097525	
10053	28799506	2-HIGH	188696.34	1992-01-11	Clerk#000101471	
10054	13288838	2-HIGH	109539.09	1995-04-29	Clerk#000018523	
10055	25487725	5-LOW	99500.32	1996-02-29	Clerk#000101558	
10080	2163970	5-LOW	255230.55	1993-02-13	Clerk#000072269	
10081	9827765	4-NOT SPECIFIED	314821.78	1993-08-08	Clerk#000019398	
10082	25241941	4-NOT SPECIFIED	207989.78	1994-08-31	Clerk#000009746	
10083	11293418	5-LOW	73335.8	1995-09-26	Clerk#000189673	
10084	28470898	1-URGENT	3495.11	1997-07-08	Clerk#000017542	

Databasedefault

TableOrder_list_Snapshot_1103

Summary300,000,000 Rows9 Columns

Size10 GB

Elapsed Time0:1:11.0

Created2017-11-17 14:19:50

Summary

Analyze order lists by customer

Dataset

Order_data [W]

Created at 2017-11-17 14:19:50

Last modified at 2017-11-17 14:19:50

Origin imported dataset

Datasource

Order_data

Query statement

SELECT * FROM tpch.orders

Created at 2017-11-17 09:42:41



Create data connection

Please set required items and complete data connection creation

DB type

Oracle ✓

MySQL

PostgreSQL

Hive

presto

APACHE PHOENIX

Tibero

Server

Host	Port	SID
http://192.10.20.85	3306	
☒ URL only		
User ID for test	Password fot test	
polaris	*****	
Security		
☒ Always connect		
☐ Connect by user's account		
☐ Connect with ID and password		
Validation CheckInvalid Connection. Please check server and account information		

Permission

1 Workspace [Edit](#)

☒ Allow all workspaces to use this datasource

Advanced setting

Socket timeout60 Sec.

Connection name

Enter connection name

Previous

Next



Admin workspace

Owner

Workspace List

Workbook 20 Notebook 0 Workbench 7

23 Datasource

Created on 2018-06-11 by Administrator

Datasource (23)

Search by datasource name

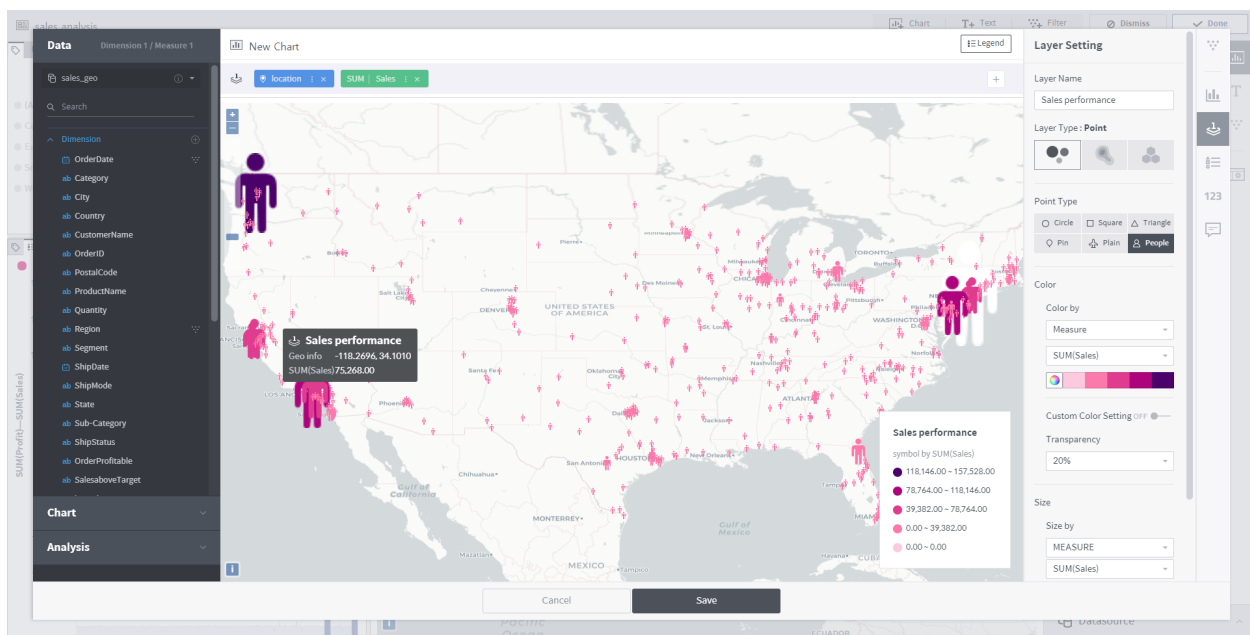
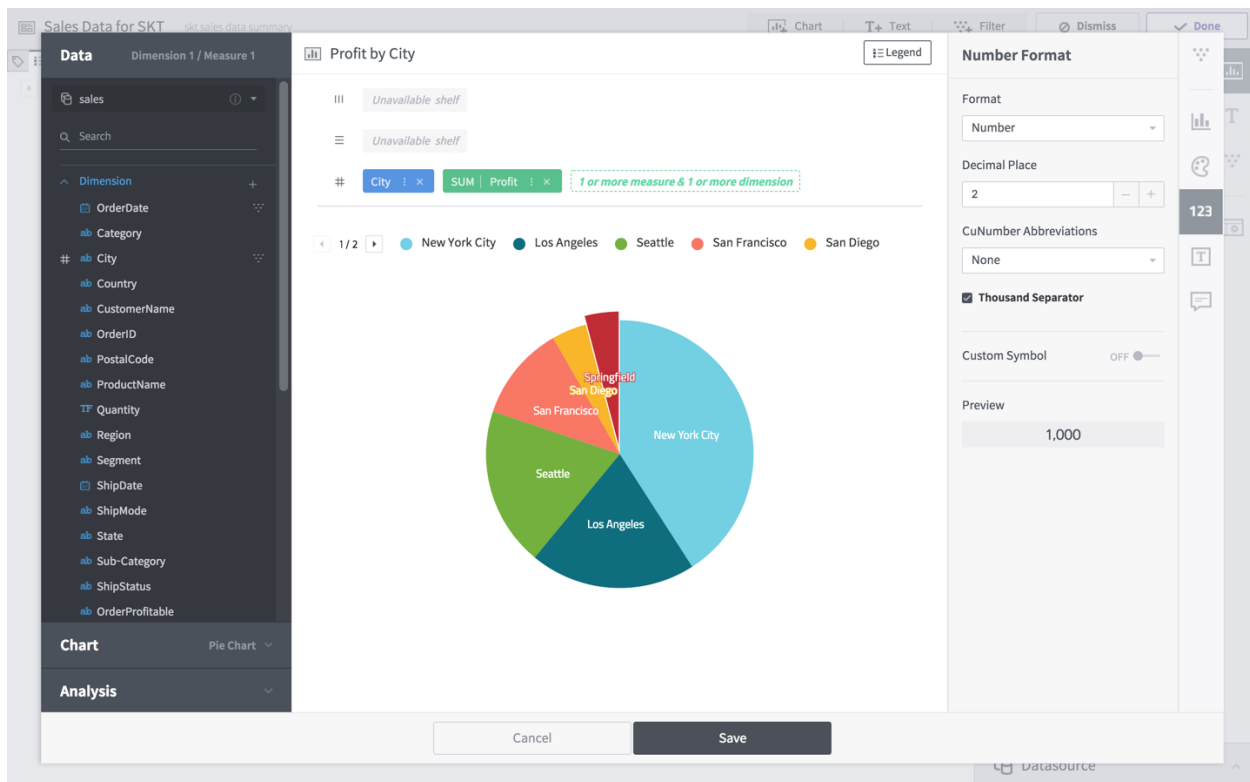
Show open data only

Type

All

No.	Datasource	Type	Used in	Full size	Updated
16	The_2014_Inc_5000 Open data	Ingested type	Open data	1.19 MB	2018-07-10
17	EMSI_JobChange_UK Open data	Ingested type	Open data	46.73 KB	2018-07-10
18	OECD_TAX_ALL_02 Open data	Ingested type	Open data	926.70 KB	2018-07-09
19	WorldCup_Matches Open data	Ingested type	Open data	69.31 KB	2018-07-06
20	oecd_test Open data	Ingested type	Open data	30.61 KB	2018-07-06
21	tour de france Open data	Ingested type	Open data	27.94 KB	2018-07-06
22	cell_1h	Ingested type	2 Workspaces	90.79 MB	2018-07-06
23	FIFA_18_Player_Ratings Open data	Ingested type	Open data	3.41 MB	2018-07-06
More					

Close



Notebook - Job

anonymous

notebook_test

▶
⌂
🔍
📄
📁
📌
📎
📏
📊
Head
🗑️
⚙️

🔒
🔧
🔒
default

```
// 1. load dataset
import app.metatron.discovery.connector._;
val conf = new MetisClientSetting();
conf.setting("host", "metatron-web-01").setting("port", "8080");
val client = new MetisClient(conf);
val dataset = client.loadData(spark, "datasources", "ds-gis-37", "1000")
```

READY ▶ ⌂ 🔍 📄 📁 📌 📎 📏 📊

```
// 2. analyze
dataset.show()
```

READY ▶ ⌂ 🔍 📄 📁 📌 📎 📏 📊

READY ▶ ⌂ 🔍 📄 📁 📌 📎 📏 📊

≡
METATRON DISCOVERY
👤

←
Workbench
테스트
테스트입니다

Hive(2.3)

🔍
default

Table list

🔍
search

- contract
- contract_part
- contract_date_non_parti_1
- contract_tmp_parti
- ttt
- contract_orc_part
- contract_orc
- employee
- test
- sample_ingestion
- sample_ingestion_time
- sample_ingestion_orc
- sample_ingestion_time_orc
- test_sample_ingestion_pa...

쿼리 01
+

```

1 SELECT A.C.ONE,
2         A.C.TWO,
3         SUM(A.C.TEN)
4 FROM TB_NUM AS A
5 WHERE A.C.ONE = 5
6 GROUP BY A.C.ONE, A.C.TWO;
7
8 USING 'GROUP BY' QUERY EXAMPLE
9 COMMENT ON TABLE USER_INFO_EX
10      IS '고객 정보 확장'; -- USER_INFO_EX 테이블에 주석 추가
11
12 SELECT *
13 FROM USER_TAB_COMMENTS
14 WHERE TABLE_NAME = 'USER_INFO_EX'; -- USER_INFO_EX 테이블의 주석 확인
15
16 -----
17 COMMENT ON COLUMN USER_INFO_EX.RNAME
18      IS '고객 실제 이름'; -- USER_INFO_EX 의 RNAME 컬럼에 주석 추가

```

쿼리 01 - 결과 1
×

🔍
Search by Column Data
데이터 전체선택

📄
🔍
📄
📄

SEQ	L_orderkey	L_partkey	L_suppkey	L_linenum	L_quantity	L_extendedprice	L_discount
1	1	31037869	1537885	1	17.0	30690.27	0.04
2	1	13461816	1461817	2	36.0	63977.04	0.09
3	1	12739956	739957	3	8.0	15962.56	0.1
4	1	426299	926300	4	28.0	34307.56	0.09

1000 Rows



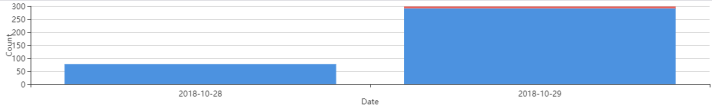
Data Monitoring

Log Statistics Job Log

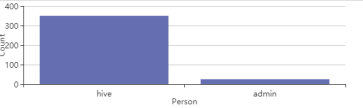
Log type All Performed Start Time Today Last 7 days 2018-10-29 00:00 - 2018-10-29 23:59 Apply

Q Search by user ID

Daily query success / failure rate



Daily query frequency by user



In order of longest

Query	Query Time	User	Result	Elapsed time	More
SELECT city, count(*) FROM default.sales group by city	2018-10-29 10:46	admin	FAIL	47 min 46--	
SELECT city, count(*) FROM default.sales group by city	2018-10-29 10:19	admin	FAIL	22 min 11--	
SELECT city, count(*) FROM default.sales group by city	2018-10-29 10:22	admin	FAIL	18 min 10--	
SELECT city, count(*) FROM default.sales group by city	2018-10-29 10:21	admin	FAIL	16 min 45--	
SELECT city, count(*) FROM default.sales group by city	2018-10-29 10:03	admin	FAIL	11 min 5 sec	

Amount of scan data

Query	Query Time	User	Result	Row Count	More
SELECT job_log FROM polaris.audit	2018-10-29 15:14	admin	SUCCESS	100,001	
SELECT * FROM polaris.audit	2018-10-29 15:14	admin	SUCCESS	100,001	
SELECT * FROM default.sales	2018-10-29 10:03	admin	SUCCESS	9,995	
SELECT city, count(*) FROM default.sales group by city	2018-10-29 14:03	admin	SUCCESS	532	
SELECT city, count(*) FROM default.sales group by city	2018-10-29 13:59	admin	SUCCESS	532	

Frequency of successful queries

Query	Count	More
show databases	77	
SHOW SCHEMAS	60	
SELECT 'time' as 'time', 'd' as 'd', 'sd' as 'sd', 'm1' as 'm1', 'm2' as 'm2' FROM default.sales	32	
SELECT 'time', 'd', 'sd', 'm1', 'm2', 'ym', 'dd' FROM (SELECT * FROM default.sales)	32	
SHOW TABLES IN default	27	

Frequency of failed queries

Query	Count	More
SELECT city, count(*) FROM default.sales group by city	4	
SELECT city, count(*) FROM default.sales group by city	2	
SELECT * FROM cazen_lee.snapshot1 where no = 'asf'	1	

Total memory usage

Query	Application ID	Queue	Memory @	More
SELECT count(*) as c1, a.owner, b.sourcedatabasename, a.sourcetablename, ...	application_1540788884--	default	323,837	
SELECT city, count(*) FROM default.sales group by city	application_1540788884--	default	130,786	
SELECT city, count(*) FROM default.sales group by city	application_1540788884--	default	128,310	
SELECT * FROM cazen_lee.snapshot1 where no = 'asf'	application_1540788884--	default	80,154	
SELECT 'time' as 'time', 'd' as 'd', 'sd' as 'sd', 'm1' as 'm1', 'm2' as 'm2' FROM default.sales	-	-	-	

Total CPU usage

Query	Application ID	Queue	CPU @	More
SELECT count(*) as c1, a.owner, b.sourcedatabasename, a.sourcetablename, ...	application_1540788884--	default	51	
SELECT city, count(*) FROM default.sales group by city	application_1540788884--	default	20	
SELECT city, count(*) FROM default.sales group by city	application_1540788884--	default	20	
SELECT * FROM cazen_lee.snapshot1 where no = 'asf'	application_1540788884--	default	12	
SELECT 'time' as 'time', 'd' as 'd', 'sd' as 'sd', 'm1' as 'm1', 'm2' as 'm2' FROM default.sales	-	-	-	

Total Queue resource usage

Queue	Memory usage @	CPU usage @	More
default	663,087	103	

Timestamp column	Dimension columns				Metric columns	
timestamp	publisher	advertiser	gender	country	click	price
2011-01-01T01:01:35Z	bieberfever.com	google.com	Male	USA	0	0.65
2011-01-01T01:03:63Z	bieberfever.com	google.com	Male	USA	0	0.62
2011-01-01T01:04:51Z	bieberfever.com	google.com	Male	USA	1	0.45
2011-01-01T01:00:00Z	ultratrifast.com	google.com	Female	UK	0	0.87
2011-01-01T02:00:00Z	ultratrifast.com	google.com	Female	UK	0	0.99
2011-01-01T02:00:00Z	ultratrifast.com	google.com	Female	UK	1	1.53

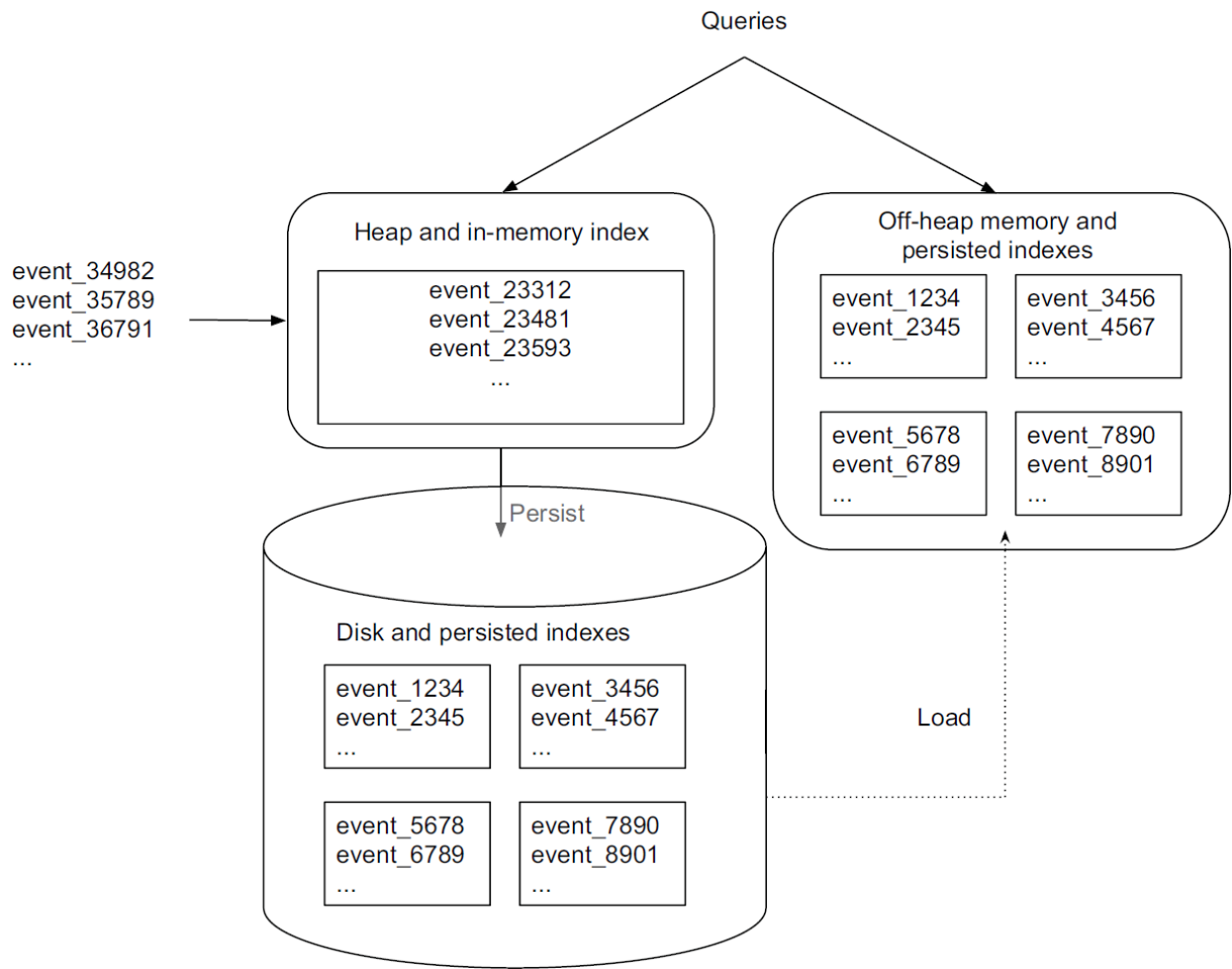
timestamp	domain	gender	clicked
2011-01-01T00:01:35Z	bieber.com	Female	1
2011-01-01T00:03:03Z	bieber.com	Female	0
2011-01-01T00:04:51Z	ultra.com	Male	1
2011-01-01T00:05:33Z	ultra.com	Male	1
2011-01-01T00:05:53Z	ultra.com	Female	0
2011-01-01T00:06:17Z	ultra.com	Female	1
2011-01-01T00:23:15Z	bieber.com	Female	0
2011-01-01T00:38:51Z	ultra.com	Male	1
2011-01-01T00:49:33Z	bieber.com	Female	1
2011-01-01T00:49:53Z	ultra.com	Female	0

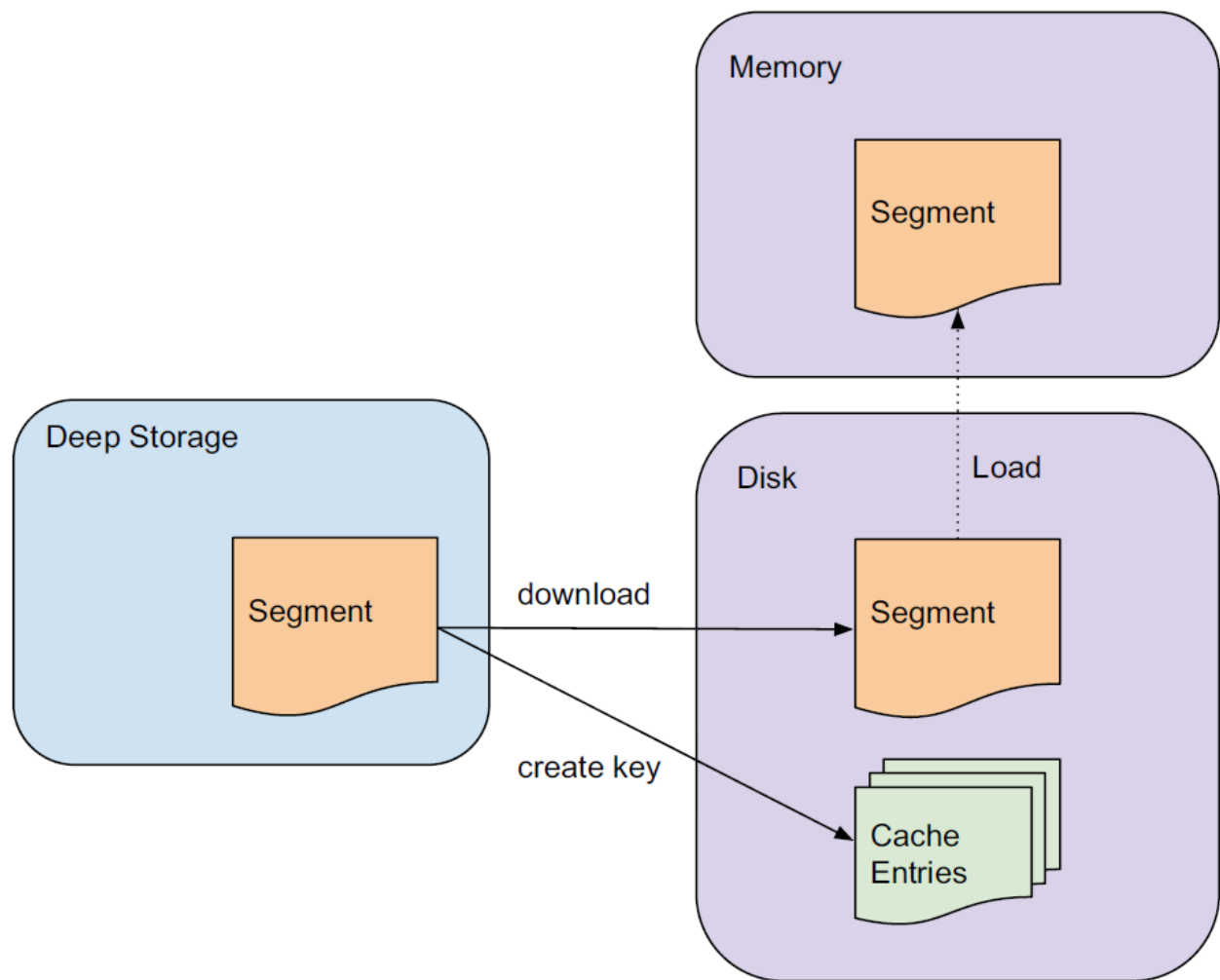


timestamp	domain	gender	clicked
2011-01-01T00:00:00Z	bieber.com	Female	1
2011-01-01T00:00:00Z	ultra.com	Female	2
2011-01-01T00:00:00Z	ultra.com	Male	3

Timestamp	Page	Username	Gender	City	Characters Added	Characters Removed
2011-01-01T01:00:00Z	Justin Bieber	Boxer	Male	San Francisco	1800	25
2011-01-01T01:00:00Z	Justin Bieber	Reach	Male	Waterloo	2912	42
2011-01-01T02:00:00Z	Ke\$ha	Helz	Male	Calgary	1953	17
2011-01-01T02:00:00Z	Ke\$ha	Xeno	Male	Taiyuan	3194	170

--





 **Approximate Algorithms**

 **Visualizations**

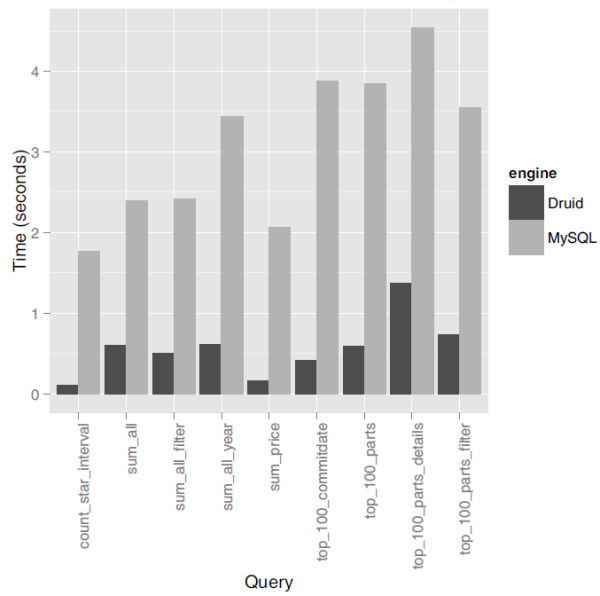
 **Machine Learning**

 **Druid**

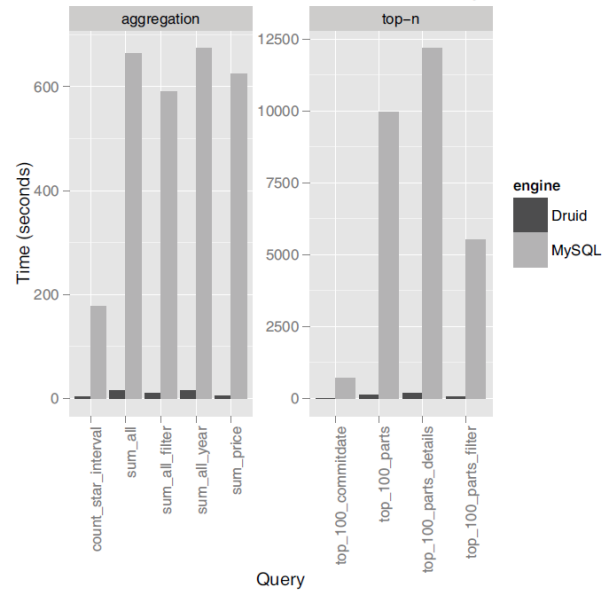
 **Batch Ingestion (Hadoop)**

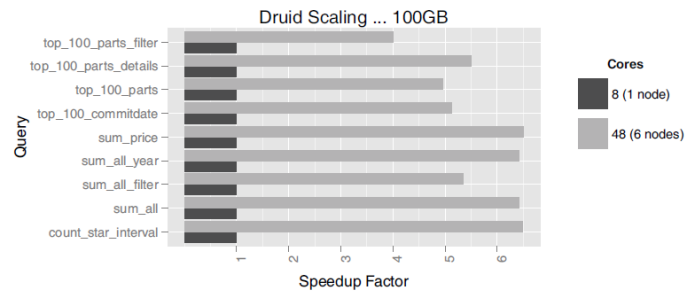
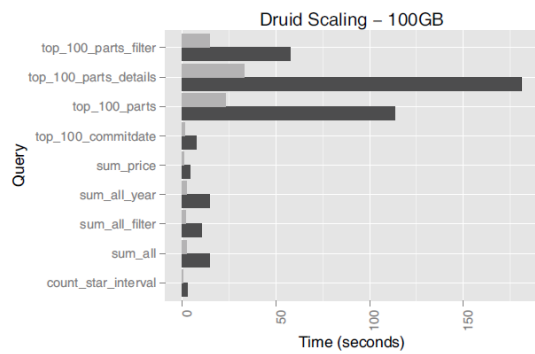
 **Straming Ingestion (Storm)**

Median query time (100 runs) – 1GB data – single node



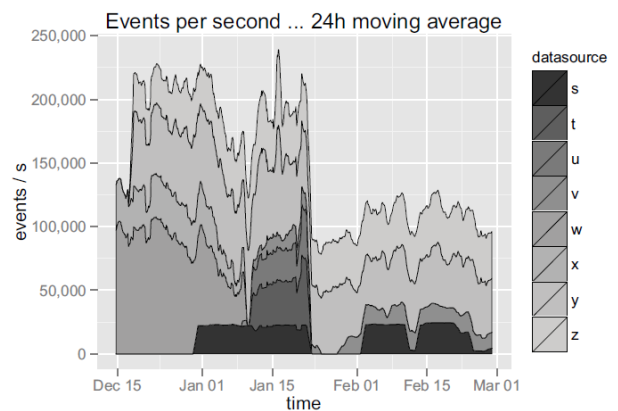
Median Query Time (3+ runs) – 100GB data – single node





Data Source	Dimensions	Metrics	Peak events/s
s	7	2	28334.60
t	10	7	68808.70
u	5	1	49933.93
v	30	10	22240.45
w	35	14	135763.17
x	28	6	46525.85
y	33	24	162462.41
z	33	24	95747.74

Ingestion characteristics of various data sources



Combined cluster ingestion rates

