



Discovery 사용자 매뉴얼

크고 복잡한 데이터를 쉽고 빠르게 연결·분석·시각화





metatron 3.0

1부

Discovery 개요

Copyright ©  SK telecom

SK Telecom 의 사전 승인 없이
본 내용의 전부 또는 일부에 대한 복사, 배포, 사용을 금합니다.

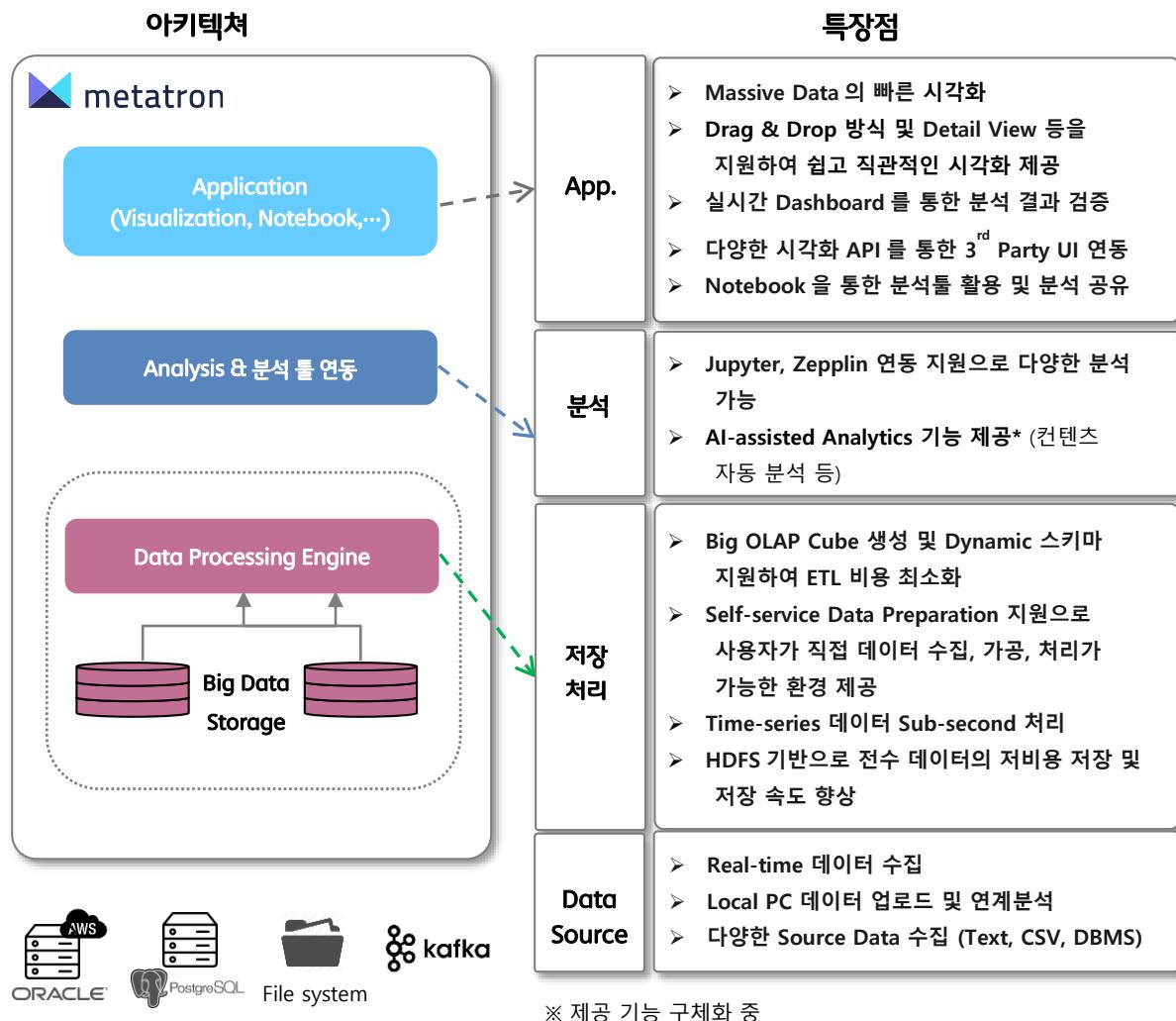
목 차

1. METATRON DISCOVERY 특장점	3
2. METATRON DISCOVERY 구조	5
3. 관계형 OLAP VS. 다차원 OLAP	7
3.1 관계형 OLAP 시스템의 특징과 한계	7
3.2 다차원 OLAP 시스템과 Druid 엔진	9

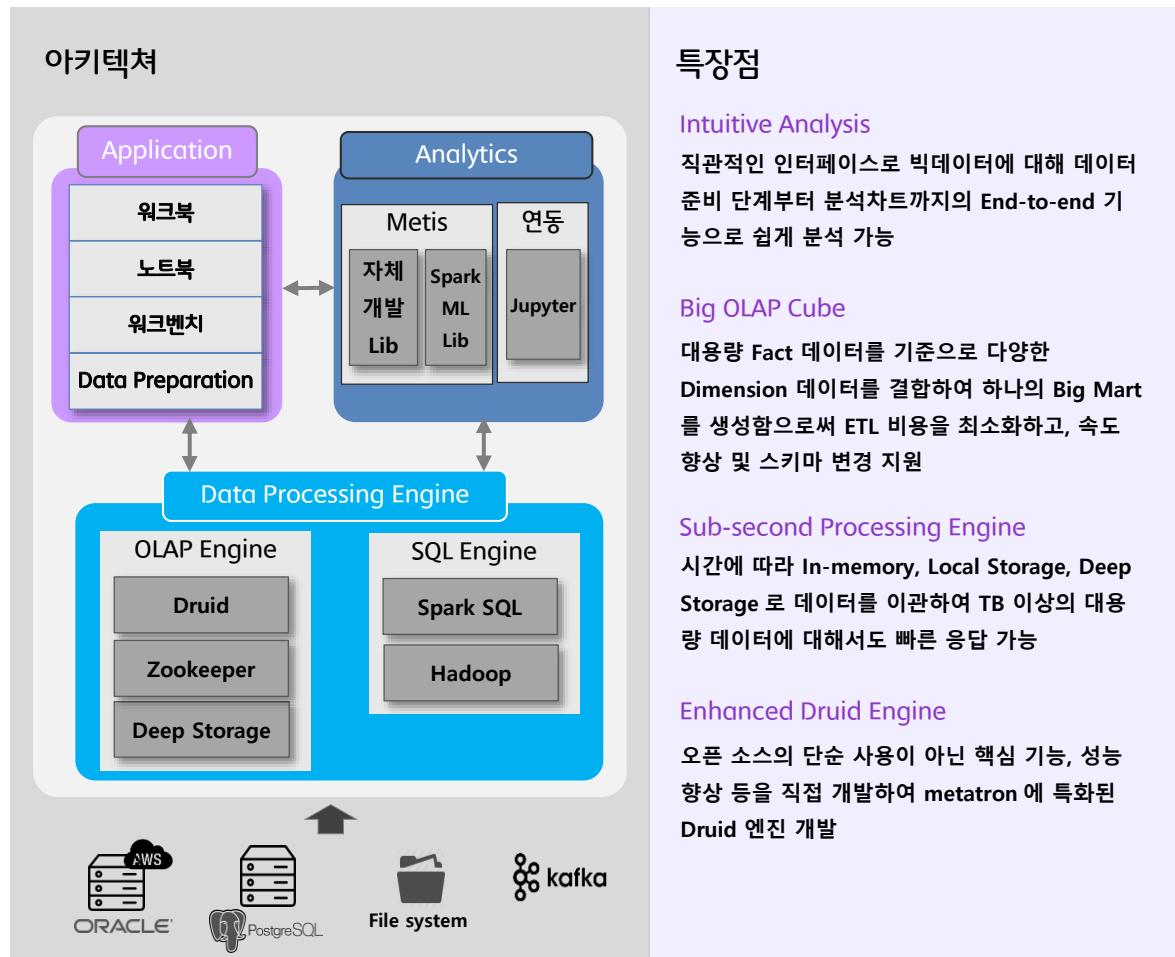
1. metatron Discovery 특장점

metatron Discovery는 대용량 데이터에 대한 전처리부터 시각화 기반 데이터 탐색, 고급분석까지 End-to-end로 지원하는 Single Solution for Data Discovery를 위한 솔루션입니다. 아래 그림은 metatron의 아키텍쳐와 주요 특장점을 정리한 것입니다.

metatron 기반 기술의 아키텍쳐와 특장점

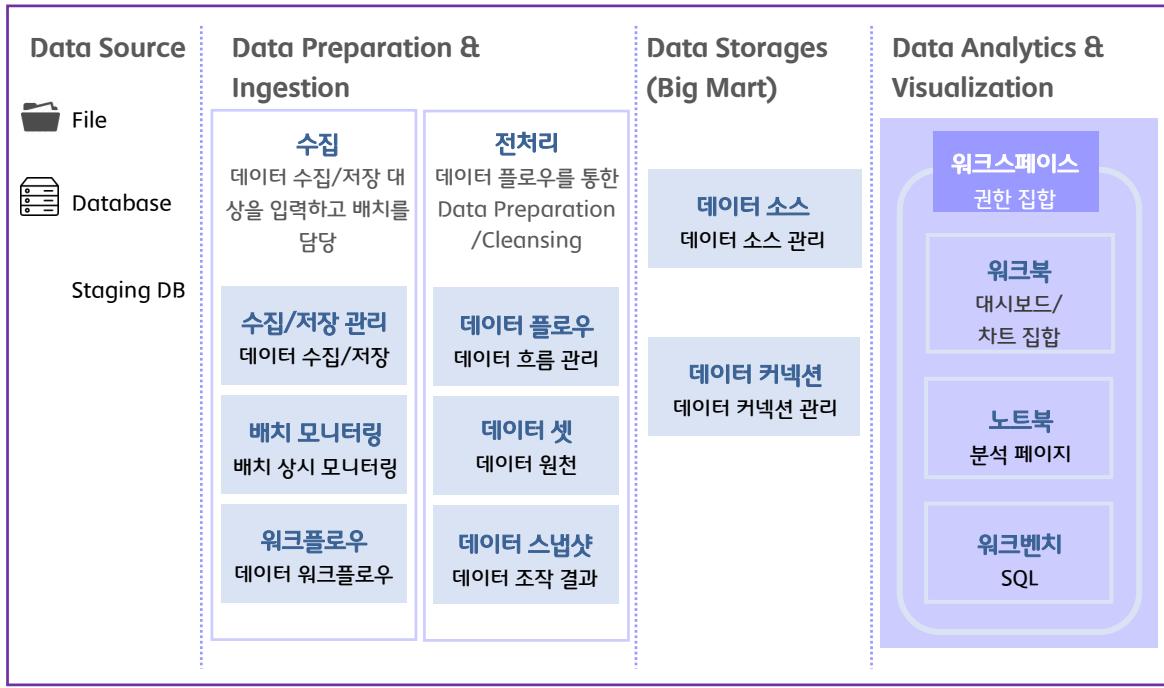


metatron Discovery 모듈 아키텍쳐와 특장점



2. metatron Discovery 구조

metatron Discovery는 metatron 운용 서버에 적재된 데이터 소스나 그 밖의 외부 데이터 소스로부터 원하는 데이터를 불러와서 각종 고급 분석 기능을 통해 분석한 후, 그 결과를 다양한 형식의 차트와 보고서로 출력하는 기능을 합니다. 본 모듈을 이용하려면 다음과 같은 전반적인 구조를 이해해야 합니다.



원천 데이터(Source Data)

데이터 소스 생성 시 가져오는 데이터를 말합니다. 현재 metatron에서는 파일 형식, 데이터 베이스, 스테이징 DB 형식을 지원합니다.

데이터 전처리, 적재(Data Preparation & Ingestion)

원천 데이터에서부터 데이터를 정제 및 가공하여 metatron의 Hive 저장소에 적재하거나 또는 로컬 파일로 저장하는 기능을 제공합니다.

데이터 스토리지(Data Storages)

데이터를 metatron의 Druid 엔진에 다차원 OLAP 기반으로 저장합니다. 이러한 형태로 저장된 데이터는 시각화·분석이 용이합니다. (자세한 사항은 3.2절 참조)

데이터 분석, 시각화(Data Analytics & Visualization)

데이터 스토리지에 저장한 데이터를 탐색·분석·시각화하는 기능을 제공합니다.

- **워크스페이스:** 워크북, 워크벤치, 노트북을 생성·관리하는 작업 공간입니다.
- **워크북:** Druid 엔진에 저장되어 있는 시계열 기반의 다차원 데이터 소스를 분석·시각화·보고할 수 있습니다.
- **노트북:** 외부 분석 도구 R, python, Scala 등을 연동하여 Deep learning, machine learning 등의 고급 분석을 수행할 수 있습니다.
- **워크벤치:** SQL 기반 분석을 수행할 수 있습니다.

데이터 모니터링

metatron의 Staging DB(내부 Hive DB)와 워크벤치(외부 DB)에서 발생하는 각종 데이터 로그를 관측하는 기능입니다.

사용자 권한 및 계정

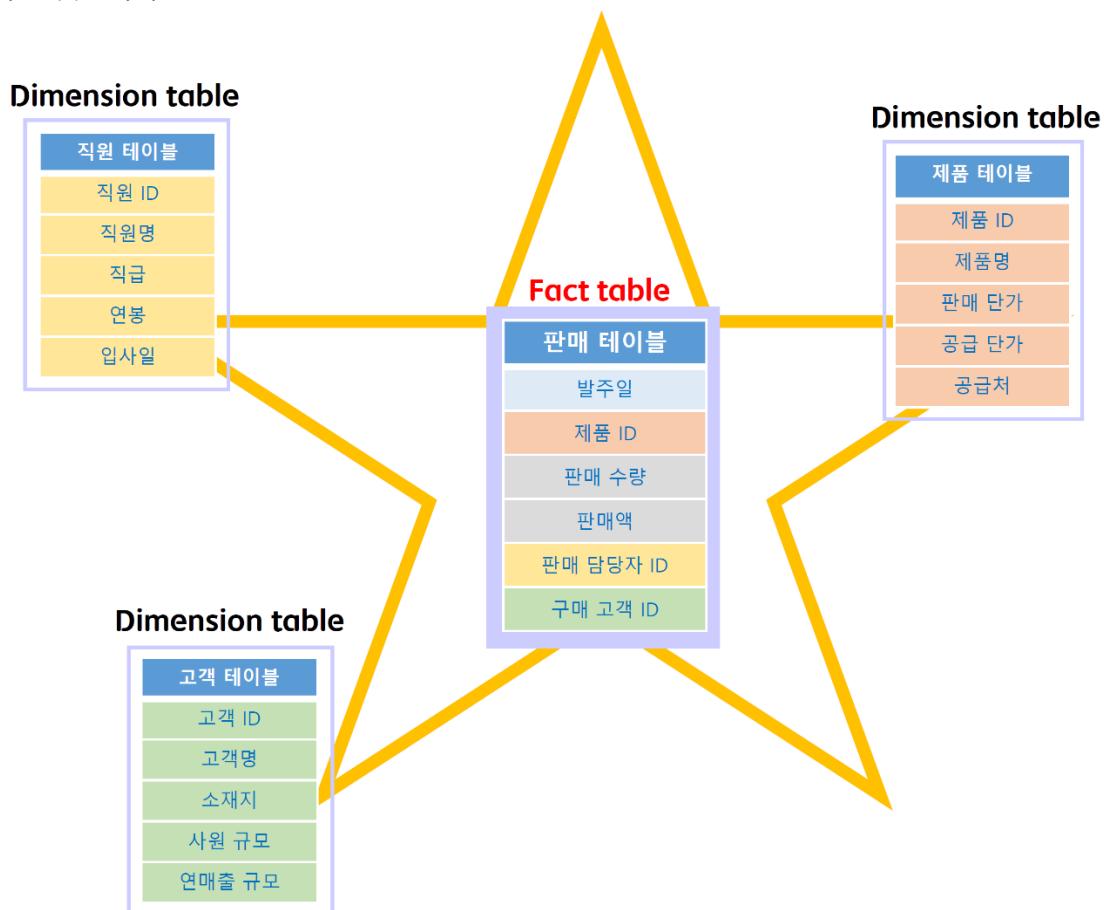
metatron Discovery의 사용자를 추가, 삭제하거나 사용자의 권한을 관리합니다.

3. 관계형 OLAP vs. 다차원 OLAP

metatron Discovery에는 빅데이터 분석을 위한 여러 가지 첨단 시스템과 방법론이 사용됩니다. 이 중에서 Druid 엔진을 이용한 '다차원 OLAP 시스템'의 개념은 Discovery를 제대로 사용하기 위해 반드시 숙지해야 하는 개념이므로 여기서 간단하게 설명하고자 합니다. 본 장에서는 우선 기존에 데이터 수집 및 관리를 위해 가장 보편적으로 사용되었던 관계형 OLAP 시스템의 특징과 한계를 간단히 설명한 후, Discovery에서 채택한 다차원 OLAP 시스템에서 이러한 문제를 어떻게 극복하는지 확인합니다.

3.1 관계형 OLAP 시스템의 특징과 한계

관계형 OLAP 시스템은 데이터를 수집 및 관리하는 데 보편적으로 사용되고 있는 방식입니다. 아래 그림은 관계형 OLAP 시스템에서 가장 흔하게 사용되는 '스타 스키마'(star schema) 구조의 예시를 나타낸 것입니다.



그림과 같이 기본적인 스타 스키마는 여러 개의 차원 테이블(dimension table)과 소수의 사실 테이블(fact table)로 구성됩니다.

- **차원 테이블**에는 스키마를 구성하는 각 차원에 대한 상세한 데이터가 보관됩니다. 위 그림에 제시된 예시에서는 총 3개의 차원 테이블이 각각 제품, 직원, 고객에 관한 데이터를 담고 있습니다. 이들 각각을 '차원'이라고 부르는 이유는 데이터를 분류하는 기준으로 사용할 수 있기 때문입니다(예: 직원별 판매량, 고객별 구매량 등).
- **사실 테이블**에는 새로운 이벤트가 발생할 때마다 그 이벤트의 내역을 담은 레코드가 하나씩 추가되는데, 이때 각 차원 테이블과는 참조 ID를 매개로 연결이 됩니다.

아래 그림은 이해를 돋기 위해 위 스키마 구조를 토대로 간단한 테이블 예시를 정리한 것입니다.

Dimension table (제품)					Dimension table (담당자)					Dimension table (고객)				
ID	제품명	판매 단가	공급 단가	공급처	ID	직원명	직급	연봉(만원)	입사일	ID	고객명	소재지	사원 규모	연매출 규모
P1	연필	200	100	동예공업	E1	홍길동	대리	3,500	2009.09.01	C1	백제기업	공주	100~499명	500~999억
P2	가위	1,000	700	부여산업	E2	장영실	과장	4,000	2006.10.10.	C2	고구려기업	평양	50~99명	100~199억
P3	지우개	100	50	동예공업	E3	이성계	사원	2,500	2016.07.08	C3	신라기업	경주	100~499명	500~999억
P4	볼펜	500	300	옥저공업	E4	김유신	차장	6,500	1999.08.01	C4	가야기업	김해	5~49명	50~99억

Fact table					
발주일	제품 ID	판매수량	판매액	판매 담당자 ID	구매 고객 ID
2015-01-02	P2	8	8,000	E4	C4
2015-01-31	P2	20	20,000	E1	C2
2015-02-03	P2	16	16,000	E3	C2
2015-02-22	P4	19	9,500	E1	C4
2015-02-28	P3	3	300	E1	C2
2015-02-28	P3	8	800	E2	C3
2015-03-02	P2	4	4,000	E4	C3
2015-03-03	P3	20	2,000	E3	C2
2015-03-05	P3	8	800	E1	C2
2015-04-04	P3	16	1,600	E3	C3
2015-04-22	P3	10	1,000	E4	C4
2015-05-05	P4	9	4,500	E2	C1
2015-06-09	P2	3	3,000	E3	C2
2015-08-11	P3	5	500	E2	C3
2015-09-21	P3	9	900	E1	C1
2015-10-30	P4	3	1,500	E4	C4
2015-11-11	P4	15	7,500	E3	C4
...
2018-10-19	P2	12	12,000	E4	C3
2018-11-02	P3	20	2,000	E2	C4

여기서 사실 테이블의 각 레코드에는 판매 제품, 판매 담당자, 구매 고객이 간단한 ID로만 기록되고 있지만, 이러한 ID 각각은 고유한 차원 테이블과 연결되어 있어 필요 시 자세한 정보를 조회하는 데 이용할 수 있습니다. 예를 들어, 첫 번째 레코드에 기록된 P2 제품은 상품 테이블을 조회하면 해당 제품 종류가 연필이고 판매 단가는 200원, 공급 단가는 100원, 공급처는 동예공업임을 알 수 있습니다. 이러한 데이터 수집 방식을 사용하면, 사실 테이블에 최소한의 데이터만 저장함으로써 저장 공간을 효율적으로 사용할 수 있다는 장점이 있습니다.

하지만 이렇게 쌓인 데이터를 분석하고자 할 때에는 관계형 OLAP의 'ID 참조' 기반이 오히려 데이터 접근성과 속도를 저하시키는 원인이 됩니다. 사실 테이블에서 분석에 사용되는 레코드 각각에 포함된 모든 참조 ID에 연결된 원본 데이터를 각각의 차원 테이블에서 일일이 로드해야 하기 때문입니다. 만일 분석 대상 레코드 수가 10만 개이고 각 레코드별 참조 ID가 4개라면, 10만 x 4회의 데이터 로드 과정을 거쳐야 하고 그만큼 데이터 처리 장치에 부하가 발생하게 됩니다. 더군다나 해당 데이터베이스의 스키마 구조를 제대로 이해하고 있는 소수의 숙련자만이 이러한 데이터 로드 쿼리를 이용할 수 있다는 한계도 있습니다.

다음 절에서 설명하는 다차원 OLAP 시스템은 1개 또는 소수의 테이블에 모든 데이터를 저장함으로써 이러한 레거시 구조의 한계를 극복합니다.

3.2 다차원 OLAP 시스템과 DRUID 엔진

앞서 설명한 스타 스키마 기반의 관계형 OLAP 시스템과 달리, 다차원 OLAP 시스템은 1개 또는 소수의 다차원 테이블을 이용합니다. metatron Discovery의 경우 Druid 엔진을 통해 이러한 OLAP 시스템을 구현합니다. 아래는 앞 절의 관계형 OLAP 시스템 설명에 사용했던 데이터와 동일한 대상을 Druid 데이터 테이블로 나타낸 것입니다.

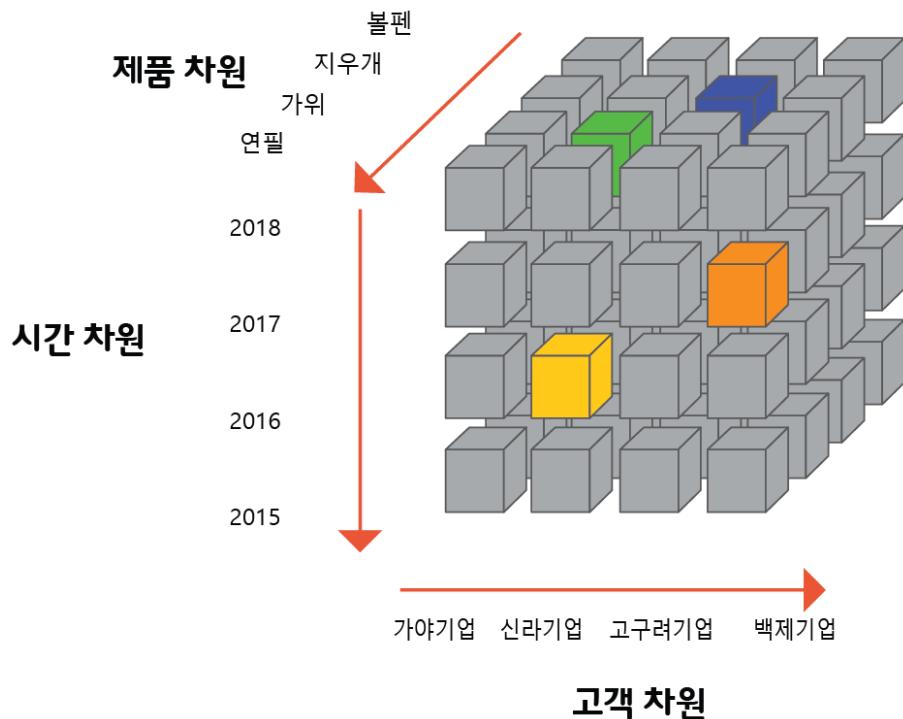
차원1 발주일	차원2				측정값1		측정값2		차원3				차원4			
	판매 제품 정보				판매량		판매액		판매 담당 직원 정보				구입 고객사 정보			
	Order_date	Product_name	Unit_price	Unit_cost	Supplier	Sales_quantity	Sales_amount	Employee_name	Employee_title	Annual_salary	Hire_date	Customer_name	Customer_location	Customer_size	Customer_revenue	
2015-01-02	가위	1,000	700	부여산업	8	8,000	김유신	차장	6,500	1999-08-01	가야기업	김해	5~49명	50~99억		
2015-01-31	가위	1,000	700	부여산업	20	20,000	홍길동	대리	3,500	2009-09-01	고구려기업	평양	50~99명	100~199억		
2015-02-03	가위	1,000	700	부여산업	16	16,000	이성계	사원	2,500	2016-07-08	고구려기업	평양	50~99명	100~199억		
2015-02-22	볼펜	500	300	옥저공업	19	9,500	홍길동	대리	3,500	2009-09-01	가야기업	김해	5~49명	50~99억		
2015-02-28	지우개	100	50	동예공업	3	300	홍길동	대리	3,500	2009-09-01	고구려기업	평양	50~99명	100~199억		
2015-02-28	지우개	100	50	동예공업	8	800	장영실	과장	4,000	2006-10-10	신라기업	경주	100~499명	500~999억		
2015-03-02	가위	1,000	700	부여산업	4	4,000	김유신	차장	6,500	1999-08-01	신라기업	경주	100~499명	500~999억		
2015-03-03	지우개	100	50	동예공업	20	2,000	이성계	사원	2,500	2016-07-08	고구려기업	평양	50~99명	100~199억		
2015-03-05	지우개	100	50	동예공업	8	800	홍길동	대리	3,500	2009-09-01	고구려기업	평양	50~99명	100~199억		
2015-04-04	지우개	100	50	동예공업	16	1,600	이성계	사원	2,500	2016-07-08	신라기업	경주	100~499명	500~999억		
2015-04-22	지우개	100	50	동예공업	10	1,000	김유신	차장	6,500	1999-08-01	가야기업	김해	5~49명	50~99억		
2015-05-05	볼펜	500	300	옥저공업	9	4,500	장영실	과장	4,000	2006-10-10	백제기업	공주	100~499명	500~999억		
2015-06-09	가위	1,000	700	부여산업	3	3,000	이성계	사원	2,500	2016-07-08	고구려기업	평양	50~99명	100~199억		
2015-08-11	지우개	100	50	동예공업	5	500	장영실	과장	4,000	2006-10-10	신라기업	경주	100~499명	500~999억		
2015-09-21	지우개	100	50	동예공업	9	900	홍길동	대리	3,500	2009-09-01	백제기업	공주	100~499명	500~999억		
2015-10-30	볼펜	500	300	옥저공업	3	1,500	김유신	차장	6,500	1999-08-01	가야기업	김해	5~49명	50~99억		
2015-11-11	볼펜	500	300	옥저공업	15	7,500	이성계	사원	2,500	2016-07-08	가야기업	김해	5~49명	50~99억		
2015-12-23	연필	200	100	동예공업	14	2,800	홍길동	대리	3,500	2009-09-01	고구려기업	평양	50~99명	100~199억		
2015-12-30	지우개	100	50	동예공업	14	1,400	김유신	차장	6,500	1999-08-01	백제기업	공주	100~499명	500~999억		
2016-01-11	가위	1,000	700	부여산업	7	7,000	이성계	사원	2,500	2016-07-08	신라기업	경주	100~499명	500~999억		
2016-02-22	지우개	100	50	동예공업	17	1,700	홍길동	대리	3,500	2009-09-01	신라기업	경주	100~499명	500~999억		
2016-03-03	지우개	100	50	동예공업	16	1,600	이성계	사원	2,500	2016-07-08	신라기업	경주	100~499명	500~999억		
2016-03-22	볼펜	500	300	옥저공업	2	1,000	이성계	사원	2,500	2016-07-08	신라기업	경주	100~499명	500~999억		
2016-03-28	지우개	100	50	동예공업	7	700	김유신	차장	6,500	1999-08-01	고구려기업	평양	50~99명	100~199억		
2016-04-01	가위	1,000	700	부여산업	8	8,000	김유신	차장	6,500	1999-08-01	가야기업	김해	5~49명	50~99억		
2016-04-04	연필	200	100	동예공업	17	3,400	김유신	차장	6,500	1999-08-01	고구려기업	평양	50~99명	100~199억		
2016-05-05	연필	200	100	동예공업	16	3,200	장영실	과장	4,000	2006-10-10	신라기업	경주	100~499명	500~999억		
2016-07-08	가위	1,000	700	부여산업	15	15,000	김유신	차장	6,500	1999-08-01	고구려기업	평양	50~99명	100~199억		
2016-08-08	가위	1,000	700	부여산업	15	15,000	이성계	사원	2,500	2016-07-08	가야기업	김해	5~49명	50~99억		
2016-09-22	가위	1,000	700	부여산업	1	1,000	김유신	차장	6,500	1999-08-01	가야기업	김해	5~49명	50~99억		
2016-10-05	가위	1,000	700	부여산업	13	13,000	장영실	과장	4,000	2006-10-10	백제기업	공주	100~499명	500~999억		
2016-11-19	지우개	100	50	동예공업	19	1,900	홍길동	대리	3,500	2009-09-01	가야기업	김해	5~49명	50~99억		
2016-12-11	가위	1,000	700	부여산업	18	18,000	홍길동	대리	3,500	2009-09-01	신라기업	경주	100~499명	500~999억		
2017-02-12	지우개	100	50	동예공업	4	400	홍길동	대리	3,500	2009-09-01	가야기업	김해	5~49명	50~99억		
2017-03-12	볼펜	500	300	옥저공업	9	4,500	장영실	과장	4,000	2006-10-10	가야기업	김해	5~49명	50~99억		
2017-03-23	연필	200	100	동예공업	9	1,800	장영실	과장	4,000	2006-10-10	백제기업	공주	100~499명	500~999억		
2017-04-04	지우개	100	50	동예공업	8	800	홍길동	대리	3,500	2009-09-01	가야기업	김해	5~49명	50~99억		
2017-04-28	가위	1,000	700	부여산업	2	2,000	홍길동	대리	3,500	2009-09-01	백제기업	공주	100~499명	500~999억		
2017-05-01	연필	200	100	동예공업	18	3,600	홍길동	대리	3,500	2009-09-01	고구려기업	평양	50~99명	100~199억		
2017-05-04	볼펜	500	300	옥저공업	5	2,500	김유신	차장	6,500	1999-08-01	가야기업	김해	5~49명	50~99억		
2017-06-06	가위	1,000	700	부여산업	8	8,000	홍길동	대리	3,500	2009-09-01	고구려기업	평양	50~99명	100~199억		
2017-07-09	볼펜	500	300	옥저공업	12	6,000	이성계	사원	2,500	2016-07-08	가야기업	김해	5~49명	50~99억		
2017-08-10	연필	200	100	동예공업	4	800	이성계	사원	2,500	2016-07-08	가야기업	김해	5~49명	50~99억		
2017-09-11	연필	200	100	동예공업	4	800	홍길동	대리	3,500	2009-09-01	가야기업	김해	5~49명	50~99억		
2017-10-12	가위	1,000	700	부여산업	3	3,000	홍길동	대리	3,500	2009-09-01	신라기업	경주	100~499명	500~999억		
2017-10-19	연필	200	100	동예공업	6	1,200	장영실	과장	4,000	2006-10-10	고구려기업	평양	50~99명	100~199억		
2017-12-22	지우개	100	50	동예공업	14	1,400	김유신	차장	6,500	1999-08-01	백제기업	공주	100~499명	500~999억		
2018-01-01	가위	1,000	700	부여산업	11	11,000	홍길동	대리	3,500	2009-09-01	가야기업	김해	5~49명	50~99억		
2018-03-07	가위	1,000	700	부여산업	2	2,000	이성계	사원	2,500	2016-07-08	가야기업	김해	5~49명	50~99억		
2018-03-11	가위	1,000	700	부여산업	3	3,000	홍길동	대리	3,500	2009-09-01	가야기업	김해	5~49명	50~99억		
2018-04-14	연필	200	100	동예공업	19	3,800	장영실	과장	4,000	2006-10-10	고구려기업	평양	50~99명	100~199억		
2018-04-18	가위	1,000	700	부여산업	5	5,000	이성계	사원	2,500	2016-07-08	가야기업	김해	5~49명	50~99억		
2018-04-22	볼펜	500	300	옥저공업	2	1,000	장영실	과장	4,000	2006-10-10	고구려기업	평양	50~99명	100~199억		
2018-05-22	가위	1,000	700	부여산업	20	20,000	홍길동	대리	3,500	2009-09-01	신라기업	경주	100~499명	500~999억		
2018-05-28	가위	1,000	700	부여산업	6	6,000	장영실	과장	4,000	2006-10-10	가야기업	김해	5~49명	50~99억		
2018-06-09	연필	200	100	동예공업	16	3,200	장영실	과장	4,000	2006-10-10	백제기업	공주	100~499명	500~999억		
2018-07-15	볼펜	500	300	옥저공업	2	1,000	이성계	사원	2,500	2016-07-08	백제기업	공주	100~499명	500~999억		
2018-10-01	연필	200	100	동예공업	4	800	이성계	사원	2,500	2016-07-08	고구려기업	평양	50~99명	100~199억		
2018-10-16	연필	200	100	동예공업	2	400	홍길동	대리	3,500	2009-09-01	백제기업	공주	100~499명	500~999억		
2018-10-19	가위	1,000	700	부여산업	12	12,000	김유신	차장	6,500	1999-08-01	신라기업	경주	100~499명	500~999억		
2018-11-02	지우개	100	50	동예공업	20	2,000	장영실	과장	4,000	2006-10-10	가야기업	김해	5~49명	50~99억		

Druid 엔진 기반의 다차원 테이블에는 스타 스키마의 사실 테이블과 같이 매 이벤트가 발생할 때마다 새로운 레코드가 추가되며, 이때의 시간이 타임스탬프로 기록됩니다. 하지만 위 예시에서 볼 수 있듯이, 별도의 차원 테이블을 참조하는 ID들을 사용하지 않고 모든 데이터 항목이 1개의 테이블에 한꺼번에 저장됩니다. 타임스탬프도 하나의 차원(시간 차원)으로 간주되기 때문에 위 데이터 마트는 총 4개의 차원을 가집니다.

이러한 저장 방식을 채택할 경우 중복되는 데이터 값이 많아지기 때문에(예를 들면 똑같은 제품명이 여러 레코드에 반복적으로 언급됨) 데이터 용량은 많이 늘어나게 되지만, 다음과 같은 이유로 분석의 측면에서는 더욱 빠르고 효율적인 작업이 가능합니다.

- 1개 또는 소수의 테이블에 모든 데이터가 저장되어 있고 여기서 분석에 필요한 컬럼만 골라서 추출하면 되기 때문에, 관계형 구조에 비해 데이터 로드로 인한 부하가 훨씬 적습니다.
- 테이블 간 참조 관계를 이해하고 그에 맞게 쿼리를 작성·실행하는 데 드는 비용이 절감됩니다.
- 특히 Druid 엔진의 경우 Hadoop 클러스터를 기반으로 하여 데이터 마트를 여러 노드에서 분산 저장·처리하기에 용이하므로, Terabyte 단위의 데이터도 매우 신속하고 효율적으로 처리·분석할 수 있습니다.

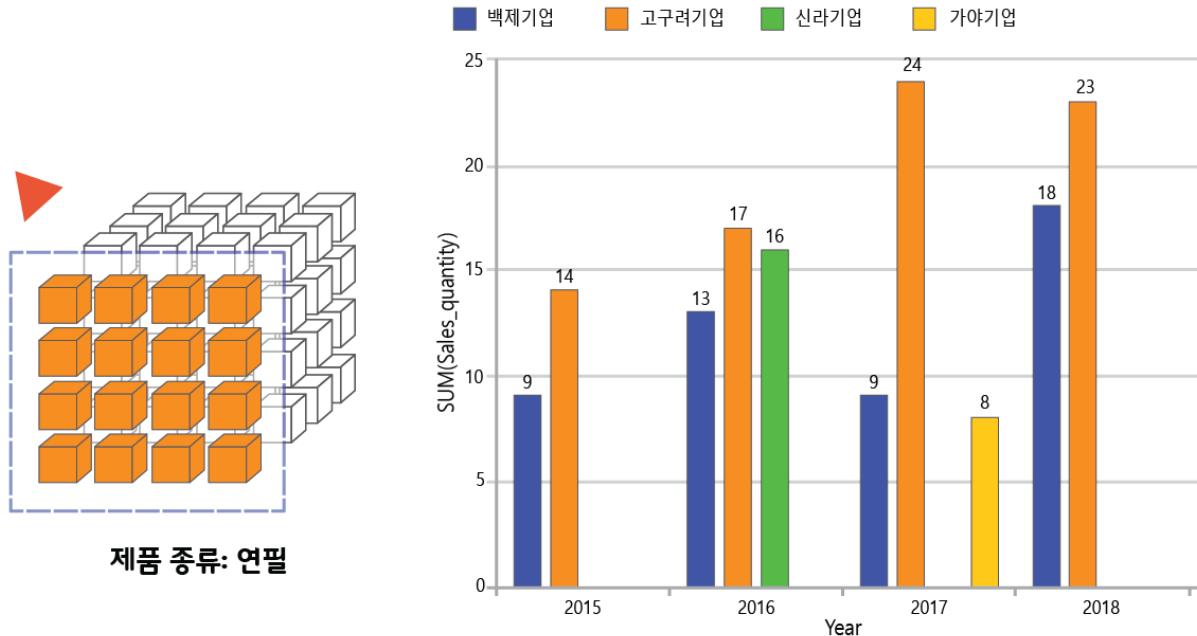
이제 다차원 OLAP 시스템에서 데이터 분석이 어떻게 이루어지는지 확인해보겠습니다. 다차원 데이터 마트 분석을 이해하는 데는 아래와 같은 3차원 큐브 모델이 사용되곤 합니다.



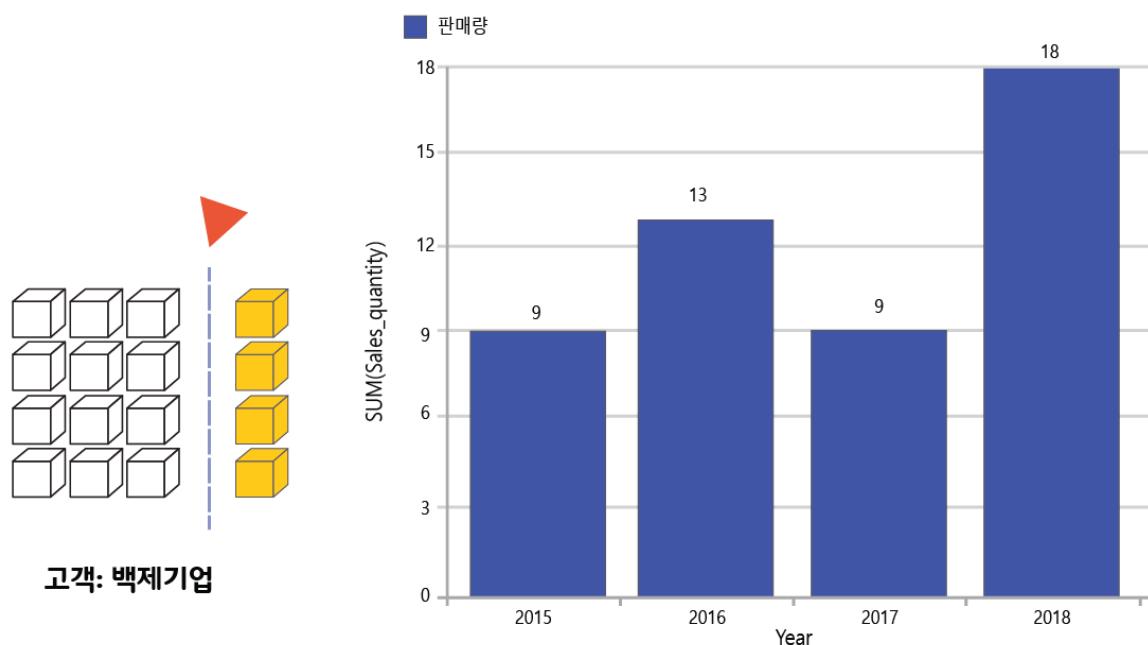
큐브의 각 축은 OLAP형 데이터 마트를 구성하는 각각의 차원을 가리키는 것입니다. 다만, 실제 데이터 마트에서 차원이 3개로 국한되는 경우는 드물기 때문에, 이러한 큐브 모델은 어디까지나 3차원으로 단순화시킨 데이터 마트를 분석하는 과정을 도식화하기 위함일 뿐 4차원 이상의 현실 데이터 마트를 표현할 수 있는 것은 아닙니다.

본 절에서 예시로 사용하는 데이터 마트 역시 총 4개 차원으로 구성되지만, 이러한 3차원 큐브 모델을 통해 분석 원리를 설명하기 위해 편의상 '직원' 차원을 무시하고 나머지 3개 차원만을 사용하겠습니다. 다음은 특정 제품이 특정 고객에게 특정 시기에 몇 개나 팔렸는지 확인하는 분석 절차입니다.

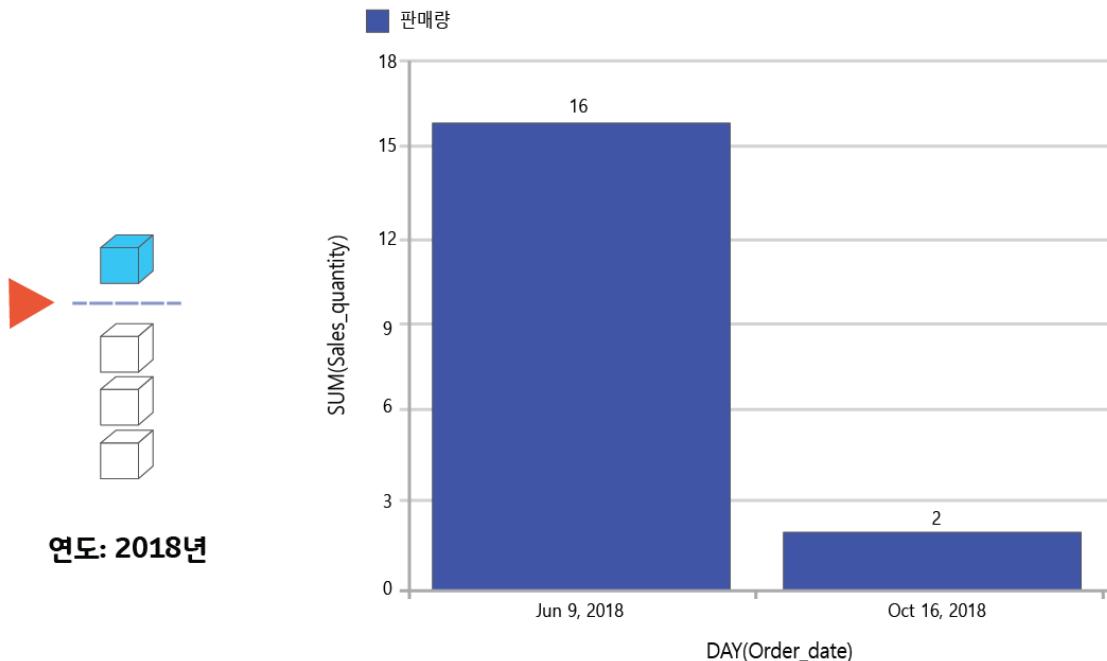
1. 우선 데이터 마트에서 제품 종류가 '연필'인 데이터만 골라냅니다.



2. 그런 다음 '백제기업'이라는 고객사가 구입한 내역만을 골라냅니다.



3. 마지막으로 2018년 내역만을 추려냅니다. 아래 예시를 보면 백제기업은 연필을 2018년 6월 9일에 16개, 2018년 10월 16일에 2개 구입했음을 알 수 있습니다.



metatron Discovery에서는 이러한 다차원 데이터 분석을 위한 수많은 기능을 직관적인 GUI 기반으로 지원합니다.



metatron 3.0

2부

운영/관리(MANAGEMENT)



Copyright © SK telecom

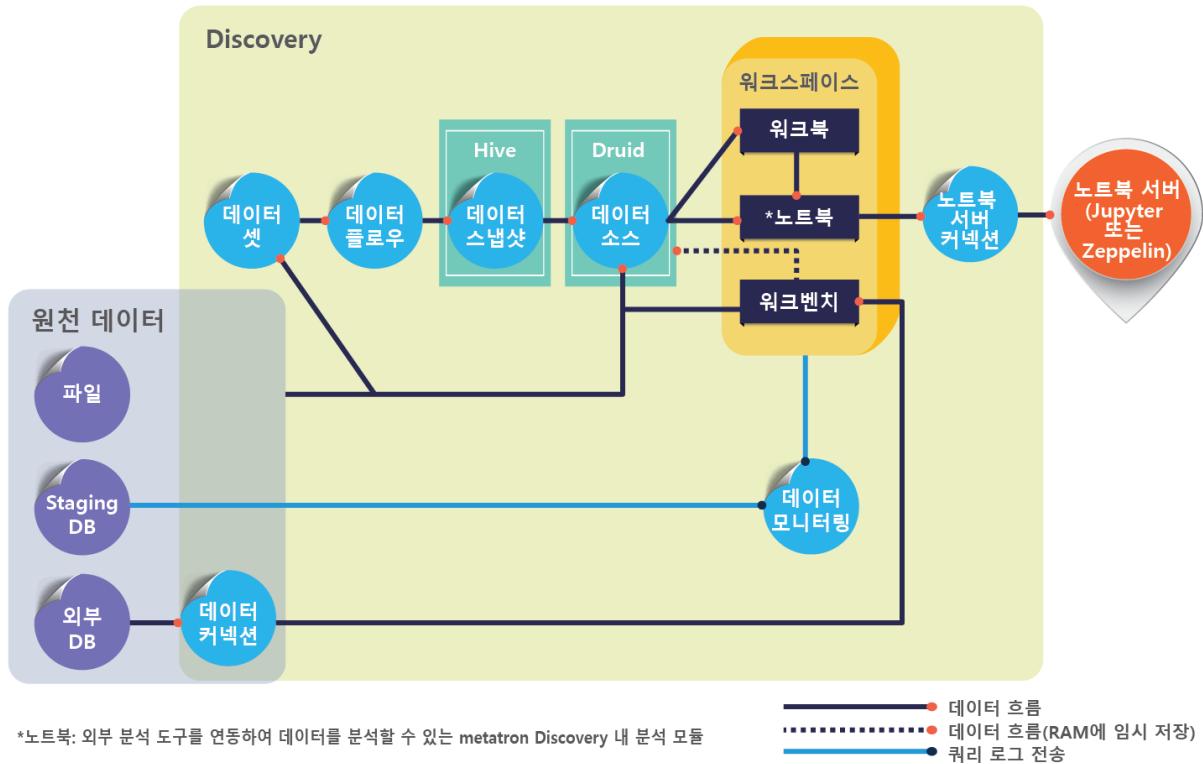
SK Telecom 의 사전 승인 없이
본 내용의 전부 또는 일부에 대한 복사, 배포, 사용을 금합니다.

목 차

1. 운영/관리(MANAGEMENT) 개요.....	4
2. 데이터 스토리지 관리.....	5
2.1 데이터 소스.....	5
2.1.1 데이터 소스 관리 홈 화면.....	5
2.1.2 데이터 소스 상세 정보.....	7
2.1.3 데이터 소스 생성하기.....	16
2.2 데이터 커넥션.....	29
2.2.1 데이터 커넥션 관리 홈 화면	29
2.2.2 데이터 커넥션 새로 만들기	31
3. 데이터 프리퍼레이션 관리	34
3.1 데이터셋.....	34
3.1.1 데이터셋 관리 홈 화면	35
3.1.2 데이터셋 상세 정보	36
3.1.3 데이터셋 새로 만들기.....	37
3.2 데이터플로우	45
3.2.1 데이터플로우 관리 홈 화면	45
3.2.2 데이터플로우 새로 만들기.....	46
3.2.3 데이터플로우 상세 정보	48
3.2.4 Wrangled 데이터셋 편집	50
3.2.5 데이터 스냅샷 생성하기	52
3.3 데이터 스냅샷	54
3.3.1 데이터 스냅샷 관리 홈 화면	54
3.3.2 데이터 스냅샷 상세 정보.....	55
4. 노트북 관리	56
4.1 노트북 서버	56
4.1.1 노트북 서버 목록 조회	56
4.1.2 노트북 서버 새로 등록하기	57

5. 데이터 모니터링 관리	58
5.1 로그 분석	58
5.1.1 로그 분석 홈 화면	58
5.2 잡 로그	60
5.2.1 잡 로그 홈 화면	60
5.2.2 잡 로그 상세 정보	61
5.3 데이터 리니지	62
5.3.1 데이터 리니지 관리 홈 화면	62
5.3.2 데이터 리니지 상세 정보	63

1. 운영/관리(Management) 개요



위 그림과 같이 metatron Discovery의 3가지 분석 모듈(워크북, 노트북, 워크벤치)에서 사용하는 데이터는 다양한 원천 데이터 유형과 여러 엔진 및 저장소를 통해 마련됩니다. 따라서 이러한 데이터 흐름을 정형화 및 관리하고 여러 원천 데이터를 연결시켜주는 작업이 반드시 필요합니다.

운영/관리(Management) 메뉴는 위 그림에서 검정색으로 표시된 공정들을 관리하며, 이러한 공정들은 다음과 같은 4개의 소메뉴로 정리됩니다.

- 2장 데이터 스토리지(데이터 소스, 데이터 커넥션):** 데이터 분석 및 시각화에 필요한 데이터 소스와 데이터베이스 커넥션을 관리하는 절차를 설명합니다.
- 3장 데이터 프리퍼레이션(데이터셋, 데이터플로우, 데이터 스냅샷):** 가져온 원천 데이터를 전처리하는 절차를 설명합니다.
- 4장 노트북 관리:** 노트북 모듈을 사용하기 위해서는 Jupyter 또는 Zeppelin이 설치된 서버와의 연동이 필요합니다. 여기서는 이러한 서버들을 설정하고 관리하는 절차를 설명합니다.
- 5장 데이터 모니터링(로그 분석, 잡 로그, 데이터 리니지):** 데이터 사용 현황을 감독하고 추적하는 절차를 설명합니다.

2. 데이터 스토리지 관리

metatron Discovery에서 사용할 데이터는 다음과 같은 두 가지 방식으로 연동시킬 수 있습니다.

- 원천 데이터를 'Druid'라는 내부 엔진에 탑재하여 '데이터 소스' 단위로 보관(2.1절 참조)
- SQL 데이터베이스와 직접 연동(2.2절 참조)

데이터 스토리지는 이러한 데이터를 metatron에 연동하고 관리하는 기능을 제공합니다.

2.1 데이터 소스

metatron Discovery에서 '데이터 소스'는 Druid 엔진에 수집되는 데이터 단위를 의미합니다. 각 데이터 소스는 Druid 데이터베이스에서 하나의 테이블로서 저장됩니다. 이러한 데이터 소스들은 '워크북'이나 '노트북'에서 분석·시각화하는 데 사용됩니다.

2.1.1 데이터 소스 관리 화면

본 화면에서는 데이터 소스의 신규 등록·편집·조회가 가능합니다.

The screenshot shows the 'Data Storage' section of the Metatron Discovery interface. At the top, there are several search and filter fields: '데이터 타입' (1), '수집 타입' (2), '상태' (3) with radio buttons for '모두', '가능', '불가능', and '준비중', and a checkbox for '허용' (4). Below these are time-related filters: 'Time' (5) with '생성일' and '수정일' radio buttons, and dropdowns for '전체', 'Today', and 'Last 7 days'. There are also date and time pickers for 'yyyy-MM-dd hh:mm' and 'yyyy-MM-dd hh:mm'. To the right of these filters are buttons for '초기화' (clear), '적용' (apply), and a search bar (6) with placeholder text '데이터 소스 이름으로 검색하세요'. On the far right, there is a message '30개 데이터가 있습니다.' (30 items found) and a button for '새로운 데이터 소스 생성' (create new data source). The main area displays a table of data sources (9) with columns: '데이터 소스' (name), '데이터 타입' (type), '수집 타입' (collection type), '상태' (status), and '생성일' (creation date). Each row shows a name, type (e.g., 파일, Staging DB, 데이터베이스), collection type (e.g., 수집형 데이터, 수집형 데이터), status (e.g., Enabled, Preparing), and creation date. The last column contains a delete icon (10) for each row. A blue box highlights the search bar and the message above it.

① 데이터 타입: 데이터 소스를 원천 데이터 타입별로 선별하여 조회합니다.

- 모두: 원천 데이터 타입과 관계 없이 모든 데이터 소스가 출력됩니다.
- 파일: 로컬 PC에서 가져온 파일로 생성한 데이터 소스들이 출력됩니다.
- 데이터베이스: 데이터베이스에서 가져온 데이터로 생성한 데이터 소스들이 출력됩니다.

- Staging DB: metatron의 내부 Hive 데이터베이스에서 가져온 데이터를 기반으로 만든 데이터 소스들이 출력됩니다.

2 수집 타입: 데이터 소스가 데이터를 어떤 방식으로 수집하는지 각각의 타입별로 선별하여 조회합니다.

- 모두: 수집 타입과 관계 없이 데이터 소스가 출력됩니다.
- 수집형 데이터(Ingested): 데이터를 metatron 서버에 직접 저장하는 방식으로 수집된 데이터 소스들이 출력됩니다.
- 연결형 데이터(Linked): 연결된 데이터베이스에서 필요한 시점마다 데이터를 가져오는 방식의 데이터 소스들이 출력됩니다.

3 상태: 현재 데이터 스토리지에 저장되어 있는 데이터 소스의 가용 여부로 선별하여 조회합니다.

- 모두: 데이터 소스의 가용 여부와 상관 없이 모든 데이터 소스가 출력됩니다.
- 가능: 데이터 ingestion을 마쳐 워크북이나 워크벤치에서 사용 가능한 데이터 소스들이 출력됩니다.
- 불가능: 데이터 ingestion을 마쳤으나 Druid 엔진의 일부 프로세스에서 제대로 진행이 되지 않아 사용이 불가능한 데이터 소스들이 출력됩니다.
- 준비중: 아직 생성된 지 얼마 되지 않아서 데이터 ingestion이 진행 중인 데이터 소스들이 출력됩니다.

4 오픈 데이터만 보기: 모든 워크스페이스에서 사용이 허용된 데이터 소스들만 선별하여 조회합니다.

5 Time: 데이터 소스 조회 시 적용되는 시간 기준입니다. 생성일과 수정일 중 원하는 기준으로 선택할 수 있으며 시간 범위는 전체/오늘/지난 1주일/특정 날짜 기간 중 선택이 가능합니다.

6 데이터 소스 이름으로 검색: 현재 등록된 데이터 소스를 이름으로 검색합니다.

7 데이터 개수: 현재 목록에 조회된 데이터 소스의 개수를 나타냅니다.

8 새로운 데이터 소스 생성: 클릭하면 새로운 데이터 소스를 생성할 수 있습니다.

9 데이터 소스 목록: 설정한 선별 조건에 맞는 데이터 소스들을 보여줍니다. 이 중 하나를 클릭하면 상세 내역을 볼 수 있습니다. (2.1.2절 참조)

10 삭제: 해당 데이터 소스에 마우스 오버 시 휴지통 아이콘이 나타납니다. 클릭하면 해당 데이터 소스를 삭제할 수 있습니다.

2.1.2 데이터 소스 상세 정보

데이터 소스 관리 홈에 열거된 데이터 소스 중 하나를 클릭하면 해당 데이터 소스에 관한 다양한 속성을 확인할 수 있습니다. 아래 각 영역을 확인할 때에는 각 데이터 소스가 metatron의 내부 Druid 데이터베이스에 하나의 테이블로 저장되며, Druid의 시계열 특성 때문에 반드시 타임스탬프 컬럼을 포함하게 됨을 유의하십시오.

A. 상단 공통 영역

The screenshot shows the Metatron Discovery interface. At the top, there is a header with the text "METATRON DISCOVERY". Below the header, there are several UI elements: a back arrow, a name input field containing "test_orc_test", a description input field containing "테스트용입니다", a timestamp indicating the last update was on "2018-04-07 22:19:20" by "Administrator", and a user profile icon. A blue box highlights the top navigation bar, which contains five tabs: "정보" (selected), "그리드 데이터", "컬럼 상세", and "모니터링".

- ① **이름:** 해당 데이터 소스의 이름입니다. 클릭 시 수정 가능합니다.
- ② **설명:** 해당 데이터 소스에 대한 설명입니다. 클릭 시 수정 가능합니다.
- ③ **마지막 수정 정보:** 해당 데이터 소스를 누가 언제 마지막으로 수정했는지 보여줍니다.
- ④ **삭제:** 이 아이콘을 누르면 해당 데이터 소스를 삭제할 수 있는 메뉴가 나타납니다.
- ⑤ **탭 선택 영역:** 각 탭은 해당 데이터 소스에 관한 특정한 속성군을 보여줍니다. 데이터 소스의 종류에 따라 4개 탭이 모두 나오지 않을 수도 있습니다. 각각에 대한 자세한 설명은 아래의 각 절에서 확인하십시오.

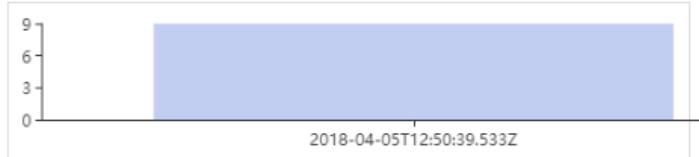
B. 정보 탭

정보 탭에서는 해당 데이터 소스의 상세 정보를 확인하고 몇 가지 기본적인 사항을 수정할 수 있습니다. 이 탭의 화면은 크게 '데이터 정보', '권한', '수집 정보'로 분류됩니다.

데이터 정보 영역

이 영역에서는 해당 데이터 소스의 기본 정보를 보여줍니다.

데이터 정보

1	데이터 타입	HIVE																				
2	상태	Enabled (엔진의 룰에 의해 데이터소스 사용 가능)																				
3	사이즈	2.04 KB																				
4	Duration	2017-04-20 09:00:00 ~ 2017-04-21 09:00:00																				
5	Granularity 설정	Segment Granularity : DAY Query Granularity : DAY																				
6	Histogram	 A histogram showing data distribution over time. The x-axis represents time from 2018-04-05T12:50:39.533Z to 2018-04-05T12:50:39.533Z. The y-axis represents frequency from 0 to 9. A single large bar covers the entire time range with a height of approximately 8.5. 2018-04-05T12:50:39.533Z																				
7	데이터 미리보기	<table border="1"><thead><tr><th>time</th><th>d</th><th>sd</th><th>m1</th><th>m2</th></tr></thead><tbody><tr><td>2017-04-20...</td><td>d1</td><td>sd1</td><td>10.0</td><td>30</td></tr><tr><td>2017-04-20...</td><td>d1</td><td>sd2</td><td>20.0</td><td>30</td></tr><tr><td>2017-04-20...</td><td>d1</td><td>sd3</td><td>30.0</td><td>10</td></tr></tbody></table> 	time	d	sd	m1	m2	2017-04-20...	d1	sd1	10.0	30	2017-04-20...	d1	sd2	20.0	30	2017-04-20...	d1	sd3	30.0	10
time	d	sd	m1	m2																		
2017-04-20...	d1	sd1	10.0	30																		
2017-04-20...	d1	sd2	20.0	30																		
2017-04-20...	d1	sd3	30.0	10																		

- ① **데이터 타입:** 해당 데이터 소스 생성 시 가져온 원천 데이터의 타입입니다.
- ② **상태:** 해당 데이터 소스의 현재 사용 가능 여부를 나타냅니다.
- ③ **사이즈:** 해당 데이터 소스의 크기를 나타냅니다.
- ④ **Duration:** 해당 데이터 소스에 포함된 타임스탬프의 시간 범위를 나타냅니다.
- ⑤ **Granularity 설정:** 해당 데이터 소스 생성 시 설정한 Granularity 주기를 나타냅니다.
 - Segment Granularity: 분산 환경에서 동작하는 Druid의 특성을 활용하기 위해 데이터를 분할하여 저장하게 되는데 이때 저장하는 단위를 결정합니다.
 - Query Granularity: 분석에서 수행하고자 하는 최소 시간 단위를 결정합니다. 이는 최소 단위까지의 결과를 미리 생성하여 이후에 보다 빠른 응답을 얻을 수 있도록 하기 위함입니다.
- ⑥ **Histogram:** 각 시간대별로 저장된 데이터의 용량을 KByte 단위로 보여주는 그래프입니다. 이러한 히스토그램은 레코드별로 반드시 타임스탬프 기록을 남겨야하는 Druid 엔진의 특성에 따른 것입니다.
- ⑦ **데이터 미리보기:** 해당 데이터 소스 생성 시 가져온 원천 데이터의 내용을 미리 볼 수 있는 창입니다. 우측 하단의  버튼을 클릭하면 새 창으로 해당 데이터의 상세 정보가 나타나는 데이터 그리드 화면을 볼 수 있습니다.

권한 영역

이 영역에서는 해당 데이터 소스를 어느 워크스페이스에서 사용할 수 있는지를 확인·설정합니다.

권한	<p>1 <input type="checkbox"/> 모든 워크스페이스에 데이터 소스를 사용하도록 허용</p> <p>2 <input type="checkbox"/> 수집</p> <p>3 <input type="checkbox"/> 0 워크스페이스</p>
----	---

- ① **모든 워크스페이스에 데이터 소스를 사용하도록 허용 확인란:** 이 확인란에 체크하면 모든 워크스페이스에서 해당 데이터 소스를 사용할 수 있습니다.
- ② **수집:** 해당 데이터 소스의 사용을 허용할 특정 워크스페이스를 지정할 때 사용합니다. 해당 데이터 소스를 오픈 데이터로 지정하면 이 버튼이 사라집니다.
- ③ **공유 워크스페이스 수:** 해당 데이터 소스를 사용할 수 있도록 허용된 워크스페이스의 수를 나타냅니다.

수집 정보 영역

이 영역에는 해당 데이터 소스의 원천이 되는 마스터 데이터에 관한 수집 정보가 나타납니다.

수집 정보	
마스터 데이터	타입 HIVE
	데이터베이스 / 테이블 default.sample_ingestion_time_orc
	Partition keys None
	Advanced setting (Opt.) None
Data range	2017-03-13 21:40:00 ~ 2018-03-13 21:40:00

C. 그리드 데이터 탭

그리드 데이터 탭에는 해당 데이터 소스의 데이터 내용이 자세히 표시됩니다.

time	ab d	ab sd	## m1	## m2
2017-04-2	d1	sd1	10.0	30
2017-04-2	d1	sd2	20.0	30
2017-04-2	d1	sd3	30.0	10
2017-04-2	d2	sd1	10.0	30
2017-04-2	d2	sd2	20.0	30
2017-04-2	d2	sd3	30.0	30
2017-04-2	d3	sd1	10.0	30
2017-04-2	d3	sd2	20.0	30
2017-04-2	d3	sd3	30.0	30

- ① **데이터 검색:** 데이터 테이블에 있는 내용으로 검색이 가능합니다.
- ② **역할:** 데이터 테이블의 컬럼을 전체/차원값/측정값으로 선별하여 조회합니다.
- ③ **타입:** 데이터 테이블의 컬럼을 필드 타입별로 선별하여 조회합니다.
- ④ **행:** 데이터 테이블에 등록된 레코드의 수를 나타냅니다.
- ⑤ **Download CSV:** 화면에 출력된 데이터 내용을 CSV 파일로 다운로드 할 수 있습니다.

D. 컬럼 상세 탭

컬럼 상세 탭에서는 해당 데이터 소스 테이블을 구성하는 각 컬럼에 대한 상세한 정보를 보여줍니다.

컬럼 조회/설정 기능

컬럼 상세 탭의 상단부는 원하는 조건으로 컬럼을 선별하는 사용자 인터페이스를 제공합니다. 조건에 맞게 선별된 컬럼들은 좌측 영역에 조회됩니다. 또한 여기서는 컬럼 설정값을 수정할 수도 있습니다.

The screenshot shows the 'Column Details' tab interface. At the top, there is a search bar (1) and filter buttons for '역할' (Role), '전체' (All), '차원값' (Dimension Value), and '측정값' (Measure Value). Below these are buttons for '타입' (Type) (3), '전체' (All) (4), and 'Configure schema' (5). On the left, a list of columns (7) is shown under the heading 'aaa'. The columns are categorized by type: Dimension Value (차원값) and Measure Value (측정값). The right side displays the '스키마 정보' (Schema Information) and '디스플레이 세팅' (Display Settings) sections. The '스키마 정보' section includes details like Logical type (TIMESTAMP), Format (yyyy-MM-dd'T'HH:mm:ss,SSSZ), and Description. The '디스플레이 세팅' section shows settings for Missing values. The '요약' (Summary) section provides statistics such as count (288), valid percentage (288 (100%)), unique values (0 (0%)), outliers (0 (0%)), and missing values (0 (0%)).

- ① **데이터 검색:** 컬럼 이름으로 검색이 가능합니다.
- ② **역할:** 데이터 테이블의 컬럼을 전체/차원값/측정값으로 선별하여 조회합니다.
- ③ **타입:** 데이터 테이블의 컬럼을 필드 타입별로 선별하여 조회합니다.
- ④ **전체 보기:** 데이터 검색, 역할, 타입 옵션에서 설정한 모든 선별 조건을 취소하고 전체 컬럼 보기로 회귀합니다.
- ⑤ **Configure schema:** 클릭하면 현재의 컬럼 설정값을 바꿀 수 있는 창이 출력됩니다.
- ⑥ **데이터 소스 이름:** 해당 데이터 소스, 즉 해당 '테이블'의 이름을 나타냅니다.
- ⑦ **컬럼 목록:** 해당 테이블을 구성하는 컬럼들을 열거합니다.

컬럼 기본 정보 영역

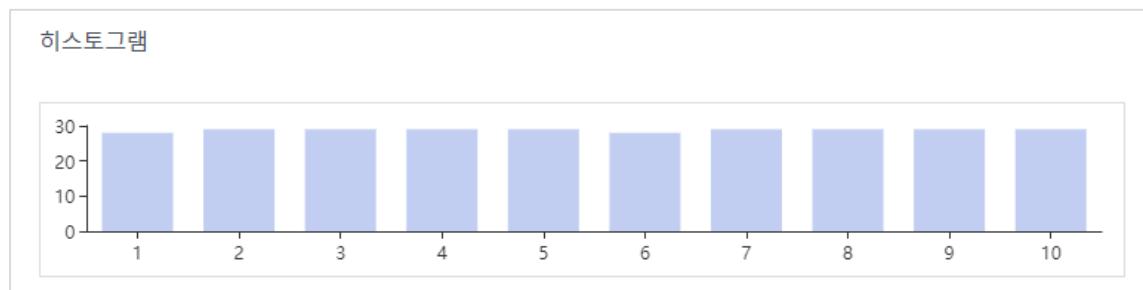
이 영역에서는 선택된 컬럼에 대한 여러 가지 기본적인 정보를 보여줍니다.

1	스키마 정보	2	디스플레이 세팅
	룰리명 AI_401_892		설정안함
	역할 MEASURE		
	논리명 AI_401_892		
	Logical type DOUBLE		
	설명		
3	요약	4	통계
	컬럼 이름 AI_401_892		최솟값 158.46163518300006
	건수 288		Lower Quartile 168.90262331099993
	Valid 288 (100%)		중간값 166.21263834199996
	Unique 288 (100%)		Upper Quartile 163.429285008
	Outliers ⓘ 2 (0.69%)		최댓값 848.0334560950001
	Missing 0 (0%)		평균 169.91626999777088
			표준편차 47.2046662967384
			Skewness 166.71408520072424

- 1** **스키마 정보:** 선택된 컬럼의 속성을 보여줍니다.
- 2** **디스플레이 세팅:** 선택된 컬럼의 메타 데이터 정보를 보여줍니다.
- 3** **요약:** 선택된 컬럼의 값 구성 정보를 보여줍니다.
- 4** **통계:** 선택된 컬럼에 입력된 값 중에서 최솟값과 최댓값을 보여줍니다.

히스토그램 영역

각 시간대별로 저장된 데이터의 용량을 KByte 단위로 보여주는 그래프입니다. 이러한 히스토그램은 레코드별로 반드시 타임스탬프 기록을 남겨야하는 Druid 엔진의 특성에 따른 것입니다.

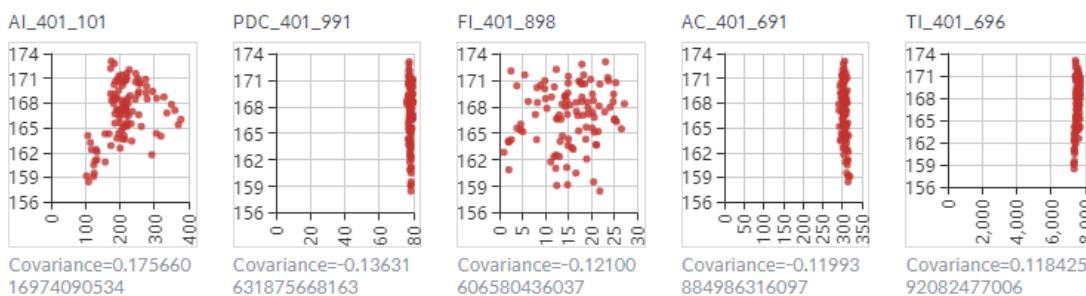


Covariance 순위 영역

구성 값들의 공분산 수치가 가장 높은 컬럼 5개를 각각의 공분산 값과 분포도와 함께 보여줍니다.

Covariance 순위 (top 5)

1	AI_401_101	0.17566016974090534
2	PDC_401_991	-0.13631631875668163
3	FI_401_898	-0.12100606580436037
4	AC_401_691	-0.11993884986316097
5	TI_401_696	0.11842592082477006



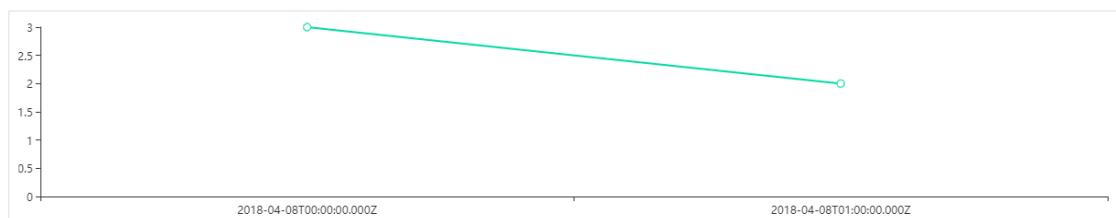
E. 모니터링 탭

모니터링 탭에서는 데이터 소스가 사용된 로그를 볼 수 있습니다.

transaction 변경 영역

해당 데이터 소스의 시간에 따른 트랜잭션량 추이를 보여줍니다.

transaction 변경



데이터 사이즈 변경 영역

해당 데이터 소스의 시간에 따른 용량 추이를 보여줍니다.

데이터 사이즈 변경



Query distribution 영역

해당 데이터 소스에 대해 수행된 쿼리에 관한 정보를 보여줍니다.

Query distribution by user (during last one week)

1

Query distribution by elapsed time (during last one week)

2



1 **Query distribution by user (during last one week):** 지난 한 주간 쿼리를 수행한 사용자별로 분류하여 그래프로 보여줍니다.

2 **Query distribution by elapsed time (during last one week):** 지난 한 주간 쿼리를 수행한 소요시간별로 분류하여 그래프로 보여줍니다.

쿼리 로그 영역

수행된 각 쿼리에 대한 상세 이력을 확인할 수 있습니다.

The screenshot shows the 'Query Log History' section of a management interface. At the top, there are search filters: 'Query date' (with '전체' selected), 'Last 7 days', date range input ('yyyy-MM-dd hh:mm ~ yyyy-MM-dd hh:mm'), an 'Apply' button, 'Query type' dropdown ('All'), and 'Result' dropdown ('전체'). Below these are two tables. The first table, labeled '4' and '쿼리 목록', lists five queries with columns: No., 쿼리 날짜, 쿼리 타입, 사용자, 작업시간, and Result. The second table, labeled '5' and '자세히', shows detailed logs for each query. The first row of the detailed log is visible, showing '작업자' (작업자), '작업내용' (작업내용), and '작업상태' (작업상태).

No.	쿼리 날짜	쿼리 타입	사용자	작업시간	Result
1	2018-04-08 10:03:06	SUMMARY	admin	19ms	성공
2	2018-04-08 10:03:01	SUMMARY	admin	19ms	성공
3	2018-04-08 09:59:48	SEARCH	admin	19ms	성공
4	2018-04-08 09:59:46	SUMMARY	admin	20ms	성공
5	2018-04-08 09:52:00	SUMMARY	admin	21ms	성공

작업자	작업내용	작업상태
작업자	작업내용	작업상태

- ① **Query date:** 확인하고자 하는 쿼리들의 실행 시간대를 설정합니다.
- ② **쿼리 타입:** 실행한 쿼리를 타입별로 선별하여 출력합니다.
- ③ **Result:** 쿼리 결과를 성공/실패로 선별하여 출력합니다.
- ④ **쿼리 목록:** 설정한 조건에 부합하는 쿼리들이 나열됩니다.
- ⑤ **자세히:** 클릭하면 해당 쿼리문을 확인할 수 있습니다.

2.1.3 데이터 소스 생성하기

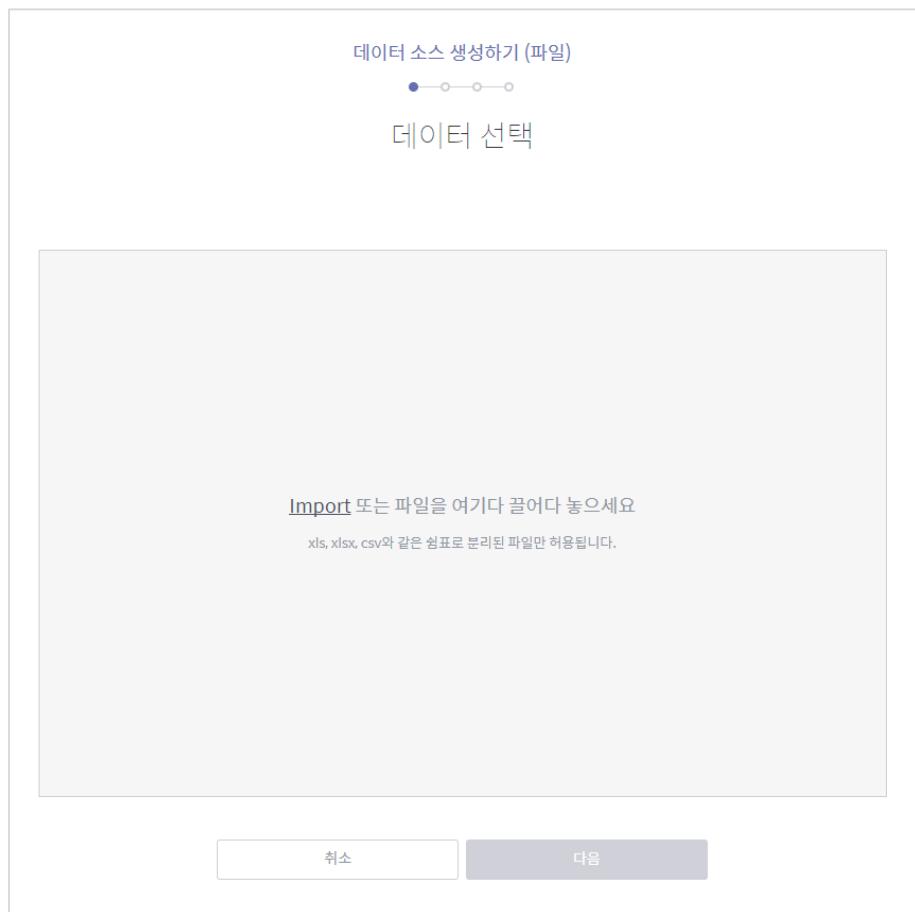
데이터 소스 관리 화면에서 목록 우측 상단의 '+ 새로운 데이터 소스 생성' 버튼을 클릭하면 데이터 소스를 새로 생성할 수 있습니다. 먼저 원천 데이터의 타입을 선택합니다.



- ① **파일**: 사용자의 로컬 PC에 저장되어 있는 파일을 가져와서 데이터 소스를 생성합니다. xls, xlsx, csv와 같이 쉼표로 분리된 파일만 허용됩니다(자세한 절차는 A절 참조).
- ② **데이터베이스**: 외부 데이터베이스에서 데이터를 가져와서 데이터 소스를 생성합니다. 현재 metatron Discovery에서 지원하는 데이터베이스로는 Oracle, MySQL, Hive, presto, TIBERO가 있습니다(자세한 절차는 B절 참조).
- ③ **Staging DB**: metatron의 내부 Hive 데이터베이스에서 가져온 데이터를 기반으로 데이터 소스를 생성합니다.
- ④ **실시간**: 현재 해당 기능은 지원하지 않습니다.
- ⑤ **데이터스냅샷**: 현재 해당 기능은 지원하지 않습니다.
- ⑥ **메타트론 엔진**: metatron 1.0 버전에 저장된 데이터 소스를 마이그레이션 합니다.

A. 파일로 데이터 소스 생성하기

- 원천 데이터 탑입 선택 화면에서 '파일'을 선택합니다.
- 사용자 로컬 PC에서 데이터 소스로 사용할 파일을 가져옵니다. Import 버튼을 클릭하여 파일을 선택할 수도 있고 화면 상으로 파일을 끌어다 놓을 수도 있습니다. 파일을 가져왔으면 다음 버튼을 누릅니다.



3. 가져온 파일에서 데이터 소스에 포함시킬 시트를 선택합니다.

데이터 소스 생성하기 (파일)

데이터 선택

ab OrderDate	ab Category	ab City	ab Country	ab CustomerNa...	ab Ord...
2011-01-04T00:...	Office Supplies	Houston	United States	Darren Powers	CA-2
2011-01-05T00:...	Office Supplies	Naperville	United States	Phillina Ober	CA-2
2011-01-05T00:...	Office Supplies	Naperville	United States	Phillina Ober	CA-2
2011-01-05T00:...	Office Supplies	Naperville	United States	Phillina Ober	CA-2
2011-01-06T00:...	Office Supplies	Philadelph...	United States	Mick Brown	CA-2
2011-01-07T00:...	Office Supplies	Athens	United States	Jack OBriant	CA-2
2011-01-07T00:...	Office Supplies	Los Angeles	United States	Lycoris Saunders	CA-2

Import 또는 파일을 여기다 끌어다 놓으세요

1 sale_20180130.xls
2 sale.txt (1) > 3 sale.txt (1)
4 4034048 byte 5 28 컬럼
6 100 7 / 9994 행 1~100
8 첫번째 행을 머리글 열로 사용합니다. (선택하지 않은 경우 새 행이 생성되고 헤드 열로 사용됨)

취소 다음

- ① **파일 이름:** 가져온 파일의 이름입니다. 다른 파일을 다시 가져올 수도 있습니다.
- ② **파일 시트 목록:** 가져온 파일에 포함된 시트들을 보여줍니다. 여기서 데이터 소스로 만들 시트를 선택합니다.
- ③ **파일 시트 이름:** 현재 선택된 시트 이름입니다.
- ④ **용량:** 가져온 파일의 용량입니다.
- ⑤ **컬럼:** 가져온 파일의 컬럼 개수입니다.
- ⑥ **행:** 가져온 파일의 행 개수입니다. 숫자를 입력하면 해당 숫자만큼의 행이 화면에 나타납니다.
- ⑦ **타입:** 각 컬럼으로부터 인식한 데이터 타입이 몇 종류인지 보여줍니다. 컬럼별 데이터 타입은 이후 화면에서 수정할 수 있습니다.
- ⑧ **첫째 행을 컬럼명으로 사용 여부 확인란:** 선택하면 파일 내의 첫번째 행의 내용이 컬럼명으로 사용됩니다. 선택하지 않을 경우 컬럼명을 기재할 행이 새로 생성됩니다.

4. 데이터 소스에서 구현하고자 하는 스키마를 설정합니다.

조건에 맞는 컬럼 조회하기

The screenshot shows the '데이터 소스 생성하기 (파일)' (Data Source Creation (File)) process at the '스키마 설정' (Schema Setting) step. The interface is divided into several sections:

- Left Panel (1):** A search bar labeled '컬럼명으로 검색하세요' (Search by column name) and a list of columns categorized by type (e.g., Sales, Segment, ShipDate, Region, ShipCategory, ShipMode, State, Sub-Category, DaystoShipActual, SalesForecast, ShipStatus, DaystoShipScheduled, OrderProfitable, SalesperCustomer).
- Top Right (2):** Filter options: 역할 (Role) with '전체' (All) selected, 차원값 (Dimension Value), 측정값 (Measure Value), and 추천 필터 (Recommended Filter).
- Top Right (3):** Filter tabs: 타입 (Type) and 전체 (All).
- Right Panel (4):** A detailed schema configuration for 'ShipCategory'. It shows a list of categories (Paper, Binders, Labels, Storage, Art, Paper, Chairs, Art, Binders) with their descriptions and a '설정' (Setting) section for each.
- Bottom Left (5):** A note: '시간 태입 또는 현재 시간 중 하나는 시간값으로 지정되어야 합니다.' (One of the time types or current time must be designated as a time value). Below it are buttons for '이전' (Previous) and '다음' (Next).
- Bottom Center (6):** A note: '타임타입 컬럼이 없습니다.' (There is no time type column).
- Bottom Right (7):** A note: '시간 태입 또는 현재 시간 중 하나는 시간값으로 지정되어야 합니다.' (One of the time types or current time must be designated as a time value). Below it are buttons for '이전' (Previous) and '다음' (Next).

- ① **컬럼명으로 검색:** 가져온 파일에 들어있는 컬럼을 이름으로 검색합니다.
- ② **역할:** 가져온 파일에 들어있는 컬럼을 전체/차원값/측정값으로 선별하여 조회합니다.
- ③ **추천 필터:** 최우선 추천 필터가 적용된 컬럼들만 선별하여 조회합니다.
- ④ **타입:** 가져온 파일에 들어있는 컬럼을 필드 타입별로 선별하여 조회합니다.
- ⑤ **컬럼 목록 영역:** 설정한 선별 조건에 맞는 컬럼들을 보여줍니다(자세한 설명은 아래 '일괄 컬럼 수정 및 삭제' 항목 참조).
- ⑥ **개별 컬럼 설정 영역:** 컬럼 목록에서 선택한 컬럼의 속성들을 설정할 수 있는 영역입니다(자세한 설명은 아래 '개별 컬럼 속성 수정' 항목 참조).
- ⑦ **타임스탬프 설정:** 데이터 소스 내에 무조건 존재해야하는 타임스탬프 설정입니다. metatron 엔진은 데이터 소스 저장 시 무조건 시간 값을 보유해야 하는 시계열 엔진입니다. 따라서, 기존 데이터가 보유하고 있는 시간 태입 컬럼을 타임스탬프로 지정할 것인지, 또는 현재 시간 값으로 이루어진 시간 태입 컬럼을 생성하여 타임스탬프로 지정할 것인지를 선택할 수 있습니다.

일괄 컬럼 수정 및 삭제

데이터 컬럼 목록 영역에서는 컬럼들의 일괄 수정 및 삭제가 가능합니다. 타입 변경을 하거나 삭제하고자 하는 컬럼들의 확인란을 체크한 후 아래 설명을 참조하여 원하는 동작을 설정한 다음 '적용' 버튼을 누르십시오.

The screenshot shows a 'Batch Column Modification' dialog box. At the top, there's a search bar labeled '컬럼명으로 검색하세요' and a '역할' button. Below is a list of columns with checkboxes:

컬럼	역할
차원값	ab ShipCategory
차원값	ab ShipMode
차원값	ab State
차원값	ab Sub-Category
차원값	ab DaystoShipActual
차원값	ab SalesForecast
차원값	ab ShipStatus
차원값	ab DaystoShipScheduled >
차원값	ab OrderProfitable >
차원값	ab SalesperCustomer

Below the list are three numbered buttons:

- 1 타입 변경/삭제: 현재 선택한 컬럼들에 대해 원하는 동작을 선택합니다.
- 2 차원값/측정값: 현재 선택한 컬럼의 역할을 차원값 혹은 측정값으로 변경할 수 있습니다.
- 3 컬럼 타입: 선택한 컬럼들을 어떤 타입으로 변경할지 선택합니다.

The '타입 변경' dropdown is set to '차원값'. The '컬럼 타입' dropdown is set to 'ab 문자'.

- 1 **타입 변경/삭제:** 현재 선택한 컬럼들에 대해 원하는 동작을 선택합니다.
- 2 **차원값/측정값:** 현재 선택한 컬럼의 역할을 차원값 혹은 측정값으로 변경할 수 있습니다.
 - 차원값으로 변경할 경우, 데이터 타입을 문자/불린/정수/소수/날짜/시간/위도/경도 중 하나로 변경이 가능합니다.
 - 측정값으로 변경할 경우, 데이터 타입을 정수/소수 중 하나로 변경이 가능합니다.
- 3 **컬럼 타입:** 선택한 컬럼들을 어떤 타입으로 변경할지 선택합니다.

개별 컬럼 속성 설정

컬럼 목록에서 선택한 컬럼의 속성을 수정하는 영역입니다.

1 Sales

2 데이터 10 행

Consumer
Home Office
Home Office
Home Office
Consumer
Corporate
Consumer
Home Office
Home Office
Home Office

3 역할
 차원값
 측정값

4 타입
ab 문자

5 추천 필터 최우선 추천 필터로 적용 [순서 설정](#)
 단일 항목만 선택할 수 있습니다

6 Missing
 대체
 버림
 설정안함

- 1 **이름:** 선택한 컬럼의 이름입니다.
- 2 **데이터 행 수:** 현재 화면에 표시된 데이터 행 수를 가리킵니다.
- 3 **역할:** 해당 컬럼에 부여된 역할인 차원값/측정값이 나타납니다. 여기서 변경 가능합니다.
- 4 **타입:** 해당 컬럼의 데이터 타입을 나타냅니다. 다른 데이터 타입으로 변경 가능하며, 선택한 역할(차원값/측정값)에 따라 선택 가능한 데이터 타입이 달라집니다.
- 5 **추천 필터:** 대용량 데이터를 포함한 컬럼이 있을 경우, 데이터 소스 로딩 시 긴 시간이 소요되어 탐색이 어려울 수 있습니다. 데이터 관리자는 이와 같은 어려움을 예상되는 컬럼을 추천 필터로 등록하여 대시보드나 차트 구성이 원활하게 이루어 질 수 있도록 도와줄 수 있습니다.
 - 최우선 추천 필터로 적용: 선택한 컬럼을 최우선 추천 필터로 적용할지 결정합니다.
 - 단일 항목만 선택할 수 있습니다: 해당 컬럼을 최우선 추천 필터로 적용하면 이 항목이 생깁니다. 데이터의 용량이 매우 큰 경우에는 체크하십시오.
 - 순서 설정: 클릭하면 현재 추천 필터를 적용한 컬럼들의 우선 순위를 변경할 수 있습니다.
- 6 **Missing:** 컬럼 내 Null 값을 처리하는 방식에 대해 설정할 수 있습니다.
 - 대체: 여기에 입력된 값으로 Null 값이 대체됩니다.
 - 버림: Null 값을 버립니다.
 - 설정안함: Null 값이 그대로 보여집니다. 단 데이터 소스의 타임스탬프의 Null 값은 무조건 버려지게 됩니다.

5. 데이터 소스 수집 설정을 하고 다음 버튼을 누릅니다.



① **Granularity 설정:** 두 가지 타입의 Granularity를 설정합니다.

- Segment granularity: 분산 노드 환경에서 동작하는 Druid의 특성을 활용하기 위해 데이터를 분할하여 저장하게 되는데 이때 저장하는 시간 단위를 결정합니다.
- Query granularity: 분석에서 수행하고자 하는 최소 시간 단위를 결정합니다. 이는 최소 단위까지의 결과를 미리 생성하여 이후에 보다 빠른 응답을 얻을 수 있도록 하기 위함입니다.

② **롤업:** 롤업은 차원값을 따라 데이터를 요약하는 작업이 포함됩니다. 요약 규칙은 계층 구조를 따라 합계를 계산하거나 'profit=sales=expenses'와 같은 수식 집합을 적용하는 것일 수 있습니다. 즉, 롤업은 수집 시에 효과적인 옵션입니다.

③ **Advanced setting(Opt.):** 적재 성능을 설정합니다. 적재 성능 설정은 수집 시 효과적인 옵션입니다. 텍스트상자에 JSON 형식의 구문을 입력하십시오. 예시:

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

6. 가져온 파일에서 설정한 데이터에 관한 정보를 확인한 뒤, 이름과 설명을 입력하고 완료 버튼을 누르면 데이터 소스가 생성됩니다. 이때, 원천 데이터에서부터 metatron 내장 엔진(Druid)으로 저장(ingestion)하기 때문에 데이터량에 따라 수 초~분의 시간이 소요될 수 있습니다.

파일	sale_20180130.xls
시트	sale.txt (1)
Granularity 설정	<input checked="" type="radio"/> Segment Granularity: 일 <input type="radio"/> Query Granularity: 일
룰업	true

이름
이름을 입력하세요

설명
설명을 입력하세요

이전 **마침**

7. 데이터 소스 관리 화면으로 이동하면 생성된 데이터 소스를 화면에서 확인할 수 있습니다. ingestion이 수행되는 중에는 아래와 같이 상태가 Disabled로 표시되게 되고 ingestion이 완료되면 Enabled로 변경됩니다. 이때부터 데이터 소스를 사용할 수 있습니다.

No	Data source	Data type	Ingestion type	Status	생성일
1	sales-sales data [2011-2014] [전체공개]	IMPORT	Ingested	Disabled	2017-10-25 Polaris

B. 데이터베이스 또는 Staging DB로 데이터 소스 생성하기

- 원천 데이터 탑재 화면에서 '데이터베이스' 또는 'Staging DB'를 선택합니다.
- '데이터베이스'를 선택했을 경우에는 아래와 같은 데이터 커넥션 세팅 화면에서 기준에 생성된 데이터 커넥션을 불러오거나 새로운 데이터베이스 접속 정보를 입력합니다. 'Staging DB' 선택 시에는 이 단계를 건너뜁니다.



- 데이터 커넥션 로드:** 현재 저장되어 있는 데이터 커넥션을 선택할 수 있습니다. 선택하면 해당 데이터 커넥션에 연결되어 있는 데이터베이스에 대한 접근 정보가 자동으로 불려와집니다. 단, 이때도 'Test' 버튼을 눌러서 연결 검증은 반드시 실시해야 합니다.
- DB 탑재:** 연결할 데이터베이스의 탑재를 선택합니다. 현재 총 5개 탑재의 데이터베이스를 지원합니다. (Oracle, MySQL, Hive, presto, TIBERO)
- 수집 타입:** 데이터 소스가 데이터를 수집하는 방식을 선택합니다.
 - 수집형 데이터(Ingested):** 데이터를 metatron 서버에 직접 저장하는 방식으로 수집된 데이터 소스들이 출력됩니다.
 - 연결형 데이터(Linked):** 연결된 데이터베이스에서 필요한 시점마다 데이터를 가져오는 방식의 데이터 소스들이 출력됩니다.
- Host:** 연결할 호스트 값을 입력합니다.
- Port:** 연결할 포트 번호를 입력합니다.

- 6 SID/Catalog:** Oracle과 TIBERO에서는 SID, presto에서는 Catalog값을 입력합니다.
- 7 Username:** 해당 데이터베이스의 username을 입력합니다.
- 8 비밀번호:** 해당 데이터베이스의 비밀번호를 입력합니다.
- 9 Test:** 모든 입력 항목을 다 작성하면 테스트 버튼이 활성화 됩니다. 클릭하면 커넥션이 정상적인지 여부가 버튼 하단에 나타납니다. 정상적이라면 '유효한 커넥션', 비정상적이라면 '잘못된 커넥션' 이라는 문구가 나타납니다.
- 10 새로운 커넥션으로 저장하기:** 기존 데이터 커넥션을 사용하지 않고 새로운 데이터베이스와 연결할 경우 이러한 연결 정보를 새로운 데이터 커넥션으로 저장해야 합니다. 이때 사용할 이름을 입력하십시오.

3. 데이터를 선택합니다. 연결된 데이터베이스 계정에서 테이블을 선택할 수도 있고 쿼리문을 직접 작성할 수도 있습니다. Staging DB로 데이터 소스 생성 시에는 테이블 지정만 가능합니다.

테이블

데이터베이스와 테이블명을 선택한 후 실제 저장될 데이터가 조회되면, 확인 후 '다음' 버튼을 누릅니다.

데이터 소스 생성하기 (DB)

데이터 선택

✓ 테이블 쿼리

1 2

OrderDate	Category	City	Country	CustomerName	OrderId	PostalCode	ProductNa
2011-01-01 00:00:00	Office Supplies	Houston	United States	Darren Powers	CA-2011-1...	77095	Message Bo...
2011-01-02 00:00:00	Office Supplies	Philadelphia	United States	Mick Brown	CA-2011-1...	19143	Avery Hi-Li...
2011-01-03 00:00:00	Furniture	Huntsville	United States	Vivek Sundaresam	CA-2011-1...	77340	Howard Mil...
2011-01-04 00:00:00	Office Supplies	Huntsville	United States	Vivek Sundaresam	CA-2011-1...	77340	Acco Four F...
2011-01-05 00:00:00	Furniture	Springfield	United States	Anthony Jacobs	CA-2011-1...	22153	Howard Mil...
2011-01-06 00:00:00	Office Supplies	Springfield	United States	Anthony Jacobs	CA-2011-1...	22153	Avery 482...
2011-01-07 00:00:00	Furniture	Philadelphia	United States	Brendan Sweed	CA-2011-1...	19140	Sauder Cor...
2011-01-08 00:00:00	Furniture	Philadelphia	United States	Delfina Latchford	CA-2011-11...	19134	DAX Black...
2011-01-09 00:00:00	Office Supplies	Philadelphia	United States	Delfina Latchford	CA-2011-11...	19134	GBC Linen...
2011-01-10 00:00:00	Office Supplies	Philadelphia	United States	Delfina Latchford	CA-2011-11...	19134	Round Spe...
2011-01-11 00:00:00	Technology	Philadelphia	United States	Delfina Latchford	CA-2011-11...	19134	AT&T 8410...
2011-01-12 00:00:00	Office Supplies	Springfield	United States	David Wiener	CA-2011-1...	97477	Fellowes M...

이전 다음

- 1 **데이터베이스 선택:** 선택한 데이터 커넥션과 연결되어 있는 데이터베이스 중 하나를 선택합니다.
- 2 **스키마 선택:** 선택한 데이터베이스에 속한 테이블 중 하나를 선택합니다.

쿼리

원하는 데이터를 가져올 수 있는 쿼리문을 직접 작성하고 '실행' 버튼을 클릭하면 하단에 데이터가 보여집니다. 데이터를 확인한 후 '다음' 버튼을 누르십시오.

The screenshot shows a user interface for querying data from 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 at the top: "Table" and "Query", with "Query" selected. A SQL query is displayed: "select * from test.emp_table". Below the query is a "Run" button. The main area shows 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, there are "Previous" and "Next" buttons.

4. 이후 절차는 파일에서 데이터를 가져오는 방식과 동일합니다. 2.1.3 A절의 3번 단계부터 참조하십시오. 단, 데이터베이스로부터 데이터 소스를 생성할 경우 수집 설정 시 다음과 같은 추가적인 설정 항목이 있습니다.

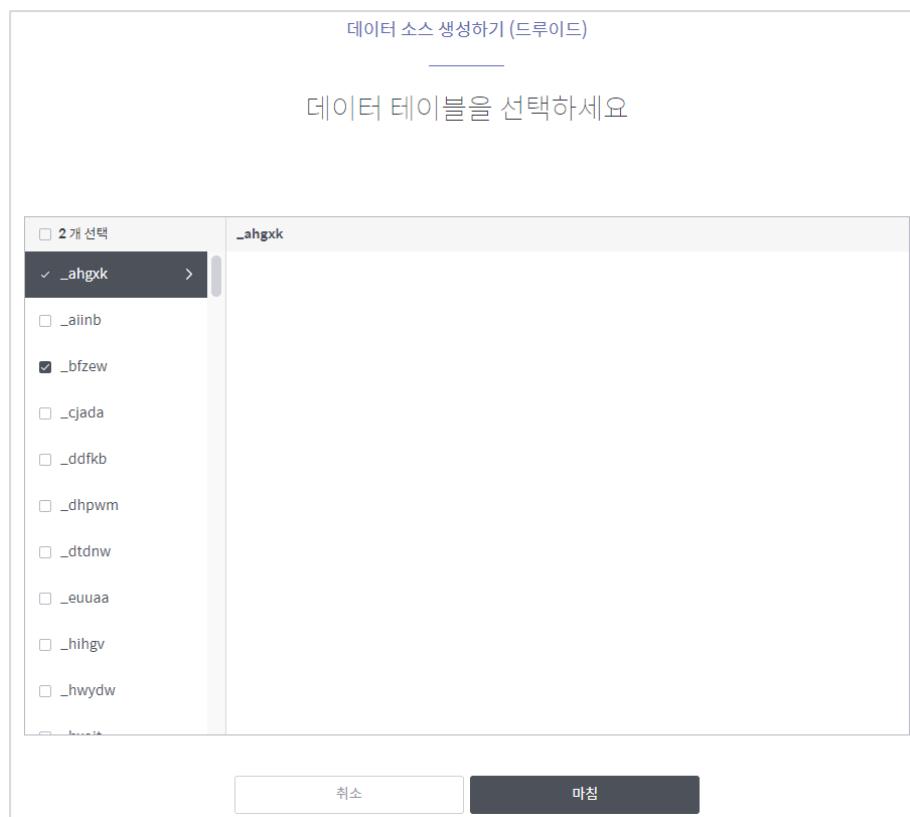
The screenshot shows a configuration dialog for collection settings. It has two main sections:

- 수집 설정**:
① 1회 수집 (radio button selected) ○ 주기적 수집
② 수집 데이터 범위:
○ Gets the entire data ○ Gets only the first 10000 rows

- ① 수집 설정: 데이터 수집에 관해 설정 합니다.
 - 1회 수집: 선택하면 데이터 저장을 이번 한번만 하게 됩니다.
 - 주기적 수집: 기간을 두어 데이터 저장을 주기적으로 실행합니다.
- ② 수집 데이터 범위: 수집 데이터의 범위를 설정합니다.
 - Gets the entire data: 데이터 수집을 주기적으로 실행할 경우, 전체 데이터를 수집합니다.
 - Gets only the first: 데이터의 시작부터 특정 몇 개 까지만 넣을 것인지를 설정합니다.

C. 메타트론 엔진(v1.0) 데이터 소스 마이그레이션

- 원천 데이터 탑색 화면에서 '데이터베이스' 또는 '메타트론 엔진'을 선택합니다.
- 아래와 같이 metatron V1.0에서 만든 데이터 소스가 좌측 화면에 나열되면, 그 중에서 현 버전으로 마이그레이션하고자 하는 데이터 소스들의 확인란에 체크합니다.



- '마침' 버튼을 누르면 선택한 데이터 소스들이 마이그레이션됩니다.

2.2 데이터 커넥션

metatron Discovery는 외부 데이터베이스를 직접 연결하는 기능을 지원합니다. 이와 같이 외부 데이터베이스를 연결하려면 해당 데이터베이스에 대한 접근 정보가 담긴 데이터 커넥션을 생성·관리해야 합니다. 이렇게 데이터 커넥션을 등록해 두면 새로운 데이터베이스 접속 정보를 다시 입력해야 하는 수고를 덜 수 있습니다. 데이터 커넥션의 용도는 다음과 같습니다.

- 일반용: 데이터베이스 기반으로 데이터 소스를 새로 생성할 때 해당 데이터베이스와 연결(2.1.3 B절 참조)
- 워크벤치용: 워크벤치에서 사용할 데이터베이스와 연결('6부 워크벤치' 참조)

2.2.1 데이터 커넥션 관리 화면

본 화면에서는 데이터베이스 커넥션의 신규 등록·편집·조회가 가능합니다.

타입	데이터 커넥션	DB 타입	Host	Port	생성일
Workbench	azure-mysql-test	MYSQL	metatron-poc-h04	3306	2018-03-22 11:05 by A...
General	Test_Connection	MYSQL	metatron-poc-h04	3306	2018-02-09 16:24 by A...
Workbench	metatron-azure	MYSQL	metatron-poc-a01	3306	2018-01-25 11:07 by A...
General	azure-local	MYSQL	metatron-poc-h04	3306	2018-01-14 21:12 by A...
Workbench	polaris	MYSQL	exntu.kr	32306	2017-12-04 10:19 by A...
Workbench	Hive(2.3)	HIVE	metatron-poc-h03	10000	2017-11-15 17:22 by A...
Workbench	Hive(1.2)	HIVE	metatron-poc-h04	10000	2017-11-15 17:21 by Ad...

① **타입:** 데이터 커넥션을 용도별로 선별하여 조회합니다.

- 모두: 용도와 상관 없이 모든 데이터 커넥션이 출력됩니다.
- 일반: 일반용 데이터 커넥션이 출력됩니다. 데이터 소스를 새로 생성할 때 사용됩니다.
- 워크벤치: 워크벤치용 데이터 커넥션이 출력됩니다. 워크벤치를 새로 생성할 때 사용됩니다.

② **DB타입:** 데이터 커넥션을 데이터베이스 타입(Oracle/MySQL/Hive/presto/Tibero)별로 선별하여 조회합니다.

③ **Time:** 데이터 커넥션 조회 시 적용되는 시간 기준입니다. 생성일과 수정일 중 원하는 기준으로 선택할 수 있으며 시간 범위는 전체/오늘/지난 1주일/특정 날짜 기간 중 선택이 가능합니다.

- ④ 데이터 커넥션 이름으로 검색:** 현재 등록된 데이터 커넥션을 이름으로 검색합니다.
- ⑤ 데이터 커넥션 개수:** 현재 목록에 조회된 데이터 커넥션의 개수를 나타냅니다.
- ⑥ 새로운 데이터 커넥션 생성:** 클릭하면 새로운 데이터 커넥션을 생성할 수 있습니다.
- ⑦ 데이터 커넥션 목록:** 설정한 선별 조건에 맞는 데이터 커넥션들을 보여줍니다. 이 중 하나를 클릭하면 설정을 변경할 수 있습니다.
- ⑧ 삭제:** 해당 데이터 커넥션에 마우스 오버 시 휴지통 아이콘이 나타납니다. 클릭하면 해당 데이터 커넥션을 삭제할 수 있습니다.

2.2.2 데이터 커넥션 새로 만들기

A. 일반용 데이터 커넥션 새로 만들기

일반용 데이터 커넥션은 metatron Discovery에서 데이터 소스를 새로 생성할 때 사용하기 위한 것으로서, 다음의 양식을 입력하면 새로 만들 수 있습니다.

데이터커넥션 생성하기

필수 항목을 설정하고 데이터 연결 생성을 완료하십시오

1 타입 ① 일반 ② 워크벤치

2 DB 타입

ORACLE DATABASE	MySQL
Hive	presto
TIBERO	

3 Host 4 Port

5 SID
데이터베이스 이름을 입력하세요

6 Username
Username을 입력하세요

7 비밀번호
비밀번호를 입력하세요

8 테스트

9 데이터커넥션 이름
새로운 데이터 커넥션 이름을 입력하세요

취소 마침

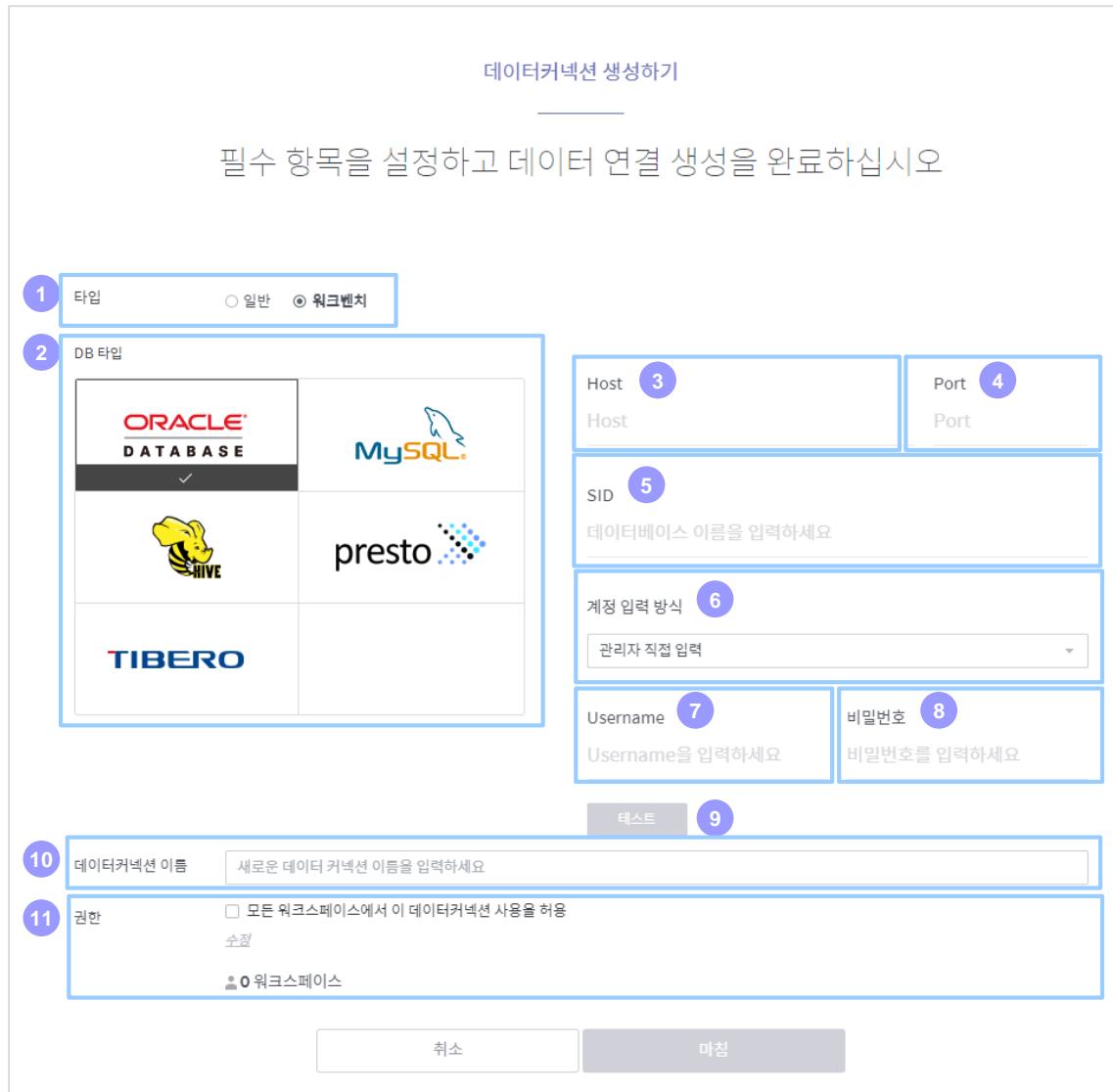
- 1 **타입:** 생성할 데이터 커넥션의 용도입니다. 일반용 데이터 커넥션을 만들려면 '일반'을 선택하십시오.
- 2 **DB 타입:** 현재 총 5개 타입의 데이터베이스를 지원합니다. (Oracle, MySQL, Hive, presto, TIBERO)
- 3 **Host:** 연결할 호스트 값을 입력합니다.
- 4 **Port:** 연결할 포트 번호를 입력합니다.
- 5 **SID/Catalog:** Oracle과 TIBERO에서는 SID, presto에서는 Catalog값을 입력합니다.
- 6 **Username:** 해당 데이터베이스의 username을 입력합니다.
- 7 **비밀번호:** 해당 데이터베이스의 비밀번호를 입력합니다.
- 8 **테스트:** 모든 입력 항목을 다 작성하면 테스트 버튼이 활성화 됩니다. 클릭하면 커넥션이 정상적인지 여부가 버튼 하단에 나타납니다. 정상적이라면 '유효한 커넥션', 비정상적이라면 '잘못된 커넥션' 이라는 문구가 나타납니다.
- 9 **데이터커넥션 이름:** 새로 생성할 데이터 커넥션의 이름을 입력합니다.

B. 워크벤치용 데이터 커넥션 새로 만들기

워크벤치용 데이터 커넥션은 워크벤치에서 사용할 데이터베이스를 연결하기 위한 것으로서, 다음의 양식을 입력하면 새로 만들 수 있습니다.

데이터커넥션 생성하기

필수 항목을 설정하고 데이터 연결 생성을 완료하십시오



1 타입 2 DB 타입 3 Host 4 Port
5 SID 6 계정 입력 방식 7 Username 8 비밀번호
9 테스트 10 데이터커넥션 이름 11 권한

- 1 **타입:** 생성할 데이터 커넥션의 용도입니다. 워크벤치용 데이터 커넥션을 만들려면 '워크벤치'를 선택하십시오.
- 2 **DB 타입:** 현재 총 5개 타입의 데이터베이스를 지원합니다. (Oracle, MySQL, Hive, presto, TIBERO)
- 3 **Host:** 연결할 호스트 값을 입력합니다.
- 4 **Port:** 연결할 포트 번호를 입력합니다.
- 5 **SID/Catalog:** Oracle과 TIBERO에서는 SID, presto에서는 Catalog값을 입력합니다.
- 6 **계정 입력 방식:** 워크벤치에서 해당 커넥션을 사용할 때, 로그인 방식을 설정합니다.
 - 관리자 직접 입력: 데이터 커넥션 생성 시 사용자가 직접 입력한 정보를 사용하여 로그인합니다.

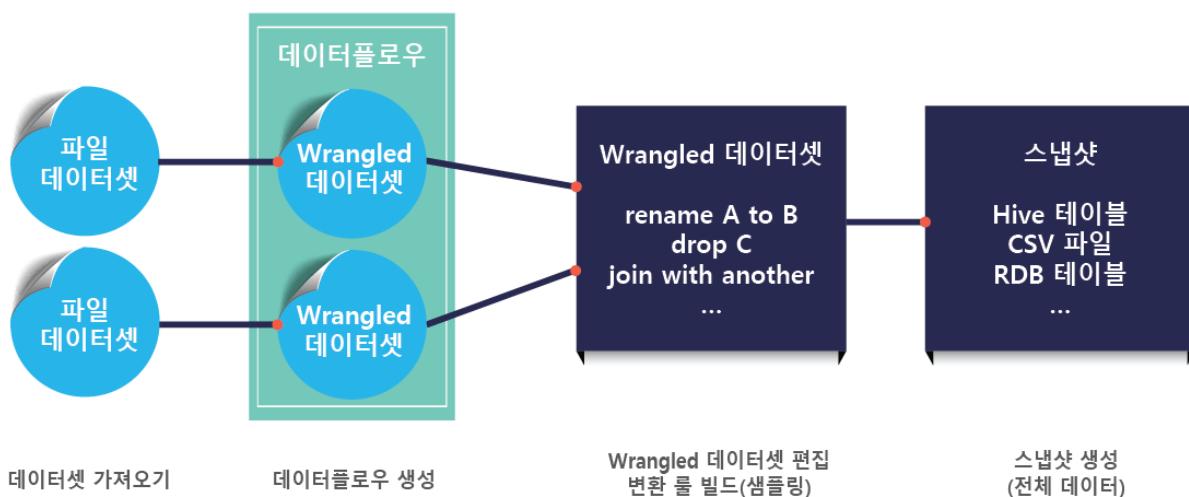
- 사용자의 계정 사용: metatron Discovery에 등록되어 있는 사용자 계정 정보를 사용하여 로그인합니다.
- 워크벤치 접속 시, 직접 입력: 워크벤치에서 접근을 수행할 때마다 입력 받아서 로그인합니다.

- ⑦ **Username:** 해당 데이터베이스의 username을 입력합니다.
- ⑧ **비밀번호:** 해당 데이터베이스의 비밀번호를 입력합니다.
- ⑨ **테스트:** 모든 입력 항목을 다 작성하면 테스트 버튼이 활성화 됩니다. 클릭하면 커넥션이 정상적인지 여부가 버튼 하단에 나타납니다. 정상적이라면 '유효한 커넥션', 비정상적이라면 '잘못된 커넥션'이라는 문구가 나타납니다.
- ⑩ **데이터커넥션 이름:** 새로 생성할 데이터 커넥션의 이름을 입력합니다.
- ⑪ **권한:** 생성할 데이터 커넥션의 사용을 허용할 워크스페이스를 지정합니다.
 - 모든 워크스페이스에 데이터 소스를 사용하도록 허용 확인란: 이 확인란에 체크하면 모든 워크스페이스에서 해당 데이터 커넥션을 사용할 수 있습니다.
 - 수정: 해당 데이터 커넥션의 사용을 허용할 특정 워크스페이스를 지정할 때 사용합니다. 해당 데이터 커넥션을 오픈 데이터로 지정할 경우 이 버튼이 사라집니다.
 - 공유 워크스페이스 수: 해당 데이터 커넥션을 사용할 수 있도록 허용된 워크스페이스의 수를 나타냅니다.

3. 데이터 프리퍼레이션 관리

데이터 프리퍼레이션은 분석에 사용할 데이터 소스를 만들기에 앞서 원천 데이터를 규격에 맞게 전처리(정제, 통합, 정리, 변환 등)하는 기능입니다. 이는 단순히 이미 갖춰져 있는 데이터를 정제하는 것 이 아니라, 앞으로 추가될 레코드들에 대해서도 미리 전처리 절차를 확립해 놓음으로써 데이터가 새로 입력될 때마다 정해놓은 규칙에 따라 자동으로 정제되게 하는 것입니다. 메뉴는 다음과 같이 구성되어 있습니다.

- **데이터셋:** 원천 데이터와의 연결 관리(3.1절 참조)
- **데이터플로우:** 데이터셋이 전처리 되는 과정을 정의(3.2절 참조)
- **스냅샷 저장:** 데이터플로우 과정에서 현재까지 입력 및 전처리된 데이터를 포착하여 파일 또는 데이터베이스 테이블로 저장(3.3절 참조)



3.1 데이터셋

여기서 데이터셋이란 데이터 프리퍼레이션을 위한 원천 데이터를 가리킵니다. 크게 2가지로 구분됩니다.

- **Imported 데이터셋:** 전처리 과정이 정의되지 않은 데이터셋.
- **Wrangled 데이터셋:** 전처리 과정이 정의된 데이터셋. 참고로 전처리 과정은 데이터플로우 화면에서 정의할 수 있습니다.

3.1.1 데이터셋 관리 홈 화면

본 화면에서는 데이터셋의 신규 등록·편집·조회가 가능합니다.

The screenshot shows the 'Data Preparation' section of the Metatron Discovery interface. At the top, there are three tabs: 'Dataflow', 'Dataset' (which is selected), and 'Data snapshot'. Below the tabs, there are several numbered callouts:

- 1**: A radio button group for selecting dataset types: '타입' (Type) with options '전체' (All), 'Imported dataset', and 'Wrangled dataset'.
- 2**: A search bar labeled '데이터셋 이름으로 검색하세요' (Search by dataset name).
- 3**: A message box indicating '74개 데이터가 있습니다' (74 datasets available).
- 4**: A button labeled '새로운 데이터셋 생성' (Create new dataset).
- 5**: A dropdown menu labeled 'Dataset' with a list of datasets:
 - 인구통계 [W] - 서울 인구 통계
 - 인구통계
 - region [W]
 - customer [W]
 - customer [W]
 - lineitem [W]
 - □ □ - □ □ □
 - aaa [W]
 - ≡ [W]
 - ≡ [W]
 - ≡ [W]
 - aaa - aaaaa
- 6**: A delete icon next to the last dataset entry.

Dataset	사용처	소스	생성일
인구통계 [W] - 서울 인구 통계	1	인구통계	2018-04-10 13:47 by a...
인구통계	1	FILE	2018-04-10 13:47 by a...
region [W]	1	region	2018-04-10 13:29 by a...
customer [W]	1	기능 테스트	2018-04-08 21:47 by a...
customer [W]	1	기능 테스트	2018-04-08 21:46 by a...
lineitem [W]	1	기능 테스트	2018-04-08 21:14 by a...
□ □ □ - □ □ □	0	FILE	2018-04-08 19:56 by a...
aaa [W]	1]]	2018-04-08 18:23 by a...
≡ [W]	1]]	2018-04-08 18:23 by a...
≡ [W]	1	#####	2018-04-08 18:21 by a...
≡ [W]	1	ㅋㅋ	2018-04-08 17:45 by a...
aaa - aaaaa	1	FILE	2018-04-08 17:44 by a...

1 타입: 데이터셋을 타입별로 선별하여 조회합니다.

- 모두: 모든 타입의 데이터셋이 출력됩니다.
- Imported dataset: Imported 데이터셋들이 출력됩니다.
- Wrangled dataset: Wrangled 데이터셋들이 출력됩니다.

2 데이터셋 이름으로 검색: 현재 등록된 데이터셋을 이름으로 검색합니다.

3 데이터 개수: 현재 목록에 조회된 데이터셋의 개수를 나타냅니다.

4 새로운 데이터셋 생성: 클릭하면 새로운 데이터셋을 등록할 수 있습니다.

5 데이터셋 목록: 설정한 선별 조건에 맞는 데이터셋들을 보여줍니다. 이 중 하나를 클릭하면 상세 내역을 볼 수 있습니다. (3.1.2절 참조)

6 삭제: 해당 데이터셋에 마우스 오버 시 휴지통 아이콘이 나타납니다. 클릭하면 해당 데이터셋을 삭제할 수 있습니다.

3.1.2 데이터셋 상세 정보

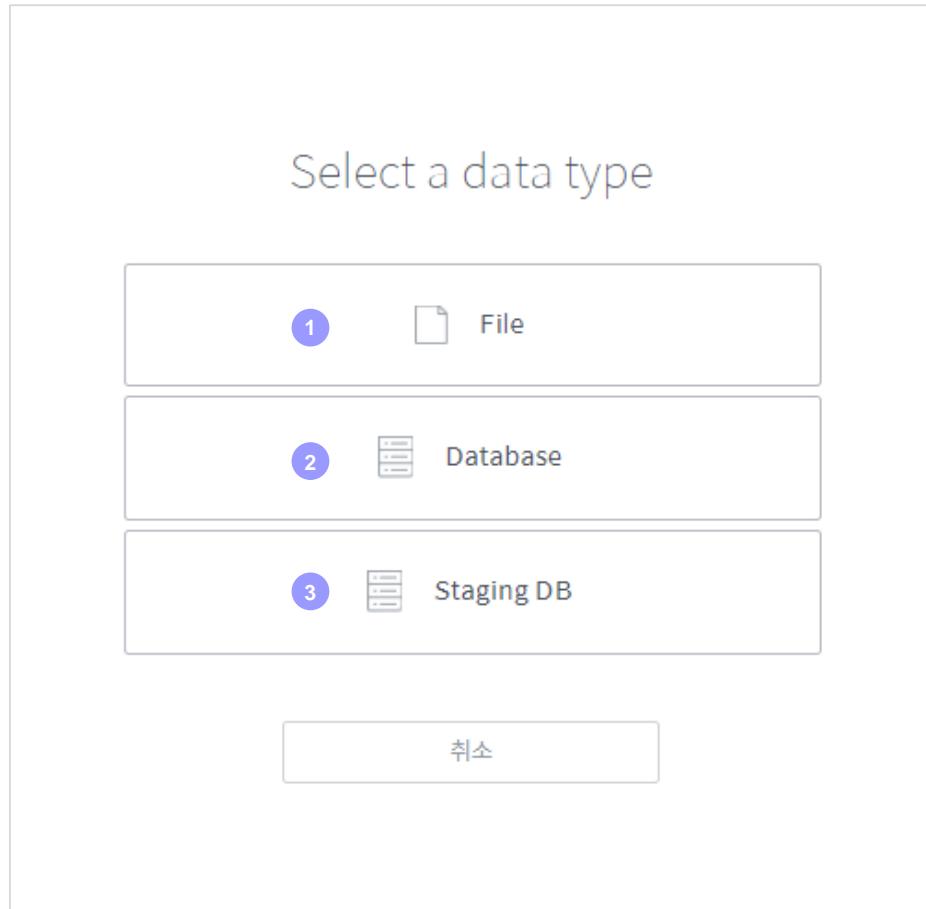
데이터셋 관리 홈에 열거된 데이터셋 중 하나를 클릭하면 해당 데이터셋에 관한 다양한 속성을 확인할 수 있습니다.

The screenshot shows the METATRON DISCOVERY interface. At the top, there are two tabs: '인구통계 [W]' (selected) and '서울 인구 통계'. The main area displays a table titled 'WRANGLED' with 35 Rows / 13 Columns. The table contains data for three countries: SC071.AT, SC071.BE, and SC071.BG. The columns include 'indic_em_geo_time', '2005', '2006', and '2007'. Below the table, there is a section for '사용처' (User) showing a single entry: '인구통계 - 인구통계'. The interface includes various buttons and icons for managing the dataset.

- ① **이름:** 해당 데이터셋의 이름입니다. 클릭 시 수정 가능합니다.
- ② **설명:** 해당 데이터셋에 대한 설명입니다. 클릭 시 수정 가능합니다.
- ③ **타입:** 해당 데이터셋의 타입(Imported/Wrangled)을 나타냅니다.
- ④ **요약:** 해당 데이터셋에 연결된 원천 데이터 테이블의 레코드 및 컬럼 수입니다.
- ⑤ **데이터:** 해당 데이터셋에 연결된 원천 데이터 테이블의 컬럼 정보입니다.
- ⑥ **사용처:** 해당 데이터셋을 사용하고 있는 데이터플로우 이름, 그 데이터플로우에서 사용하는 Imported 및 Wrangled 데이터셋의 개수, 그 데이터플로우를 마지막으로 업데이트한 시간과 사용자 ID가 순서대로 나타냅니다. Import 타입 데이터셋의 상세 정보 화면에는 그 데이터셋에 연결된 모든 데이터플로우가 표시됩니다.

3.1.3 데이터셋 새로 만들기

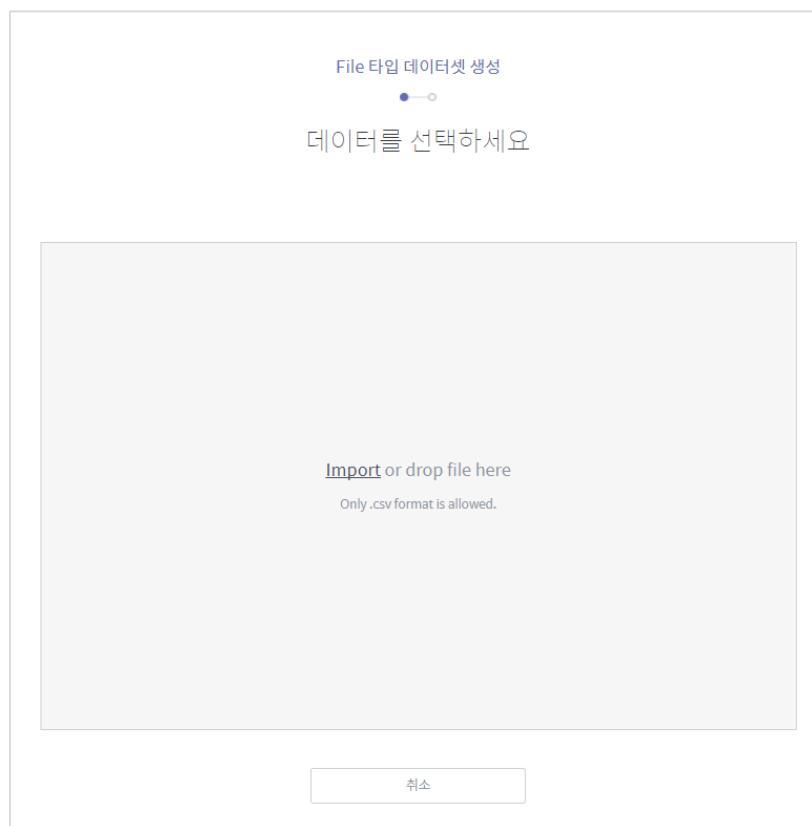
데이터셋 관리 홈 우측 상단에서 '+ 새로운 데이터셋 생성' 버튼을 누르면 새로운 데이터셋을 생성함으로써, 전처리에 사용할 데이터를 등록할 수 있습니다.



- ① **파일:** 사용자의 로컬 PC에 저장되어 있는 파일을 가져와서 데이터셋으로 등록합니다. xls, xlsx, csv와 같은 쉼표로 분리된 파일만 허용됩니다.
- ② **데이터베이스:** 외부 데이터베이스에서 데이터를 가져와서 데이터셋으로 등록합니다.
- ③ **Staging DB:** metatron의 내부 Hive 데이터베이스에서 가져온 데이터를 데이터셋으로 등록합니다.

A. 파일로 데이터셋 생성하기

- 원천 데이터 탑입 선택 화면에서 '파일'을 선택합니다.
- 사용자 로컬 PC에서 데이터셋으로 사용할 파일을 가져옵니다. Import 버튼을 클릭하여 파일을 선택할 수도 있고 화면 상으로 파일을 끌어다 놓을 수도 있습니다. 파일을 가져왔으면 다음 버튼을 누릅니다.



3. 가져온 파일 안의 데이터가 아래 그림과 같이 화면에 나타나게 됩니다. 의도한 데이터가 맞다면 구분자 확인 후에 '다음'을 누르고, 그렇지 않다면 파일을 다시 가져오십시오.

File 타입 데이터셋 생성

데이터를 선택하세요

● ○

1 2 Import or drop file here

"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 Shelves,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

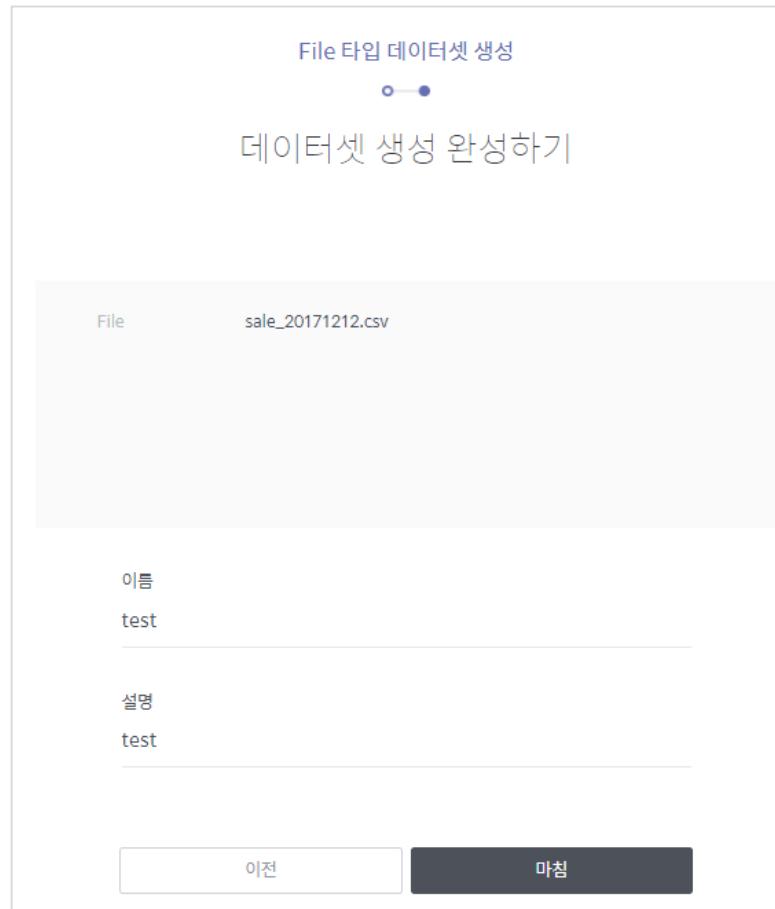
2011-01-06T00:00:00Z,Office Supplies,Philadelphia,United States,Mick Brown,CA-2011-141817,19143,"Avery Hi-Liter EverBold Pen Style Fluorescent Highlighters, 4/Pack",3,East,Consumer,2011. 1. 13.,Standard Class,Pennsylvania,Art,Shipped Late,null,null,

Column delimiter 4

취소다음

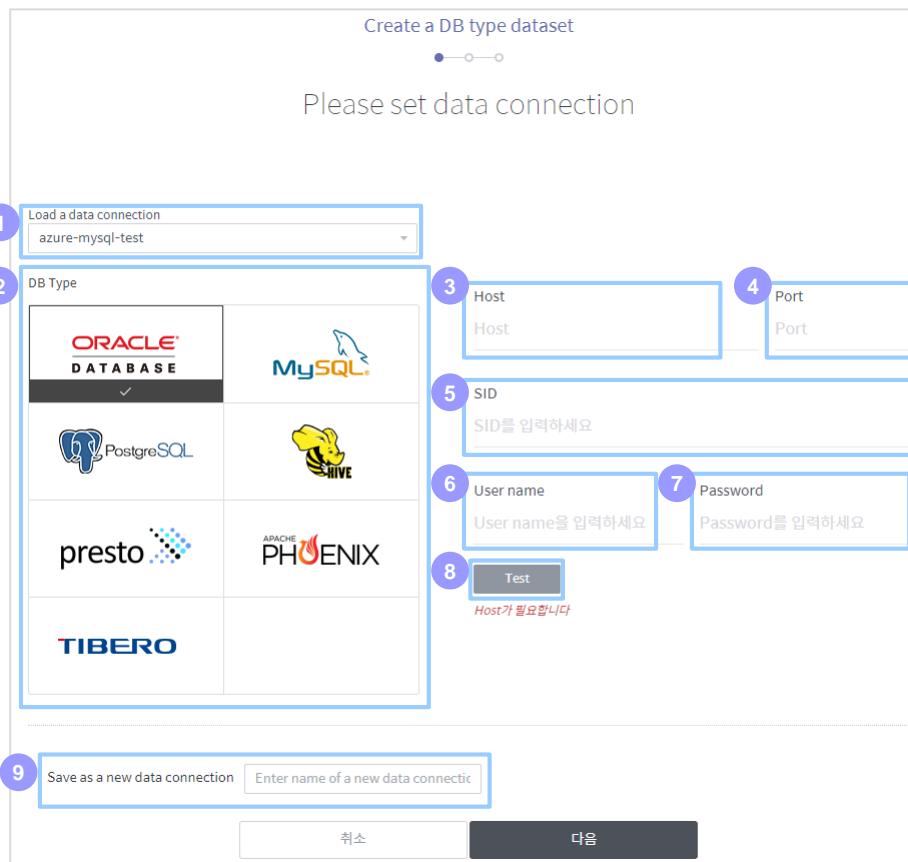
- 1 **이름:** 해당 파일의 이름입니다.
- 2 **Import or drop file here:** 새로운 파일을 가져올 수 있습니다.
- 3 **데이터 미리 보기:** 해당 파일의 데이터 내용을 미리 볼 수 있습니다.
- 4 **Column delimiter:** 컬럼 구분 기호를 입력합니다. 기본값은 마침표(.)으로 설정되어 있습니다.

4. 데이터셋을 생성하기 위해 불러온 파일 형식의 데이터에 관한 정보를 확인한 뒤, 이름과 설명을 입력하고 완료 버튼을 누르면 데이터셋이 생성됩니다.



B. 데이터베이스 또는 Staging DB로 데이터셋 생성하기

- 원천 데이터 탑입 화면에서 '데이터베이스' 또는 'Staging DB'를 선택합니다.
- '데이터베이스'를 선택했을 경우에는 아래와 같은 데이터커넥션 세팅 화면에서 기존에 생성된 데이터 커넥션을 불러오거나 새로운 데이터베이스 접속 정보를 입력합니다. 'Staging DB' 선택 시에는 이 단계를 건너뜁니다.



- 1 Load a data connection:** 현재 저장되어 있는 데이터 커넥션을 선택할 수 있습니다. 선택하면 해당 데이터 커넥션에 연결되어 있는 데이터베이스에 대한 접근 정보가 자동으로 불려와집니다. 단, 이때도 'Test' 버튼을 눌러서 연결 검증은 반드시 실시해야 합니다.
- 2 DB Type:** 연결할 데이터베이스의 탑입을 선택합니다. 현재 총 7개 탑입의 데이터베이스를 지원합니다. (Oracle, MySQL, PostgreSQL, Hive, presto, Phoenix, TIBERO)
- 3 Host:** 연결할 호스트 값을 입력합니다.
- 4 Port:** 연결할 포트 번호를 입력합니다.
- 5 SID/Catalog/DB 이름:** Oracle과 TIBERO에서는 SID, presto에서는 Catalog, PostgreSQL에서는 DB 이름을 입력합니다.
- 6 Username:** 해당 데이터베이스의 username을 입력합니다.
- 7 Password:** 해당 데이터베이스의 비밀번호를 입력합니다.
- 8 Test:** 모든 입력 항목을 다 작성하면 테스트 버튼이 활성화 됩니다. 클릭하면 커넥션이 정상적인지 여부가 버튼 하단에 나타납니다. 정상적이라면 '유효한 커넥션', 비정상적이라면 '잘못된 커넥션'이라는 문구가 나타납니다.

- 9 **Save as a new data connection:** 기존 데이터 커넥션을 사용하지 않고 새로운 데이터베이스와 연결할 경우 이러한 연결 정보를 새로운 데이터 커넥션으로 저장해야 합니다. 이때 사용할 이름을 입력하십시오.

3. 데이터를 선택합니다. 연결된 데이터베이스 계정에서 테이블을 선택할 수도 있고 쿼리문을 직접 작성할 수도 있습니다. Staging DB로 데이터 소스 생성 시에는 테이블 지정만 가능합니다.

테이블

데이터베이스와 테이블명을 선택한 후 실제 저장될 데이터가 조회되면, 확인 후 '다음' 버튼을 누릅니다.

JDBC 탑 데이터셋 생성

데이터를 선택하세요

✓ 테이블 쿼리

default 1 sale_head 2

orderdate	category	city	country	customername	orderid	postalcode	productname
2011-01-04T...	Office Supp...	Houston	United States	Darren Powers	CA-2011-1...	77095	Message Book
2011-01-05T...	Office Supp...	Naperville	United States	Phillina Ober	CA-2011-1...	60540	GBC Standard
2011-01-05T...	Office Supp...	Naperville	United States	Phillina Ober	CA-2011-1...	60540	Avery 508
2011-01-05T...	Office Supp...	Naperville	United States	Phillina Ober	CA-2011-1...	60540	SAFCO Botle...
2011-01-06T...	Office Supp...	Philadelphia	United States	Mick Brown	CA-2011-1...	19143	Avery Hi-Liter
2011-01-07T...	Office Supp...	Athens	United States	Jack OBriant	CA-2011-1...	30605	Dixon Prang W...
2011-01-07T...	Office Supp...	Los Angeles	United States	Lycoris Saunders	CA-2011-1...	90049	Xerox 225
2011-01-07T...	Furniture	Henderson	United States	Maria Etezadi	CA-2011-1...	42420	Global Deluxe
2011-01-07T...	Office Supp...	Henderson	United States	Maria Etezadi	CA-2011-1...	42420	Rogers Handh...
2011-01-07T...	Office Supp...	Henderson	United States	Maria Etezadi	CA-2011-1...	42420	Ibico Hi-Tech
2011-01-07T...	Office Supp...	Henderson	United States	Maria Etezadi	CA-2011-1...	42420	Alliance Super...
2011-01-07T...	Office Supp...	Henderson	United States	Maria Etezadi	CA-2011-1...	42420	Southworth 2...

취소 다음

- 1 **데이터베이스 선택:** 선택한 데이터 커넥션과 연결되어 있는 데이터베이스 중 하나를 선택합니다.
- 2 **스키마 선택:** 선택한 데이터베이스에 속한 테이블 중 하나를 선택합니다.

쿼리

원하는 데이터를 가져올 수 있는 쿼리문을 직접 작성하고 '실행' 버튼을 클릭하면 하단에 데이터가 보여집니다. 데이터를 확인한 후 '다음' 버튼을 누르십시오.

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. 마지막으로 지금까지 설정한 데이터를 확인하고 만들고자 하는 데이터셋 이름과 설명을 넣은 후 '마침'버튼을 누르면 데이터셋이 완성됩니다. 이렇게 만들어진 데이터셋은 Imported 데이터셋입니다.

DB 타입 데이터셋 생성

데이터셋 생성 완성하기

Type	HIVE
Host	metatron-poc-h03
Port	10000
Database / Table	tpch / region

이름

이름을 입력하세요

설명

설명을 입력하세요

이전

마침

3.2 데이터플로우

데이터플로우는 데이터셋을 전처리하여 데이터 품질을 확보하고 분석 또는 시각화 목적으로 최적화된 형태로 만드는 과정입니다.

3.2.1 데이터플로우 관리 홈 화면

본 화면에서는 원천 데이터를 관리하고 전처리하는 기능을 수행할 수 있습니다.

The screenshot shows the 'Data Preparation' section of the METATRON DISCOVERY interface. At the top, there are tabs for 'Dataflow' (which is selected), 'Dataset', and 'Data snapshot'. Below the tabs, there is a search bar labeled 'Q. 데이터플로우 이름으로 검색하세요.' (Search by Dataflow name). To the right of the search bar, it says '16개 데이터가 있습니다' (16 datasets) and has a '+' button for '데이터플로우 추가' (Add Dataflow). A vertical sidebar on the left lists categories: 'Dataflow' (selected), '인구통계' (Demographic), 'region', ']]-:::', 'LLLLL-LLL', '33-333', '000-0000', 'o-o', '0Lo-0Lo', 'adfa-adfaf', 'nullTest', 'test', '기능 테스트 - jhkim', and 'dataflow_KB'. The main area displays a table of datasets with columns for 'Dataset' (number of rows, number of columns), '생성일' (Created Date), and a small icon. A blue box highlights the search bar. Numbered circles 1 through 5 point to specific UI elements: 1 points to the search bar; 2 points to the dataset count and add button; 3 points to the sidebar category list; 4 points to the table header; 5 points to the delete icon in the table row.

- ① **검색:** 현재 저장되어 있는 데이터플로우를 이름으로 검색합니다.
- ② **데이터 개수:** 현재 목록에 조회된 데이터플로우의 개수를 나타냅니다.
- ③ **새로운 플로우 추가:** 클릭하면 새로운 데이터플로우를 추가할 수 있습니다.
- ④ **데이터플로우 목록:** 설정한 선별 조건에 맞는 데이터플로우들을 보여줍니다. 'Dataset' 컬럼에서 각 데이터플로우별로 두 개씩 할당된 아이콘 옆 숫자들은 각각 해당 데이터플로우에서 사용하는 Imported 및 Wrangled 데이터셋의 개수를 가리킵니다. 이 중 하나를 클릭하면 상세 내역을 볼 수 있습니다. (3.2.3절 참조)
- ⑤ **삭제:** 해당 데이터플로우에 마우스 오버 시 휴지통 아이콘이 나타납니다. 클릭하면 해당 데이터플로우를 삭제할 수 있습니다.

3.2.2 데이터플로우 새로 만들기

데이터플로우 관리 홈 화면에서 '데이터플로우 추가' 버튼을 누르면 데이터플로우 생성 화면이 나타납니다.

1. 나열된 데이터셋 중에 데이터플로우 생성에 쓸 데이터셋들을 선택하십시오.

The screenshot shows the 'Dataset Selection' step of the Data Pipeline creation wizard. At the top, there's a title '데이터플로우 생성하기' with a subtitle '데이터셋 선택하기'. Below this is a search bar labeled '1. 데이터셋 이름으로 검색하세요' and a dropdown menu labeled '2. Dataset' with a list of datasets. The list includes 'sale', 'return', 'region', 'orders', 'nullTrst', 'lineitem', 'gm', 'customer', and another 'customer' entry. The 'return', 'nullTrst', and 'lineitem' datasets are checked. At the bottom of the list are buttons for '더보기' (More) and '3개 선택' (Select 3). On the right side, there's a link '+ 데이터셋 생성하기' (Create new dataset). At the very bottom are '취소' (Cancel) and '다음' (Next) buttons, with the '다음' button highlighted with a blue circle containing the number '3'.

선택 여부	Dataset	Type	Last Updated
<input type="checkbox"/>	sale	Imported dataset(FILE)	2018-03-27 10:03:11
<input checked="" type="checkbox"/>	return	Imported dataset(FILE)	2018-03-30 17:47:47
<input type="checkbox"/>	region	Imported dataset(HIVE)	2018-03-27 09:58:19
<input type="checkbox"/>	orders	Imported dataset(HIVE)	2018-03-26 13:10:25
<input checked="" type="checkbox"/>	nullTrst	Imported dataset(FILE)	2018-04-02 14:31:28
<input checked="" type="checkbox"/>	lineitem	Imported dataset(HIVE)	2018-03-26 12:51:29
<input type="checkbox"/>	gm	Imported dataset(FILE)	2018-02-23 19:07:32
<input type="checkbox"/>	customer	Imported dataset(HIVE)	2018-02-23 19:01:37
<input type="checkbox"/>	customer	Imported dataset(HIVE)	2018-03-26 13:12:31

- 1 **데이터셋 이름으로 검색:** 현재 등록되어 있는 데이터셋을 이름으로 검색합니다.
- 2 **데이터셋 목록:** 등록된 데이터셋들을 보여줍니다. 데이터셋 이름으로 검색했을 경우에는 이름에 해당 문자열을 포함하는 데이터셋만 조회됩니다.
- 3 **데이터셋 생성하기:** 해당 버튼을 클릭하면 새로운 데이터셋을 생성할 수 있도록 데이터셋 생성 화면으로 이동할 수 있는 대화창이 나타납니다.

2. 선택한 데이터셋들을 확인한 다음, 생성하려는 데이터플로우의 이름과 설명을 넣은 후 '마침' 버튼을 누르면 데이터셋이 완성됩니다

데이터플로우 생성하기

Please complete dataflow creation

데이터셋

return
nullTrst
lineitem

이름

이름을 입력하세요

설명

설명을 입력하세요

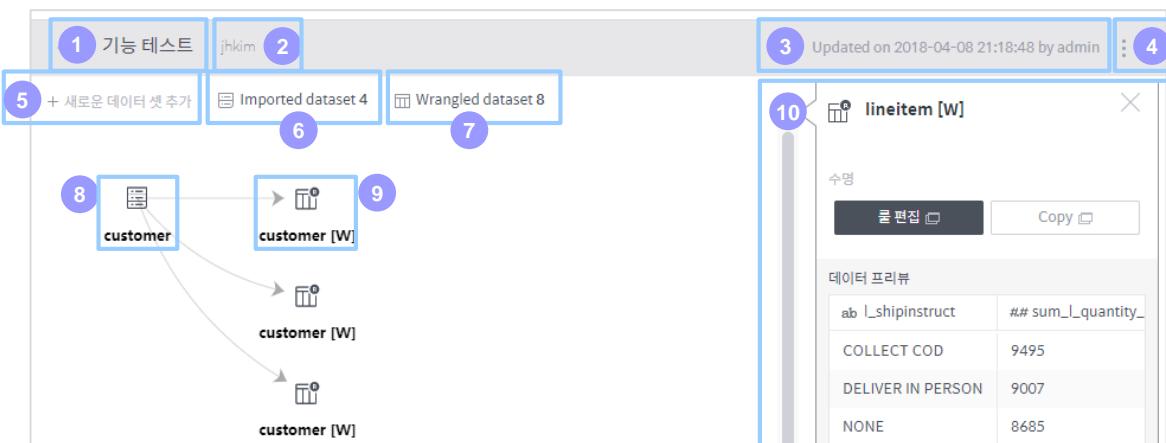
이전

마침

3.2.3 데이터플로우 상세 정보

데이터플로우에 포함된 모든 데이터셋의 연관 관계를 나타내고, 각 데이터셋을 편집할 수 있게 보여주는 화면입니다. JOIN/UNION과 같은 기능을 통해 어떤 데이터셋이 어떤 데이터셋에 영향을 주는지의 관계가 그림으로 표현됩니다.

데이터플로우 상세 정보 화면 개관



- ① **이름:** 해당 데이터플로우의 이름입니다.
- ② **설명:** 해당 데이터플로우에 대한 설명입니다.
- ③ **마지막 수정 정보:** 해당 데이터플로우를 누가 언제 마지막으로 수정했는지 보여줍니다.
- ④ **삭제:** 이 아이콘을 누르면 해당 데이터플로우를 삭제할 수 있는 메뉴가 나타납니다.
- ⑤ **새로운 데이터셋 추가:** 클릭하면 해당 데이터플로우에 새로운 데이터셋을 추가할 수 있는 새 창이 뜨게 됩니다. 추가하고자 하는 데이터셋을 클릭 후 '마침' 버튼을 누르면 데이터셋이 추가됩니다.
- ⑥ **Imported dataset:** 해당 데이터플로우에 포함된 Imported 데이터셋의 개수를 나타냅니다.
- ⑦ **Wrangled dataset:** 해당 데이터플로우에 포함된 Wrangled 데이터셋의 개수를 나타냅니다.
- ⑧ : 해당 데이터플로우에 포함된 각각의 Imported 데이터셋을 가리킵니다.
- ⑨ : 각각의 Imported 데이터셋을 랭글링한 Wrangled 데이터셋을 나타냅니다. 여기서 '랭글링'은 여러 룰을 이용하여(3.2.4절 참조) 정제되지 않은 Imported 데이터셋을 사용하기 쉬운 형태의 데이터로 전환하는 과정을 의미합니다. Wrangled 데이터셋은 데이터 소스로 등록되어 데이터 분석 및 시각화 등에 사용할 수 있습니다.
- ⑩ **선택한 데이터셋 정보:** 화면에 표시된 데이터셋 아이콘 중 하나를 클릭하면 선택한 데이터셋에 간단한 정보가 이 영역에 표시됩니다. 이 영역에서는 선택한 데이터셋을 확인·복사·삭제할 수 있으며 이 데이터셋에 적용되는 룰을 편집할 수 있습니다. 이 영역에 대한 자세한 사항은 아래를 참조하십시오.

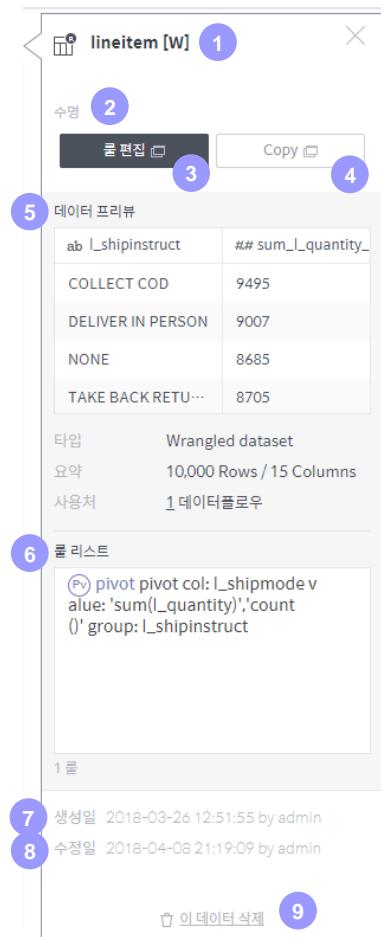
Imported 데이터셋 선택 시 출력되는 정보

- ① **이름:** 선택한 데이터셋의 이름입니다.
- ② **설명:** 선택한 데이터셋에 대한 설명입니다.
- ③ **새로운 데이터셋 생성:** 클릭하면 선택한 Imported 데이터셋에 속한 Wrangled 데이터셋이 하나 더 생성됩니다.
- ④ **데이터 프리뷰:** 선택한 데이터셋에 관한 간단한 기본적인 정보를 보여줍니다.
- ⑤ **생성일:** 선택한 데이터셋을 언제 누가 생성했는지 보여줍니다.
- ⑥ **수정일:** 선택한 데이터셋을 마지막으로 언제 누가 수정했는지 보여줍니다.
- ⑦ **이 데이터 삭제:** 선택한 데이터셋을 삭제합니다.



Wrangled 데이터셋 선택 시 출력되는 정보

- ① **이름:** 선택한 데이터셋의 이름입니다.
- ② **설명:** 선택한 데이터셋에 대한 설명입니다.
- ③ **룰 편집:** 클릭하면 선택한 데이터셋에 대한 룰을 편집할 수 있는 창이 열립니다(이 창에 관한 설명은 다음 절 참조).
- ④ **Copy:** 클릭하면 선택한 데이터셋이 복사되어 동일한 Wrangled 데이터셋이 하나 더 생성됩니다.
- ⑤ **데이터 프리뷰:** 선택한 데이터셋에 관한 간단한 기본적인 정보를 보여줍니다.
- ⑥ **룰 리스트:** 선택한 데이터셋에 현재 적용 중인 룰들을 보여줍니다.
- ⑦ **생성일:** 선택한 데이터셋을 언제 누가 생성했는지 보여줍니다.
- ⑧ **수정일:** 선택한 데이터셋을 마지막으로 언제 누가 수정했는지 보여줍니다.
- ⑨ **이 데이터 삭제:** 선택한 데이터셋을 삭제합니다.



3.2.4 Wrangled 데이터셋 편집

Wrangled 데이터셋을 편집하는 것은 데이터 프리퍼레이션의 가장 근간이 되는 작업이라고 할 수 있습니다. 데이터셋을 변환하는 룰을 정의하고, 그것이 적용된 결과를 봄면서 추가로 룰을 적용하거나(APPEND), 수정하거나(UPDATE), 지우거나(DELETE), 방금한 일을 되돌리거나(UNDO), 다시 적용하면서(REDO) 데이터를 변형하는 일련의 규칙을 빠른 속도로 만들어 갈 수 있습니다.

데이터셋 변환 룰 및 적용 결과

화면 상단부 우측에는 현재 적용된 변환 룰들이 열거되고 좌측에는 그러한 룰들이 반영된 결과를 보여줍니다.

The screenshot shows the Metatron Data Preparation interface. At the top, there's a title bar with 'sale - transform #1' and a search bar labeled 'Search data'. Below the search bar is a table with several columns: 'State', 'Category', 'Shipping', 'Price_1', 'Margin', and 'Rebate'. Each column has a count of categories: 22, 13, 2, 12, 11 respectively. The table contains data rows for states like Idaho, Michigan, New Hampshire, etc., and categories like Phones, Labels, Fasteners, Paper, Furnishings, Binders, Envelopes, Chairs, Storage, and Accessories. To the right of the table is a sidebar titled '룰 리스트' (Rule List) which lists various transformation rules such as 'settype', 'drop', 'sort', 'replace', 'rename', etc., with their corresponding parameters. At the bottom left, there's an 'Add Rule' button and a 'Command' dropdown. On the far right, there are buttons for '취소' (Cancel) and 'Add'.

- ① **이름:** 선택한 Wrangled 데이터셋의 이름입니다.
- ② **검색:** 선택한 Imported 데이터셋 내의 데이터를 검색하여 검색한 문자열이 포함된 데이터만 나타나게 할 수 있습니다.
- ③ **냅샷:** 선택한 Wrangled 데이터셋의 결과로 수집 후 전처리된 전체 데이터를 파일 또는 데이터베이스 테이블 형식으로 저장하여 데이터 소스로 사용될 수 있도록 합니다(3.3절 참조).
- ④ **랭글링 결과 데이터:** 데이터셋에 룰들을 적용하여 랭글링된 결과 데이터를 보여줍니다.
- ⑤ **룰 리스트:** 현재까지 데이터셋에 적용된 룰들을 보여줍니다.

데이터셋 변환 룰 편집 및 추가

화면 하단부에는 데이터셋 변환 룰을 편집하거나 추가할 수 있는 영역이 있습니다.

각 명령어별 기능은 다음과 같습니다.

명령어	내용
rename	컬럼의 이름을 변경
drop	컬럼 제거
delete	조건식에 해당하는 행을 제거
keep	조건식에 해당하는 행들만 남기고, 나머지 행들을 제거
set	주어진 표현식의 결과를 해당 컬럼값으로 설정
derive	위의 set 과 비슷하지만 원본 컬럼은 그대로 두고 새로운 컬럼을 생성
header	지정한 행의 컬럼값들을 각 컬럼의 이름으로 일괄 지정
replace	해당 컬럼값들에 대해 치환식을 일괄 적용함
settype	컬럼의 type 을 변경 (Integer, Float, String, etc.)
extract	주어진 패턴을 찾아 새로운 컬럼으로 생성
split	한 컬럼을 여러 컬럼으로 분할
merge	여러 컬럼을 하나의 컬럼으로 통합
join	다른 데이터셋을 이용해서 join 된 결과로 변형
union	같은 스키마를 갖는 다른 데이터셋들의 내용을 통합
sort	선택한 컬럼값으로 정렬
move	컬럼을 앞뒤로 이동
aggregation	그룹화 연산의 결과로 새로운 컬럼 생성
flatten	배열 컬럼을 지정하였을 때, 각 원소를 새로운 컬럼으로 생성
pivot	지정된 컬럼의 각 값을 이름으로 갖는 컬럼들을 생성하고 그룹화 연산 수행
unpivot	컬럼이름들을 컬럼값으로 갖도록 하여 여러 개의 행으로 전환
countpattern	지정된 컬럼에서 주어진 패턴이 나온 횟수로 새 컬럼 생성
nest	여러 컬럼을 지정해서 그 값들로 배열 또는 맵을 만들어서 새 컬럼 생성
unnest	배열 또는 맵에서 특정하는 요소를 뽑아서 새 컬럼 생성

3.2.5 데이터 스냅샷 생성하기

Wrangled 데이터셋 편집 화면에서 '스냅샷' 버튼을 누르면 데이터 스냅샷을 저장할 수 있습니다.

A. 데이터 스냅샷 타입: File

데이터 스냅샷 타입을 'File'로 선택하면 생성된 스냅샷이 서버에 파일 형태로 저장됩니다.

데이터 스냅샷 생성

1 데이터 스냅샷 타입 파일 HIVE

2 파일 포맷 CSV

3 압축 NONE

4 Full Data ETL engine Embedded Engine

5 파티션 키로 사용할 열 선택

취소 마침

- 1 **데이터 스냅샷 타입:** 데이터 스냅샷을 저장할 데이터베이스 타입을 선택합니다.
- 2 **파일 포맷:** 해당 데이터 스냅샷을 어떤 파일 포맷으로 저장할 것인지 선택합니다.
- 3 **압축:** 해당 데이터 스냅샷을 저장할 때 어떤 압축 방식을 사용할 것인지 선택합니다.
- 4 **Full Data ETL engine:** 데이터 ETL(추출, 변환, 적재)에 사용되는 엔진을 선택합니다.
 - Embedded engine: 소량의 데이터를 포착하는 데 적합합니다.
 - Spark: 대량의 데이터를 포착하는 데 적합합니다.
- 5 **파티션 키로 사용할 열 선택:** 파티션 키 기능을 사용할지 여부를 선택합니다.

B. 데이터 스냅샷 타입: HIVE

데이터 스냅샷 타입을 'HIVE'로 선택하면 생성된 스냅샷이 metatron의 내부 Hive 저장소에 데이터베이스 형태로 저장됩니다.

데이터 스냅샷 생성

1 데이터 스냅샷 타입 파일 HIVE

2 DB 이름 default

3 테이블 이름 snapshot1

4 파일 포맷 CSV

5 압축 SNAPPY

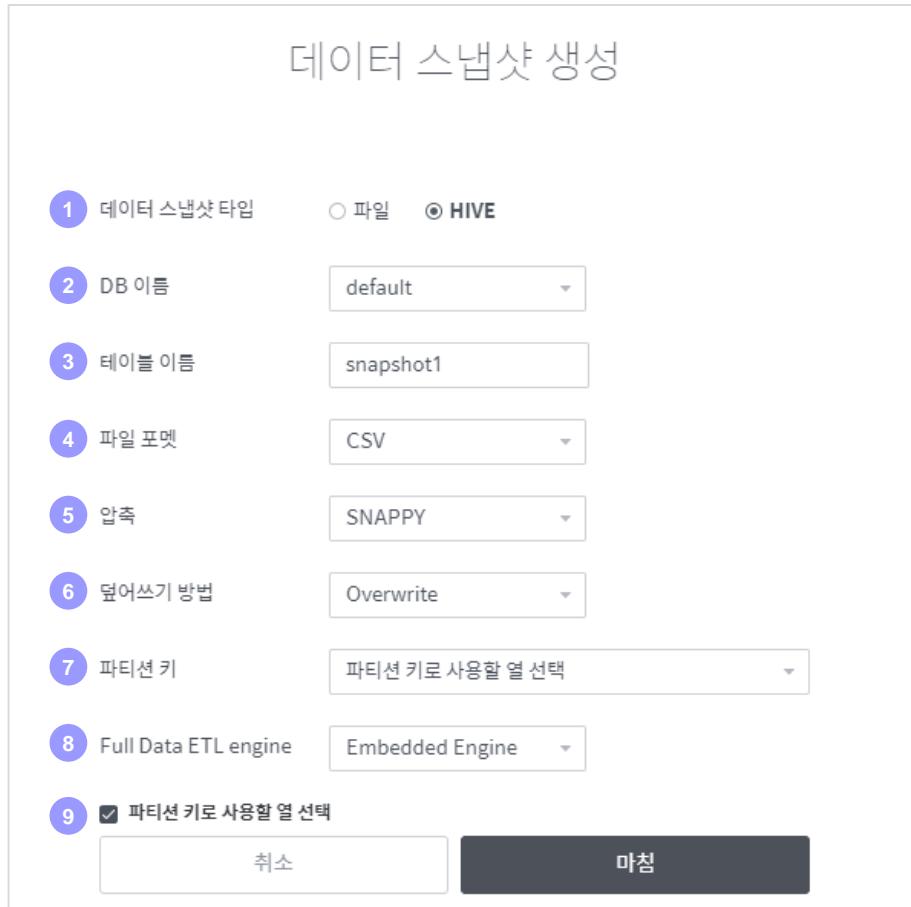
6 덮어쓰기 방법 Overwrite

7 파티션 키 파티션 키로 사용할 열 선택

8 Full Data ETL engine Embedded Engine

9 파티션 키로 사용할 열 선택

취소 마침



- 1 **데이터 스냅샷 타입:** 데이터 스냅샷을 저장할 데이터베이스 타입을 선택합니다.
- 2 **DB 이름:** 데이터 스냅샷을 저장할 데이터베이스를 선택합니다.
- 3 **테이블 이름:** 해당 데이터 스냅샷이 저장되는 테이블의 이름을 설정합니다.
- 4 **파일 포맷:** 해당 데이터 스냅샷을 어떤 파일 포맷으로 저장할 것인지 선택합니다. Hive에서 레코드 값이 저장되는 파일의 포맷을 선택합니다.
- 5 **압축:** 해당 데이터 스냅샷을 저장할 때 어떤 압축 방식을 사용할 것인지 선택합니다.
- 6 **덮어쓰기 방법:** 동일한 이름의 테이블이 존재할 경우 어떻게 저장할 것인지 선택합니다.
 - Overwrite: 기존 테이블 위에 본 스냅샷의 데이터를 덮어씁니다.
 - Append: 기존 테이블에 본 스냅샷의 데이터를 추가합니다.
- 7 **파티션 키:** 파티션 키로 사용할 컬럼을 지정합니다.
- 8 **Full Data ETL engine:** 데이터 ETL(추출, 변환, 적재)에 사용되는 엔진을 선택합니다.
 - Embedded engine: 소량의 데이터를 포착하는 데 적합합니다.
 - Spark: 대량의 데이터를 포착하는 데 적합합니다.
- 9 **파티션 키로 사용할 열 선택:** 파티션 키 기능을 사용할지 여부를 선택합니다.

3.3 데이터 스냅샷

데이터 스냅샷은 Wrangled 데이터셋에 정의된 전처리 룰을 현재까지 수집된 데이터에 적용한 후 그 결과를 저장하는 기능입니다. 이렇게 저장된 데이터는 데이터 소스로 활용하여 분석 및 시각화할 수 있습니다. 이렇게 함으로써 원천 데이터가 데이터 프리퍼레이션 과정을 거쳐 분석에 최적화된 형태로 저장되는 것입니다.

3.3.1 데이터 스냅샷 관리 홈 화면

이 화면에는 만들어진 데이터 스냅샷들의 이름, 생성일시, 만들어낸 데이터플로우 및 Wrangled 데이터셋 이름, 성공 여부, 소요 시간 등이 표시됩니다.

데이터스냅샷	Dataflow Dataset	상태	경과 시간	생성일	操作
superstore [W] 20180405_063201	test superstore [W]	성공	00:00:28.00	2018-04-05 15:32 by a...	
superstore [W] 20180403_080927	test superstore [W]	성공	00:00:22.00	2018-04-03 17:09 by a...	
superstore [W] 20180403_080917	test superstore [W]	처리중	--:--:--	2018-04-03 17:09 by a...	
lineitem [W] 20180327_005911	기능 테스트 lineitem [W]	성공	00:00:10.00	2018-03-27 09:59 by a...	
credit_card [W] 20180323_090931	dataflow_KB credit_card [W]	성공	00:00:13.00	2018-03-23 18:09 by a...	
credit_card [W] 20180323_083923	dataflow_KB credit_card [W]	성공	00:00:14.00	2018-03-23 17:39 by a...	

- ① **상태:** 데이터 스냅샷을 성패 기준으로 선별하여 조회할 수 있습니다.
 - 전체: 데이터 포착 성패와 관계 없이 모든 데이터 스냅샷이 출력됩니다.
 - 성공: 데이터 포착에 성공한 스냅샷들이 출력됩니다.
 - 실패: 데이터 포착에 실패한 스냅샷들이 출력됩니다.
- ② **이름으로 검색:** 생성된 데이터 스냅샷을 이름으로 검색합니다.
- ③ **데이터 개수:** 현재 목록에 조회된 데이터 스냅샷의 개수를 나타냅니다.
- ④ **데이터 스냅샷 목록:** 설정한 선별 조건에 맞는 데이터 스냅샷들을 보여줍니다. 이 중 하나를 클릭하면 상세 내역을 볼 수 있습니다. (3.1.2절 참조)
- ⑤ **삭제:** 해당 데이터 스냅샷에 마우스 오버 시 휴지통 아이콘이 나타납니다. 클릭하면 해당 데이터 스냅샷을 삭제할 수 있습니다.

3.3.2 데이터 스냅샷 상세 정보

데이터 스냅샷 관리 화면에 나열된 스냅샷 중 하나를 누르면 다음과 같이 해당 데이터 스냅샷의 Hive 테이블 이름, 라인수, 차지하는 용량 등의 정보를 볼 수 있고, 실제 내용도 확인해볼 수 있습니다.

The screenshot shows the METATRON DISCOVERY interface with a data snapshot named 'credit_card [W]'. The main area displays a summary bar with 'Valid' at 100%, 'Mismatched' at 0%, and 'Missing' at 0%. Below this is a preview table with 7 rows and columns labeled ab_time, ab_v1, ab_v2, ab_v3, ab_v4, and ab_v5. The sidebar provides detailed metadata:

- 데이터베이스:** default
- 테이블:** snapshot_kb
- 요약:** 99,999rows, 32 Columns
- 경과 시간:** 00:00:10.00
- 생성일:** 2018-03-23 15:15:32
- Dataflow:** dataflow_KB
- Dataset:** credit_card [W]
- Origin Imported dataset:** credit_card

- ① **이름:** 해당 스냅샷이 사용한 데이터셋입니다.
- ② **성공율:** Hive 테이블에 성공적으로 저장된 레코드의 비율
 - Valid: 모든 컬럼 형식이 일치하여 성공적으로 저장된 레코드의 비율
 - Mismatched: 컬럼 형식이 불일치하여 저장에 실패한 레코드의 비율
 - Missing: 누락된 레코드의 비율
- ③ **스냅샷 데이터 내역:** 해당 스냅샷에 포함된 데이터 내역을 보여줍니다.
- ④ **데이터베이스:** 해당 스냅샷이 저장된 Hive 데이터베이스의 이름입니다.
- ⑤ **테이블:** 해당 스냅샷이 저장된 Hive 테이블의 이름입니다.
- ⑥ **요약:** 해당 스냅샷에 포함된 데이터의 열(레코드) 수와 컬럼 수를 나타냅니다.
- ⑦ **경과 시간:** 데이터를 포함하는 데 소요된 시간을 나타냅니다.
- ⑧ **생성일:** 해당 스냅샷이 생성된 시간을 나타냅니다.
- ⑨ **Dataflow:** 해당 스냅샷이 어느 데이터플로우를 거쳐 생성되었는지 나타냅니다. 클릭하면 데이터플로우를 조회·수정할 수 있습니다.
- ⑩ **Dataset:** 해당 스냅샷이 어느 Wrangled 데이터셋으로부터 생성되었는지 나타냅니다. 클릭하면 데이터셋을 조회·수정할 수 있습니다.
- ⑪ **Origin imported dataset:** 해당 스냅샷의 어느 Imported 데이터셋(원천 데이터)에서 비롯되었는지 나타냅니다.

4. 노트북 관리

metatron Discovery는 노트북이라는 모듈을 통해 Jupyter, Zeppelin 등의 외부 분석 도구를 활용할 수 있는 기능을 지원합니다. 이러한 기능을 활용하기 위해서는 먼저 사용하고자 하는 외부 분석 도구가 설치되어 있는 서버와의 연동이 필요한데, 이러한 서버를 '노트북 서버'라고 합니다.

4.1 노트북 서버

이 메뉴에서는 노트북 서버를 새로 등록하거나 기존에 등록된 노트북 서버를 조회·수정할 수 있습니다.

4.1.1 노트북 서버 목록 조회

현재 사용 가능한 노트북 서버 목록을 확인할 수 있습니다. 노트북 서버를 추가하거나 접속 정보 변경 혹은 삭제할 수 있습니다.

Notebook Management

Notebook Server

선택	Type	Host	Port	Creator (ID)	Last updated
<input type="checkbox"/>	zeppelin	zeppelin.mcloud.sktelecom.com	80	Administrator (admin)	2017-11-23 by Administrator
<input type="checkbox"/>	jupyter	jupyter.mcloud.sktelecom.com	80	Administrator (admin)	2017-11-22 by Administrator
<input type="checkbox"/>	zeppelin	test DS host	9000	Administrator (admin)	2017-11-21 by Administrator
<input type="checkbox"/>	zeppelin	52.231.201.148	8080	Administrator (admin)	2017-11-21 by Administrator
<input type="checkbox"/>	jupyter	52.231.201.148	8888	Administrator (admin)	2017-11-21 by Administrator

1 타입: ALL
2 Search by server name
3 5개 데이터가 있습니다.
4 Add a server
5 Delete selections
6 노트북 서버 목록

1 타입: 등록된 노트북 서버를 연동된 외부 분석 도구(Jupyter/Zeppelin)별로 선별하여 조회합니다.

- 모두: 외부 분석 도구 종류와 관계 없이 모든 노트북 서버가 출력됩니다.
- Jupyter: Jupyter와 연동된 노트북 서버들이 출력됩니다.
- Zeppelin: Zeppelin과 연동된 노트북 서버들이 출력됩니다.

2 Search by server name: 현재 등록된 노트북 서버를 이름으로 검색합니다.

3 데이터 개수: 현재 목록에 조회된 데이터 소스의 개수를 나타냅니다.

4 Add a server: 클릭하면 새로운 노트북 서버를 생성할 수 있습니다.

5 Delete selections: 노트북 서버 목록에서 좌측 확인란에 체크된 노트북 서버를 삭제합니다.

6 노트북 서버 목록: 설정한 선별 조건에 맞는 노트북 서버들을 보여줍니다. 이 중 하나를 클릭하면 서버 설정을 변경할 수 있습니다.

4.1.2 노트북 서버 새로 등록하기

노트북 관리 화면에서 'Add a server' 버튼을 클릭하면 다음과 같은 노트북 서버 등록 화면이 팝업이 됩니다.

Add a notebook server

1 Type: jupyter

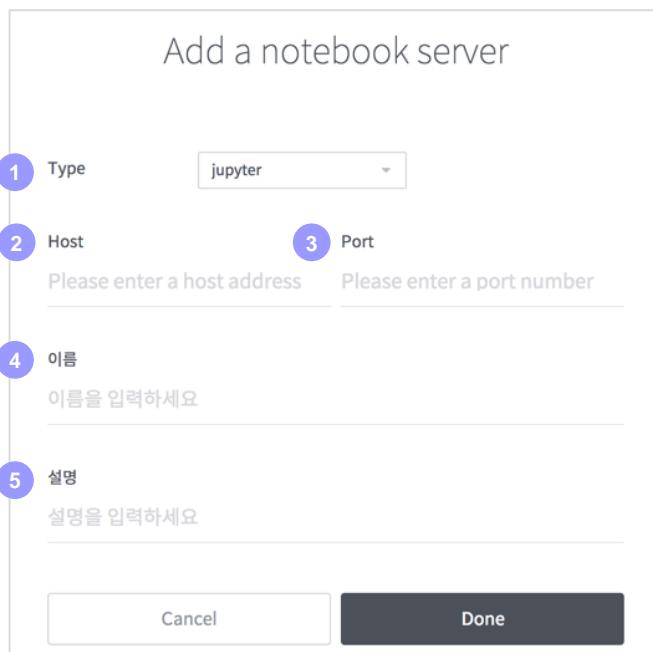
2 Host: Please enter a host address

3 Port: Please enter a port number

4 이름: 이름을 입력하세요

5 설명: 설명을 입력하세요

Cancel Done



- ① **Type:** 등록하려는 노트북 서버에 설치된 외부 분석 도구를 선택합니다.
- ② **Host:** 등록하려는 노트북 서버의 호스트 값을 입력합니다.
- ③ **Port:** 등록하려는 노트북 서버의 포트 번호를 입력합니다.
- ④ **이름:** 등록하려는 노트북 서버의 이름을 입력합니다.
- ⑤ **설명:** 등록하려는 노트북 서버에 대한 설명을 입력합니다.

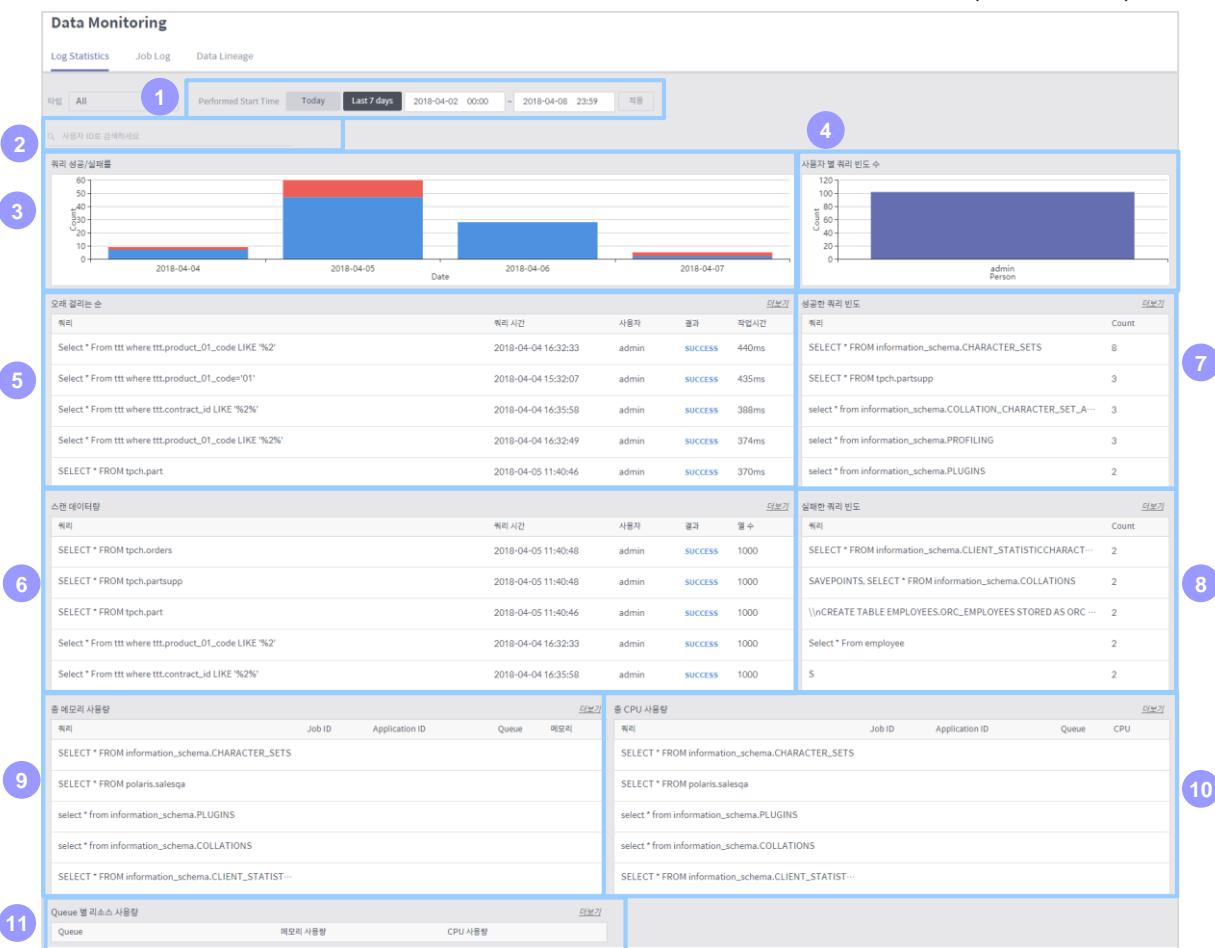
5. 데이터 모니터링 관리

데이터 모니터링은 metatron의 Staging DB(내부 Hive DB)와 워크벤치(외부 DB)에서 발생하는 각종 데이터 로그를 관측하는 기능입니다.

5.1 로그 분석

5.1.1 로그 분석 홈 화면

이 화면에서는 metatron Discovery 내 쿼리 수행과 관련된 각종 통계치를 보여줍니다. 여기에 조회된 쿼리 중 하나를 클릭하면 해당 쿼리에 대한 상세 보기 화면으로 넘어갑니다. (5.2.2절 참조)



- 1 Performed Start Time:** 쿼리 분석에 포함시킬 데이터 로그 기간을 입력합니다.
- 2 사용자 ID 검색:** 쿼리를 수행한 사용자 ID로 검색이 가능합니다.
- 3 쿼리 성공/실패율:** metatron에서 실행된 쿼리들의 성공률과 실패율이 나타납니다.
- 4 사용자 별 쿼리 빈도 수:** 쿼리를 수행한 사용자별 빈도수를 나타낸 그래프입니다. 표시된 막대 중 하나를 클릭하면 해당 사용자가 실행한 Job Log를 볼 수 있습니다.
- 5 오래 걸리는 순:** 수행한 쿼리들이 작업시간이 긴 순서대로 정렬되어 나타납니다.
- 6 스캔 데이터량:** 수행한 쿼리들이 데이터를 제일 많이 스캔한 순서대로 정렬되어 나타납니다.

- 7 성공한 쿼리 빈도:** 수행한 쿼리들이 성공한 빈도가 높은 순서대로 정렬되어 나타납니다.
- 8 실패한 쿼리 빈도:** 수행한 쿼리들이 실패한 빈도가 높은 순서대로 정렬되어 나타납니다.
- 9 총 메모리 사용량:** 수행한 쿼리들이 총 메모리 사용량이 큰 순서대로 정렬되어 나타납니다.
- 10 총 CPU 사용량:** 수행한 쿼리들이 총 CPU 사용량이 큰 순서대로 정렬되어 나타납니다.
- 11 Queue별 리소스 사용량:** Hadoop 환경의 각 YARN queue에서 소모되는 리소스량을 보여줍니다.

5.2 잡 로그

본 메뉴에서는 metatron에서 수행된 모든 쿼리의 내역을 조회할 수 있습니다.

5.2.1 잡 로그 홍 화면

이 화면에서는 쿼리 이력을 원하는 조건대로 검색할 수 있습니다.

The screenshot shows the 'Job Log' tab of the Data Monitoring interface. At the top, there are three tabs: 'Log Statistics', 'Job Log' (which is selected), and 'Data Lineage'. Below the tabs are several filter buttons: '타입 All' (Type All), '실패 All Success Fail' (Failure All Success Fail), 'Limited Elapsed time 전체 Over 10 ms Over 30 ms Over 60 ms Base elaps 적용' (Limited Elapsed time All Over 10 ms Over 30 ms Over 60 ms Base elaps Apply), 'Performed Start Time 전체 Today Last 7 days yyyy-MM-dd hhmm yyyy-MM-dd hhmm 적용' (Performed Start Time All Today Last 7 days yyyy-MM-dd hhmm yyyy-MM-dd hhmm Apply), and a search bar 'Q Job 또는 어플리케이션으로 검색하세요' (Search for Job or Application). To the right of the search bar is a blue circle labeled '5' with the text '337개 데이터가 있습니다' (337 items available). The main area displays a table of query logs with columns: 상태 (Status), Job name, Job ID, Application ID, Queue, 사용자 (ID) (User (ID)), 시작시간 (Start Time), and 작업시간 (Work Time). The table contains several rows of data, each with a status indicator (FAIL, SUCCESS), a job name, and timestamp information.

- ① **상태:** 수행된 쿼리들을 성패 기준으로 선별하여 조회합니다.
- ② **Limited Elapsed time:** 수행 시간이 오래 소요된 쿼리들을 선별하여 조회합니다. 기준 시간은 원하는 대로 설정 가능합니다.
- ③ **Performed Start Time:** 쿼리 조회 시 적용되는 시간 기준입니다. 여기서의 시간은 각 쿼리가 수행을 시작하는 시간을 기준으로 합니다.
- ④ **Job 또는 어플리케이션으로 검색:** 현재 이력으로 남은 쿼리들을 쿼리문 또는 Application ID로 검색합니다.
- ⑤ **데이터 개수:** 현재 목록에 조회된 쿼리의 개수를 나타냅니다.
- ⑥ **Job 목록:** 설정한 선별 조건에 맞는 쿼리들을 보여줍니다. 이 중 하나를 클릭하면 상세 내역을 볼 수 있습니다. (5.2.2절 참조)

5.2.2 잡 로그 상세 정보

잡 로그 홈에 열거된 쿼리 중 하나를 클릭하면 해당 쿼리에 관한 다양한 정보와 이력을 확인할 수 있습니다.

The screenshot shows the Metatron Management interface with the following details:

- Job Information:** (Select * From ttt where ttt.product_01_code LIKE '%2')
 - 1 상태:** SUCCESS
 - 2 Job name:** Select * From ttt where ttt.product_01_code LIKE '%2'
 - 3 시작시간:** 2018-04-04 16:32:33
 - 4 작업시간:** 440ms
 - 5 User:** admin
- 6 Connection:** (highlighted)
 - 타입: HIVE
 - Host: metatron-poc-h03
 - Port: 10000
 - JDBC URL: jdbc:hive2://metatron-poc-h03:10000/default
- 7 Query History:** (highlighted)

Query date	User	작업시간	Result	Action
2018-04-04 16:35:58	Administrator	388 ms	SUCCESS	Detail >
2018-04-04 16:32:49	Administrator	374 ms	SUCCESS	Detail >
2018-04-04 16:32:33	Administrator	440 ms	SUCCESS	Detail >
2018-04-04 15:32:07	Administrator	435 ms	SUCCESS	Detail >
2018-04-04 15:27:10	Administrator	357 ms	SUCCESS	Detail >
- 8 Plan:** (highlighted) Run Plan Query

- 1 상태:** 해당 쿼리의 성공 여부를 나타냅니다.
- 2 Job name:** 수행된 쿼리문입니다.
- 3 시작시간:** 해당 쿼리가 수행되기 시작한 시간을 나타냅니다.
- 4 작업시간:** 해당 쿼리가 수행되는 데 걸린 시간을 나타냅니다.
- 5 User:** 해당 쿼리를 수행한 사용자 ID입니다.
- 6 Connection:** 워크벤치에서 실행된 쿼리일 경우, 대상 데이터 커넥션의 정보를 나타냅니다.
- 7 Query History:** 워크벤치에서 실행된 쿼리일 경우, 해당 데이터베이스에서 수행된 최근 5건의 쿼리 내역과 그 결과가 나타납니다. Detail을 클릭하면 해당 쿼리문이 새 창에 출력됩니다.
- 8 Plan:** 쿼리 수행 계획을 실행합니다.

5.3 데이터 리니지

데이터 리니지에서는 metatron의 내부 Hive 데이터베이스에서 실행된 Hive 로그를 분석하여 데이터 ETL(추출, 변환, 적재)이 어떠한 흐름으로 이루어지는지를 확인할 수 있습니다. 이와 같은 워크플로우는 쿼리문을 통해 실행되기 때문에 데이터 리니지의 추적 기능은 쿼리 로그를 기반으로 합니다.

5.3.1 데이터 리니지 관리 홈 화면

The screenshot shows the 'Data Monitoring' section of the Metatron Discovery interface. At the top, there are tabs for 'Log Statistics', 'Job Log', and 'Data Lineage'. A search bar labeled '리니지 키워드를 검색하세요.' is highlighted with a blue box and a circled '1'. Below the search bar is a filter section with radio buttons for '테이블' (selected), '컬럼', 'SQL', and '워크플로우'. A third numbered callout '3' points to a table titled 'Table (7)' which lists data lineage entries:

데이터베이스	테이블
default	sales
default	sales_category
default	sales_joined
default	sales_joined3
default	sales_region
default	test
default	test333

- ① **검색:** 검색어를 입력하여 데이터를 검색할 수 있습니다. 검색어는 우측에서 선택하는 엔터티 타입에 따라 달라집니다.
- ② **엔터티 타입:** 어떠한 엔터티 타입을 기준으로 워크플로우를 추적할 것인지 선택합니다.
 - 테이블: 실행된 워크플로우에 관여된 테이블들을 보여줍니다.
 - 컬럼: 실행된 워크플로우에 관여된 컬럼들을 보여줍니다.
 - SQL: 실행된 워크플로우에 관여된 SQL 쿼리문을 보여줍니다.
 - 워크플로우: 실행된 워크플로우들을 보여줍니다.
- ③ **엔터티 목록:** 설정한 선별 조건에 맞는 엔터티들을 보여줍니다. 이 중 하나를 클릭하면 상세 내역을 볼 수 있습니다. (5.3.2절 참조)

5.3.2 데이터 리니지 상세 정보

데이터 리니지 관리 화면에 열거된 엔티티 중 하나를 클릭하면 해당 엔티티와 연계된 워크플로우 정보를 확인할 수 있습니다.

The screenshot shows the Metatron Discovery interface with the following components:

- Left Panel (1): 원천 데이터 테이블 상자**: Displays tables from the 'default' database: sales, sales_category, and sales_region. The 'sales' table has columns cal_orderdate, ab_category, ab_region, and ab_orderid. The 'sales_category' table has ab_categoryname and ab_category. The 'sales_region' table has ab_regionname and ab_region.
- Middle Panel (2): SQL 쿼리 목록 상자**: Shows a query: create table sales_joined AS select t1.OrderDate, t1.cate... with conditions (t1.category = t2.category) and (t1._col9 = t3.region). It also shows a partial query: ((t1.category = t2.categ...).
- Middle Panel (3): 결과 데이터 테이블 상자**: Displays the joined table sales_joined with columns cal_orderdate, ab_category, ab_region, ab_orderid, ab_categoryname, and ab_regionname.
- Right Panel (4): 상세 보기 팝업**: Provides detailed information for the table 'default.sales_joined':

타입	TABLE
데이터베이스	default
테이블	sales_joined
컬럼 보기	
ab_orderdate	
ab_category	
ab_city	
ab_country	
ab_customername	
ab_orderid	
ab_postalcode	
ab_productname	
Detail	
Database:	default
Owner:	sohnw
CreateTime:	Mon Jul 24 09:51:17 KST 2017
LastAccessTime:	UNKNOWN
Protect Mode:	None
Retention:	0
Location:	hdfs://localhost:9000/user/hive/warehouse/sales
Table Type:	MANAGED_TABLE
COLUMN_STATS_ACCURATE	false
comment	123
last_modified_by	sohnw
last_modified_time	1508289564
numFiles	1
numRows	-1
rawDataSize	-1
totalSize	2731122
transient_lastDdlTime	1508289564
SerDe Library:	org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
- Bottom Center (5): 미니맵**: A small map icon for navigating the interface.

- ① **원천 데이터 테이블 상자**: 해당 워크플로우 내 각 쿼리를 수행하는 데 사용된 원천 데이터가 담긴 테이블의 이름이 기본적으로 표시되고, ▾ 버튼을 클릭하면 쿼리문에 언급된 컬럼들이 조회됩니다. 상자를 선택하면 해당 테이블에 관한 정보가 조회되고, 컬럼을 선택하면 그와 연계된 쿼리와 다른 테이블의 컬럼들이 색상 표시됩니다.
- ② **SQL 쿼리 목록 상자**: 해당 워크플로우를 위해 실행된 SQL 쿼리문이 기본적으로 표시되고, ▾ 버튼을 클릭하면 원천 데이터 컬럼과 결과 데이터 컬럼 간 관계를 정의하는 where 절 내용이 컬럼별로 조회됩니다. 상자를 선택하면 해당 SQL에 관한 정보가 조회되고, 컬럼을 선택하면 그와 연계된 컬럼들이 색상 표시됩니다.
- ③ **결과 데이터 테이블 상자**: 워크플로우 쿼리들을 거친 결과 데이터가 담긴 테이블의 이름이 기본적으로 표시되고, ▾ 버튼을 클릭하면 쿼리문에 언급된 컬럼들이 조회됩니다. 상자를 선택하면 해당 테이블에 관한 정보가 조회되고, 컬럼을 선택하면 그와 연계된 쿼리와 다른 테이블의 컬럼들이 색상 표시됩니다.
- ④ **상세 보기**: 워크플로우 상에서 상자를 하나 클릭하면 해당 테이블 또는 SQL에 대한 자세한 정보가 화면 우측에 표시됩니다. 테이블 상자를 선택하면 해당 테이블에 관한 메타 정보와 소속 컬럼들이 조회되고, SQL 상자를 선택하면 쿼리문 정보가 조회됩니다.
- ⑤ **미니맵**: 화면을 확대/축소/이동할 수 있습니다.



metatron 3.0

3부

워크스페이스 이용하기



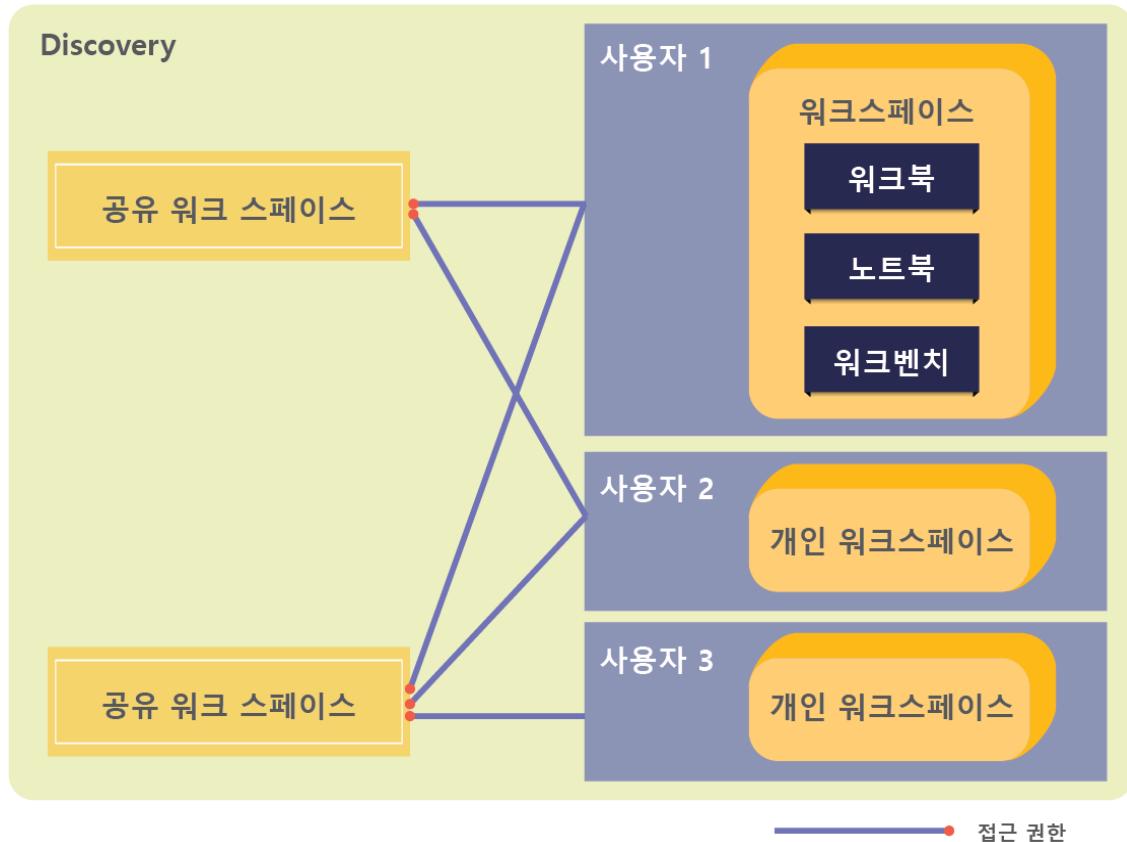
Copyright © SK telecom

SK Telecom 의 사전 승인 없이
본 내용의 전부 또는 일부에 대한 복사, 배포, 사용을 금합니다.

목 차

1. 워크스페이스 개요	3
2. 워크스페이스 관리 홈	4
2.1 워크스페이스 관리 홈 화면 구성	4
2.2 폴더 항목	6
2.3 개체 항목	6
2.4 폴더 및 개체 복사/이동/삭제	7
3. 공유 워크스페이스 리스트	8
4. 공유 워크스페이스 생성	9
5. 공유 워크스페이스 접근 권한 설정	10
5.1 권한 스키마	10
5.1.1 권한 스키마 조회하기	10
5.1.2 권한 스키마 설정 바꾸기	11
5.2 공유 회원 및 그룹 설정	12

1. 워크스페이스 개요



워크스페이스는 metatron Discovery의 분석 모듈인 워크북, 노트북, 워크벤치를 보관하는 작업 공간입니다. 워크스페이스는 개인 워크스페이스와 공유 스페이스로 분류됩니다.

- 개인 워크스페이스:** Discovery 회원별로 하나씩 할당되는 개인용 워크스페이스입니다. 이 워크스페이스는 본인만 접근 가능합니다.
- 공유 워크스페이스:** 여러 사용자가 함께 사용할 수 있는 워크스페이스입니다. 분석 과정과 결과를 다른 사용자들과 공유하기 위한 공간입니다. 각 공유 워크스페이스의 소유자 또는 관리자는 Discovery 회원들에게 다양한 수준의 접근 권한을 부여할 수 있습니다.

본 단원은 다음과 같이 구성되어 있습니다.

2장 워크스페이스 관리 홀: 워크스페이스 홈 화면 구성과 UI에 대해 설명합니다.

3장 공유 워크스페이스 리스트: '공유 워크스페이스 리스트'는 해당 사용자가 접근 가능한 공유 워크스페이스들을 나열한 페이지입니다. 여기에서는 원하는 워크리스트를 선별해서 조회할 수 있습니다.

4장 공유 워크스페이스 생성: 새로운 공유 워크스페이스를 생성하는 절차를 설명합니다.

5장 공유 워크스페이스 접근 권한 설정: 공유 워크스페이스에서 여러 사용자가 협업하는 데는 사용자별 다양한 역할과 권한을 필요로 합니다. 여기서는 역할별 권한 수준을 설정하고 이렇게 설정한 다양한 권한 수준을 각 사용자에게 할당할 수 있습니다.

2. 워크스페이스 관리 홈

본 화면에서는 metatron Discovery의 분석 모듈(워크북, 노트북, 워크벤치)의 관리 기능을 수행할 수 있습니다.

2.1 워크스페이스 관리 홈 화면 구성

아래는 워크스페이스 관리 홈의 전반적인 구성을 설명한 것입니다.

The screenshot shows the 'METATRON DISCOVERY' workspace management home page. The interface includes a top navigation bar with a main menu icon (1), workspace name 'Admin workspace' (2), ownership information (3), and a 'Workspace List' button (5). Below this is a search bar (10) and a toolbar with various icons. The main content area displays a grid of workspace items (11) and detailed views of specific workbooks (12). At the bottom are buttons for selecting all items (13) and adding new workbooks (14).

- ① **메인 메뉴 버튼:** 해당 버튼을 클릭하면 다른 워크스페이스로 접근할 수 있는 패널이 열립니다.
- ② **워크스페이스 정보:** 해당 워크스페이스의 이름과 설명을 보여줍니다. 로그인한 사용자가 소유하는 워크스페이스일 경우, 워크스페이스 이름 옆에 '소유자'라는 아이콘이 보여집니다.
- ③ **등록 개체 현황:** 워크스페이스내에 등록된 개체 타입별 개수를 보여줍니다.
- ④ **데이터 소스:** 해당 워크스페이스에서 사용 중인 데이터 소스의 개수를 보여주며, 이 영역을 클릭하면 그러한 데이터 소스들의 목록이 나타납니다.
- ⑤ **Workspace List:** 이 버튼을 클릭하면 공유 워크스페이스의 목록이 나타납니다. (자세한 내용은 3장 참조)
- ⑥ **생성 정보:** 해당 워크스페이스의 생성일과 만든 사용자 이름을 보여줍니다.

7 더 보기: 해당 워크스페이스의 생성일과 만든 사용자 이름을 보여줍니다.

- 이름 및 설명 수정: 해당 워크스페이스의 이름과 설명을 수정합니다.
- 공유 회원 및 그룹 설정: 해당 워크스페이스에 접근할 수 있는 사용자와 그룹을 지정합니다. (자세한 사항은 5.2절 참조)
- 노트북 서버 설정: 노트북 모듈에 사용되는 외부 분석 도구 서버 접근 정보를 설정합니다.
- 권한 스키마 설정: 해당 워크스페이스에서의 사용자 역할별 접근 권한을 설정합니다. (자세한 사항은 5.1절 참조)
- 소유자 변경: 해당 워크스페이스의 소유자를 바꿉니다.
- 워크스페이스 삭제: 해당 워크스페이스를 삭제합니다.

8 워크스페이스 경로: 워크스페이스 내에서의 현재 위치를 확인합니다. 경로에 나열된 상위 폴더 중 하나를 클릭하면 해당 폴더로 이동합니다.

9 폴더 생성: 클릭하면 현재 위치에서 새 폴더가 생성됩니다.

10 개체 목록 선별/정렬:

- 검색: 해당 워크스페이스 내에서 개체 또는 폴더를 이름으로 검색합니다.
- 개체 탑입 콤보박스: 워크북, 노트북, 워크벤치 중 원하는 개체 탑입만을 선별해서 조회합니다.
- 정렬 순서: 폴더 및 개체를 이름 또는 업데이트 시간 순서로 정렬합니다.
- 뷰 형식: 워크스페이스 내 개체들의 열거 형식을 그리드 뷰와 리스트 뷰 중에서 선택합니다.

11 폴더 목록: 현재 위치에서 검색 조건에 부합하는 폴더들을 보여줍니다. 이중 하나를 클릭하면 해당 폴더 안으로 이동합니다. (개별 폴더 항목에 대한 자세한 설명은 2.2절 참조)

12 개체 목록: 현재 위치에서 검색 또는 선별 조건에 부합하는 개체들을 보여줍니다. 이 중 하나를 클릭하면 해당 개체의 홈 화면으로 이동합니다. (개별 개체 항목에 대한 자세한 설명은 2.3절 참조)

13 개체 선택/복사/이동/삭제: 개체 전체 선택, 복사, 이동, 삭제를 합니다. (자세한 내용은 2.4절 참조)

14 개체 생성: 해당 워크스페이스에서 원하는 탑입의 개체를 만드는 데 사용되는 버튼들입니다. (구체적인 절차는 각각 제4부 2장, 제5부 3장, 제6부 2장 참조)

2.2 폴더 항목

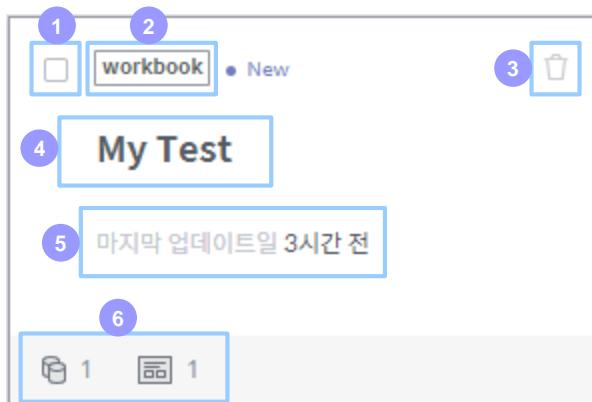
각 폴더 항목은 다음과 같이 표시됩니다.



- ① **확인란:** 해당 폴더를 선택할 때 사용됩니다. 선택한 폴더는 복제, 이동, 삭제할 수 있습니다.
- ② **이름:** 해당 폴더의 이름입니다.
- ③ **수정:** 클릭하면 폴더 이름을 수정할 수 있습니다. 이 버튼은 해당 폴더 항목에 마우스를 오버할 때만 나타납니다.
- ④ **삭제:** 클릭하면 해당 폴더가 삭제됩니다. 이 버튼은 해당 폴더 항목에 마우스를 오버할 때만 나타납니다.

2.3 개체 항목

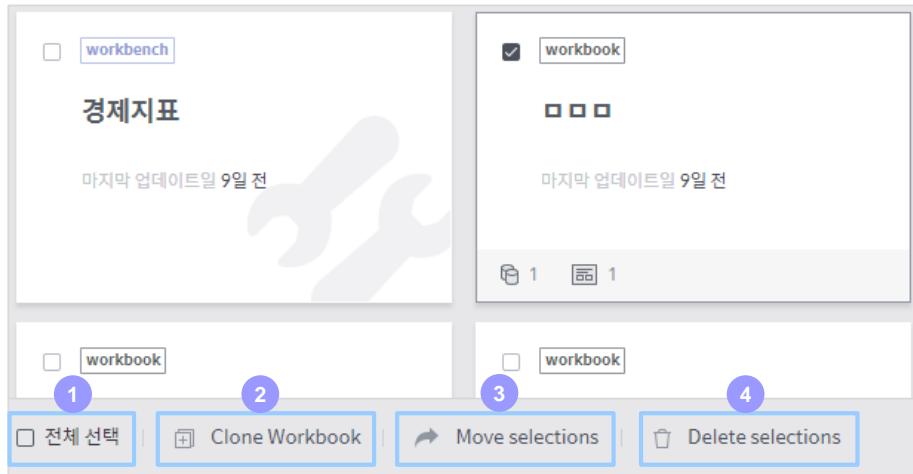
각 개체 항목은 다음과 같이 표시됩니다.



- ① **확인란:** 해당 개체를 선택할 때 사용됩니다. 선택한 개체는 복제, 이동, 삭제할 수 있습니다.
- ② **개체 타입:** 해당 개체가 워크북/노트북/워크벤치 중에서 어느 타입인지 보여줍니다.
- ③ **삭제:** 클릭하면 해당 개체가 삭제됩니다. 이 버튼은 해당 개체 항목에 마우스를 오버할 때만 나타납니다.
- ④ **이름:** 해당 개체의 이름입니다.
- ⑤ **업데이트 시점:** 해당 개체가 마지막으로 업데이트된 시점을 보여줍니다.
- ⑥ **데이터 소스/대시보드 개수:** 이 영역은 워크북 개체에만 고유합니다.
 - 아이콘 옆 숫자는 해당 워크북에 연동된 데이터 소스의 개수를 가리킵니다.
 - 아이콘 옆 숫자는 해당 워크북에 등록된 대시보드의 개수를 가리킵니다.

2.4 폴더 및 개체 복사/이동/삭제

워크스페이스 내 폴더와 개체는 복사/이동/삭제가 가능합니다. 복사/이동/삭제를 원하는 폴더 또는 개체를 선택하면 워크스페이스 홈 화면 좌측 하단의 동작 버튼들이 활성화됩니다.



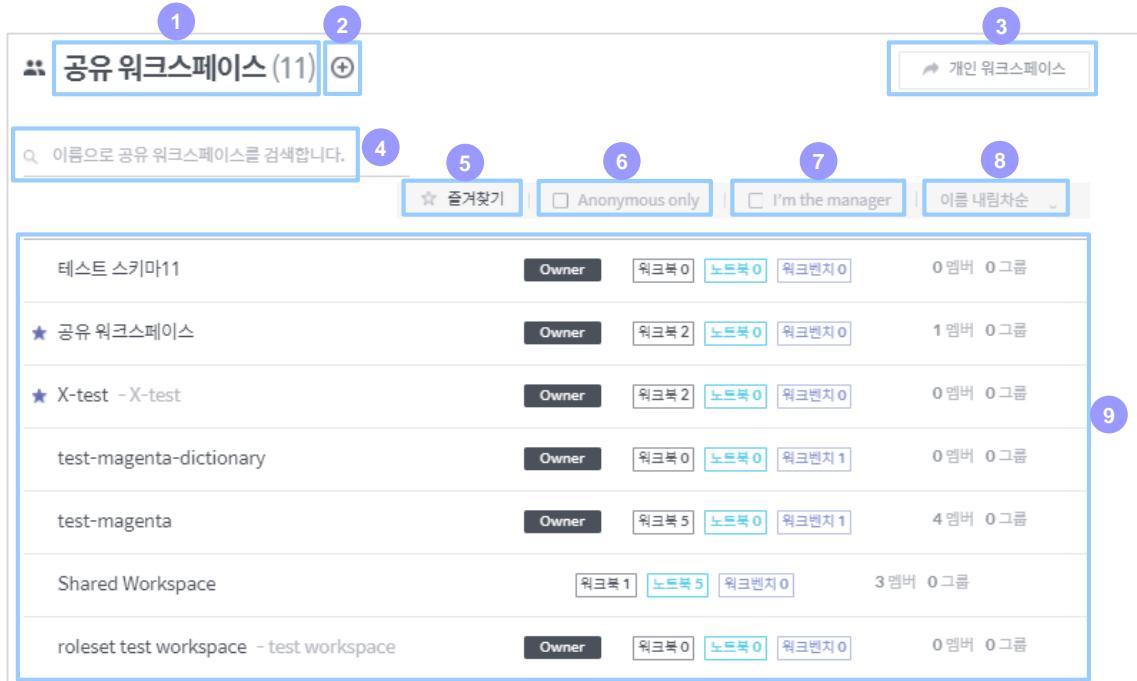
- ① **전체 선택:** 현재 폴더 목록과 개체 목록에 나타난 모든 항목을 선택합니다.
- ② **Clone Workbook:** 워크북에만 유효한 기능입니다. 이 버튼을 클릭하면 선택한 워크북이 복제됩니다.
- ③ **Move selections:** 선택한 폴더 및 개체를 이동합니다. 워크북의 경우에는 다른 워크스페이스로 이동이 가능하고, 그 외의 항목들은 동일 워크스페이스 내 다른 폴더로 이동이 가능합니다. 워크북이 다른 개체와 함께 선택 되어있을 때는 이동을 실행할 수 없습니다.
- ④ **삭제:** 선택한 폴더 및 개체를 삭제합니다.

3. 공유 워크스페이스 리스트

공유 워크스페이스 리스트 화면에서는 로그인한 사용자가 접근할 수 있는 모든 공유 워크스페이스의 목록을 열람하고 원하는 워크스페이스로 이동할 수 있습니다. 이 화면은 다음과 같은 두 가지 방식으로 접근할 수 있습니다.

- Discovery 화면 좌측 상단에서  버튼을 클릭하여 메인 패널을 연 후 'Workspace list >>'를 클릭
- 워크스페이스 홈 화면 우측 상단에서 'Workspace List' 버튼을 클릭

공유 워크스페이스 리스트 화면은 다음과 같이 구성됩니다.



The screenshot shows the 'Shared Workspace List' page. At the top left, there's a header '공유 워크스페이스 (11)' with a plus sign button. To the right is a button labeled '개인 워크스페이스'. Below the header is a search bar with placeholder text '이름으로 공유 워크스페이스를 검색합니다.' and several filter buttons: '즐겨찾기' (with a star icon), 'Anonymous only', 'I'm the manager', and '이름 내림차순'. The main content area displays a list of 11 shared workspaces:

.workspace-name	Owner	워크북	노트북	워크벤치	멤버	그룹
테스트 스키마11	Owner	0	0	0	0	0 그룹
★ 공유 워크스페이스	Owner	2	0	0	1	0 그룹
★ X-test - X-test	Owner	2	0	0	0	0 그룹
test-magenta-dictionary	Owner	0	0	1	0	0 그룹
test-magenta	Owner	5	0	1	4	0 그룹
Shared Workspace		1	5	0	3	0 그룹
roleset test workspace - test workspace	Owner	0	0	0	0	0 그룹

- ① **공유 워크스페이스 개수:** 목록에 나열된 공유 워크스페이스의 개수가 나타납니다.
- ② **공유 워크스페이스 추가:** 이 버튼을 클릭하면 공유 워크스페이스를 추가하는 화면으로 이동합니다. (구체적인 절차는 4장 참조)
- ③ **개인 워크스페이스로 이동:** 이 버튼을 클릭하면 로그인한 사용자의 개인 워크스페이스로 이동합니다.
- ④ **검색:** 공유 워크스페이스를 이름으로 검색합니다.
- ⑤ **즐겨찾기:** 즐겨찾기로 지정된 워크스페이스만 선별합니다.
- ⑥ **Anonymous only:** 전체 공개 설정된 워크스페이스만 선별합니다.
- ⑦ **I'm the manager:** 로그인한 사용자가 관리자로 설정되어 있는 워크스페이스 목록이 나타납니다.
- ⑧ **이름 내림차순/오름차순:** 공유 워크스페이스 이름을 내림차순/오름차순으로 정렬합니다.
- ⑨ **워크스페이스 목록:** 설정한 선별 조건에 맞는 워크스페이스들을 보여줍니다. 이 중 하나를 클릭하면 해당 워크스페이스에 입장합니다.

4. 공유 워크스페이스 생성

새 공유 워크스페이스를 생성하는 절차는 다음과 같습니다.

1. 공유 워크스페이스 목록에서 버튼을 클릭하면 새 공유 워크스페이스를 생성하는 화면이 나타납니다.
2. 아래 설명을 참조하여 입력 화면을 작성합니다.

Create shared workspace

1 이름
이름을 입력하세요

2 설명
설명을 입력하세요

3 Workspace permission schema
 Use a schema preset Load a schema preset Use a schema preset

User roles
Please select a schema

취소 마침

- 1 **이름:** 해당 공유 워크스페이스의 이름을 입력합니다.
- 2 **설명:** 해당 공유 워크스페이스에 대한 설명을 입력합니다.
- 3 **Workspace permission schema:** 해당 공유 워크스페이스의 역할별 권한 스키마를 설정합니다.
 - Use a schema preset: 관리자가 기존에 정의해놓은 권한 스키마를 불러옵니다.
 - Use a custom schema: 새 권한 스키마를 정의합니다. (새 권한 스키마를 정의하는 방식은 5.1.1절을 참조하십시오.)
3. '마침' 버튼을 눌러 워크스페이스 생성을 완료합니다.

5. 공유 워크스페이스 접근 권한 설정

공유 워크스페이스 접근 권한 설정은 기본적으로 다음과 같은 두 단계로 이루어집니다.

- 사용자 역할별 접근 권한 정의 (5.1절 '권한 스키마 설정' 참조)
- 개별 사용자 또는 사용자 그룹 각각에게 적합한 사용자 역할 부여 (5.2절 '공유 회원 및 그룹 설정' 참조)

5.1 권한 스키마

5.1.1 권한 스키마 조회하기

공유 워크스페이스 홈 화면 우측 상단에 있는 아이콘을 클릭한 후 '권한 스키마 설정'을 누르면 아래와 같이 현재 정의된 권한 스키마를 보여줍니다.

The screenshot shows a table titled 'User roles of 공유 워크스페이스'. The columns represent different workspace components: Workbook, Notebook, Workbench, and Workspace. The rows list three user roles: Manager, Editor, and Watcher. Each row has a 'Default role' column where a blue dot indicates the current default role for each user. The permissions are indicated by checkmarks or dashes in the respective columns.

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		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor	●	✓	✓	-	✓	✓	-	✓	✓	-	✓	-
Watcher		✓	-	-	✓	-	-	✓	-	-	-	-

위 예시 그림에서는 사용자 역할(user role)로서 Manager, Editor, Watcher가 정의되어 있습니다. '권한 스키마'란, 이와 같이 각각 고유한 접근 권한이 정의된 사용자 역할들의 집합을 일컫는 말입니다.

각 사용자 역할에 대한 컬럼별 속성은 다음과 같습니다.

Default role

새롭게 추가되는 개별 사용자나 사용자 그룹에게는 default role로 지정된 사용자 역할이 기본적으로 부여됩니다.

워크북/노트북/워크벤치 개체 타입별 권한

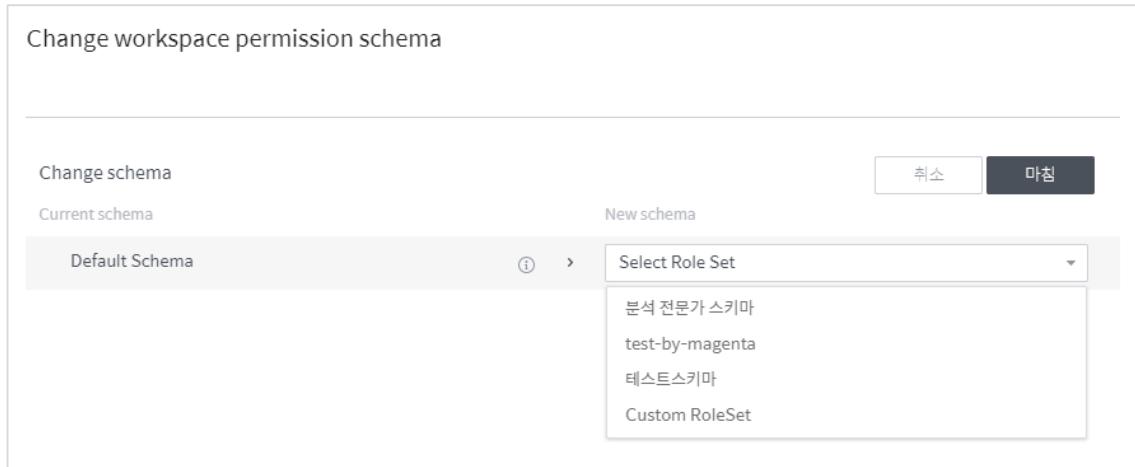
- View: 해당 타입의 개체에 접근하여 데이터를 열람할 수 있음
- Create: 해당 타입의 개체를 생성, 수정, 삭제할 수 있음
- Edit any: 다른 사용자가 생성한 해당 타입의 개체를 수정, 삭제할 수 있음

워크스페이스 권한

- Create folders: 워크스페이스 내 폴더를 생성, 수정, 삭제할 수 있음
- Set config.: 워크스페이스의 이름과 설명을 수정하고 워크스페이스 권한 스키마를 바꿀 수 있음.

5.1.2 권한 스키마 설정 바꾸기

권한 스키마 조회 화면에서 'Change schema' 버튼을 클릭하면 아래와 같이 기존에 정의된 권한 스키마를 변경할 수 있는 화면이 나타납니다.



화면 우측에 있는 'Select Role Set' 콤보박스를 클릭하면 관리자가 정의해놓은 스키마들이 제시되고, 맨 목록 맨 아래에는 새로운 사용자 역할을 정의할 수 있는 'Custom RoleSet' 항목이 있습니다. 이중 하나를 선택하면 아래와 같은 화면이 나타납니다. ('Custom RoleSet' 항목을 선택했을 경우 사용자 역할별 권한부터 정의해야 합니다. 수정 버튼을 눌러 권한 설정 화면으로 이동한 후 5.1.1 절의 설명을 참조하여 사용자 역할별로 권한을 설정하십시오.)

Change workspace permission schema

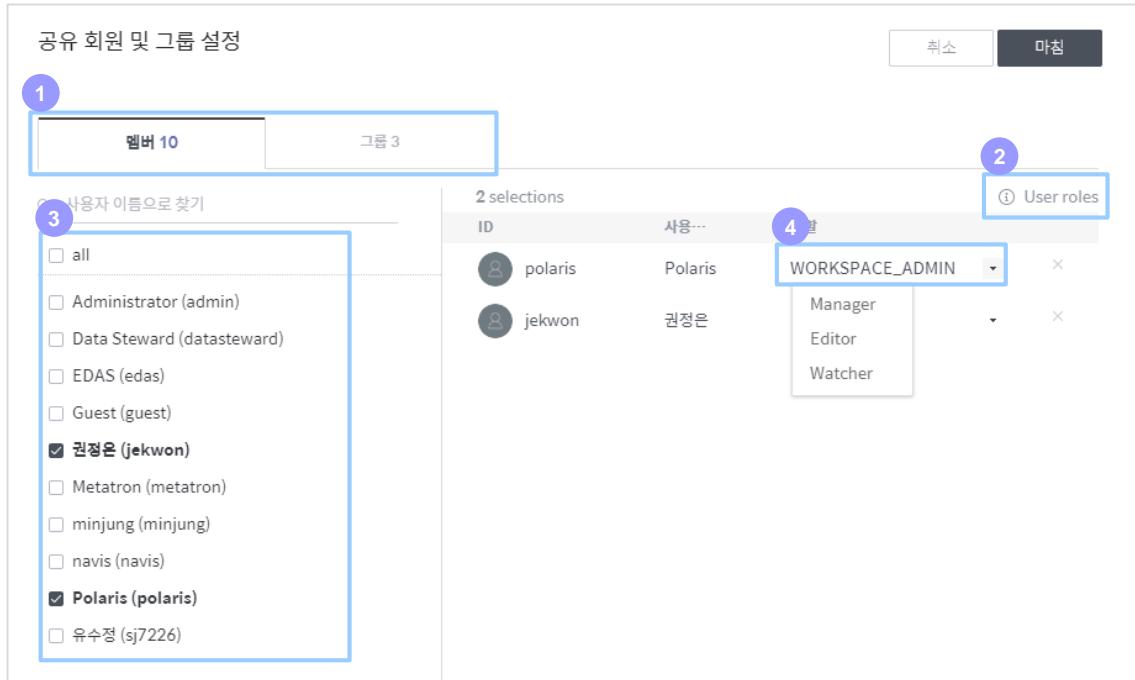
현재의 권한 스키마	새로운 권한 스키마
Current schema	New schema
Default Schema	test-by-magenta

여기서는 현 권한 스키마의 각 사용자 역할을 새로운 권한 스키마에 정의된 사용자 역할로 치환하는 작업을 합니다. 각 사용자 역할 이름 옆에 있는 아이콘에 마우스를 오버하면 해당 사용자 역할에 할당된 권한이 나타납니다.

'마침'을 누르면 권한 스키마 설정이 완료됩니다.

5.2 공유 회원 및 그룹 설정

공유 워크스페이스 홈 화면 우측 상단에 있는 아이콘을 클릭한 후 '공유 회원 및 그룹 설정'을 누르면 아래와 같이 공유 회원 및 그룹 설정 화면이 나타납니다. 여기서는 권한 스키마에서 정의된 각 사용자 역할을 개별 사용자 또는 사용자 그룹에 할당하는 작업을 합니다. 아래 설명을 참조하여 사용자 역할을 할당한 후 '마침'을 누르면 워크스페이스 접근 권한 설정이 완료됩니다.



1 사용자 역할 할당 단위 선택

- 멤버 탭: 사용자 역할을 개별 사용자 단위로 할당합니다.
- 그룹 탭: 사용자 역할을 사용자 그룹 단위로 할당합니다. (사용자 그룹은 관리자 권한으로 지정할 수 있습니다.)

2 User roles: 클릭하면 권한 스키마(사용자 역할별 권한 정의) 정보가 팝업 창으로 나타납니다.

3 멤버/그룹 목록: Discovery에 등록된 사용자들(그룹 탭에서는 그룹들)이 나열됩니다. 목록에 있는 사용자(그룹) 중 하나를 클릭하면 우측 역할 부여 영역에 추가됩니다. 이미 추가된 사용자(그룹)을 클릭하면 해당 사용자(그룹)가 우측 영역에서 제거됩니다.

4 사용자 역할 할당: 이 콤보박스를 클릭하면 현재 적용 중인 권한 스키마에 정의된 사용자 역할들이 나타납니다. 이중에서 해당 사용자(그룹)에게 할당하고자 하는 역할을 선택하면 됩니다.



metatron 3.0

4부

워크북 이용하기



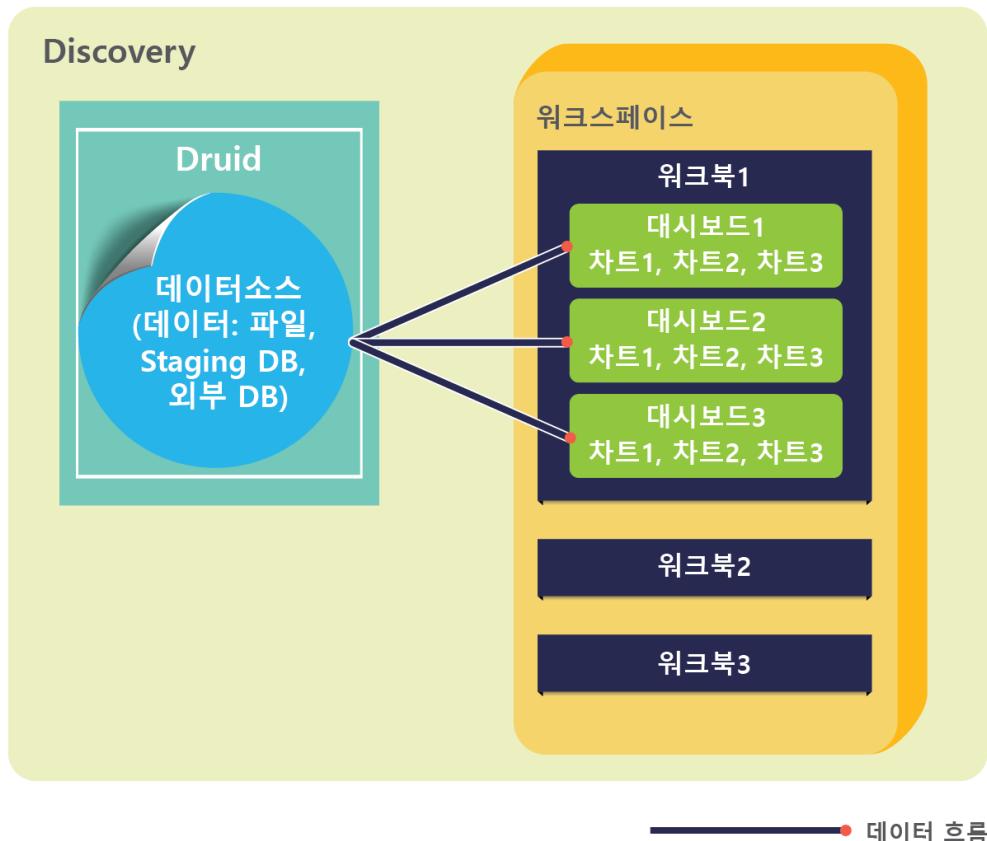
Copyright © SK telecom

SK Telecom 의 사전 승인 없이
본 내용의 전부 또는 일부에 대한 복사, 배포, 사용을 금합니다.

목 차

1. 워크북 개요	3
2. 워크북 생성하기	4
3. 대시보드 생성하기	6
4. 워크북 관리하기	13
4.1 대시보드 리스트.....	14
4.2 대시보드 상세 뷰	16
4.2.1 대시보드 - 기본 화면.....	16
4.2.2 대시보드 - 데이터 소스 정보 대화 상자	19
4.2.3 대시보드 - 프레젠테이션 뷰 화면	21
4.2.4 대시보드 - 편집 화면.....	22
5. 차트 생성·관리.....	30
5.1 차트 홈 화면 개관	30
5.2 데이터 컬럼 목록	31
5.2.1 데이터 컬럼 목록 구성	31
5.2.2 사용자 컬럼 추가	32
5.2.3 차원값과 측정값	33
5.3 피봇팅	34
5.3.1 '피봇팅'이란?	34
5.3.2 열/행/교차 선반의 개념	35
5.4 차트의 종류	37
5.5 차트 필터	40
5.5.1 자동으로 포함되는 필터들.....	40
5.5.2 차트 필터 패널.....	41
5.5.3 차트 필터 대화 상자.....	42
5.6 차트 스타일 설정	47
5.6.1 차트 스타일 설정 메뉴	47
5.6.2 차트 유형별 '공통 설정' 항목	50

1. 워크북 개요



워크북은 metatron Discovery의 Druid 엔진을 기반으로 하는 비주얼 데이터 분석 모듈입니다. 기본적인 특징은 다음과 같습니다.

- 시계열 기반의 다차원 데이터 소스를 이용하여 신속하고 유연한 데이터 분석 가능
- 각 대시보드에서 각종 차트와 텍스트 등의 시각화 위젯을 배치함으로써 프레젠테이션 형식의 보고서로 활용 가능
- 클러스터링, 예측선, 추세선 등의 자주 쓰는 알고리즘을 GUI로 구현 가능

본 단원은 다음과 같이 구성되어 있습니다.

2장 워크북 생성하기: '워크북'은 하나의 독립된 보고서로서 기능을 합니다. 여기서는 새로운 워크북을 생성하는 절차에 대해 설명합니다.

3장 대시보드 생성하기: 대시보드는 워크북을 구성하는 각 슬라이드에 해당합니다. 여기서는 워크북 내에서 새로운 대시보드를 생성하고 그 대시보드에 데이터 소스를 연결하는 절차에 대해 설명합니다.

4장 워크북 관리하기: 워크북을 이루는 대시보드 목록을 정리하고, 각 대시보드에서 각종 위젯을 구성하는 절차에 대해 설명합니다.

5장 차트 생성·관리: 차트도 대시보드를 구성하는 위젯 중 한 유형이지만, 별도로 알아야 할 개념과 다루어야 할 UI 메뉴가 많아서 본 절에서 따로 설명합니다.

2. 워크북 생성하기

metatron Discovery에서 '워크북'은 하나의 독립적인 데이터 분석 보고서로서 기능을 합니다. 하나의 워크북을 생성하면 그 안에 여러 '대시보드' 슬라이드를 담아서 적절한 순서대로 보여줄 수 있습니다.

워크북 생성 절차는 다음과 같습니다.

- 워크스페이스 하단에 있는 '+워크북' 버튼을 클릭하면 워크북을 생성할 수 있는 화면이 나타납니다.



- 생성하고자 하는 워크북의 이름(필수 사항)과 설명을 입력하고 마침 버튼을 누릅니다.
'Continue to create a dashboard of a new workbook' 박스에 체크하면 워크북 생성과 동시에 '대시보드 생성하기' 화면으로 넘어갑니다. 워크북은 그 안에 대시보드가 있어야 비로소 기능을 할 수 있기 때문에 이렇게 연결되는 것입니다.

워크북 생성하기

이름

설명

Continue to create a dashboard of a new workbook

취소 마침

3. 대시보드를 생성하는 절차는 3장을 참조하십시오.

No.	데이터 소스	타입	사용처	수집일
20	캐릭터eee - asdfsdfsdf [오픈 데이터]	수집형	모든 워크스페이스	2018-04-13
19	w.h - w.h [오픈 데이터]	수집형	모든 워크스페이스	2018-04-04
18	SuperstoreKR [오픈 데이터]	수집형	모든 워크스페이스	2018-03-29
17	sales - sales data (2011~2014) [오픈 데이터]	수집형	모든 워크스페이스	2018-03-28
16	o [오픈 데이터]	수집형	모든 워크스페이스	2018-03-21
15	ignoreInvalidRows_sample [오픈 데이터]	수집형	모든 워크스페이스	2018-03-14
14	test_missing - test [오픈 데이터]	수집형	모든 워크스페이스	2018-03-13
13	test_int_dim [오픈 데이터]	수집형	모든 워크스페이스	2018-03-13

Please select a data source

□ 오픈 데이터만 보기 타입 전체 ▾

No. 데이터 소스 타입 사용처 수집일

더보기 ▾

취소 다음

4. 새로 생성된 워크북은 아래와 같이 워크스페이스 화면에서 확인할 수 있습니다. 클릭하면 해당 워크북을 사용할 수 있는 화면이 나타납니다.

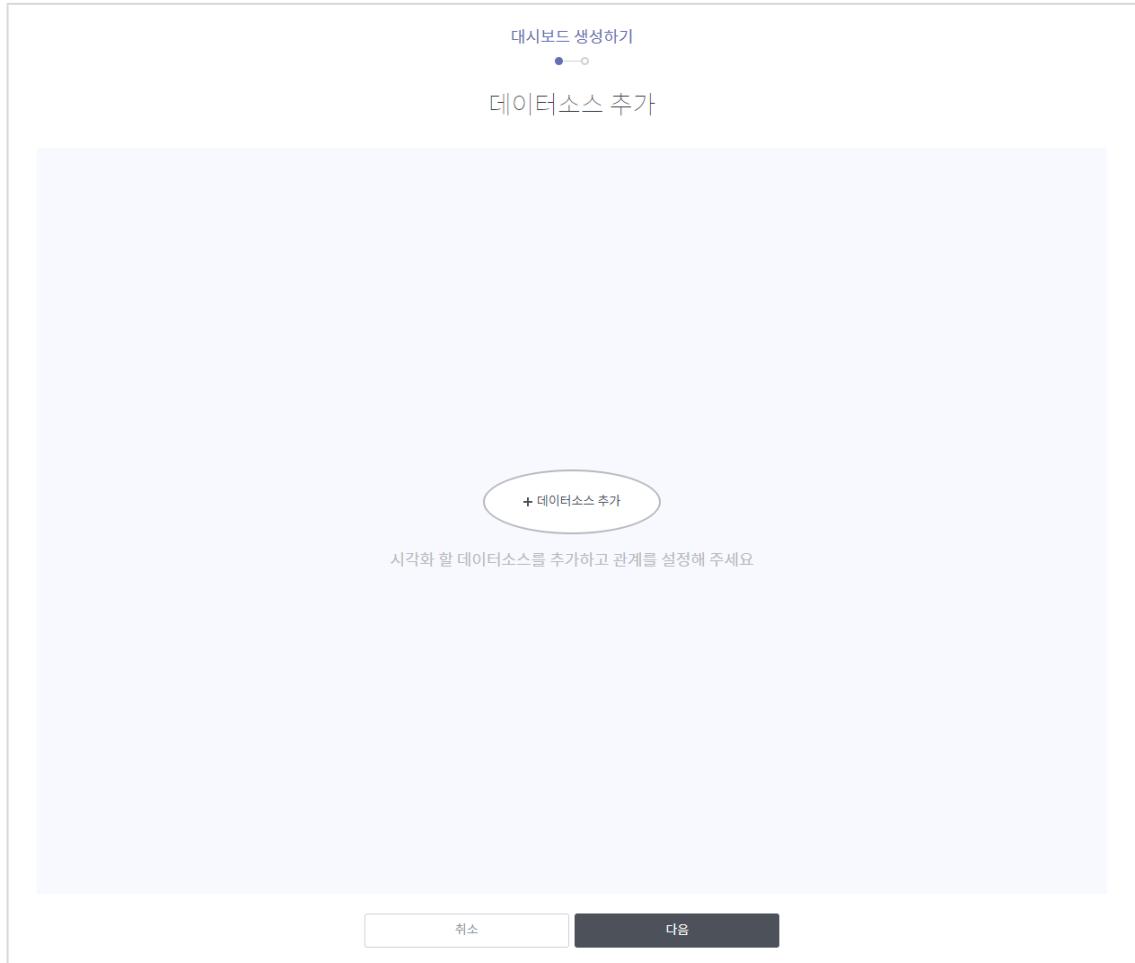
작업 44	노트북 9	워크벤치 34	20 데이터 소스	Created on 2017-11-16 by Administrator	⋮
Admin workspace					
Workspace List					
Search by content or folder name					
전체 ▾					
<input type="checkbox"/> Admin workspace	<input type="checkbox"/> new folder • New	<input type="checkbox"/> J			
<input type="checkbox"/> notebook	<input type="checkbox"/> workbook • New	<input type="checkbox"/> workbook			
크라임	dd	Sales Workbook			
마지막 업데이트일 몇 초 전	마지막 업데이트일 11시간 전	마지막 업데이트일 14시간 전			
<input type="checkbox"/> 0 <input type="checkbox"/> 0	<input type="checkbox"/> 0 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 7			
<input type="checkbox"/> workbook	<input type="checkbox"/> workbook	<input type="checkbox"/> workbook			
□ □ □	test-exntu	test			
마지막 업데이트일 2일 전	마지막 업데이트일 3일 전	마지막 업데이트일 3일 전			
<input type="checkbox"/> 0 <input type="checkbox"/> 0	<input type="checkbox"/> 2 <input type="checkbox"/> 2	<input type="checkbox"/> 1 <input type="checkbox"/> 1			
전체 선택 Clone Workbook Move selections Delete selections					
+ 워크북 + 노트북 + 워크벤치					

3. 대시보드 생성하기

대시보드는 워크북 안에 보관되며 특정한 데이터 소스를 사용자의 필요에 맞게 분석하고 시각화하는 기능을 제공합니다. 따라서 대시보드 생성 절차는 주로 데이터 소스를 연결하는 것부터 시작됩니다.

대시보드 생성 절차는 다음과 같습니다.

- 데이터 소스를 추가할 수 있도록 '+ 데이터소스 추가'를 클릭합니다.



2. 해당 워크스페이스에 공개된 데이터 소스 목록에서 대시보드에 연결할 마스터 데이터 소스들을 선택합니다. 이후 단계에서는 여기서 선택하는 마스터 데이터 소스들에 join시킬 다른 데이터 소스를 추가적으로 선택할 수 있습니다.

데이터소스를 선택해 주세요

취소 마침

1 데이터소스 이름 검색

2 오픈 데이터만 보기

3 타입 전체

4

No.	데이터소스	타입
<input checked="" type="checkbox"/>	8 Sales [오픈 데이터]	수집형
<input type="checkbox"/>	7 tour_de_france_geo [오픈 데이터]	수집형
<input type="checkbox"/>	6 WorldCupMatches [오픈 데이터]	수집형
<input type="checkbox"/>	5 WA-test-data [오픈 데이터]	수집형
<input checked="" type="checkbox"/>	4 filter test [오픈 데이터]	수집형 ✓
<input type="checkbox"/>	3 tour de france [오픈 데이터]	수집형
<input type="checkbox"/>	2 MDM-test-DT [오픈 데이터]	수집형
<input type="checkbox"/>	1 LinkedTest [오픈 데이터]	연결형

5

filter test

메타데이터 이름

설명

타입 수집형

공개설정 공개

생성일 2018-09-06

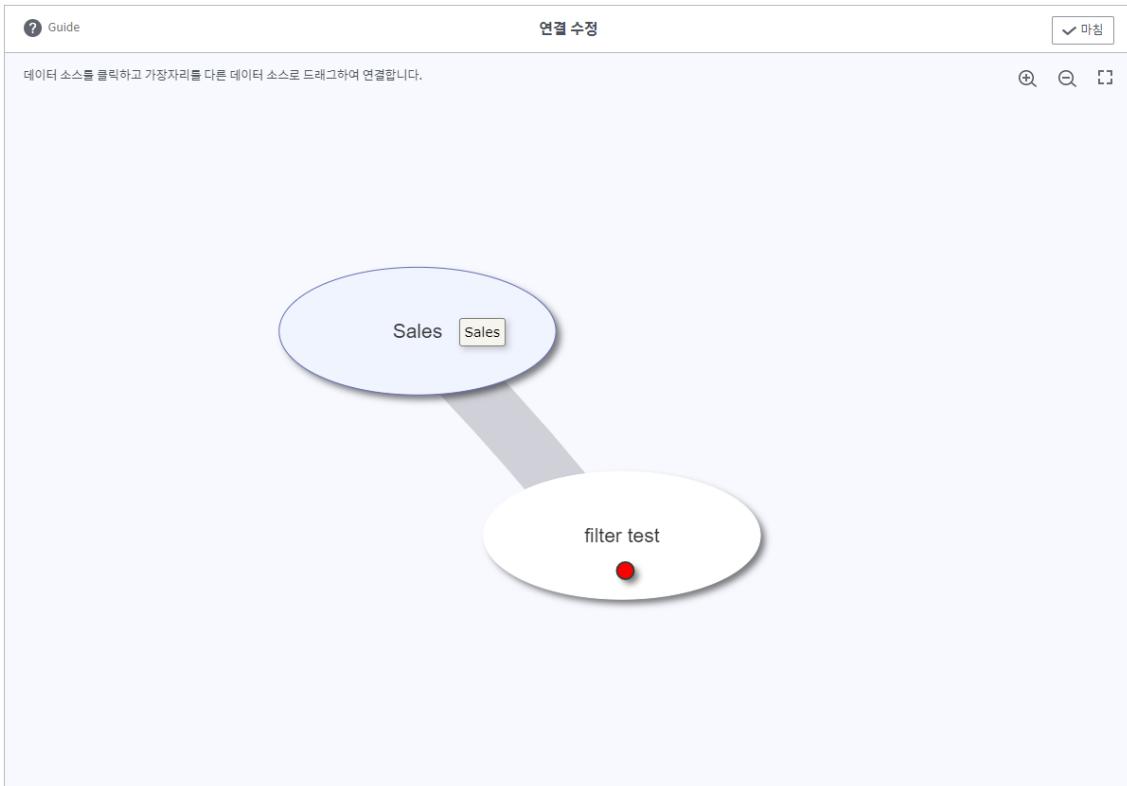
사이즈 528.90 KB

Rows 1,952

차원값	ab Row ID
차원값	ab Order Priority
측정값	# Discount
측정값	# Unit Price
측정값	# Shipping Cost
차원값	ab Customer ID
차원값	ab Customer Name

- 1 **데이터 소스 이름 검색:** 해당 워크스페이스에 허용된 데이터 소스를 이름으로 검색합니다.
- 2 **오픈 데이터만 보기:** '오픈 데이터 소스'로 지정된 데이터 소스만 선별하여 볼 수 있습니다.
- 3 **타입:** 데이터 소스를 연결형 또는 수집형으로 선별하여 볼 수 있습니다.
- 4 **데이터 소스 목록:** 설정한 선별 조건에 맞는 데이터 소스들을 보여줍니다.
- 5 **데이터 소스 정보:** 목록에서 선택한 데이터 소스의 정보를 간략하게 보여줍니다.

3. 둘 이상의 데이터 소스를 선택할 경우에는 한 데이터 소스를 다른 데이터 소스로 드래그함으로써 둘을 연결시킬 수 있습니다. 연결된 데이터 소스끼리는 상호간 필터링이 가능합니다. 데이터 소스 연결이 필요 없으시다면 '마침' 버튼을 클릭하십시오.



4. 한 데이터 소스를 다른 데이터 소스로 드래그하면 데이터 소스 연결을 설정할 수 있는 새 창이 열리게 됩니다. 양 테이블에서 상호간 필터링을 할 수 있는 연결 키로 사용할 컬럼을 하나씩 선택한 후 '마침'을 클릭하십시오.

Sales				filter test			
State				State or Province			
<input type="text"/> 검색				<input type="text"/> 검색			
Country	City	State	Postal Code		Region	State or Province	City
United States	Jonesb...	Arkansas	72401		East	New York	New Yor...
United States	Jonesb...	Arkansas	72401		East	Massachusetts	Boston
United States	Jonesb...	Arkansas	72401		East	Massachusetts	Boston
United States	Jonesb...	Arkansas	72401		East	Massachusetts	Boston
United States	Jonesb...	Arkansas	72401		Central	Texas	Dallas
United States	Jonesb...	Arkansas	72401		West	Washington	Seattle
United States	Philadel...	Pennsylvania	19143		West	California	Los Ang...
United States	Roswell	Georgia	30076		West	California	Los Ang...
United States	Alexand...	Virginia	22304		East	New York	New Yor...
United States	Alexand...	Virginia	22304		West	Washington	Seattle
United States	Alexand...	Virginia	22304		West	California	Los Ang...
United States	Alexand...	Virginia	22304		East	New York	New Yor...

5. 아래 설명을 참조하여 데이터 소스 연결을 재설정하거나, 앞에서 선택한 최상위 데이터 소스에 join시킬 다른 데이터 소스들을 추가합니다. 원하는 결과물을 도출하였으면 '다음' 버튼을 클릭하십시오.

기본 화면



- ① **마스터 데이터 소스 추가:** 클릭하면 새 마스터 데이터 소스를 추가할 수 있습니다.
- ② **연결 수정:** 클릭하면 데이터 소스 연결 설정을 수정할 수 있습니다.
- ③ **데이터 소스 연결:** 현재 설정된 데이터 연결을 보여줍니다. 데이터 소스 하나를 클릭하면 패널이 열리는데, 여기서는 해당 데이터 소스를 미리 보거나 삭제하거나 join 관계를 설정할 수 있습니다.
- ④ **탭**
 - 데이터 미리보기: 데이터 소스 join에 따른 결과 테이블을 보여줍니다.
 - 스키마 관리: 선택한 데이터 소스의 join 관계를 관리할 수 있습니다.
- ⑤ **데이터 소스 삭제:** 클릭하면 선택한 데이터 소스가 삭제됩니다.
- ⑥ **패널 닫기:** 클릭하면 선택한 데이터 소스에 대한 패널이 닫힙니다.
- ⑦ **데이터 소스 트리:** 데이터 소스끼리의 join 관계를 트리 구조로 보여줍니다.
 - : 해당 join 관계를 수정합니다.
 - : 해당 join 관계를 수정합니다. (아이콘과 동일한 기능)
 - : 해당 join 관계를 삭제합니다.
 - '+ 조인을 위하여 데이터소스를 추가해 주세요' 버튼을 클릭하면 새로운 데이터 소스를 join시키는 대화 상자가 나타납니다. (이 대화 상자에 관해서는 다음 페이지 설명 참조)

데이터 join 창

The screenshot shows the 'Join' dialog box with the following steps:

- Master datasource:** sales table with columns OrderDate, Category, City, Country. Rows: 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:** safe_lookup_sample table with columns City, user_id, current_datetime. Rows: Henders..., test03, 2018-01-24 00:00:00; Naperville, test01, 2018-01-24 00:00:00.
- Join keys:** Column selection dropdowns for both tables.
- Join type:** Radio buttons for Inner, Left, Right, Full outer join types.
- Preview results:** A table showing joined data with 31 columns and 1000 rows. Columns include sales.OrderDate, sales.Category, sales.City, sales.Country, sales.CustomerName, sales.OrderID, and others from the joined tables.

- 마스터 데이터 소스:** 새로 join시키고자 하는 데이터 소스의 마스터 데이터 소스에 관한 정보를 보여줍니다.
- 조인할 데이터 소스:** 마스터 데이터 소스에 조인할 데이터 소스를 선택합니다.
- 조인 키 추가:** '조인 키'는 마스터 데이터 소스와 조인할 데이터 소스 간의 컬럼별 조인 관계를 정의하기 위한 키입니다. 두 데이터 소스에서 서로 연결시킬 컬럼을 하나씩 선택한 후 이 버튼을 클릭하면 새로운 조인 키가 추가됩니다. 이때 각각의 데이터 소스의 컬럼에 정의된 데이터 타입이 일치해야 합니다.
- 조인 타입:** 데이터 소스를 어떻게 조인하여 변형할 것인지를 선택합니다. 이해를 돋기 위해 다음과 같은 예시를 이용하여 각 조인 타입을 설명하겠습니다.

마스터 데이터 소스

제품명(조인 키)	가격
A	\$22.11
B	\$9.23
C	\$8.99
D	\$10.10

조인할 데이터 소스

제품명(조인 키)	판매량
B	100
D	200
E	50

- Inner: 조인 키 컬럼 내 데이터 값을 기준으로 마스터 데이터 소스와 조인할 데이터 소스에 공통적으로 해당하는 레코드만 결과 테이블에 반영합니다. (두 데이터 소스의 교집합)

제품명(조인 키)	가격	판매량
B	\$9.23	100
D	\$10.10	200

- Left: 왼쪽 데이터 소스(마스터 데이터 소스)의 조인 키 컬럼 내 데이터 값을 기준으로 오른쪽 데이터 소스(조인할 데이터 소스)의 데이터를 가져와 조인한 뒤, 결과 테이블에 반영합니다. 오른쪽 데이터 소스의 레코드 중에서 왼쪽 데이터 소스에 없는 조인 키 컬럼 데이터 값을 가진 레코드는 버려집니다.

제품명(조인 키)	가격	판매량
A	\$22.11	null
B	\$9.23	100
C	\$8.99	null
D	\$10.10	200

- Right: 오른쪽 데이터 소스(조인할 데이터 소스)의 조인 키 컬럼 내 데이터 값을 기준으로 왼쪽 데이터 소스(마스터 데이터 소스)의 데이터를 가져와 조인한 뒤, 결과 테이블에 반영합니다. 왼쪽 데이터 소스의 레코드 중에서 오른쪽 데이터 소스에 없는 조인 키 컬럼 데이터 값을 가진 레코드는 버려집니다.

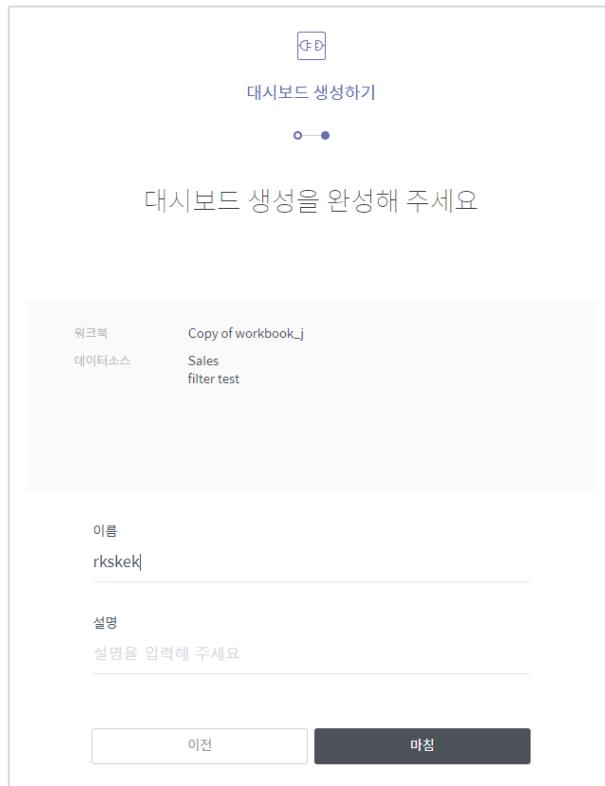
제품명(조인 키)	가격	판매량
B	\$9.23	100
D	\$10.10	200
E	null	50

- Full Outer: 조인 키 컬럼 내 데이터 값을 기준으로 양쪽 데이터 소스의 모든 데이터를 가져와 조인한 뒤, 결과 테이블에 반영합니다. (두 데이터 소스의 합집합)

제품명(조인 키)	가격	판매량
A	\$22.11	null
B	\$9.23	100
C	\$8.99	null
D	\$10.10	200
E	null	50

- ⑤ 결과 미리보기: 데이터 소스를 조인한 결과값이 나타납니다.

6. 대시보드를 생성하기 위해 불러온 데이터 소스들에 관한 정보를 확인한 뒤, 이름과 설명을 입력하고 '마침' 버튼을 누르면 새로운 대시보드가 생성됩니다.

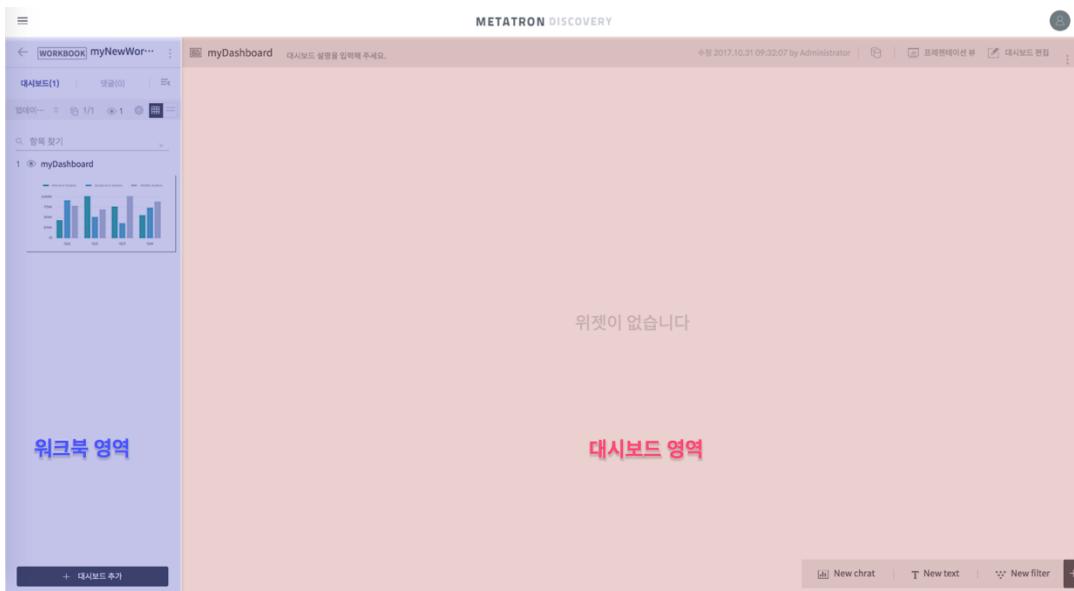


7. 새로 생성된 대시보드는 워크북 홈 화면에 추가됩니다. 클릭하면 해당 대시보드에 관한 화면이 표시됩니다.

The screenshot shows the Metatron Discovery home screen. On the left, there is a sidebar with 'WORKBOOK myNewWor...', 'Dashboard(1)', and a search bar. The main area shows a dashboard titled 'myDashboard' with a subtitle 'Dashboard description input required'. It contains several bar charts. Below the dashboard, a message says 'No items yet'. At the bottom, there are buttons for '+ Add new dashboard' and a toolbar with 'New chart', 'New text', 'New filter', and other icons.

4. 워크북 관리하기

워크스페이스 화면에 열거된 워크북 중 하나를 선택하면 해당 워크북의 홈 화면이 나타납니다. 워크북 홈 화면에서는 대시보드의 열람, 생성, 편집이 가능합니다. 또한, 프레젠테이션 모드를 통해 여러 대시보드를 슬라이드 쇼 모드로 디스플레이할 수 있습니다. 워크북 홈 화면은 다음과 같이 두 가지 영역으로 나뉩니다.



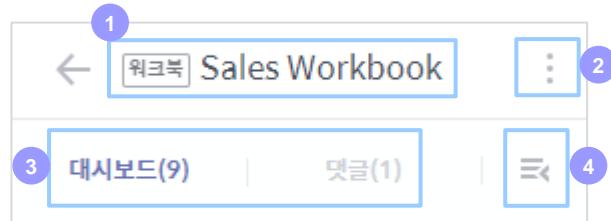
- **대시보드 리스트:** 해당 워크북을 구성하고 대시보드 목록을 관리하는 기능을 제공합니다. (4.1절 참조)
- **대시보드 상세 뷰:** 대시보드 리스트에서 선택한 대시보드에 관한 상세 화면이 나타납니다. 여기서는 차트, 텍스트, 필터 위젯을 통한 데이터의 분석, 시각화, 선별이 가능합니다. (4.2절 참조)

4.1 대시보드 리스트

대시보드 리스트는 사용자 선택에 따라 '대시보드 모드'와 '댓글 모드'로 전환이 가능하며, 각 모드의 기능은 다음과 같습니다.

- 대시보드: 새로운 대시보드를 추가하거나 기존에 등록된 대시보드를 나열합니다.
- 댓글: 새로운 댓글을 추가하거나 기존에 등록된 대시보드를 나열합니다. 댓글은 해당 워크북에 접근 가능한 모든 사용자가 작성·열람할 수 있습니다.

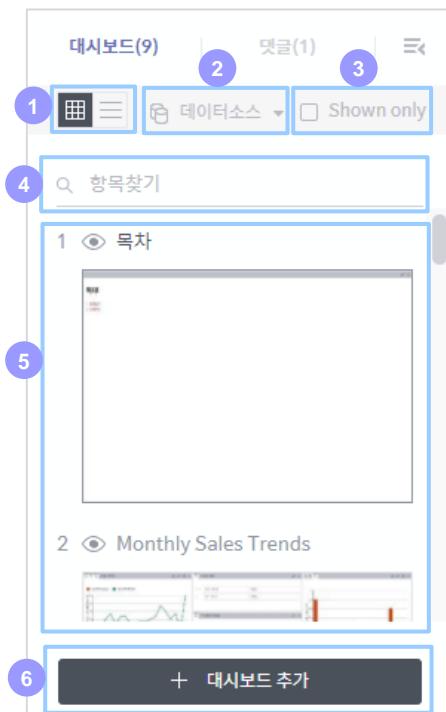
대시보드/댓글 공통 영역



- ① **이름:** 해당 워크북의 이름입니다.
- ② **더 보기:** 클릭하면 해당 워크북의 이름/설명 수정 및 워크북 삭제가 가능합니다. 또한 워크북의 최근 업데이트 시각과 생성 시각을 확인할 수 있습니다.
- ③ **대시보드/댓글 모드 선택:** 대시보드와 댓글 모드 중 어느 것을 표시할지 선택합니다. 각각의 선택지 옆에 괄호로 표시된 숫자는 워크북에 등록된 대시보드/댓글의 개수를 가리킵니다.
- ④ **접기:** 이 버튼을 클릭하면 대시보드 리스트를 접거나 펼 수 있습니다.

대시보드 모드

- ① **뷰 형식:** 대시보드 목록을 보여주는 방식을 설정합니다. 썸네일 뷰/리스트 뷰 형식 중 하나를 선택할 수 있습니다.
- ② **데이터 소스:** 현재 워크북 내에서 사용되고 있는 데이터 소스들을 확인합니다. 각 데이터 소스의 확인란을 체크하거나 체크를 해제하여 특정 데이터 소스를 사용하고 있는 대시보드만 선별하여 조회할 수 있습니다.
- ③ **Shown only:** 숨김 설정을 하지 않은 대시보드만 선별하여 조회합니다.
- ④ **항목 찾기:** 대시보드를 이름으로 검색합니다.
- ⑤ **대시보드 목록:** 검색 결과에 맞는 대시보드 목록이 보여집니다. 대시보드에 마우스 오버 시, 해당 대시보드의 복제 및 삭제가 가능합니다.
- ⑥ **대시보드 추가:** 해당 버튼을 클릭하면 새로운 대시보드를 생성하는 화면으로 넘어갑니다. (생성 절차는 3장 참조)



댓글 모드

1 **댓글 목록:** 해당 워크북에 등록된 댓글들을 최근 것부터 보여줍니다.

2 **댓글 추가:** 새로운 댓글을 작성할 수 있습니다.
Enter를 누르면 기재한 댓글이 등록되고 Shift + Enter를 누르면 텍스트 줄이 바뀝니다.

← 워크북 Sales Workbook :

대시보드(9) | 댓글(2) | ☰

나 21:44:40
고객에 관련된 설명은 이 정도면 충분합니다.

나 21:44:40
매출 측면에서 보다 상세한 설명이 필요해보입니다.

워크북에 대한 의견을 나누어 보세요.
(Enter 입력 | Shift + Enter 개행)

4.2 대시보드 상세 뷰

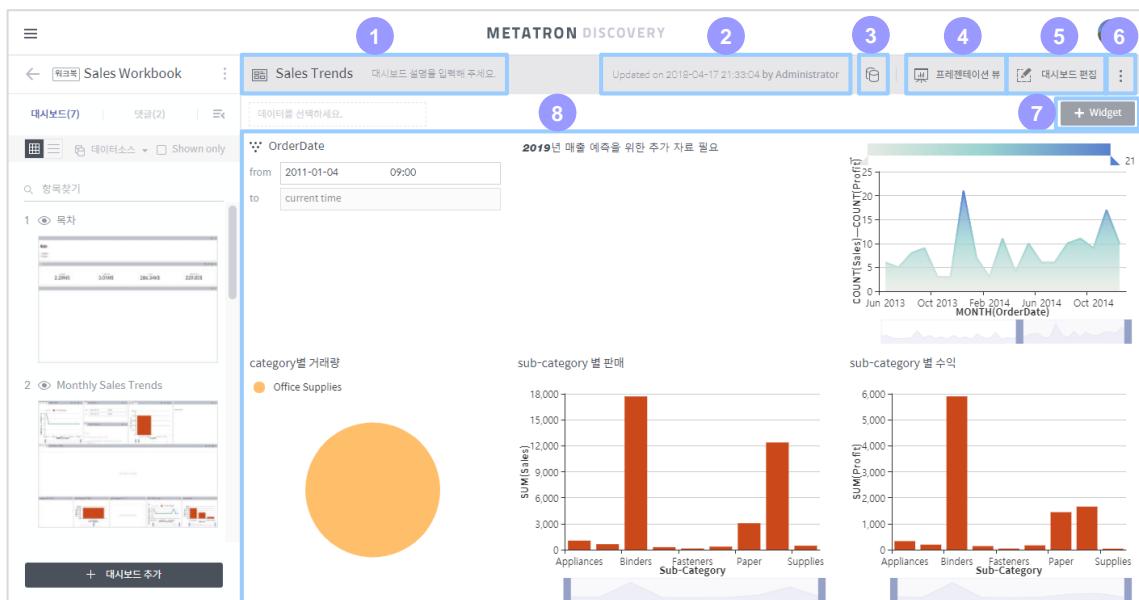
대시보드 상세 뷰에는 대시보드 리스트에서 선택한 대시보드에 관한 상세 화면이 나타납니다. 여기서는 차트, 텍스트, 필터 위젯을 통한 데이터의 분석, 시각화, 선별이 가능합니다.

4.2.1 대시보드 - 기본 화면

대시보드 기본 화면에서는 해당 대시보드에 등록된 차트들을 열람할 수 있으며, 해당 대시보드의 구성을 설정하기 위한 다양한 화면으로 이동할 수 있습니다.

A. 대시보드 기본 화면 전체 구성

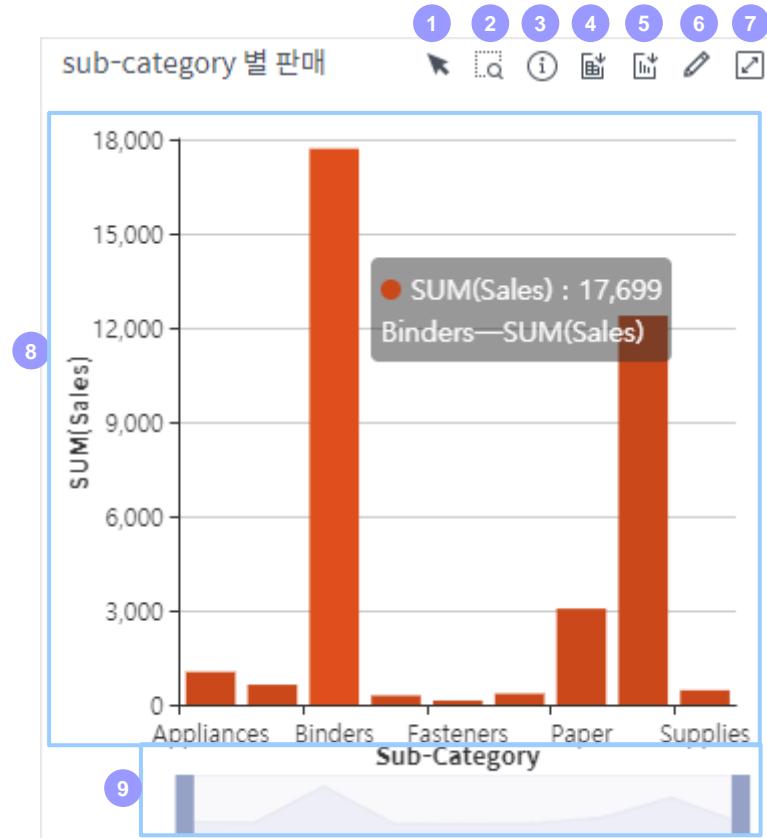
아래는 대시보드 기본 화면의 전반적인 구성에 관한 설명입니다.



- ① **이름 및 설명:** 해당 대시보드의 이름과 설명입니다. 마우스 오버 시 수정 가능합니다.
- ② **업데이트 정보:** 해당 대시보드를 마지막으로 업데이트한 시각과 사용자 이름이 나타납니다.
- ③ **데이터 소스:** 클릭하면 해당 대시보드에서 사용되고 있는 데이터 소스 정보와 그에 관한 통계/스키마 뷰를 볼 수 있습니다. (상세한 설명은 2.1.3절의 A 참조)
- ④ **프레젠테이션 뷰:** 워크북을 프레젠테이션에 적합한 UI로 열람할 수 있습니다. 클릭하면 현재 선택한 대시보드에 해당하는 슬라이드가 나타나며, 다른 대시보드로 이동할 수도 있습니다. (4.2.3절 참조)
- ⑤ **대시보드 편집:** 해당 버튼을 클릭하면 대시보드 상의 위젯을 편집할 수 있는 화면이 나타납니다. (4.2.4절 참조)
- ⑥ **더 보기:** 클릭하면 해당 대시보드를 복제 및 삭제할 수 있으며, 또한 대시보드의 최근 수정 정보와 생성 정보를 확인할 수 있습니다.
- ⑦ **위젯 추가:** 대시보드에 위젯을 추가할 수 있는 메뉴입니다. 현재 차트, 텍스트, 필터 위젯을 제공합니다. (4.2.4 D절 참조)
- ⑧ **위젯 레이아웃 영역:** 대시보드 레이아웃 화면에 표시되도록 설정한 위젯들이 보여집니다.

B. 대시보드 기본 화면 내 차트 위젯 상자

아래는 위젯 영역에 있는 차트 위젯 상자에 관한 설명입니다. 상자에 마우스를 올리면 오른쪽 상단에 해당 위젯과 관련된 설정 아이콘들이 나타납니다.



- ① **데이터 영역 선택:** 차트 그래프에서 마우스 커서로 데이터 항목을 선택하는 방식을 선택합니다. 특정한 데이터 항목(들)을 선택하면 해당 차원값 범주를 기준으로 대시보드 내 모든 차트가 필터링되어 표시됩니다.
- ② **차트 축소/확대:** 차트 화면을 축소/확대 할 수 있습니다. ⌂ 버튼을 누르면 초기화됩니다.
- ③ **차트 정보:** 해당 차트를 만들 때 피벗팅한 데이터 정보가 나타납니다.
- ④ **데이터 테이블 저장:** 해당 차트의 데이터 정보를 로컬 파일로 저장합니다.
- ⑤ **차트 이미지 저장:** 해당 차트의 이미지를 jpg파일로 저장합니다.
- ⑥ **수정:** 클릭하면 해당 위젯을 수정하기 위한 대화 상자가 열립니다.
- ⑦ **전체화면 확대:** 해당 차트를 대시보드 상세 뷰 전체로 확대하여 나타냅니다.
- ⑧ **차트 그래프 영역:** 해당 차트 그래프가 표시되는 영역입니다. 이 중에서 특정한 데이터 항목(들)을 선택하면 해당 차원값 범주를 기준으로 대시보드 내 모든 차트가 필터링되어 표시됩니다.
- ⑨ **차트 미니맵:** 해당 차트의 데이터 범주별 값 분포도를 보여줍니다. 필터 적용 시 미니맵의 범위도 그에 맞게 축소됩니다.

C. 대시보드 기본 화면 내 텍스트 위젯 상자

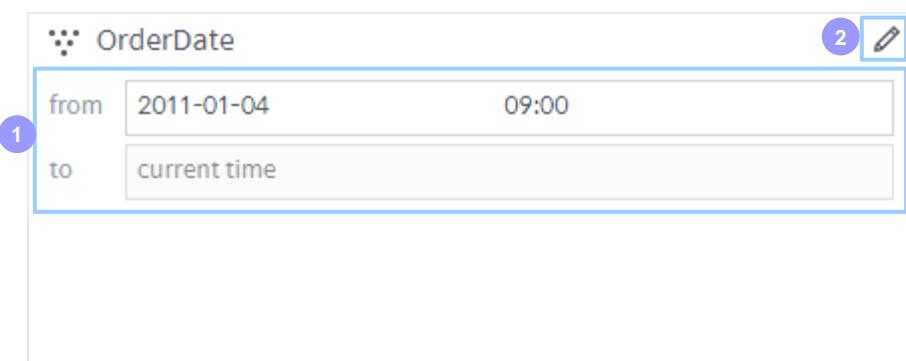
아래는 위젯 영역에 있는 텍스트 위젯 상자에 관한 설명입니다. 상자에 마우스를 올리면 오른쪽 상단에 해당 위젯을 수정할 수 있는 아이콘이 나타납니다.



- ① **텍스트 영역:** 작성한 텍스트가 표시되는 영역입니다.
- ② **수정:** 클릭하면 해당 위젯을 수정하기 위한 대화 상자가 열립니다.

D. 대시보드 기본 화면 내 필터 위젯 상자

아래는 위젯 영역에 있는 필터 위젯 상자에 관한 설명입니다. 상자에 마우스를 올리면 오른쪽 상단에 해당 위젯을 수정할 수 있는 아이콘이 나타납니다.



- ① **필터 영역:** 구성한 필터가 표시되는 영역입니다. 필터 범위를 수정할 수 있습니다.
- ② **수정:** 클릭하면 해당 위젯을 세부적으로 수정하기 위한 대화 상자가 열립니다.

4.2.2 대시보드 - 데이터 소스 정보 대화 상자

대시보드 기본 화면에서  버튼을 클릭하면 해당 대시보드에서 사용하는 데이터 소스들의 정보를 보여주는 대화 상자가 나타납니다. 좌측 상단에서 확인하고자 하는 데이터 소스를 선택할 수 있습니다. 이 대화 상자는 크게 3가지 탭(데이터 그리드, 컬럼 상세 탭, 대시보드 데이터 정보)으로 구성됩니다.

데이터 그리드 탭

해당 데이터 소스의 레코드 값이 모두 표시됩니다.

데이터 그리드	컬럼 상세						
데이터 검색	의뢰	<input checked="" type="radio"/> 전체	<input type="radio"/> 차원값	<input type="radio"/> 측정값	타입	전체	
ab Row ID	ab Order ID	ab Order Date	ab Ship Date	ab Ship Mode	ab Customer ID	ab Customer Name	ab Segment
1122	US-2014-1...	2014-01-01 00...	2014-01-26T...	Standard Class	HL-15040	Hunter Lopez	Consumer
1123	US-2014-1...	2014-01-01 00...	2014-01-26T...	Standard Class	HL-15040	Hunter Lopez	Consumer
1124	US-2014-1...	2014-01-01 00...	2014-01-26T...	Standard Class	HL-15040	Hunter Lopez	Consumer
1125	US-2014-1...	2014-01-01 00...	2014-01-26T...	Standard Class	HL-15040	Hunter Lopez	Consumer
1126	US-2014-1...	2014-01-01 00...	2014-01-26T...	Standard Class	HL-15040	Hunter Lopez	Consumer
1127	US-2014-1...	2014-01-01 00...	2014-01-26T...	Standard Class	HL-15040	Hunter Lopez	Consumer
1760	CA-2014-1...	2014-01-01 00...	2014-01-12T...	Standard Class	MB-18085	Mick Brown	Consumer
1914	CA-2014-1...	2014-01-01 00...	2014-01-17T...	First Class	EH-13990	Erica Hackney	Consumer
2579	US-2014-1...	2014-01-01 00...	2014-01-31T...	Standard Class	SD-20485	Shirley Daniels	Home Office
2580	US-2014-1...	2014-01-01 00...	2014-01-31T...	Standard Class	SD-20485	Shirley Daniels	Home Office
2581	US-2014-1...	2014-01-01 00...	2014-01-31T...	Standard Class	SD-20485	Shirley Daniels	Home Office
2582	US-2014-1...	2014-01-01 00...	2014-01-31T...	Standard Class	SD-20485	Shirley Daniels	Home Office
2583	US-2014-1...	2014-01-01 00...	2014-01-31T...	Standard Class	SD-20485	Shirley Daniels	Home Office
2584	US-2014-1...	2014-01-01 00...	2014-01-31T...	Standard Class	SD-20485	Shirley Daniels	Home Office
2585	US-2014-1...	2014-01-01 00...	2014-01-31T...	Standard Class	SD-20485	Shirley Daniels	Home Office
2979	CA-2014-1...	2014-01-01 00...	2014-01-16T...	Second Class	ND-18370	Natalie DeCherney	Consumer
3084	CA-2014-1...	2014-01-01 00...	2014-02-03T...	Standard Class	JC-15340	Jasper Cacioppo	Consumer
3366	CA-2014-1...	2014-01-01 00...	2014-02-02T...	First Class	LC-17050	Liz Carlisle	Consumer
3795	US-2014-1...	2014-01-01 00...	2014-02-02T...	Standard Class	EJ-13720	Ed Jacobs	Consumer
3796	US-2014-1...	2014-01-01 00...	2014-02-02T...	Standard Class	EJ-13720	Ed Jacobs	Consumer
3797	US-2014-1...	2014-01-01 00...	2014-02-02T...	Standard Class	EJ-13720	Ed Jacobs	Consumer
4938	CA-2014-1...	2014-01-01 00...	2014-01-18T...	Standard Class	BD-11605	Brian Dahlen	Consumer
4939	CA-2014-1...	2014-01-01 00...	2014-01-18T...	Standard Class	BD-11605	Brian Dahlen	Consumer
4940	CA-2014-1...	2014-01-01 00...	2014-01-18T...	Standard Class	BD-11605	Brian Dahlen	Consumer
5328	CA-2014-1...	2014-01-01 00...	2014-01-08T...	Second Class	LS-17230	Lycoris Saunders	Consumer
5365	CA-2014-1...	2014-01-01 00...	2014-01-21T...	Second Class	DW-13195	David Wiener	Corporate
5463	CA-2014-1...	2014-01-01 00...	2014-01-20T...	First Class	TS-21340	Toby Swindell	Consumer

컬럼 상세 탭

해당 데이터 소스를 구성하는 각 컬럼에 관한 정보를 상세히 보여줍니다.

Sales		컬럼 상세 - used in sales																																													
데이터 그리드		컬럼 상세																																													
데이터 검색		의뢰		<input checked="" type="radio"/> 전체																																											
설명																																															
<p>물리 컬럼명</p> <table border="1"> <tr><td>차원값</td><td>ab Row ID</td></tr> <tr><td>차원값</td><td>ab Order ID</td></tr> <tr><td>차원값</td><td>ab Order Date</td></tr> <tr><td>차원값</td><td>ab Ship Date</td></tr> <tr><td>차원값</td><td>ab Ship Mode</td></tr> <tr><td>차원값</td><td>ab Customer ID</td></tr> <tr><td>차원값</td><td>ab Customer Name</td></tr> <tr><td>차원값</td><td>ab Segment</td></tr> <tr><td>차원값</td><td>ab Country</td></tr> <tr><td>차원값</td><td>ab City</td></tr> <tr><td>차원값</td><td>ab State</td></tr> <tr><td>차원값</td><td># Postal Code</td></tr> <tr><td>차원값</td><td>ab Region</td></tr> <tr><td>차원값</td><td>ab Product ID</td></tr> <tr><td>차원값</td><td>ab Category</td></tr> <tr><td>차원값</td><td>ab Sub-Category</td></tr> <tr><td>차원값</td><td>ab Product Name</td></tr> <tr><td>측정값</td><td># Sales</td></tr> <tr><td>측정값</td><td># Quantity</td></tr> <tr><td>측정값</td><td># Discount</td></tr> <tr><td>측정값</td><td># Profit</td></tr> </table>						차원값	ab Row ID	차원값	ab Order ID	차원값	ab Order Date	차원값	ab Ship Date	차원값	ab Ship Mode	차원값	ab Customer ID	차원값	ab Customer Name	차원값	ab Segment	차원값	ab Country	차원값	ab City	차원값	ab State	차원값	# Postal Code	차원값	ab Region	차원값	ab Product ID	차원값	ab Category	차원값	ab Sub-Category	차원값	ab Product Name	측정값	# Sales	측정값	# Quantity	측정값	# Discount	측정값	# Profit
차원값	ab Row ID																																														
차원값	ab Order ID																																														
차원값	ab Order Date																																														
차원값	ab Ship Date																																														
차원값	ab Ship Mode																																														
차원값	ab Customer ID																																														
차원값	ab Customer Name																																														
차원값	ab Segment																																														
차원값	ab Country																																														
차원값	ab City																																														
차원값	ab State																																														
차원값	# Postal Code																																														
차원값	ab Region																																														
차원값	ab Product ID																																														
차원값	ab Category																																														
차원값	ab Sub-Category																																														
차원값	ab Product Name																																														
측정값	# Sales																																														
측정값	# Quantity																																														
측정값	# Discount																																														
측정값	# Profit																																														
<p>물리 컬럼명</p> <p>Row ID</p> <p>물리 풀 DIMENSION</p> <p>Missing 설정안함</p>																																															
통계																																															
<table border="1"> <tr><td>건 수</td><td>9994</td></tr> <tr><td>Valid</td><td>9994 (100%)</td></tr> <tr><td>Unique</td><td>9994 (100%)</td></tr> <tr><td>Missing</td><td>0 (0%)</td></tr> </table>						건 수	9994	Valid	9994 (100%)	Unique	9994 (100%)	Missing	0 (0%)																																		
건 수	9994																																														
Valid	9994 (100%)																																														
Unique	9994 (100%)																																														
Missing	0 (0%)																																														
값 목록																																															
<table border="1"> <tr><td>1047</td><td>1</td></tr> <tr><td>1048</td><td>1</td></tr> <tr><td>1050</td><td>1</td></tr> <tr><td>1053</td><td>1</td></tr> <tr><td>1054</td><td>1</td></tr> <tr><td>1055</td><td>1</td></tr> <tr><td>1056</td><td>1</td></tr> <tr><td>1172</td><td>1</td></tr> <tr><td>1173</td><td>1</td></tr> <tr><td>1174</td><td>1</td></tr> </table>						1047	1	1048	1	1050	1	1053	1	1054	1	1055	1	1056	1	1172	1	1173	1	1174	1																						
1047	1																																														
1048	1																																														
1050	1																																														
1053	1																																														
1054	1																																														
1055	1																																														
1056	1																																														
1172	1																																														
1173	1																																														
1174	1																																														

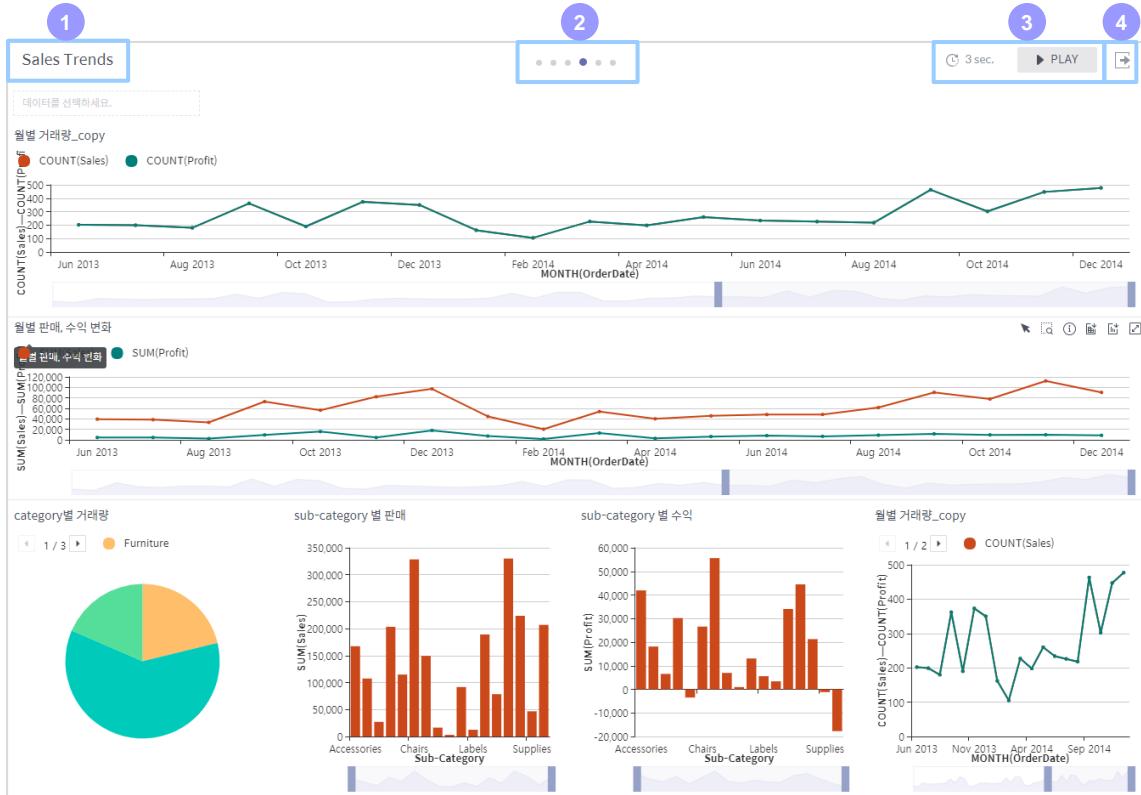
대시보드 데이터 정보 탭

해당 데이터 소스 전체에 대한 요약 정보를 보여줍니다.



4.2.3 대시보드 - 프레젠테이션 뷰 화면

대시보드 기본 화면에서 '프레젠테이션 뷰' 버튼을 클릭하면 워크북의 대시보드들을 프레젠테이션에 적합한 UI로 열람할 수 있습니다. 이를 통해 사용자는 데이터 분석 결과를 쉽게 보고하고 공유할 수 있습니다.

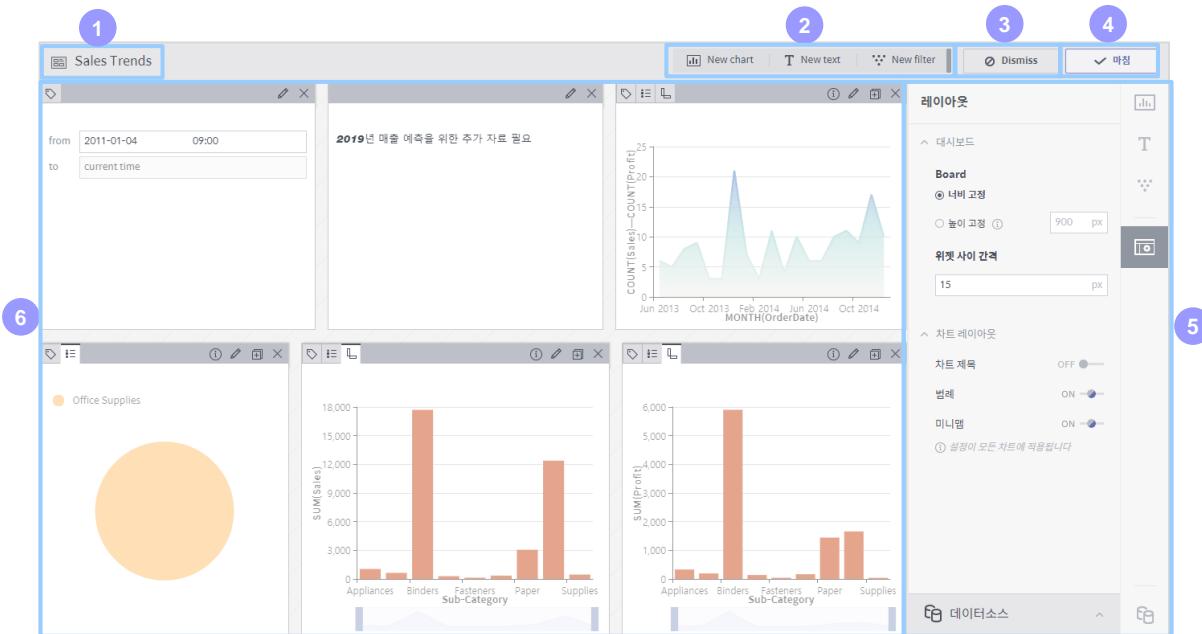


1. 이름: 현재 표시 중인 대시보드의 이름입니다.
2. 슬라이드 네비게이션: 각 동그라미는 워크북 내 대시보드 각각을 가리킵니다. 예를 들어 네 번째 동그라미를 클릭하면 네 번째 대시보드 슬라이드로 이동하고 해당 동그라미가 강조됩니다.
3. 자동 슬라이드 쇼 설정: 시간 간격 선택 후 PLAY 버튼을 클릭하면, 자동 슬라이드 쇼가 시작되어 선택한 시간 간격을 주기로 슬라이드가 넘어갑니다.
4. 나가기: 프레젠테이션 뷰를 종료하고 워크북/대시보드 기본 화면으로 돌아갑니다.

4.2.4 대시보드 - 편집 화면

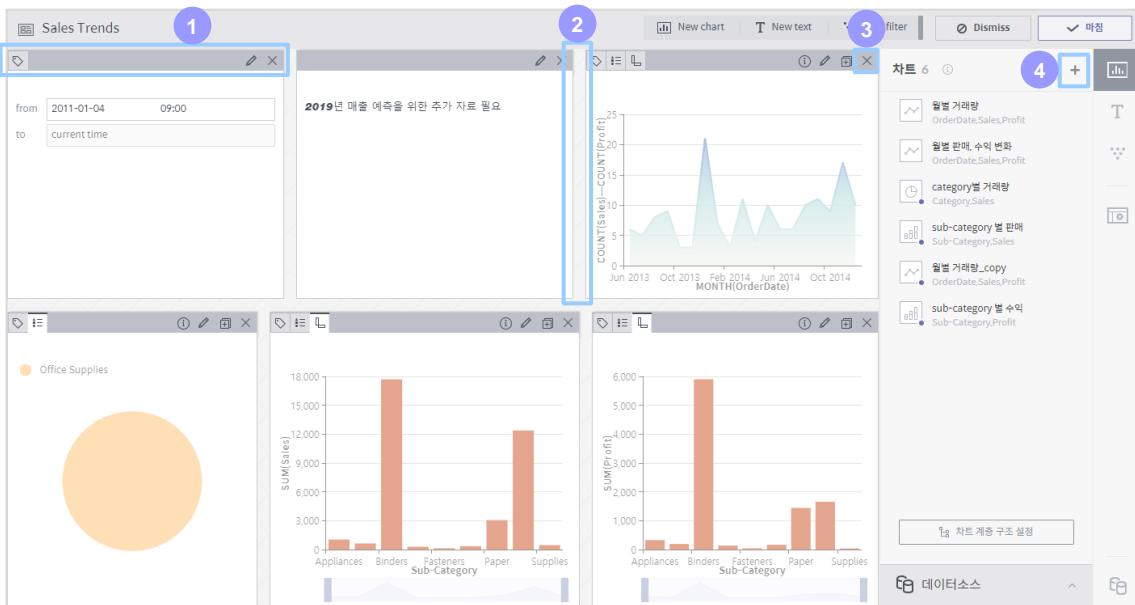
대시보드 기본 화면에서 '대시보드 편집' 버튼을 클릭하면 해당 대시보드의 구성을 편집할 수 있는 화면으로 이동합니다. 여기서는 위젯 추가 기능을 포함하여 대시보드의 편집 및 계층 구조 설정, 레이아웃 변경 기능을 사용할 수 있습니다.

A. 대시보드 편집 화면 전체 구성



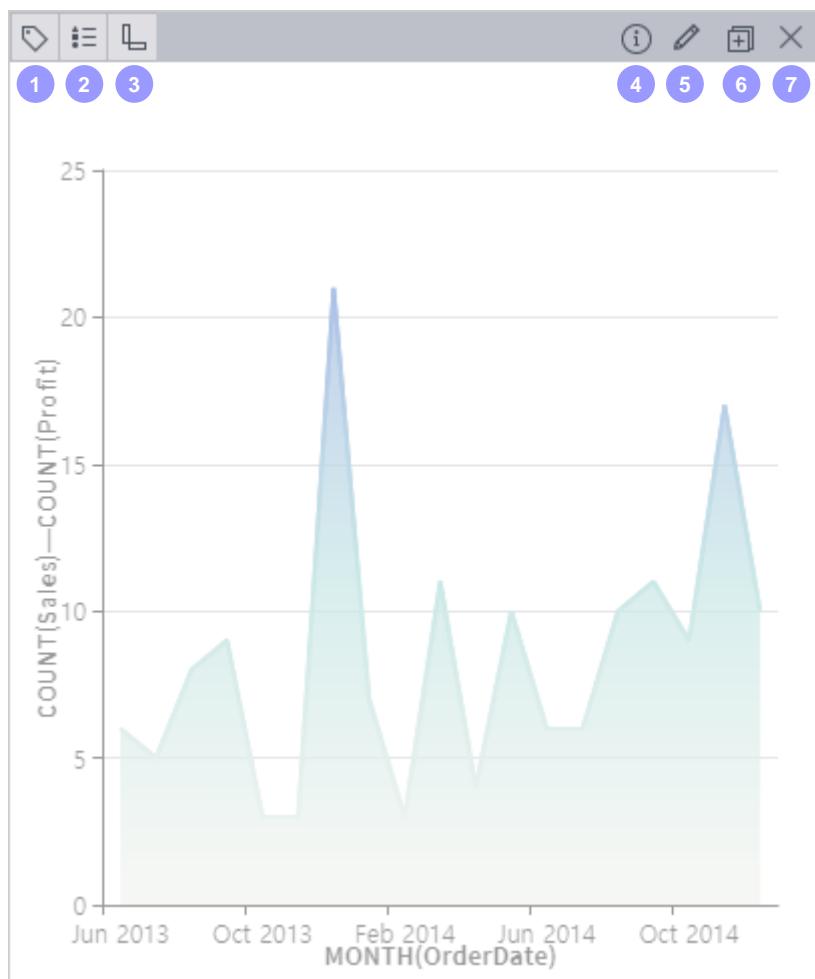
- 1** 이름: 해당 대시보드의 이름을 보여줍니다.
- 2** 위젯 추가 메뉴: 차트, 텍스트, 필터 등의 위젯을 추가할 수 있는 메뉴입니다.
- 3** Dismiss: 클릭하면 현재 화면에서 변경사항을 저장하지 않고 나갑니다.
- 4** 마침: 클릭하면 현재 화면에서 변경사항을 저장한 후 나갑니다.
- 5** 위젯 레이아웃 영역: 대시보드 레이아웃 화면에 표시되도록 설정한 위젯들의 배열, 표시 방법, 속성 등을 편집합니다. (자세한 사항은 B~C절 참조)
- 6** 패널 영역: 각종 위젯을 추가·수정·삭제하고 대시보드 표시 방식에 관련된 몇 가지 설정값을 편집합니다. (자세한 사항은 D절 참조)

B. 위젯 배열 설정



- ① **위젯 위치 변경:** 위젯 제목 부분을 드래그하여 위젯의 위치를 변경할 수 있습니다.
- ② **위젯 너비 조정:** 위젯과 위젯의 사이를 움직여 위젯의 너비를 조정할 수 있습니다.
- ③ **화면에 위젯 추가:** 우측 패널에 있는 위젯 목록에서 원하는 위젯을 좌측 위젯 레이아웃 영역으로 드래그하면 해당 위젯이 레이아웃 영역에 추가됩니다.
- ④ **화면에서 위젯 삭제:** 위젯 레이아웃 영역에 표시된 각 위젯에서 ✕ 버튼을 클릭하면 해당 위젯이 레이아웃 영역에서 제거됩니다.

C. 개별 위젯 편집 영역

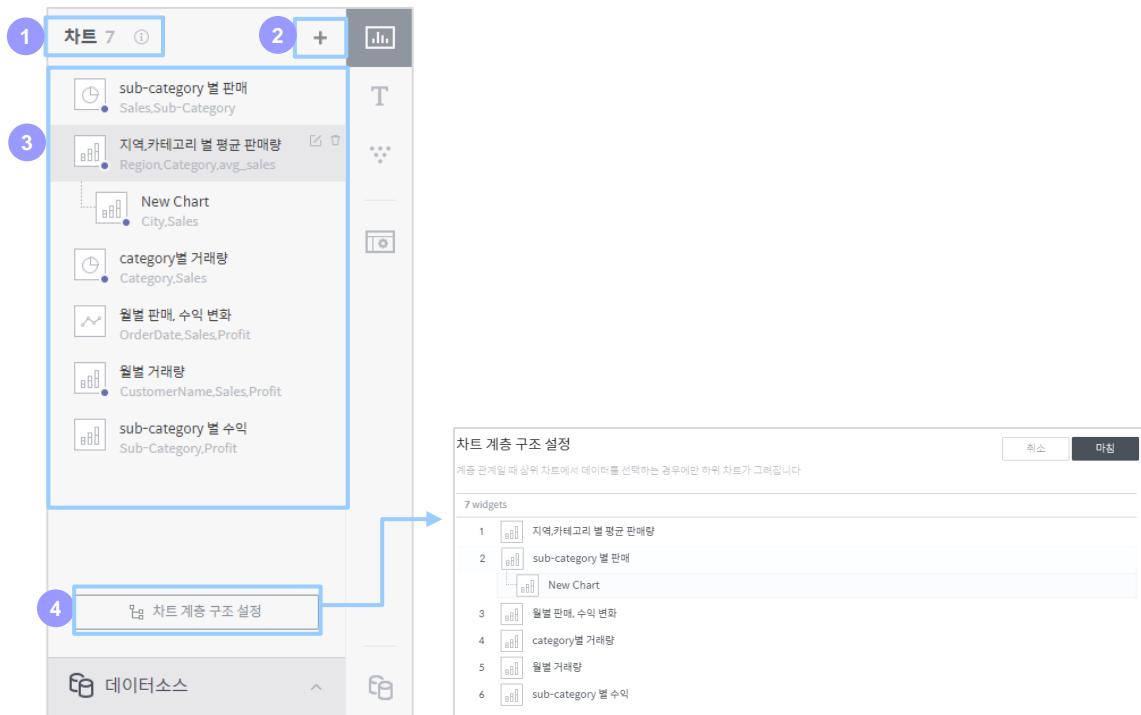


- ① **위젯 이름:** 해당 위젯의 이름을 표시하거나 숨길 수 있습니다.
- ② **범례:** 범례를 표시하거나 숨길 수 있습니다.
- ③ **미니맵:** 미니맵을 표시하거나 숨길 수 있습니다.
- ④ **차트 정보:** 해당 차트를 만들 때 피벗팅한 데이터 정보가 나타납니다.
- ⑤ **수정:** 클릭하면 해당 위젯을 수정하기 위한 대화 상자가 열립니다.
- ⑥ **복사:** 클릭하면 해당 차트와 똑같은 차트가 복제되어 화면에 나타납니다.
- ⑦ **삭제:** 클릭하면 해당 차트가 대시보드 화면에서 제거됩니다.

D. 대시보드 편집 화면 내 패널 영역

차트 위젯 패널

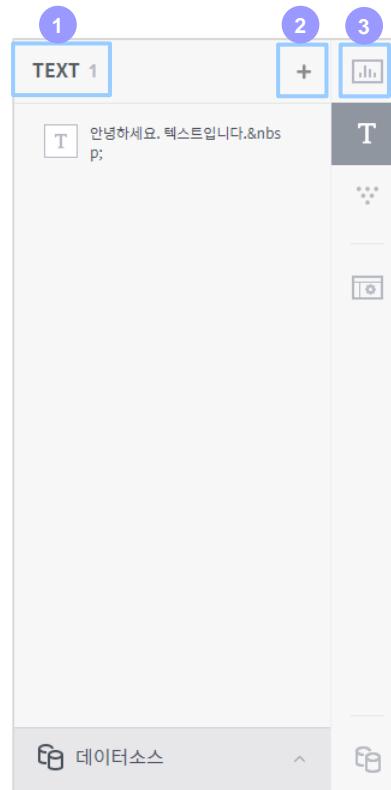
차트 위젯 패널에서는 대시보드 내 차트의 추가, 수정, 삭제 등이 가능합니다.



- ① **차트 위젯 수:** 현재 대시보드에 등록된 차트 위젯의 개수를 나타냅니다.
- ② **차트 위젯 추가:** 대시보드 내에 새로운 차트 위젯을 생성할 수 있습니다. (자세한 절차는 5장 참조)
- ③ **차트 위젯 목록:** 현재 대시보드에 등록된 차트 위젯들이 열거됩니다. 수정 또는 삭제를 원하는 위젯 항목에 마우스를 오버하면 이를 위한 아이콘이 나타납니다. 또한 위젯 항목을 위젯 레이아웃 영역으로 드래그하면 해당 위젯이 레이아웃 영역에 표시됩니다.
- ④ **차트 계층 구조 설정:** 대시보드 내 차트 간 상하 관계를 설정할 수 있습니다. 부모 차트에서 데이터 항목을 하나 선택하면 자식 차트가 그 항목을 기준으로 필터링됩니다. 계층 구조를 설정하려면 하위 관계로 설정할 차트를 드래그하여 원하는 상위 관계 차트 밑으로 옮기면 됩니다. 차트 계층 구조 설정이 완료되면 차트 메뉴 상에서도 구조가 변경된 것을 확인할 수 있습니다.

텍스트 위젯 패널

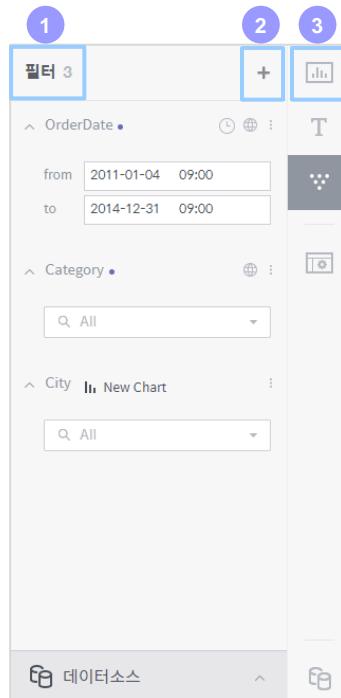
텍스트 위젯 패널에서는 대시보드 내 텍스트 위젯의 추가, 수정, 삭제 등이 가능합니다.



- ① **텍스트 위젯 수:** 현재 대시보드에 등록된 텍스트 위젯의 개수를 나타냅니다.
- ② **텍스트 위젯 추가:** 대시보드 내에 새로운 텍스트 위젯을 생성할 수 있습니다.
- ③ **텍스트 위젯 목록:** 현재 대시보드에 등록된 텍스트 위젯들이 열거됩니다. 수정 또는 삭제를 원하는 위젯 항목에 마우스를 오버하면 이를 위한 아이콘이 나타납니다. 또한 위젯 항목을 위젯 레이아웃 영역으로 드래그하면 해당 위젯이 레이아웃 영역에 표시됩니다.

필터 위젯 패널

필터 위젯 패널에서는 대시보드 내 필터 위젯의 추가, 수정, 삭제 등이 가능합니다.

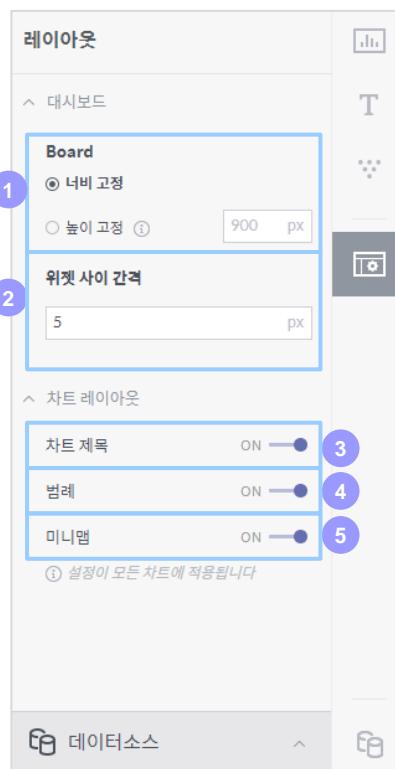


- ① **필터 위젯 수:** 현재 대시보드에 등록된 필터 위젯의 개수를 나타냅니다.
- ② **필터 위젯 추가:** 대시보드 내에 새로운 필터 위젯을 생성할 수 있습니다.
- ③ **필터 위젯 목록:** 현재 대시보드에 등록된 필터 위젯들이 열거됩니다. 수정 또는 삭제를 원하는 위젯 항목에 마우스를 오버하면 이를 위한 아이콘이 나타납니다. 또한 위젯 항목을 위젯 레이아웃 영역으로 드래그하면 위젯이 위젯 레이아웃 영역에 표시됩니다.

이 패널의 기능과 구조는 모두 차트 흘의 차트 필터 패널과 동일하기 때문에 구체적인 설명은 5.5.2절을 참조하십시오. 단, 여기서 설정하는 필터에는 반드시 '글로벌' 속성이 적용되어 모든 차트에 일괄 적용됩니다.

레이아웃 패널

레이아웃 패널에서는 위젯 레이아웃 영역에서 위젯의 배열과 개별 위젯 표시 방법에 관하여 몇 가지 설정을 설정합니다.



① 보드 높이 설정

- 화면 고정: 보드의 높이를 화면에 맞춥니다.
- 높이 고정: 보드의 높이를 고유한 픽셀 값으로 설정합니다.

② 위젯 사이 간격: 위젯 레이아웃 영역에 표시되는 위젯 간의 간격을 지정합니다.

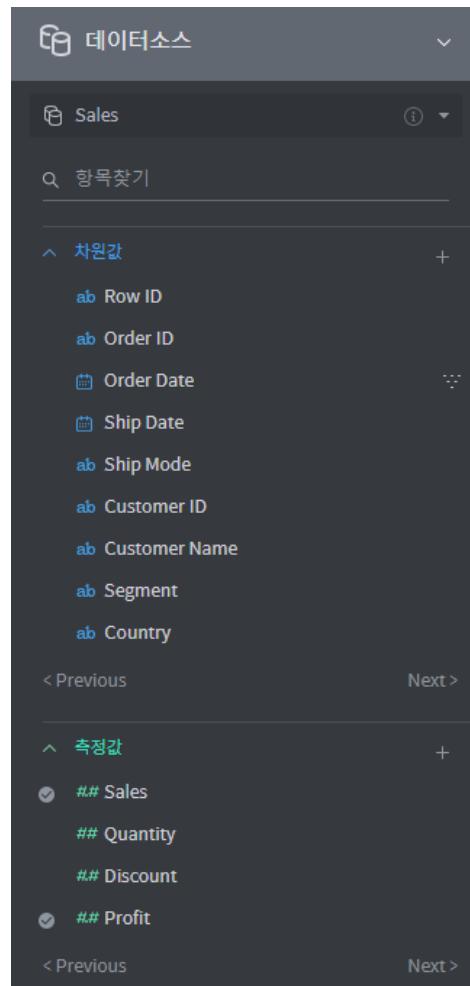
③ 차트 제목: 위젯 레이아웃 영역 내 차트 및 필터 위젯들의 제목 표시 여부를 일괄 설정합니다.

④ 범례: 위젯 레이아웃 영역 내 차트 위젯들의 범례 표시 여부를 일괄 설정합니다.

⑤ 미니맵: 위젯 레이아웃 영역 내 차트 위젯들의 미니맵 표시 여부를 일괄 설정합니다.

데이터 소스 패널

데이터 소스 패널에서는 연동된 데이터 소스의 정보를 열람·수정하고, 컬럼 필터를 간편하게 추가하거나 제거할 수 있습니다.



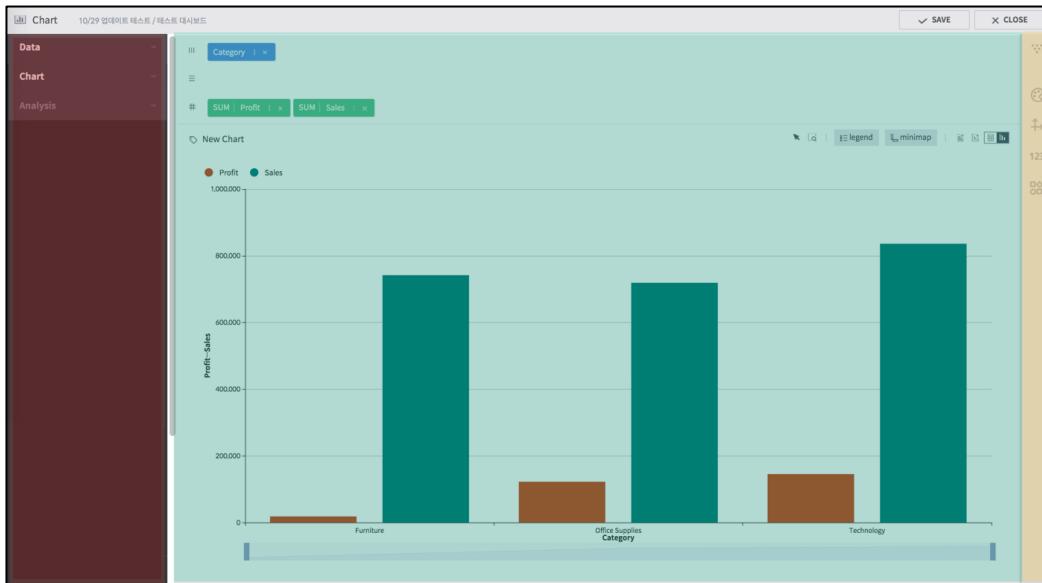
이 패널은 차트 홈 화면의 데이터 컬럼 목록과 동일한 기능과 구조를 가지므로, 이에 대한 자세한 설명은 5.2절을 참조하십시오. 단, 여기서 지정/해제하는 필터는 대시보드 필터이고, 차트 홈에서 지정/해제하는 필터는 차트 필터임을 유의하십시오.

5. 차트 생성·관리

기본적으로 워크북 내 각 대시보드는 분석한 데이터를 시각화하여 보여주는 다양한 차트로 구성됩니다. 본 절에서는 데이터 분석 목적에 맞게 차트를 만들기 위해 필수적으로 알아야 할 몇 가지 개념과 Discovery의 차트 구성 UI에 대해 설명합니다.

5.1 차트 홈 화면 개관

차트 홈 화면은 다음과 같이 세 영역으로 구분됩니다.

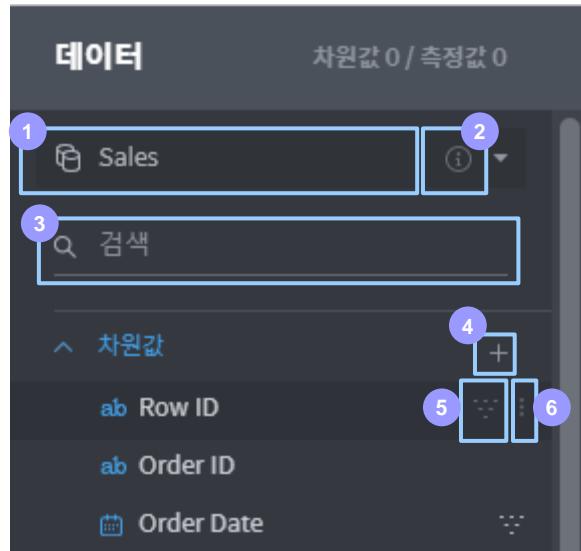


- ① **컬럼/차트 선택 영역:** 차트를 만들기 위해 필수적으로 행해야 하는 액션의 순서대로 UI가 구성되어 있습니다. Data(데이터 컬럼 리스트)선택을 통해 차트를 피벗팅할 수 있으며, Chart(차트 종류 리스트)를 선택하여 데이터를 시각화할 수 있습니다. 또한 Analytics(분석)을 통해 원하는 분석 조건을 차트에 탑재할 수 있습니다.
- ② **시각화 영역:** 피벗팅할 수 있는 선반영역과 실제 차트가 그려지는 시각화 영역으로 구성됩니다. 컬럼/차트 선택 영역에서 차트를 그릴 수 있는 데이터와 차트가 선택되면 이곳에서 차트가 나타납니다.
- ③ **옵션 영역:** 차트를 보기 좋게 꾸미고, 차트가 보여지는 방식을 사용자의 기호에 맞게 선택할 수 있습니다. 옵션 영역은 필터, 파레트, 축, 숫자표현, 차트표현으로 이루어져 있습니다.

5.2 데이터 컬럼 목록

5.2.1 데이터 컬럼 목록 구성

데이터 컬럼 목록에서는 연동된 데이터 소스들의 정보를 열람·수정하고, 컬럼 필터를 간편하게 추가하거나 제거할 수 있습니다.



- ① **데이터 소스 선택/설정:** 데이터 소스를 선택하거나 해당 데이터 소스의 연결/join 관계를 설정합니다.
- ② **데이터 상세:** 클릭하면 새 대화 상자를 통해 선택된 데이터 소스에 관한 정보를 보여줍니다(자세한 설명은 4.2.2절 참조).
- ③ **컬럼 이름으로 검색:** 데이터 소스 내 컬럼을 이름으로 검색합니다.
- ④ **사용자 컬럼 추가:** 클릭하면 데이터 소스에 있는 컬럼들을 조합·가공하여 새로운 컬럼을 만들 수 있는 대화 상자가 열립니다. 추가된 사용자 컬럼은 대시보드 전체에서 사용할 수 있습니다.
- ⑤ **필터 지정/해제:** 이 버튼은 해당 컬럼에 마우스를 오버하면 생기며, 클릭 시 해당 컬럼을 차트 필터로 지정하고 다시 한번 클릭하면 지정된 차트 필터가 해제됩니다. 필터로 지정된 컬럼 항목에는 📈 아이콘이 마우스 오버와 상관 없이 표시됩니다.
- ⑥ **더 보기:** 이 버튼은 해당 컬럼에 마우스를 오버하면 생기며, 컬럼에 대한 추가적인 정보를 확인하고 별칭을 지정할 수 있습니다.
 - ⓘ: 클릭하면 새 대화 상자가 나타나면서 해당 컬럼의 요약 정보와 데이터 값들을 보여줍니다.
 - Logical column name: 해당 컬럼의 논리적 컬럼명을 보여줍니다.
 - Logical type: 해당 컬럼의 논리적 데이터 타입을 보여줍니다.
 - Alias: 해당 컬럼에 대한 별칭을 지정할 수 있습니다. 정식 컬럼명은 영숫자와 몇 가지 특수문자로 제한되며 공백도 포함할 수 없기 때문에 보다 구분하기 편한 별칭을 등록하면 분석 시 편의를 도모할 수 있습니다. 지정된 별칭은 대시보드 전체에 적용됩니다.
 - 값 별칭: 해당 컬럼에 포함된 각 데이터 값에 대해서도 별칭을 지정할 수 있습니다. 지정된 별칭은 대시보드 전체에 적용됩니다.

5.2.2 사용자 컬럼 추가

데이터 소스 컬럼 목록에서 **+** 버튼을 클릭하면 사용자 컬럼 추가를 위한 대화 상자가 열립니다. 여기서는 데이터 소스에 있는 기존 컬럼들에 각종 공식을 적용하여 차트를 만드는 데 필요한 새로운 컬럼을 만들 수 있습니다.

1 컬럼 명: DIMENSION_1

2 CAST([CHARACTER_SET_NAME], 'text')

3 컬럼 추가: CHARACTER_SET_NAME, DEFAULT_COLLATE_NAME, DESCRIPTION, MAXLEN, current_datetime

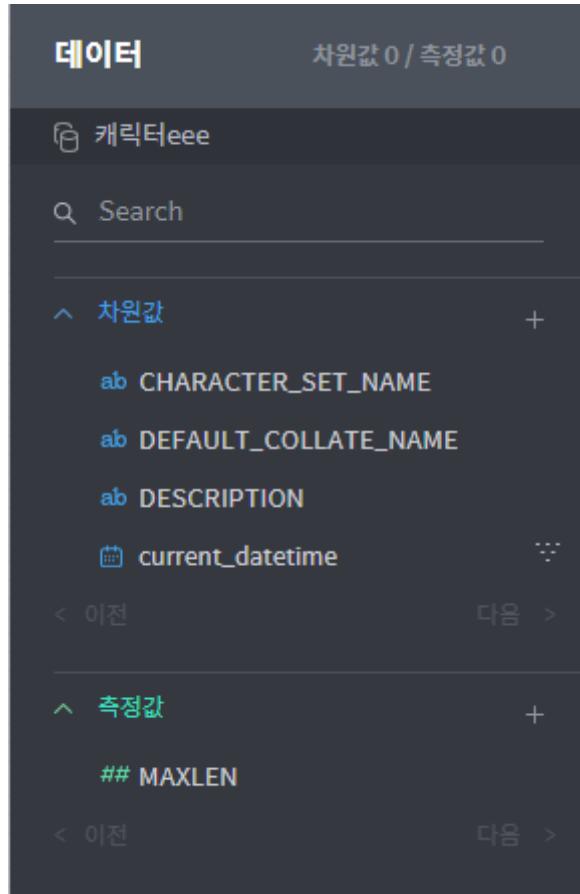
4 공식 추가: ALL, LOGICAL FUNCTION, IF, NVL, CASE, IN, TYPE_CONVERT FUNCTION, CAST, TIMESTAMP, UNIX_TIMESTAMP, TIME FUNCTION, DATEDIFF

✓ 계산식에 이상이 없습니다 유효성 체크

- ① **컬럼 명:** 사용자 컬럼의 이름을 적는 란입니다.
- ② **코드 영역:** 사용자 컬럼을 만들기 위한 코드를 적는 란입니다. 아래의 컬럼 및 공식 목록에서 원하는 항목을 클릭하면 이 영역에 자동으로 태이핑됩니다.
- ③ **컬럼 추가:** 데이터 소스에 주어진 기존 컬럼 목록입니다. 목록에 제시된 컬럼 항목 중 하나를 클릭하면 해당 컬럼이 코드 영역에 자동으로 태이핑됩니다.
- ④ **공식 추가:** metatron에서 지원하는 공식 목록입니다. 목록에 제시된 공식 중 하나를 클릭하면 해당 공식이 코드 영역에 자동으로 태이핑되고 태이핑 커서가 파라미터를 입력하는 부분으로 자동 이동됩니다. 각 공식의 용도와 사용법, 예제에 관해서는 화면 우측의 도움말 상자를 참조하십시오.

5.2.3 차원값과 측정값

대시보드에 연동된 데이터 소스의 컬럼들은 아래와 같이 차원값 컬럼과 측정값 컬럼으로 구분됩니다. Discovery의 차트 기능을 온전히 활용하기 위해서는 차원값과 측정값의 개념을 명확하게 이해해야 합니다.



차원값(Dimension) 컬럼

범주형 데이터 컬럼을 가리키며, 특징은 아래와 같습니다.

- 집계(aggregated)보다는 분류(categorical)에 의미가 있는 데이터 필드(예: Category, Region, Organization 등)
- 측정값을 표시하는 데 기준이 됨.

측정값(Measure) 컬럼

수량적 데이터 필드를 가리키며, 특징은 아래와 같습니다.

- 집계(aggregated)할 수 있거나 양적인(quantitative) 정보를 포함하는 필드(예: Sales 등)
- 차원값에 의해 제시된 기준을 토대로 차트에 표현되는 데이터

5.3 피봇팅

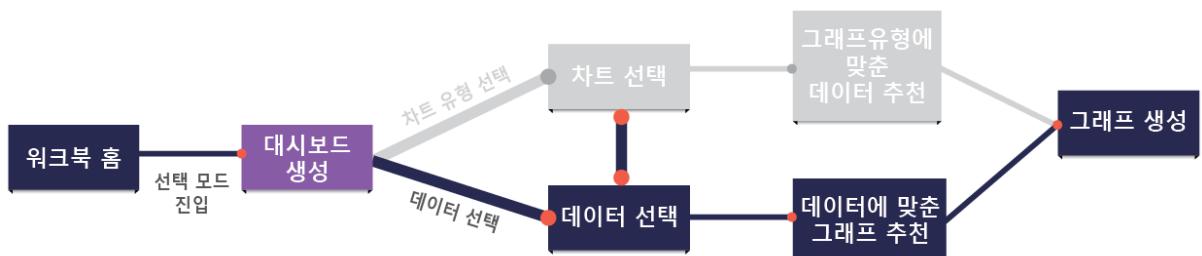
5.3.1 ‘피봇팅’이란?

‘피봇팅’이란 주어진 테이블을 특정 컬럼들을 기준으로 그룹화하는 과정을 의미하며, 이를 통해 분석가는 원천 데이터의 특정한 측면을 그래픽 또는 도표로 확인할 수 있습니다. 이러한 과정에는 의미 있는 데이터를 포함하는 컬럼들을 열/행/교차 선반에 배치하는 것을 포함합니다.



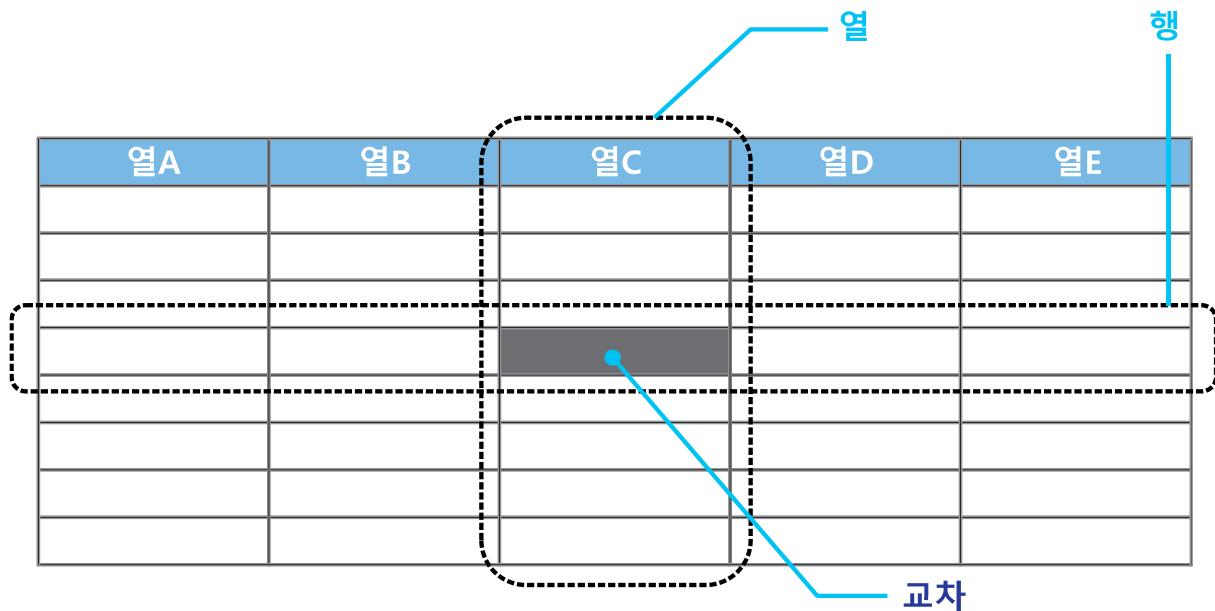
위 예시 그림은 두 개의 차원값 컬럼을 열선반에 배치하고 하나의 측정값 컬럼을 교차선반에 배치한 상태를 보여주고 있습니다. 차트에는 이렇게 선반에 올려놓은 컬럼들의 데이터가 표시됩니다.

차트 유형별로 선반별 필수/권장 컬럼 유형이 다르며, 컬럼들을 선반에 올려놓기 전에 먼저 차트 유형을 선택하면 선반에 필요한 컬럼 유형이 제시됩니다.

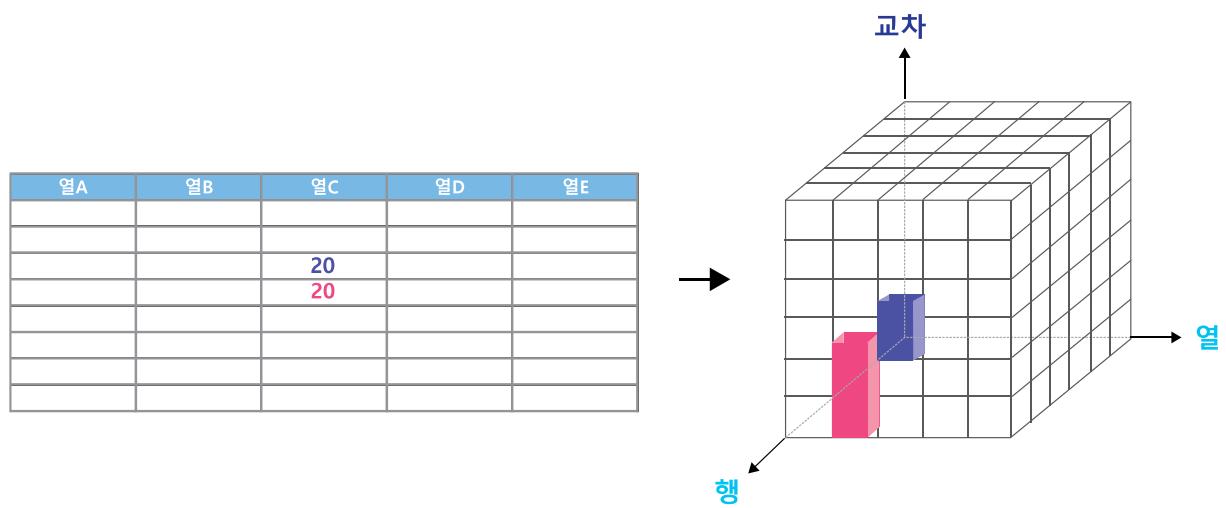


5.3.2 열/행/교차 선반의 개념

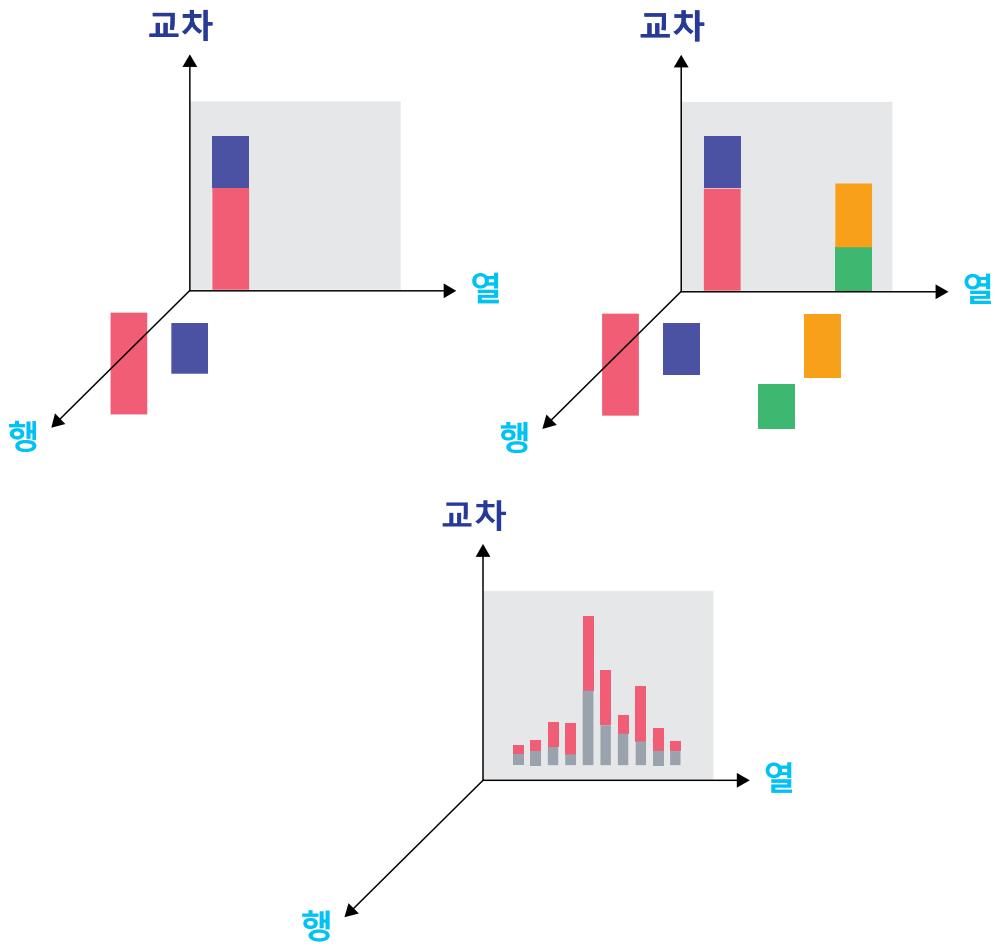
열/행/교차 선반의 개념은 Excel의 구조를 생각하면 쉽게 이해할 수 있습니다. 아래 그림과 같이 열/행은 블록을 정의하는 역할을 하고, 교차는 블록 안에 들어갈 값을 정하는 역할을 합니다.



metatron의 열/행/교차도 이와 비슷합니다. Excel에서 열/행/교차를 2차원 값인 그리드에 표현한다면, metatron은 열/행/교차를 3차원 큐브에서 표현합니다. metatron은 OLAP Data Discovery 도구로서, OLAP Cube를 통해 3차원에서 데이터를 조회합니다(OLAP에 대한 자세한 설명은 제1부 metatron 개요 참조). 아래의 차트는 metatron에서 3차원 큐브로 나타낸 열/행/교차 값의 축 그림입니다.

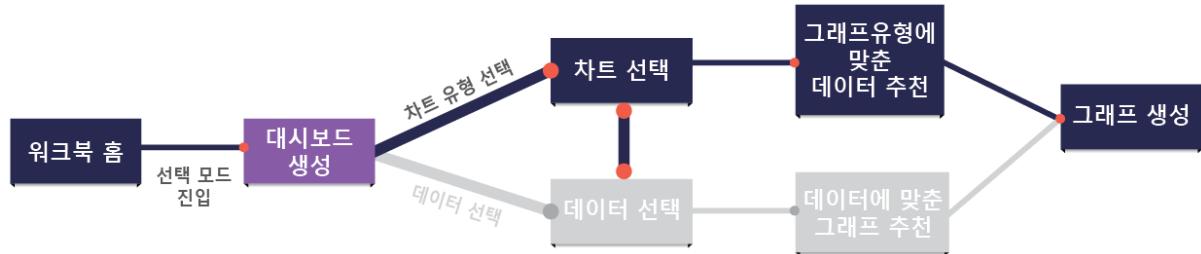


Excel 그리드의 값을 3차원 차트에서 표현한다면 그리드의 교차값이 막대바 형태로 여러 개 세워지게 될 수 있음을 유추할 수 있습니다. 하지만, metatron에서는 2차원 단면으로 차트가 보여지기 때문에 열과 행 기준으로 막대 바를 쌓아 올려서 표현하게 됩니다. 결국 metatron에서는 아래 그림의 회색 부분과 같은 2차원 형태의 차트로 나타납니다.



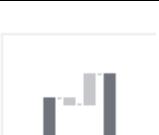
5.4 차트의 종류

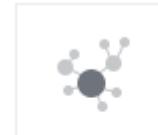
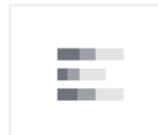
metatron Discovery에서는 20여가지의 차트를 제공하고 있습니다. 차트를 선택하기 전에 먼저 컬럼들을 선반에 올려놓으면 그에 어울리는 차트들이 보라색으로 강조됩니다.



아래 표는 각 차트별 생성조건과 사용 속성, 사용 유형, 사용 예시에 대해 정리한 것입니다.

차트명/아이콘	생성 조건	사용 속성	사용 유형	사용 예시
막대형 차트	열: 차원값 1 개 이상 & 교차: 측정값 1 개 이상	개별 항목의 값 비교	그룹들을 비교할 때 사용하거나 시간에 따른 변화 추이를 보고 싶을 때 사용합니다. 변화 추이가 클 때 사용하면 효과적입니다.	제품별 매출 및 수익 비교
표	열 또는 행: 차원값 1 개 이상, 교차: 측정값 1 개 이상	항목별 교차 데이터를 텍스트로 표시	특정 기준에 따른 측정 값을 보고 싶을 때 사용합니다. 시각화보다는 자세한 데이터와 정확한 값을 보려는 경우에 사용합니다.	연도별 매출 상세
선형 차트	열: 차원값 1 개 이상 & 교차: 측정값 1 개 이상	시간의 흐름에 따른 데이터 변화	시간에 따른 변화 추이를 보고 싶을 때 사용합니다. 변화 추이가 작을 때는 막대형 차트보다 선형 차트를 사용하는 것이 효과적입니다.	월별 매출 추이
분산형 차트	열: 측정값 1 개 & 행: 측정값 1 개 & 교차: 차원값 1 개 이상	관련된 여러 항목의 연관관계 표시	두 변수 간의 관계를 정의하고 싶을 때 사용합니다.	제품의 매출과 수익의 관계

차트명/아이콘	생성 조건	사용 속성	사용 유형	사용 예시
 히트맵	열 또는 행: 차원값 1 개 이상 & 교차: 측정값 1 개 이상	항목별 교차 데이터를 색상 분포 형태로 표시	색과 크기를 이용해 두 변수를 직관적으로 비교할 때 사용합니다. 표 차트에서 시각적 요소를 강조하기 위해 사용합니다.	지역별 각 제품 판매량
 원형 차트	교차: 차원값 1 개 이상, 측정값 1 개 이상	전체 대비 각 항목이 차지하는 비율	전체를 이루는 부분들을 비교할 때 사용합니다.	웹 브라우저의 마켓 쉐어 비교
 콘트롤 차트	열: 시간속성의 차원값 1 개 & 교차: 측정값 1 개 이상	공정 상태를 나타내는 특성치를 표시	생산 공정 품질 관리를 위해 사용	시간에 따른 제품 품질 추이 및 평균의 변동 관찰
 핵심지표	교차: 측정값 1 개 이상	주요 지표를 추세와 함께 표시	조직의 현재 성과에 대한 정보를 빠르게 전달하기 위해 사용합니다.	올 해 이입된 고객 수 혹은 조직의 성과 지표
 박스플롯	열: 차원값 1 개 이상, 행: 차원값 1 개, 교차: 측정값 1 개	값의 상승과 하락을 표시	각 그룹들의 분포를 비교할 때 사용하거나 특정 값의 목표치를 나타낼 때 사용합니다.	비행기 기종 별 지연 시간 분포 비교
 폭포 차트	열: 시간속성의 차원값 1 개 & 교차: 측정값 1 개	값을 더하거나 뺄 때의 누계를 표시	시간에 따른 값의 증가와 감소를 강조할 때 사용합니다.	일정 기간 동안 조직의 팀원 수 변화 관찰 혹은 주식
 워드클라우드	교차: 차원값 1 개 이상, 측정값 1 개	출현 빈도에 비례한 크기로 텍스트를 표시	중요한 단어들을 강조해서 요약하고 싶을 때 사용합니다.	고객들의 코멘트 요약

차트명/아이콘	생성 조건	사용 속성	사용 유형	사용 예시
 결합차트	열: 차원값 1 개 이상 & 교차: 측정값 2 개 이상, 4 개 이하	막대와 라인 차트를 결합하여 데이터를 비교	서로 다른 종류의 정보를 강조하기 위해 사용합니다.	제품별 가격과 판매량을 동시 관찰
 트리맵	열: 차원값 1 개 & 행: 차원값 1 개 이상 & 교차: 측정값 1 개	계층 데이터를 중첩된 사각형 집합으로 표시	계층 구조를 갖는 데이터를 시각화하기 위해 사용합니다.	제품별(대분류-중분류-소분류) 매출 관찰
 레이더 차트	교차: 차원값 1 개, 측정값 1 개 이상	중심점을 기준으로 다양한 평가요소를 표시	여러 측정 목표를 직관적으로 비교할 때 사용합니다.	품질을 5 요인으로 나누어 평가한 제품들 비교
 네트워크 다이어그램	주체 선반: 차원값 1 개 & 대상 선반: 차원값 1 개 & 연결 선반: 측정값 1 개	의존관계가 있는 요소들을 표현하는 연결 다이어그램	데이터의 발생 순서 흐름을 보고 싶을 때 사용합니다.	프로젝트 태스크 흐름을 관찰
 생기 다이어그램	열: 차원값 3 개 이상 & 교차: 측정값 1 개	연결선의 너비로 흐름의 양을 비율적으로 표시	데이터의 양적인 흐름을 관찰할 때 사용합니다.	공장의 에너지 흐름 관찰
 측정 차트	열: 행: 차원값 1 개 이상 & 교차: 측정값 1 개	설정된 목표에 대한 성능을 시각화하여 표시	데이터가 어떤 비율로 구성되어 있는지 보고 싶을 때 사용합니다.	지역 별 수익 관찰

5.5 차트 필터

차트 필터는 각 컬럼별로 차트에 나타낼 데이터의 범위를 한정 짓는 기능을 합니다. 본 절에서는 차트 필터를 지정하고 활용하는 방법에 대해 설명하며, 세부 구성은 다음과 같습니다.

- **5.5.1 자동으로 포함되는 필터들:** 별도의 차트 필터 추가 없이도 기본적으로 포함되는 필터들을 설명합니다.
- **5.5.2 차트 필터 패널:** 차트 홈 화면 우측에 나타나는 차트 필터 패널에 대해 설명합니다. 이 패널에서는 등록된 필터를 간단하게 조회·설정할 수 있습니다.
- **5.5.3 차트 필터 대화 상자:** 차트 필터 패널에서 열 수 있는 차트 필터 대화 상자에 대해 설명합니다. 이 대화 상자에서는 차트 필터를 추가하거나 세부적으로 설정할 수 있습니다.

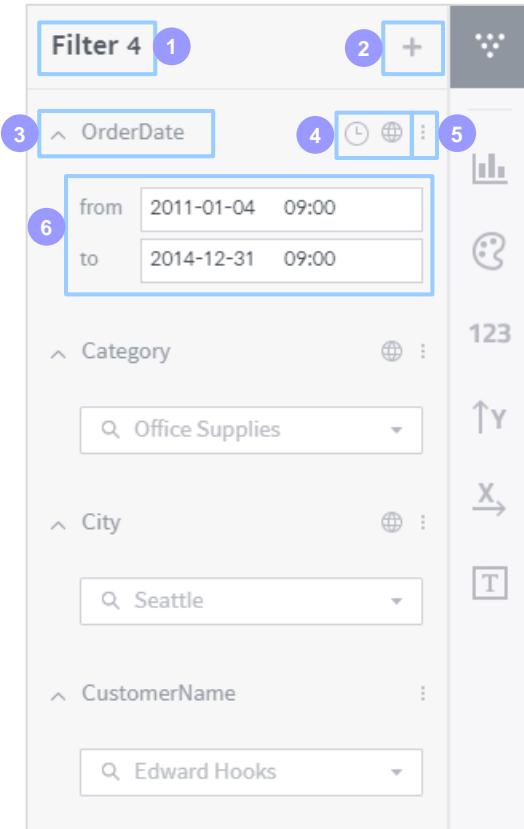
5.5.1 자동으로 포함되는 필터들

다음에 해당하는 컬럼들의 필터는 별도의 차트 필터 추가를 하지 않아도 자동으로 포함됩니다.

- **타임스탬프 컬럼 필터:** metatron Discovery의 Druid 엔진의 시계열 특성 때문에 시간 조건 필터링이 필수적으로 사용됩니다.
- **추천 필터:** 데이터 소스에서 '추천 필터'로 지정된 컬럼 필터들입니다.
- **'글로벌' 속성이 적용된 대시보드 필터:** 해당 대시보드에 등록된 모든 차트에 공통적으로 적용되는 필터입니다.

5.5.2 차트 필터 패널

옵션 영역에서 맨 위에 있는 아이콘을 클릭하면 아래와 같이 차트 필터 패널이 나타나서 현재 등록된 필터들의 기본적인 사항을 조회·수정할 수 있습니다.



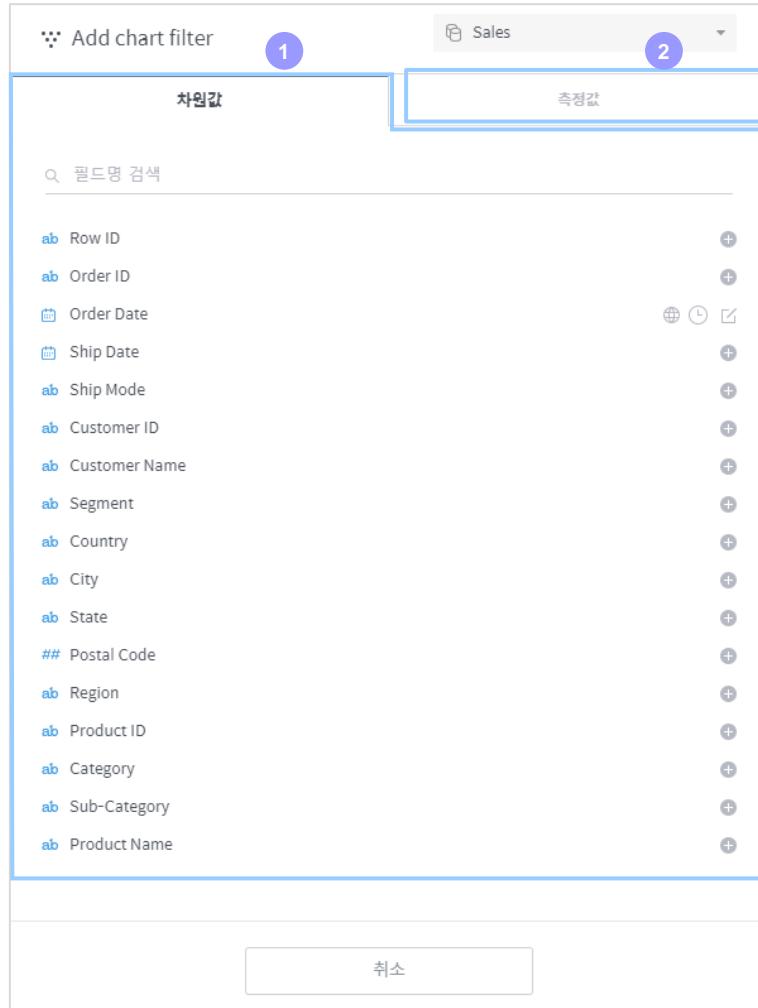
- ① **필터 수:** 현재 등록된 필터의 개수를 나타냅니다.
- ② **필터 추가/변경:** 클릭하면 새로운 필터를 추가하거나 기존 필터를 세부적으로 설정할 수 있는 대화 상자가 나타납니다.
- ③ **필터 이름:** 해당 필터의 컬럼 이름을 보여줍니다.
- ④ **필터 속성:** 해당 필터의 기본 속성을 아이콘으로 간단하게 보여줍니다.
 - : 해당 필터의 컬럼이 타임스탬프 컬럼임을 나타냅니다.
 - : 해당 필터가 '글로벌' 속성이 적용된 대시보드 필터임을 가리킵니다.
- ⑤ **더 보기:** 해당 필터를 초기화하거나 상세하게 설정할 수 있습니다.
- ⑥ **필터링 범위:** 차트에 표시할 데이터 값의 범위를 설정합니다.

5.5.3 차트 필터 대화 상자

차트 필터 패널 상단에서 버튼을 클릭하거나 각 필터 영역에서 버튼을 클릭하면 차트 필터 대화 상자를 열 수 있습니다. 이 대화 상자에서는 새로운 필터를 추가하거나 기존 필터를 세부적으로 설정할 수 있습니다.

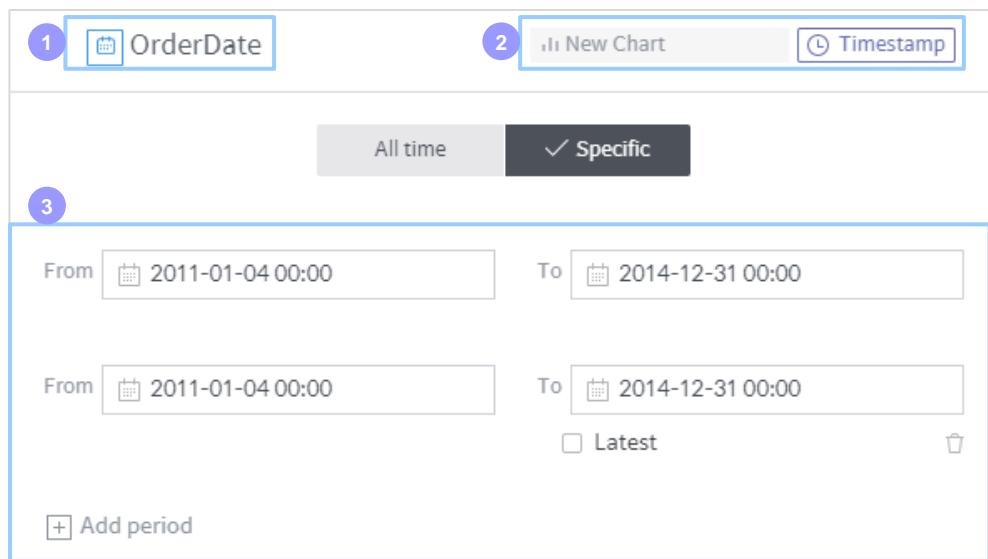
A. 차트 필터 대화 상자 구성

차트 필터 대화 상자는 다음과 같이 차원값과 측정값 영역으로 구성됩니다.



- ① **차원값:** 해당 차트와 연동된 데이터 소스의 차원값을 필터로 지정할 수 있습니다. 차원값 필터 설정 방식은 아래 C절을 참조하십시오.
 - 표시가 있는 차원값은 타임스탬프 속성임을 나타내고, 해당 차원값은 타임스탬프 필터를 설정할 수 있습니다. 타임스탬프 필터 설정 방식은 아래 B절을 참조하십시오.
- ② **측정값:** 해당 차트와 연동된 데이터 소스의 측정값을 필터로 지정할 수 있습니다. 측정값 필터 설정 방식은 아래 D절을 참조하십시오.

B. 타임스탬프 컬럼 필터 설정



① **이름:** 해당 필터가 적용된 컬럼 이름을 보여줍니다.

② **필터 속성:** 해당 필터의 기본 속성을 보여줍니다.

- **Timestamp:** 해당 필터의 컬럼이 타임스탬프 컬럼임을 나타냅니다.
- **차트 이름:** 해당 필터가 적용되는 차트의 이름을 나타냅니다.

③ **기본 상태 설정:** 차트에 나타낼 시간 범위를 결정합니다.

- **모두:** 시간 필터링을 하지 않은 전체 기간의 컬럼 내용을 차트에서 볼 수 있습니다.
- **기간:** 특정 기간과 시간 동안으로 필터링한 컬럼 내용만 차트에서 볼 수 있습니다.
‘현재시간으로 설정’버튼에 체크하면 기준 시간 중 최근 시간이 현재시간으로 설정됩니다.
- **시간 범위 추가/삭제:** 우측 하단에 있는 **[+ Add period]** 및 **[trash icon]** 아이콘을 이용하여 시간 범위를 추가하거나 삭제할 수 있습니다.

C. 차원값 컬럼 필터 설정

The screenshot shows the metatron workspace interface. On the left, a list of names with their corresponding values is displayed. On the right, the filter configuration for the 'CustomerName' column is shown.

Left Panel:

- Header: CustomerName
- Buttons: Single (selected), Multiple
- List: 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
- Bottom Buttons: Search, Filter, Sort, Refresh

Right Panel:

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

- ① **이름:** 해당 필터가 적용된 컬럼 이름을 보여줍니다.
- ② **차트이름:** 해당 필터가 적용되는 차트의 이름을 보여줍니다.
- ③ **범위 선택:** 선택한 필터의 컬럼에 들어있는 데이터 범주 중 필터링하여 차트에 표시할 범위를 선택합니다.
 - 단건: 하나의 데이터 범주만 선택하여 차트에서 표시할 수 있습니다.
 - 다건: 여러 개의 데이터 범주를 선택하여 차트에 표시할 수 있습니다.
- ④ **이름으로 검색:** 선택한 필터에 들어있는 데이터 값을 이름으로 검색합니다.
- ⑤ **필터링:** 목록에 표시할 데이터 범주들을 선별하여 나타낼 수 있습니다.
 - Wild card: 선택한 필터에 들어있는 데이터를 특정 문자로 검색하여 선별하여 볼 수 있습니다. 예를 들어 L이라는 글자로 시작되는 데이터만 선별해서 보고 싶을 때는 '시작 단어'에 L을 입력하여 적용 버튼을 누릅니다. 89로 끝나는 데이터만 선별해서 보고 싶을 때는 '종료 단어'에 89를 입력하여 적용 버튼을 누릅니다. Cart라는 단어가 들어간 데이터만 선별해서 보고 싶을 때는 '포함'에 Cart를 입력하여 적용 버튼을 누릅니다.
 - Condition: 해당 컬럼의 측정값을 설정한 조건대로 선별하여 볼 수 있습니다. 맨 왼쪽 칸에서 선별하여 보기 원하는 측정값을 선택합니다. 해당 측정값의 합계/평균/카운트/최소값/최대값이 사용자가 선별하여 보길 원하는 숫자 값보다 같거나/초과되거나/미만이거나/이상이거나/이하인 데이터가 선별되어 나타납니다.

- Limitation: 해당 컬럼의 측정값을 설정한 조건의 상위 또는 하위 랭킹으로 선별하여 볼 수 있습니다. 선별하여 보기를 원하는 측정값을 선택한 뒤, 해당 측정값의 합계/평균/카운트/최소값/최대값이 사용자가 선별하여 보길 원하는 숫자만큼 상위/하위 랭크로 선별되어 나타납니다.

6 정렬: 목록에 열거된 데이터를 정렬하는 방식을 선택합니다.

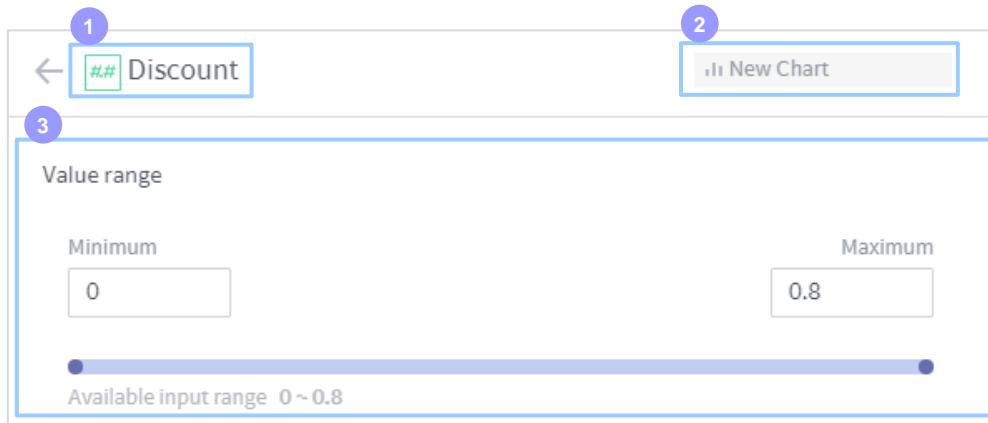
- 빈도수 오름차순/내림차순: 해당 데이터를 빈도수의 오름차순/내림차순으로 정렬합니다.
- 영숫자 오름차순/내림차순: 해당 데이터 이름의 오름차순/내림차순으로 정렬합니다.

7  아이콘: 목록에 열거된 데이터 범주 중에서 특정 항목만을 표시할 때 사용합니다.

표시를 원하는 항목 옆 아이콘만 활성화 시킨 후 맨 위에 있는  아이콘을 클릭하면 선택한 항목만 표시됩니다.

8 **Defined value:** 컬럼 들어있지 않은 데이터 범주를 필터 조건으로 추가하는 데 사용됩니다.
현재 데이터 소스에는 없지만 나중에 추가적으로 들어올 수 있는 데이터 범주를 미리 예측하여 필터를 만들기 위한 것입니다.

D. 측정값 컬럼 필터 설정



- ① **이름:** 해당 필터가 적용된 컬럼 이름을 보여줍니다.
- ② **차트이름:** 해당 필터가 적용되는 차트의 이름을 보여줍니다.
- ③ **범위 선택:** 선택한 필터의 컬럼에 들어있는 데이터 값 중에서 차트에 표시할 최댓값과 최솟값을 지정합니다.

5.6 차트 스타일 설정

데이터를 피봇팅하고 나면 우측에 차트 스타일을 설정할 수 있는 옵션 메뉴가 나타납니다. 메뉴의 구성은 각 차트의 종류별로 다르게 나타납니다. 5.6.1절에서는 모든 차트 유형에 일반적으로 적용되는 설정 항목을 설명하고, 5.6.2절에서는 차트 유형별로 고유한 '공통 설정' 유형에 대해서 설명합니다.

5.6.1 차트 스타일 설정 메뉴

본 절에서는 차트 스타일 설정 메뉴를 구성하는 각 항목별로 설정 방식을 설명합니다. 사용하는 차트 유형에 따라 아래 제시된 항목 중 일부가 사용되지 않을 수 있음을 유의하십시오.

공통 설정

여기서는 해당 차트의 모양을 정의합니다. 공통 설정 항목은 차트마다 다르므로 5.6.2절에서 차트별 공통 설정에 관한 설명을 참조하십시오.



색상 설정

여기서는 차트에 들어가는 각종 색상을 정의합니다.

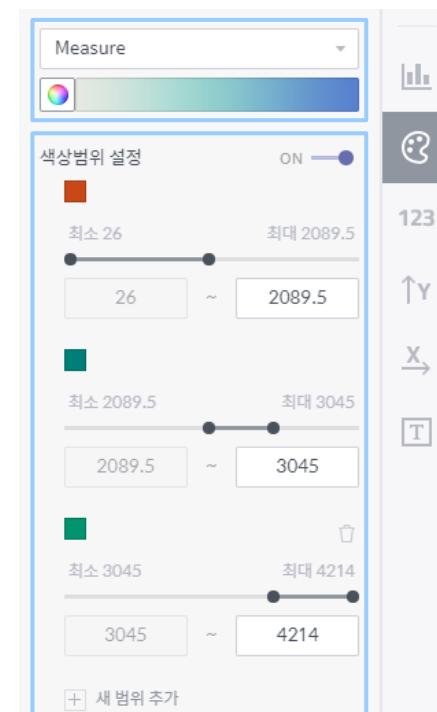
① **그래프 색상 설정:** 그래프에서 데이터 표시 항목별 색상을 구분하는 기준을 정한 후 색상 테마를 선택합니다.

- Series: 측정값의 종류에 따라 색상을 구분합니다.
- Dimension: 차원값의 종류에 따라 색상을 구분합니다.
- Measure: 측정값의 크기에 따라 색상을 구분합니다.

② **글자색:** 차원값의 텍스트 색상을 선택합니다.

③ **영역색:** 표 배경 색상을 선택합니다.

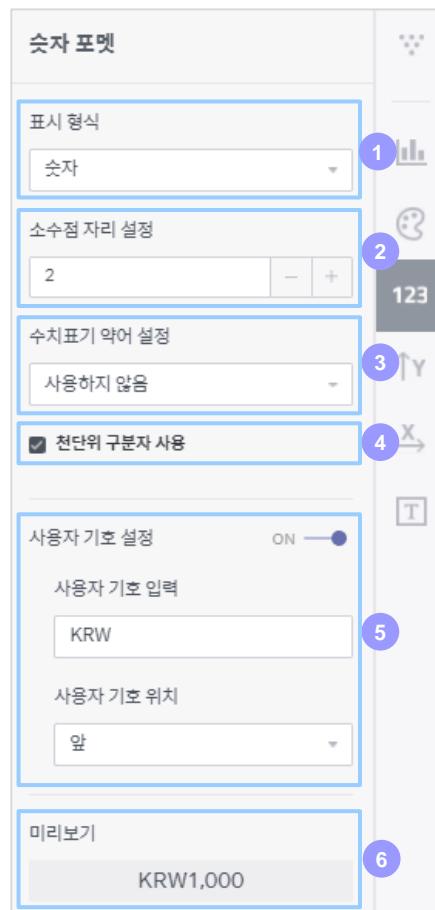
④ **색상범위 설정:** 데이터 표시 색상 구분 기준을 'Measure'로 선택할 때 나오는 항목으로서, ON으로 설정하면 측정값의 범위에 따라 색상을 다르게 나타낼 수 있습니다. 색상범위는 최저 구간부터 시작해서 원하는 개수만큼 세분화할 수 있는데, 새로운 구간을 추가하려면 현재의 마지막 구간의 최댓값을 먼저 조정한 다음 '새 범위 추가' 버튼을 클릭해야 합니다.



숫자 포맷

여기서는 차트 내 그래프에 텍스트로 나타나는 데이터 값의 표시 형식을 정의합니다. 이 기능을 사용하려면 데이터 레이블 설정 메뉴에서 레이블 표시 기능을 먼저 켜 주십시오.

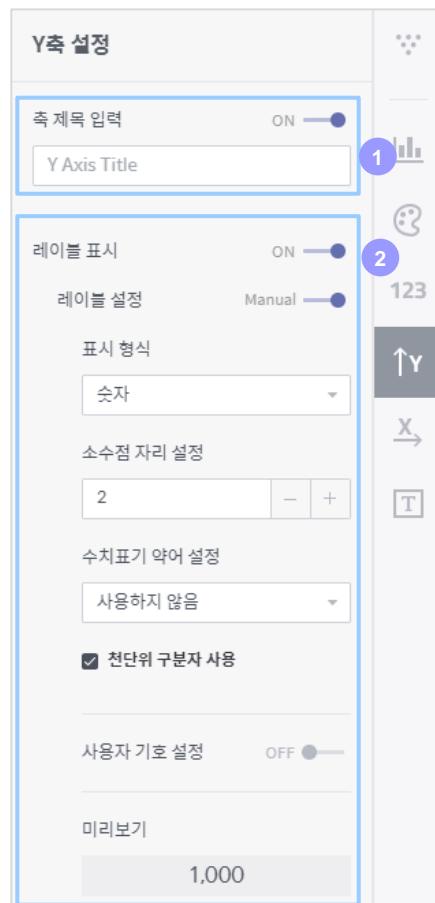
- ① **표시 형식:** 데이터 값을 숫자, 통화, 퍼센트, 지수 중 어떤 형식으로 표시할지 선택합니다.
- ② **소수점 자리 설정:** 데이터 값을 소수점 몇 번째 자리까지 표시할지 결정합니다.
- ③ **수치표시 약어 설정:** 데이터 값의 자릿수가 클 경우 천 단위(K), 백만 단위(M), 십억 단위(B) 중 하나를 약어로 설정할 수 있습니다. '자동 조정'을 선택하면, 데이터 값들의 자릿수에 가장 적절한 단위가 자동으로 결정됩니다.
- ④ **천단위 구분자 사용:** 데이터 값을 천단위 구분자를 사용하여 표시할 것인지 선택합니다.
- ⑤ **사용자 기호 설정:** 데이터 값들의 앞/뒤에 사용자가 원하는 텍스트를 삽입하여 표시할 수 있습니다.
- ⑥ **미리 보기:** 정의한 숫자 형식에 따른 결과를 예시로 보여줍니다.



Y축 설정(차트 유형 세로형 기준)

여기서는 차트의 Y축 표시 방식을 정의합니다. 공통 설정 항목에서 차트 유형을 '가로형'으로 바꾸면 X축과 Y축 설정 항목이 뒤바뀝니다.

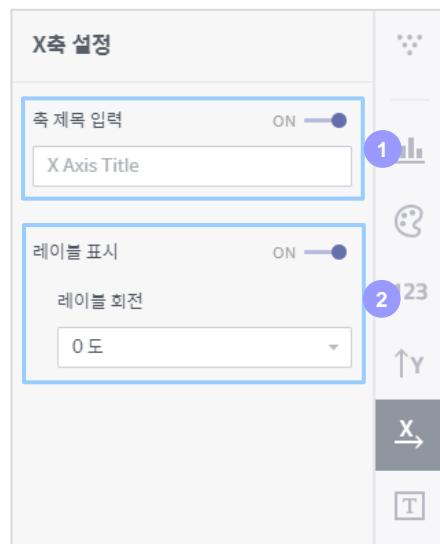
- ① **축 제목 입력:** 차트의 Y축에 제목을 입력할 수 있습니다. 해당 기능을 사용하지 않으면 Y축에 제목이 나타나지 않게 됩니다.
- ② **레이블 표시:** 차트의 Y축에 데이터 레이블을 표시할 것인지 선택합니다. 해당 기능을 사용하지 않으면 Y축에 데이터 레이블은 나타나지 않게 됩니다.
 - **레이블 설정:** Y축 데이터 레이블에 표시되는 숫자의 형식을 지정합니다. 자동으로 설정하면 '숫자 포맷' 항목의 설정값이 동일하게 반영되며, 수동으로 설정하면 Y축 데이터 레이블만의 고유한 형식을 지정할 수 있습니다.



X축 설정(차트 유형 세로형 기준)

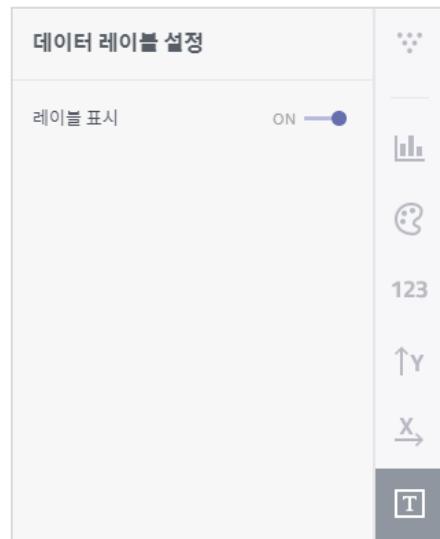
여기서는 차트의 X축 표시 방식을 정의합니다. 공통 설정 항목에서 차트 유형을 '가로형'으로 바꾸면 X축과 Y축 설정 항목이 뒤바뀝니다.

- ① **축 제목 입력:** 차트의 X축에 제목을 입력할 수 있습니다. 해당 기능을 사용하지 않으면 X축에 제목이 나타나지 않게 됩니다.
- ② **레이블 표시:** 차트의 X축에 데이터 레이블을 표시할 것인지 선택합니다. 해당 기능을 사용하지 않으면 X축에 데이터 레이블은 나타나지 않게 됩니다.
 - **레이블 회전:** 차트의 X축에 나타나는 데이터 레이블을 0도/45도/90도 중 어떤 각도로 표시할 것인지 선택합니다.



데이터 레이블 설정

차트 내 그래프에 데이터 값을 표시할 것인지 여부를 선택합니다.

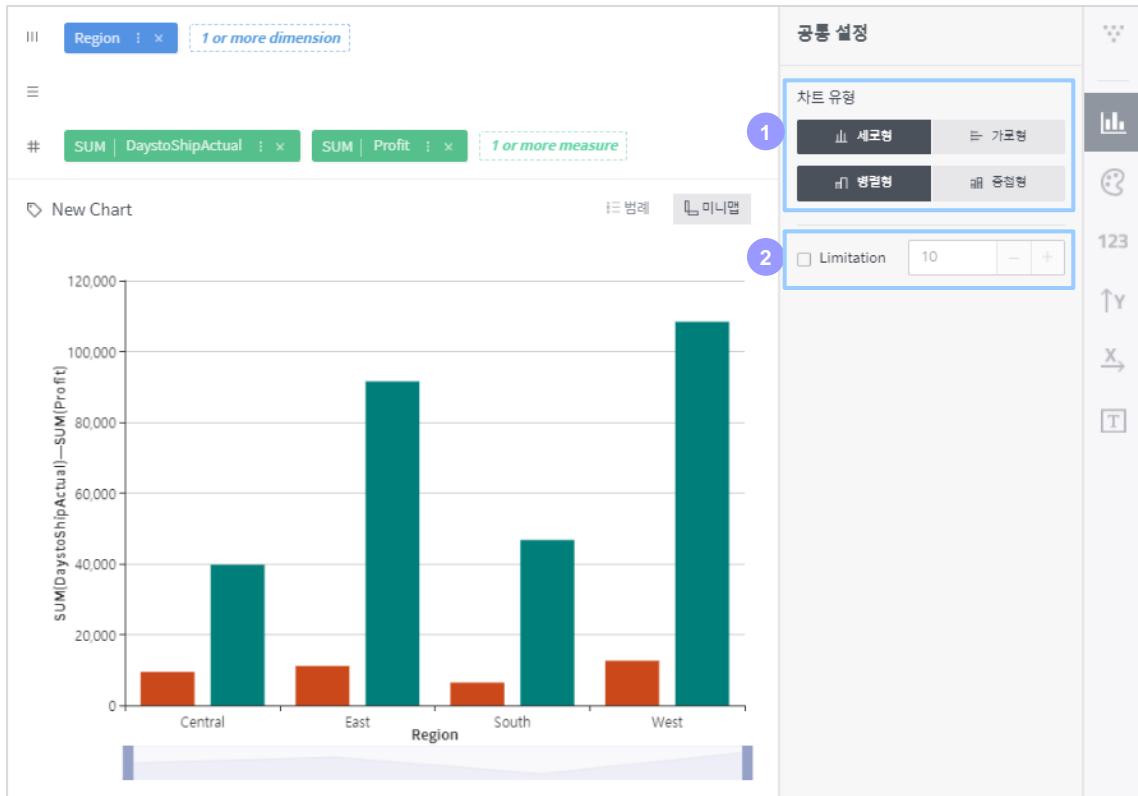


5.6.2 차트 유형별 ‘공통 설정’ 항목

본 절에서는 가장 보편적으로 사용되는 상위 6개 차트(막대형 차트, 표, 선형 차트, 분산형 차트, 히트맵, 원형 차트)의 스타일 설정 방법에 대해 설명합니다.

A. 막대형 차트

차원값 컬럼을 구성하는 각 범주 항목에 속한 데이터 값이 막대 모양으로 표시됩니다.



1 차트 유형: 차트의 모양을 정의합니다.

- 세로형: 차원값 축이 세로를 기준으로 하여 데이터 값 막대가 세로로 나타납니다.
- 가로형: 차원값 축이 가로를 기준으로 하여 데이터 값 막대가 가로로 나타납니다.
- 병렬형: 측정값을 2개 이상 선택했을 때 측정값별로 각각 다른 막대로 병렬하여 나타냅니다.
- 중첩형: 측정값을 2개 이상 선택했을 때 모든 측정값을 한 막대에 중첩시켜 나타냅니다.

2 Limitation: 차트에 나타나는 컬럼의 개수를 결정합니다.

B. 표

열/행 선반에 올려놓은 차원값 컬럼들의 범주 항목을 토대로 표 블록이 형성되며 그에 상응하는 측정값이 교차 영역에 텍스트로 표시됩니다.

	SalesaboveTa...
null	
SUM(Profit)	286,347
AVG(Profit)	28.65
SUM(Sales)	2,297,354

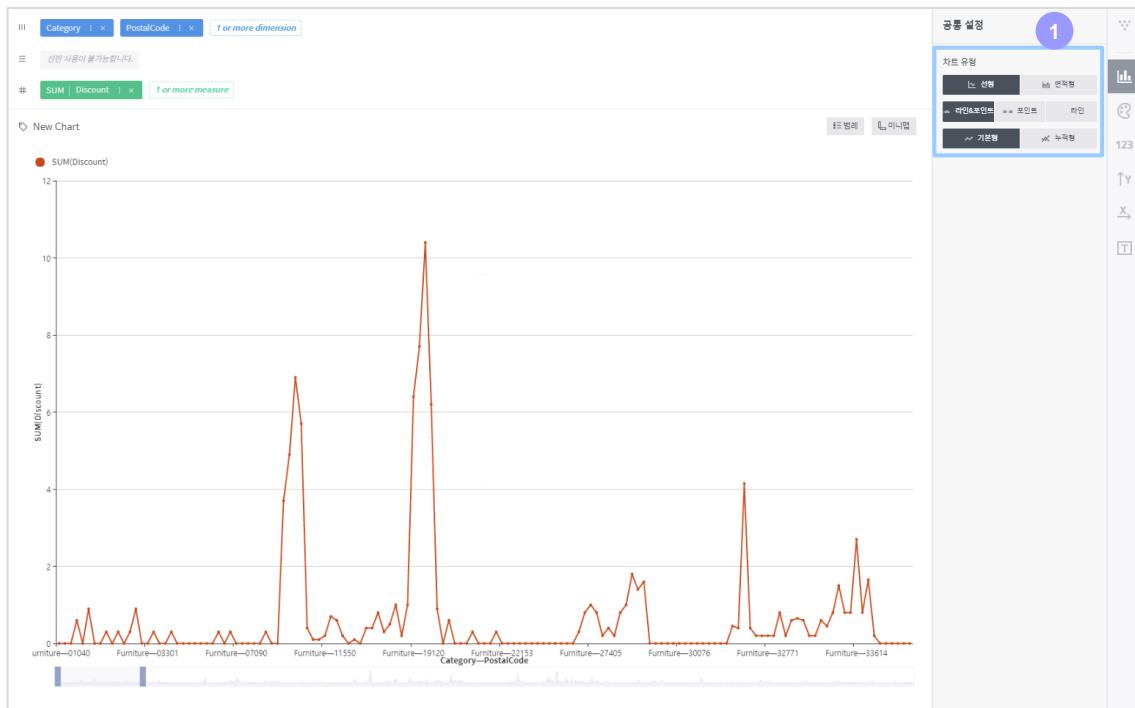
1 차트 유형: 차트의 모양을 정의합니다.

- **피봇 데이터:** 분류 기준이 되는 차원값 범주가 동일한 측정값끼리는 한 셀에 집계(SUM, MIN, MAX 등)되는 방식으로 보여집니다.
- **원본 데이터:** 원본 측정값이 집계되지 않은 채로 특정한 차원값 컬럼을 기준으로 전부 출력됩니다.
- **세로보기:** 측정값의 데이터를 표에서 세로로 표시합니다. 원본 데이터 유형으로 표를 나타낼 때는 사용할 수 없습니다.
- **가로보기:** 피봇 데이터 유형으로 표를 나타낼 경우 표를 가로보기 할 수 있습니다. 측정값의 데이터를 표에서 가로로 표시합니다.

2 Show Head Column: 헤드 컬럼에 표시되는 텍스트 정렬 방식을 가로와 세로 별로 설정할 수 있습니다. 원본데이터로 표시할 경우에는 헤드 컬럼은 필수로 표시됩니다. 피봇 데이터 유형일 경우에 헤드 컬럼을 사용하지 않을 수 있습니다.

C. 선형 차트

차원값 컬럼을 구성하는 각 범주 항목에 속한 데이터 값이 점 모양으로 표시됩니다. 인접하는 범주 항목의 점끼리는 서로 연결되어 변화 추이를 확인할 수 있습니다.

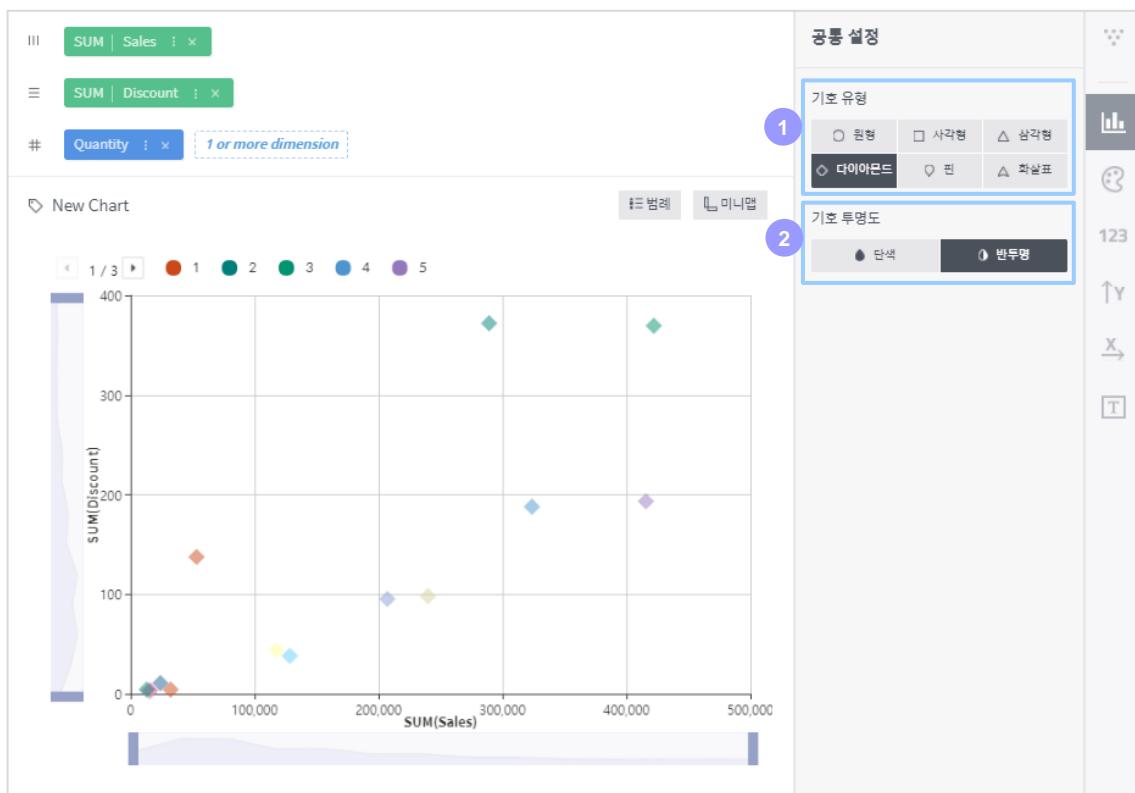


1 차트 유형: 차트의 모양을 정의합니다.

- 선형: 측정값을 기준점으로 선을 이어서 차트를 나타냅니다.
- 면적형: 선으로 이어진 면적에 색상을 입혀 차트를 나타냅니다.
- 라인&포인트: 측정값을 기준점으로 한 점과 그 점을 연결한 선을 함께 나타냅니다.
- 포인트: 포인트는 측정값을 기준으로 한 점만 나타냅니다.
- 라인: 선의 연결만을 나타냅니다.
- 기본형: 측정값을 그대로 차트에 나타냅니다.
- 누적형: 측정값을 누적한 값을 차트에 나타냅니다.

D. 분산형 차트

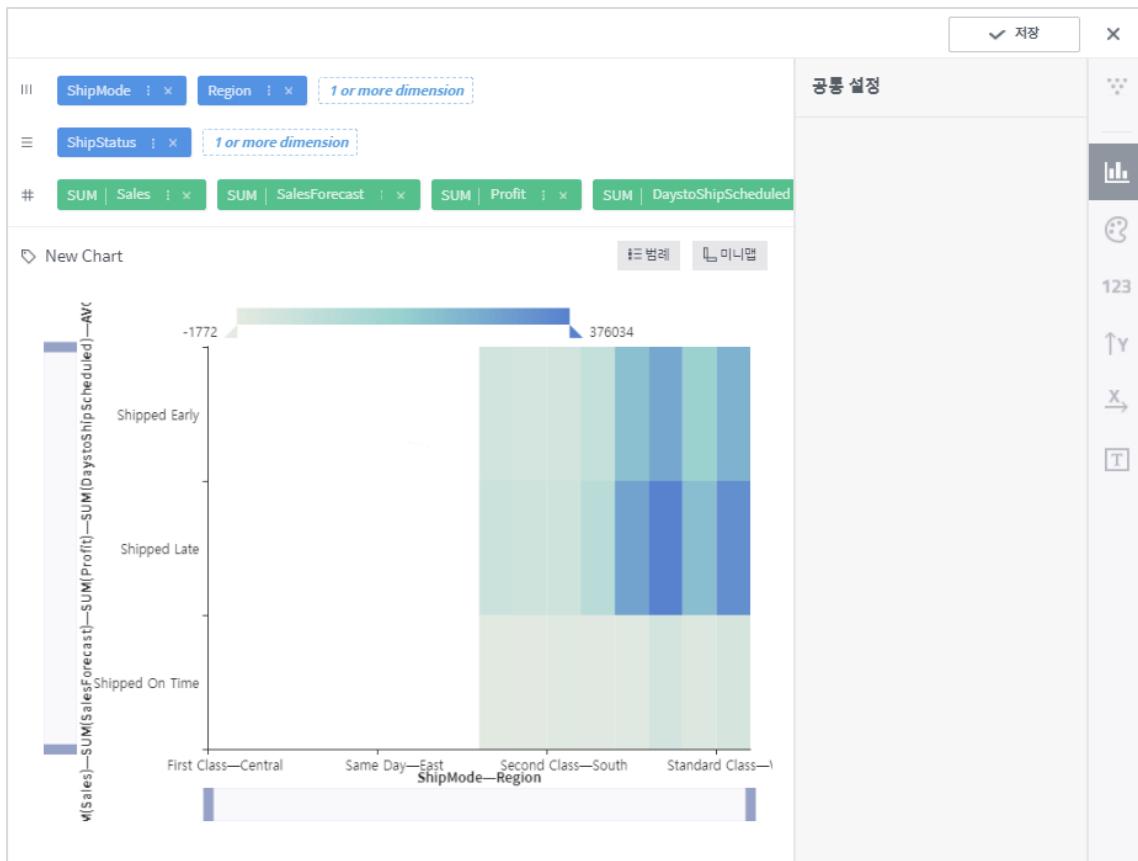
차원값 컬럼을 구성하는 각 범주 항목에 속한 데이터 값이 정의된 기호 모양으로 표시됩니다.



- 1 **기호 유형:** 차트에 표시되는 기호의 모양을 설정합니다.
- 2 **기호 투명도:** 차트에 표시되는 기호의 투명도를 설정합니다. 단색/반투명 중 선택하여 나타낼 수 있습니다.

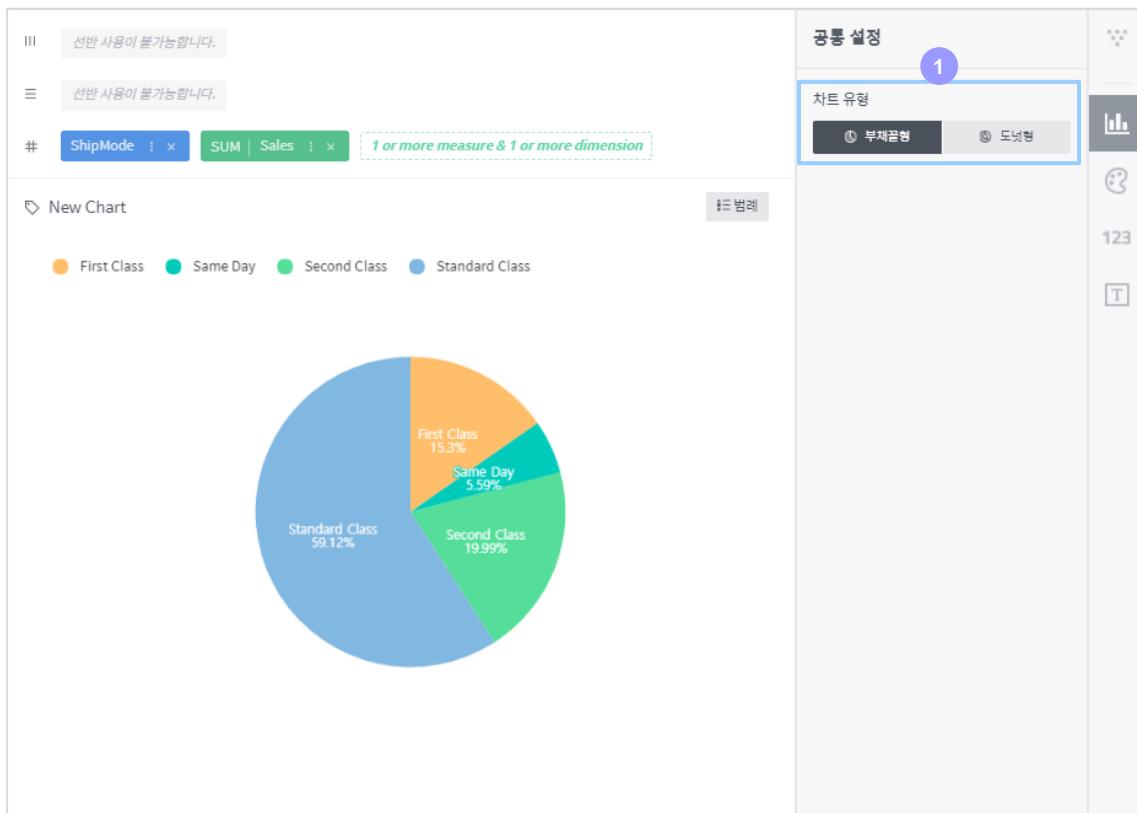
E. 히트맵

교차선반에 올려진 측정값 컬럼의 각 데이터 값이 색상으로 표시됩니다. 데이터 값이 클수록 색상 농도가 짙어집니다. 히트맵의 공통 설정 항목에는 설정할 사항이 없습니다.



F. 원형 차트

차원값 컬럼의 각 범주 항목별로 차지하는 비중을 시각화하는 차트입니다.



① **차트 유형:** 차트의 모양을 정의합니다.

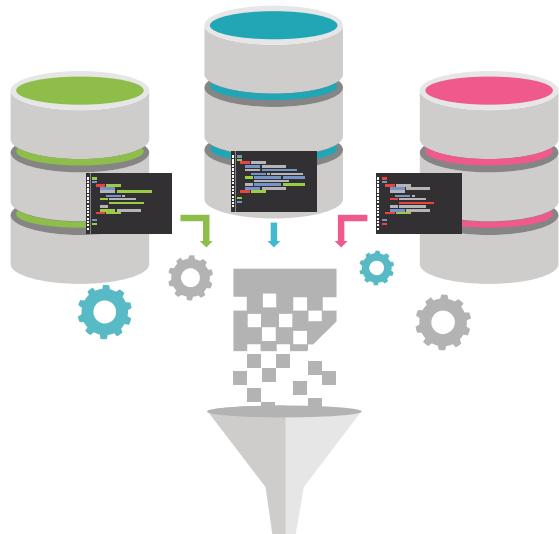
- **부채꼴형:** 차트가 원형으로 나타납니다.
- **도넛형:** 차트가 도넛형으로 나타납니다.



metatron 3.0

5부

노트북 이용하기



Copyright © SK telecom

SK Telecom 의 사전 승인 없이
본 내용의 전부 또는 일부에 대한 복사, 배포, 사용을 금합니다.

목 차

1. 노트북 개요.....	3
2. 노트북 서버 초기 설정.....	4
3. 노트북 생성하기	5
4. 노트북 기능 사용하기.....	10
4.1 노트북 상세 조회.....	10
4.2 노트북 코딩	11
4.3 노트북 API 등록	12

1. 노트북 개요



metatron Discovery 노트북은 **머신러닝을 기반으로** 하는 외부 분석 도구를 통한 데이터 분석 환경을 제공합니다. 기본 특징은 다음과 같습니다.

- 분석을 위한 데이터는 Druid 엔진에 저장된 데이터 소스를 불러오거나 워크북에 삽입된 대시보드 또는 차트에서 추출할 수 있습니다.
- 가용한 외부 분석 도구와 언어는 다음과 같습니다.

외부 분석 도구	사용 언어
Jupyter	R
	Python
Zeppelin	Spark

이러한 도구들을 이용하여 분석가는 데이터를 원하는 대로 분석·정리·예측하고 분석 로직을 다른 사람들과 공유할 수 있습니다.

본 단원은 다음과 같이 구성되어 있습니다.

2장 노트북 서버 초기 설정: 위에서 설명한 Jupyter 또는 Zeppelin을 연동해서 사용하려면 이러한 분석 도구를 호스팅하는 서버를 metatron Discovery에 등록해야 합니다. 이 장에서는 Jupyter 및 Zeppelin 서버를 등록하는 절차에 대해 소개합니다.

3장 노트북 생성하기: 머신러닝 기반 분석 코드를 보관하는 각각의 문서를 '노트북'이라고 부릅니다. 이 장에서는 노트북을 생성하고 분석할 데이터를 연계하고 분석툴과 언어를 결정하는 절차에 대해 소개합니다.

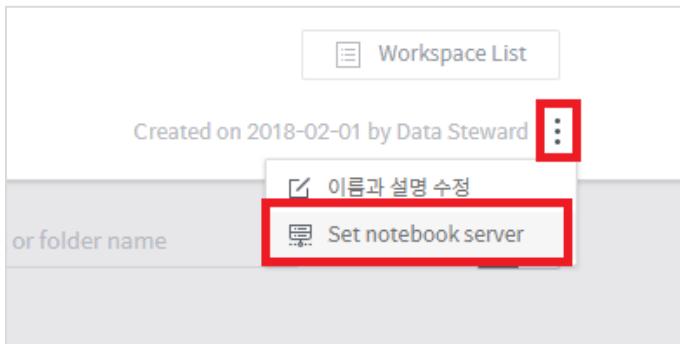
4장 노트북 활용하기: 생성된 노트북에서 코딩을 실시하고 그 결과물을 실행 및 공유하는 절차에 대해 소개합니다.

2. 노트북 서버 초기 설정

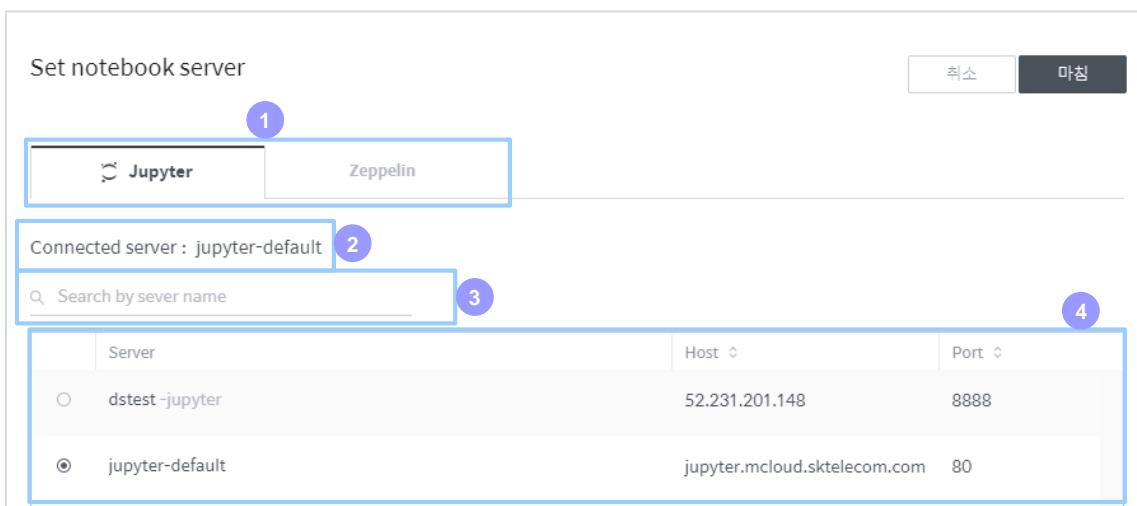
해당 워크스페이스에서 노트북 기능을 이용하여 데이터를 분석하기 위해서는 노트북 서버 초기 설정이 필요합니다. 노트북 서버 초기 설정을 하려면 노트북 서버가 등록이 되어있어야 하는데, 이에 관한 자세한 내용은 '2부 운영/관리'에서 '4. 노트북 관리'를 참조하시기 바랍니다.

노트북 서버 초기 설정 절차는 다음과 같습니다.

- 워크스페이스의 우측 상단에 있는 **:** 버튼을 클릭한 후 'Set notebook server'를 선택합니다.



- 아래와 같은 노트북 서버 설정 화면이 나타나면, 아래 항목별 설명을 참조하여 사용하고자 하는 노트북 서버를 선택한 후 '마침'을 클릭합니다.



- 서버 타입 선택:** 사용자가 등록해 놓은 Jupyter 또는 Zeppelin 서버 중 해당 워크스페이스에서 연결하여 사용하고자 하는 노트북 서버를 클릭합니다.
- Connected server:** 서버 목록에서 현재 선택된 노트북 서버의 이름입니다.
- Search by server name:** 현재 등록되어있는 노트북 서버를 이름으로 검색합니다.
- 서버 목록:** 설정한 선별 조건에 맞는 노트북 서버들을 보여줍니다.

3. 노트북 생성하기

최초 노트북 서버 설정이 완료되면 노트북을 생성할 수 있습니다. 노트북 생성 절차는 기본적으로 '어떤 데이터'를 '어떤 분석 도구와 언어'로 분석할지를 지정하는 과정입니다.

노트북 생성 절차는 다음과 같습니다.

- 워크스페이스 하단에 있는 '+ 노트북' 버튼을 클릭하면 노트북 생성에 필요한 데이터 타입을 선택할 수 있는 화면이 나타납니다.



- 노트북 생성 화면이 나타나면 노트북에서 분석하고자 하는 데이터 타입을 선택합니다.



- 데이터 소스: 해당 워크스페이스에 허용된 데이터 소스에서 데이터를 가져와서 분석합니다.
- 대시보드: 해당 워크스페이스에 저장되어 있는 워크북 내의 대시보드에서 데이터를 가져와서 분석합니다.
- 차트: 해당 워크스페이스에 저장되어 있는 기존 차트에서 데이터를 가져와서 분석합니다.
- 선택하지 않음: Zeppelin에서 SPARK 언어를 통해 분석하고자 한다면 이 항목을 선택하십시오.

3. 분석하기 원하는 데이터 소스, 대시보드, 또는 차트를 선택한 뒤, '다음' 버튼을 누르십시오.

데이터 탑입을 '데이터 소스'로 지정했을 경우

Please select a data source

1 데이터 소스 이름 검색

2 □ 오픈 데이터만 보기

3 타입 전체

4 15 salesql_link - 링크형 세일즈 데이터 [오픈 데이터]

14 salesqa_noorderdate [오픈 데이터]

13 employee_001 [오픈 데이터] **선택됨**

12 test_int_dim [오픈 데이터]

11 test_dist [오픈 데이터]

10 Economy - Tableau example data [오픈 데이터]

9 test_missing - test [오픈 데이터]

8 대학교정보 [오픈 데이터]

5 employee_001

생성일 2017-12-19

타입 수집형

공개설정 공개

사이즈 29.64 MB

Rows 300,024

차원값 ab_emp_no
차원값 ab_birth_date
차원값 ab_first_name
차원값 ab_last_name
차원값 ab_gender
차원값 ab_hire_date

취소 다음

- 1 **데이터 소스 이름 검색:** 해당 워크스페이스에 허용된 데이터 소스를 이름으로 검색합니다.
- 2 **오픈 데이터만 보기:** '오픈 데이터 소스'로 지정된 데이터 소스만 선별하여 볼 수 있습니다.
- 3 **타입:** 데이터 소스를 연결형 또는 수집형으로 선별하여 볼 수 있습니다.
- 4 **데이터 소스 목록:** 설정한 선별 조건에 맞는 데이터 소스들을 보여줍니다.
- 5 **데이터 소스 정보:** 목록에서 선택한 데이터 소스의 정보를 간략하게 보여줍니다.

데이터 탑입을 '대시보드'로 지정했을 경우

The screenshot shows the Metatron interface with the following elements:

- Left Panel:** A sidebar listing workbooks: jungwookchoi, sohnkw, test, DTT, TestWorkbook, and MyWorkbook.
- Middle Panel:** A grid view showing three items: "sales" (selected), "sales by city", and "sales by date".
- Right Panel:** A detailed view of the selected "sales" item, including:
 - Creation Date: 2018-02-05
 - Type: 수집형 (Collection Type)
 - Publication Setting: 공개 (Public)
 - Size: 1.32 MB
 - Rows: 9,993
 - Column Headers:
 - 차원값 (Dimension): OrderDate, Category, City, Country, CustomerName, Discount, OrderID, PostalCode
 - 측정값 (Measure): #
- Bottom Buttons:** 취소 (Cancel) and 다음 (Next).

1 대시보드 탐색 영역: 해당 워크스페이스 내 워크북 하나를 선택하면 그 안에 속한 대시보드들이 조회됩니다. 이 중에서 분석하고자하는 데이터가 담긴 대시보드를 선택하십시오.

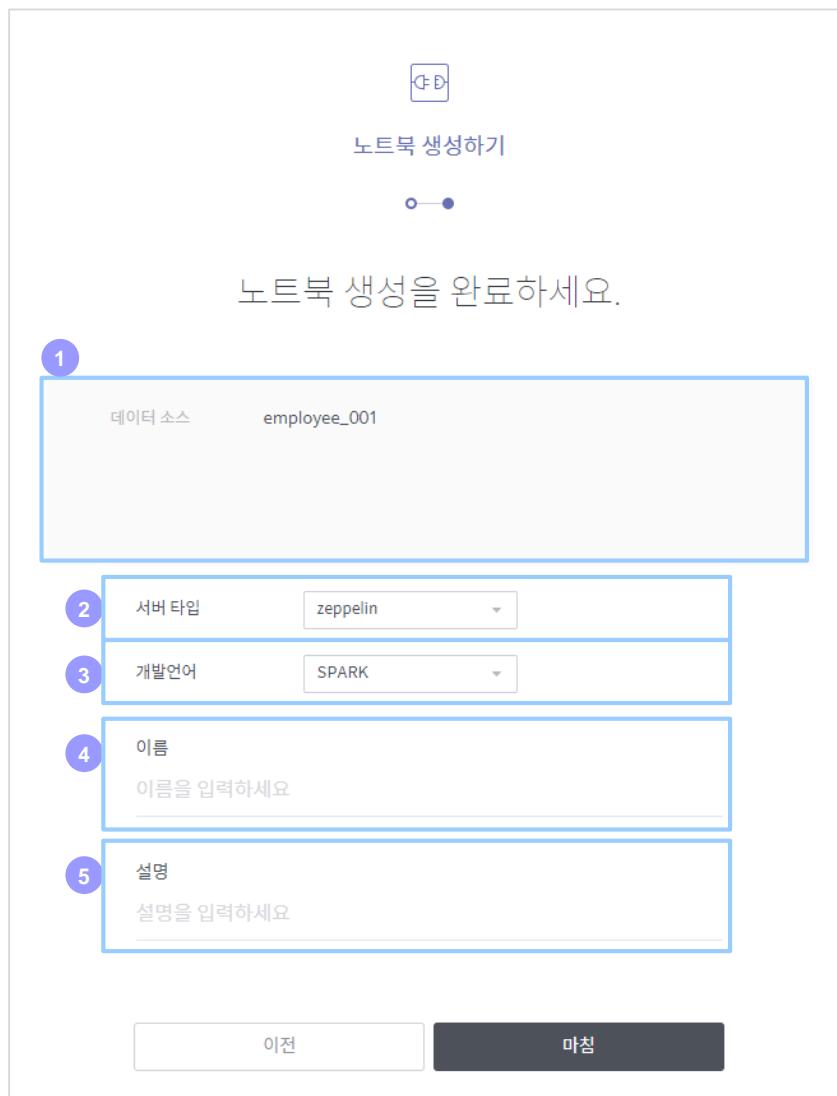
2 대시보드 정보 영역: 탐색 영역에서 선택한 대시보드의 정보를 간략하게 보여줍니다.

데이터 탑입을 '차트'로 지정했을 경우

The screenshot shows the Metatron interface with the following elements:

- Left Panel:** A sidebar listing workbooks: jungwookchoi, sohnkw, DTT, MyWorkbook, TestWorkbook, and test.
- Middle Panel:** A grid view showing three items: "sales" (selected), "sales by city", and "New Chart".
- Right Panel:** A detailed view of the selected "New Chart" item, including:
 - Creation Date: 2018-03-08
 - Column Headers:
 - 차원값 (Dimension): Category, Sub-Category, Segment, Region, State, City
 - 측정값 (Measure): #
- Bottom Buttons:** 취소 (Cancel) and 다음 (Next).

4. 노트북에 연동시킬 외부 분석 도구 및 언어 정보와 노트북의 이름과 설명을 입력한 후 '마침' 버튼을 클릭합니다.



- 1 **데이터 소스:** 앞 단계에서 선택한 데이터 소스입니다.
- 2 **서버 타입:** Jupyter와 Zeppelin 중 원하는 서버 타입을 선택합니다. 단, 초기 노트북 서버 설정에서 해당 서버 타입에 대한 설정이 선행되어야 합니다.
- 3 **개발 언어:** 분석에 사용할 개발 언어를 선택합니다. Jupyter를 선택하면 R 또는 PYTHON, 제플린을 선택하면 SPARK로 데이터 분석이 가능합니다.
- 4 **이름:** 해당 노트북의 이름을 입력합니다.
- 5 **설명:** 해당 노트북에 대한 설명을 입력합니다.

5. 노트북 생성이 완료되면, 다음과 같이 워크스페이스에서 새로 생성된 노트북을 확인할 수 있습니다. 클릭하면 해당 노트북을 사용할 수 있는 화면이 나타납니다.

The screenshot shows the 'Admin workspace' in Metatron Discovery. At the top, there are navigation links for 'Workspace List', 'Created on 2017-11-16 by Administrator', and a three-dot menu. Below the header, there's a search bar and a filter section with '전체' (All) selected. The main area displays a grid of six recently created notebooks:

Category	Name	Last Update	File Count
notebook	크라임	마지막 업데이트일 몇 초 전	0 0
workbook	dd	마지막 업데이트일 11시간 전	0 0
workbook	Sales Workbook	마지막 업데이트일 14시간 전	3 7
workbook	□□□	마지막 업데이트일 2일 전	0 0
workbook	test-exntu	마지막 업데이트일 3일 전	2 2
workbook	test	마지막 업데이트일 3일 전	1 1

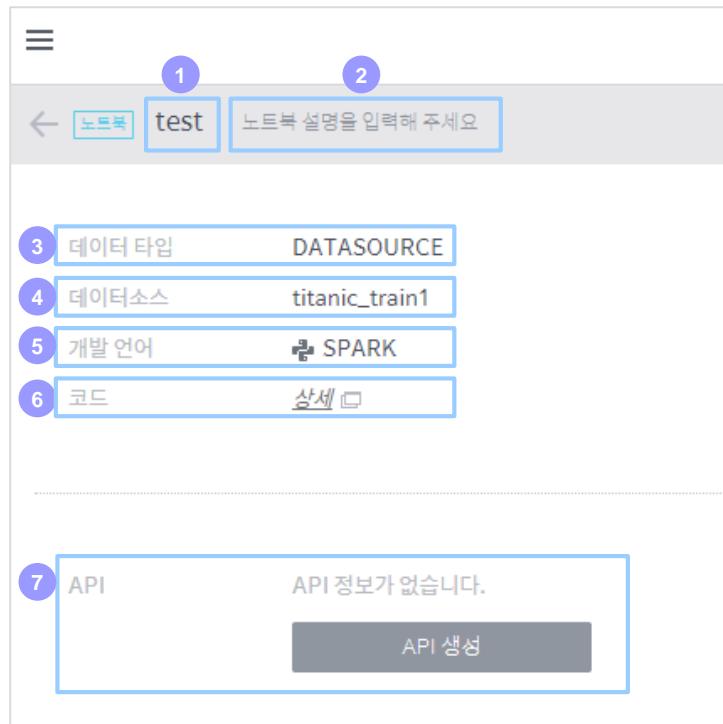
At the bottom, there are buttons for '전체 선택' (Select All), 'Clone Workbook', 'Move selections', 'Delete selections', and three buttons labeled '+ 워크북' (New Workbook), '+ 노트북' (New Notebook), and '+ 워크벤치' (New Workbench).

4. 노트북 기능 사용하기

생성된 노트북에서는 스크립트를 작성하고 그 결과를 출력하여 확인할 수 있습니다. 그리고 REST API를 통해 다른 사용자 및 시스템과 개발한 분석 코드 및 실행 결과를 공유할 수 있습니다.

4.1 노트북 상세 조회

워크스페이스에서 분석하고자 하는 노트북을 선택하면 아래와 같은 상세 화면이 나타납니다. 노트북 생성 시 입력했던 데이터 타입과 데이터 소스 이름, 개발 언어, 코드 등이 조회됩니다.



- ① **이름:** 해당 노트북의 이름입니다.
- ② **설명:** 해당 노트북에 대한 설명입니다.
- ③ **데이터 타입:** 해당 노트북에서 사용하는 데이터 타입(데이터 소스/대시보드/차트)입니다.
- ④ **데이터 소스:** 해당 노트북에서 가져오는 데이터 소스/대시보드/차트의 이름입니다.
- ⑤ **개발 언어:** 해당 노트북에서 사용하는 개발 언어입니다.
- ⑥ **코드:** 클릭하면 사용자가 해당 노트북에 적용되는 외부 도구(Jupyter/Zeppelin)의 웹 기반 노트북이 새 창으로 실행됩니다. 해당 노트북에서 사용자는 프로그래밍 언어를 활용하여 코딩 작업을 진행할 수 있습니다. (4.2절 참조)
- ⑦ **API:** 해당 노트북의 스크립트를 실행할 수 있는 API 정보입니다. 이 기능을 이용하면 사용자가 원하는 주기에 따른 스크립트 실행도 가능합니다. 'API 생성'을 클릭하면 API를 생성할 수 있는 화면이 나타납니다. (4.3절 참조)

4.2 노트북 코딩

노트북 상세 조회 화면에서 코드 란의 '상세'를 클릭하면 노트북 코딩을 위한 새 창이 나타납니다. 이 창의 상단에는 데이터셋을 로딩하는 코드가 삽입되어 있으며, 해당 셀을 실행하면 dataset 객체에 JSON 포맷의 데이터셋이 로딩됩니다.

```
1. load dataset
In [ ]: library(RMetis)
In [ ]: dataset <- datasources.get(client.url('metatron.mcloud.sktelecom.com', 80), 'ds-37')

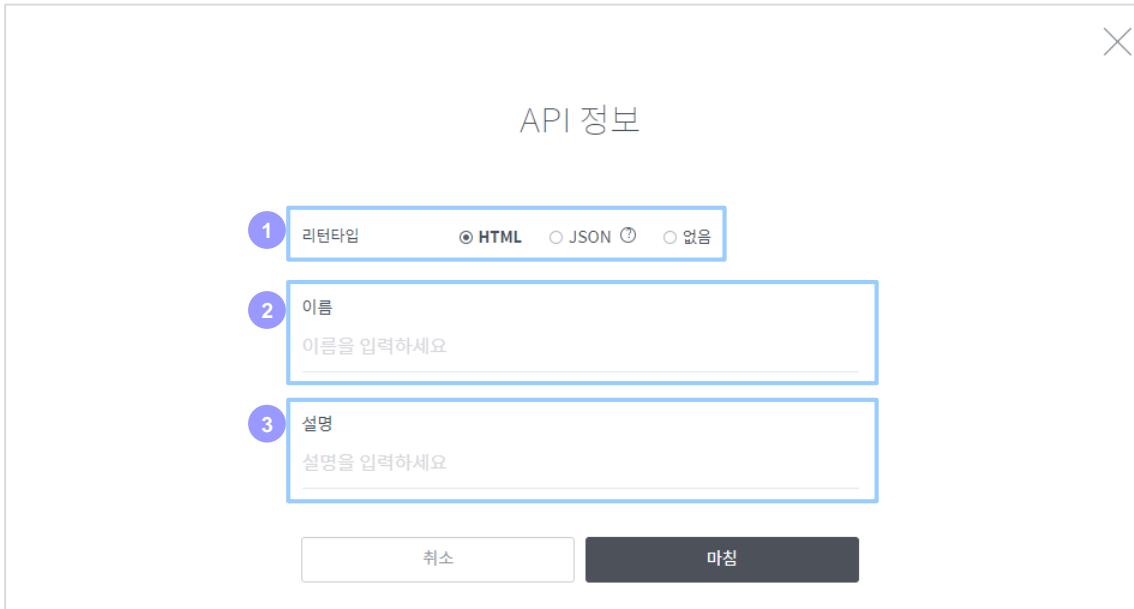
2. analyze
In [ ]:
```

위 화면은 개발 언어를 'R'로 선택하였을 때 나타나며, 'Python' 역시 위와 같은 데이터셋 로딩 셀이 삽입되어 있습니다. 코딩 작업 중에 '실행' 버튼을 클릭하면 노트북 모듈 자체 내에서 코딩 결과를 확인할 수 있습니다. 코딩 작업이 완료되면 '저장' 버튼을 클릭하십시오.

metatron Discovery에서는 이러한 오픈소스형 외부 분석 도구와의 연동 기능을 구현하는 것뿐이므로 이에 대한 자세한 설명을 제공하지 않습니다. 노트북 코딩 언어 및 인터페이스 등에 관한 자세한 정보는 해당 분석 도구와 언어에 관한 문서를 참조하십시오.

4.3 노트북 API 등록

노트북을 작성하고 나면 그 결과값을 REST API를 호출하여 반환할 수 있습니다. 이와 같이 API를 등록하면 해당 결과값을 조회할 수 있는 URL이 생성되기 때문에 분석가는 이를 통해 다른 사람들과 분석 로직을 공유할 수 있습니다.



1 리턴타입: 노트북의 결과값을 반환할 때 사용하는 타입을 선택합니다.

- HTML: 노트북 스크립트의 전체 실행 결과를 보여주는 화면을 HTML로 반환합니다.
- JSON: 노트북 스크립트에 작성된 사용자 정의 포맷의 JSON 객체를 결과로 반환합니다. 이때 metatron Discovery에서 제공하는 `response.write(...)` 함수를 사용합니다. `response.write` 함수 사용 예시 코드는 다음과 같습니다.
 - R 기반 노트북: `response.write(list(coefficient = 2, intercept = 0))`
 - Python기반 노트북: `response.write({'coefficient' : 2.5, 'intercept' : 0})`
- 없음: 노트북 스크립트를 실행하되 반환값은 제공하지 않습니다.

2 이름: 등록할 API의 이름입니다.

3 설명: 등록할 API에 대한 설명입니다.

API 정보를 모두 입력한 후 '마침' 버튼을 클릭하면 API생성이 완료되고 아래와 같은 REST API URL을 확인할 수 있습니다. 'Result' 버튼을 클릭하면 URL 실행 결과값을 팝업으로 조회할 수 있습니다.

Name	RESTful API
Description	
URL	http://metatron.mcloud.sktelecom.com/api/notebooks/rest/354599d4-444a-43ab-b966-fdadffd12e7e
Return type	HTML
API result	Result
<input checked="" type="checkbox"/> Edit API	<input type="button" value="Delete API"/>



metatron 3.0

6부

워크벤치 이용하기



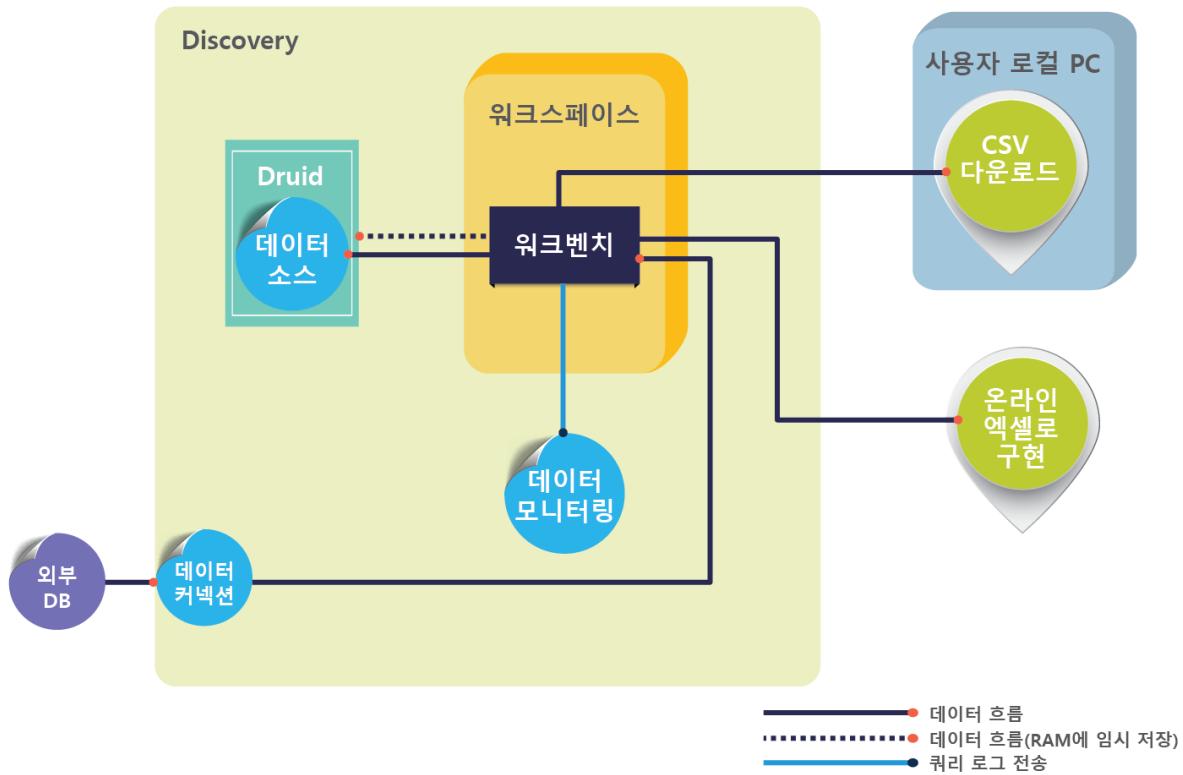
Copyright © SK telecom

SK Telecom 의 사전 승인 없이
본 내용의 전부 또는 일부에 대한 복사, 배포, 사용을 금합니다.

목 차

1. 워크벤치 개요	3
2. 워크벤치 생성하기	4
3. 워크벤치 활용하기	6
3.1 기본 정보 영역	6
3.2 스키마 및 테이블 영역	7
3.3 쿼리 에디터 영역	8
3.4 쿼리 결과 영역	9
3.5 추가 도구 영역	10
3.5.1 전역변수(Global variable) 편집 기능	10
3.5.2 Query History 목록 조회	11
3.5.3 워크벤치 네비게이션	12

1. 워크벤치 개요



metatron Discovery 워크벤치는 SQL을 기반으로 하는 데이터 전처리 및 분석 환경을 제공합니다. 주요 기능은 다음과 같습니다.

- 다양한 외부 데이터베이스를 한꺼번에 작업공간에 조회 가능
- 연동된 데이터베이스의 테이블과 컬럼을 쉽게 조회/선택하며 상세 정보 열람 가능
- 쿼리 편집 도구가 내장되어 있으며 쿼리 결과를 실시간으로 확인하고 다양하게 활용 가능
 - 쿼리 결과를 로컬 파일로 다운로드 또는 온라인 엑셀로 출력
 - 쿼리 결과를 즉시 시각화함으로써 쿼리 결과로 출력된 데이터 형태를 쉽게 파악
 - 쿼리 결과를 데이터 소스로 저장하여 워크북이나 노트북에서 분석에 활용 가능

본 단원은 다음과 같이 구성되어 있습니다.

2장 워크벤치 생성하기: SQL 기반 분석 쿼리를 보관하는 각각의 문서를 '워크벤치'라고 부릅니다. 이 장에서는 워크벤치를 생성하고 분석할 데이터 커넥션을 연계하는 절차에 대해 소개합니다.

3장 워크벤치 활용하기: 생성된 노트북에서 쿼리를 작성·실행하고 그 결과물을 시각화하는 절차에 대해 소개합니다.

2. 워크벤치 생성하기

해당 워크스페이스에서 워크벤치를 사용하기 위해서는 워크벤치용 데이터 커넥션이 설정되어 있어야 합니다. 이에 관한 자세한 내용은 '2부 운영관리'에서 '2.2 데이터 커넥션'을 참조하시기 바랍니다.

워크벤치 생성 절차는 다음과 같습니다.

- 워크스페이스 하단에 있는 '+ 워크벤치' 버튼을 클릭하면 워크벤치에서 데이터 분석에 사용할 데이터 커넥션을 연결할 수 있는 화면이 나타납니다.



- 사용자가 연결하여 사용하고자하는 워크벤치용 데이터 커넥션을 선택한 후 '다음' 버튼을 클릭합니다.

워크벤치 생성하기

데이터 커넥션을 선택해 주세요

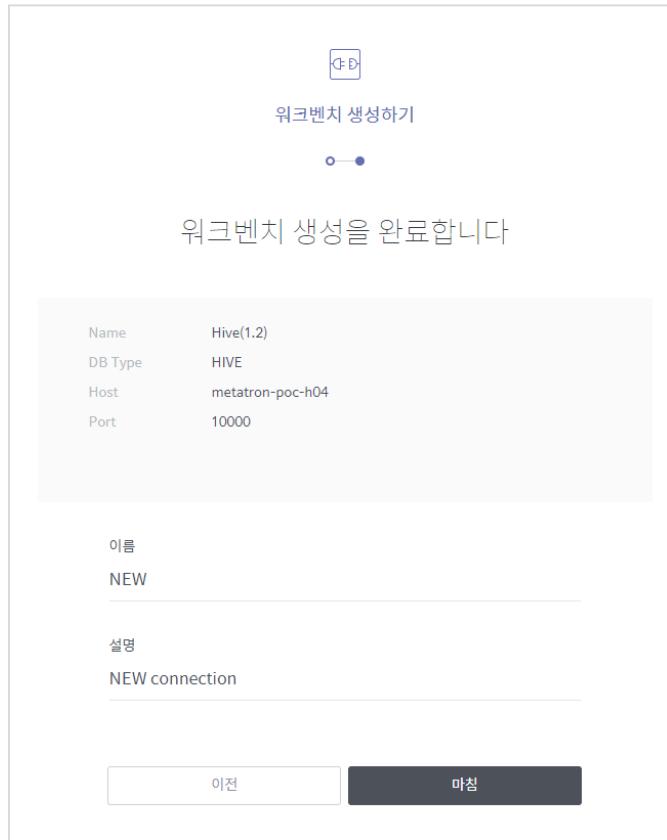
No.	데이터 커넥션	타입	Host	Port	계정 타입	수정일
2	Hive(2.3)	HIVE	metatron-po...	10000	관리자 직접 입력	2018-01-10
1	Hive(1.2)	HIVE	metatron-po...	10000	관리자 직접 입력	2017-11-23

더보기 ▾

취소 다음

- 데이터 커넥션 이름으로 검색:** 해당 워크스페이스에 허용된 데이터 커넥션을 이름으로 검색합니다.
- DB Type:** 데이터 커넥션을 데이터베이스 타입(Oracle/MySQL/Hive/presto/Tibero)별로 선별하여 볼 수 있습니다. '모두'로 선택하면 모든 DB타입의 데이터 커넥션을 볼 수 있습니다.
- 계정 타입:** 데이터 커넥션을 설정된 계정 타입(관리자 직접 입력/사용자의 계정 사용/워크벤치 접속 시 직접 입력)별로 선별하여 볼 수 있습니다. 'All'로 선택하면 모든 계정 타입의 데이터 커넥션을 볼 수 있습니다.
- 데이터 커넥션:** 설정한 선별 조건에 맞는 데이터 커넥션들을 보여줍니다.

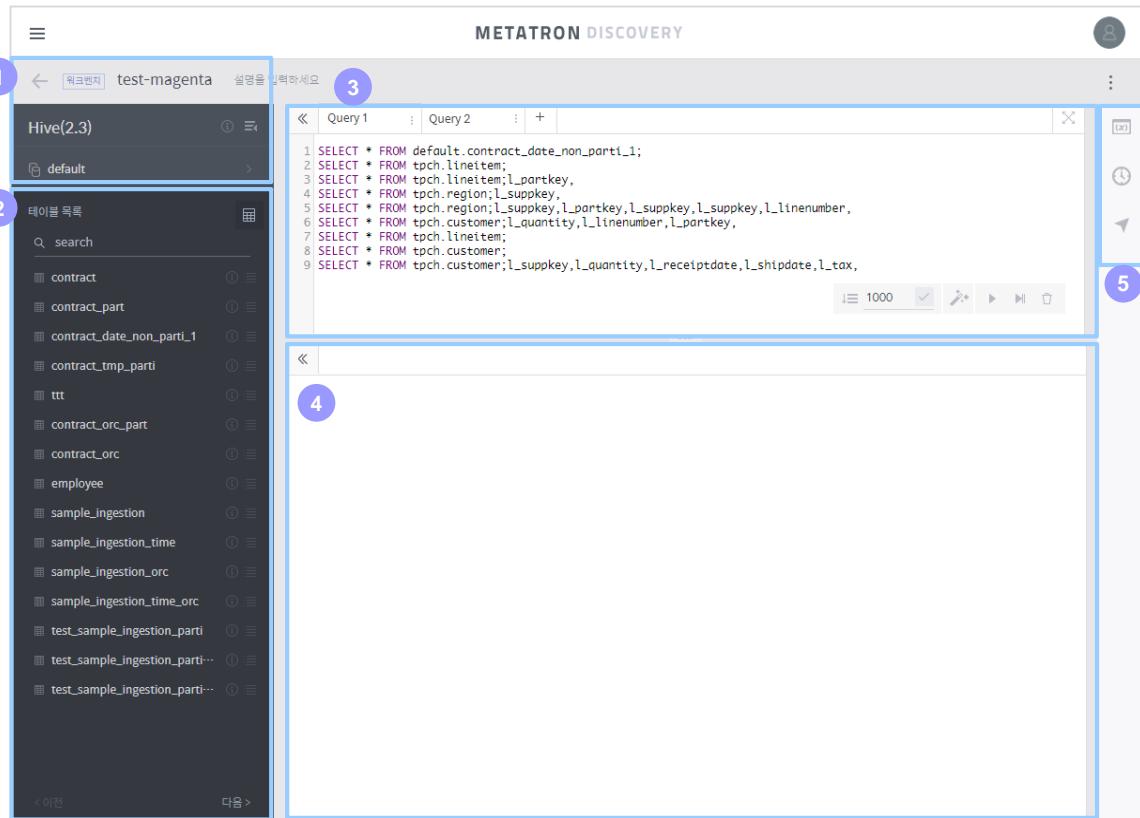
3. 선택한 데이터 커넥션의 정보를 확인하고 이름과 설명을 입력하면 워크벤치가 생성됩니다.



4. 워크벤치 생성이 완료되면, 워크스페이스 화면에서 다음과 같이 새로 생성된 워크벤치와 기존 워크벤치 목록을 확인할 수 있습니다. 클릭하면 워크벤치 기능을 사용할 수 있는 화면이 나타납니다.

3. 워크벤치 활용하기

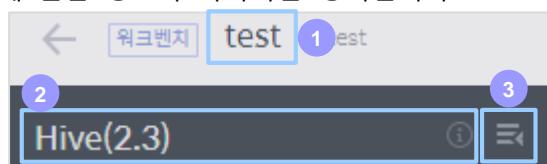
워크벤치에서는 SQL 데이터베이스 편집 및 관리가 용이하며 쿼리 결과를 다양한 형태로 시각화하여 저장할 수 있습니다. 워크벤치의 화면은 다음과 같은 5개의 영역으로 나누어 집니다.



- ① 기본 정보 영역(3.1절 참조)
- ② 스키마 및 테이블 영역(3.2절 참조)
- ③ 쿼리 에디터 영역(3.3절 참조)
- ④ 쿼리 결과 영역(3.4절 참조)
- ⑤ 추가 도구 영역(3.5절 참조)

3.1 기본 정보 영역

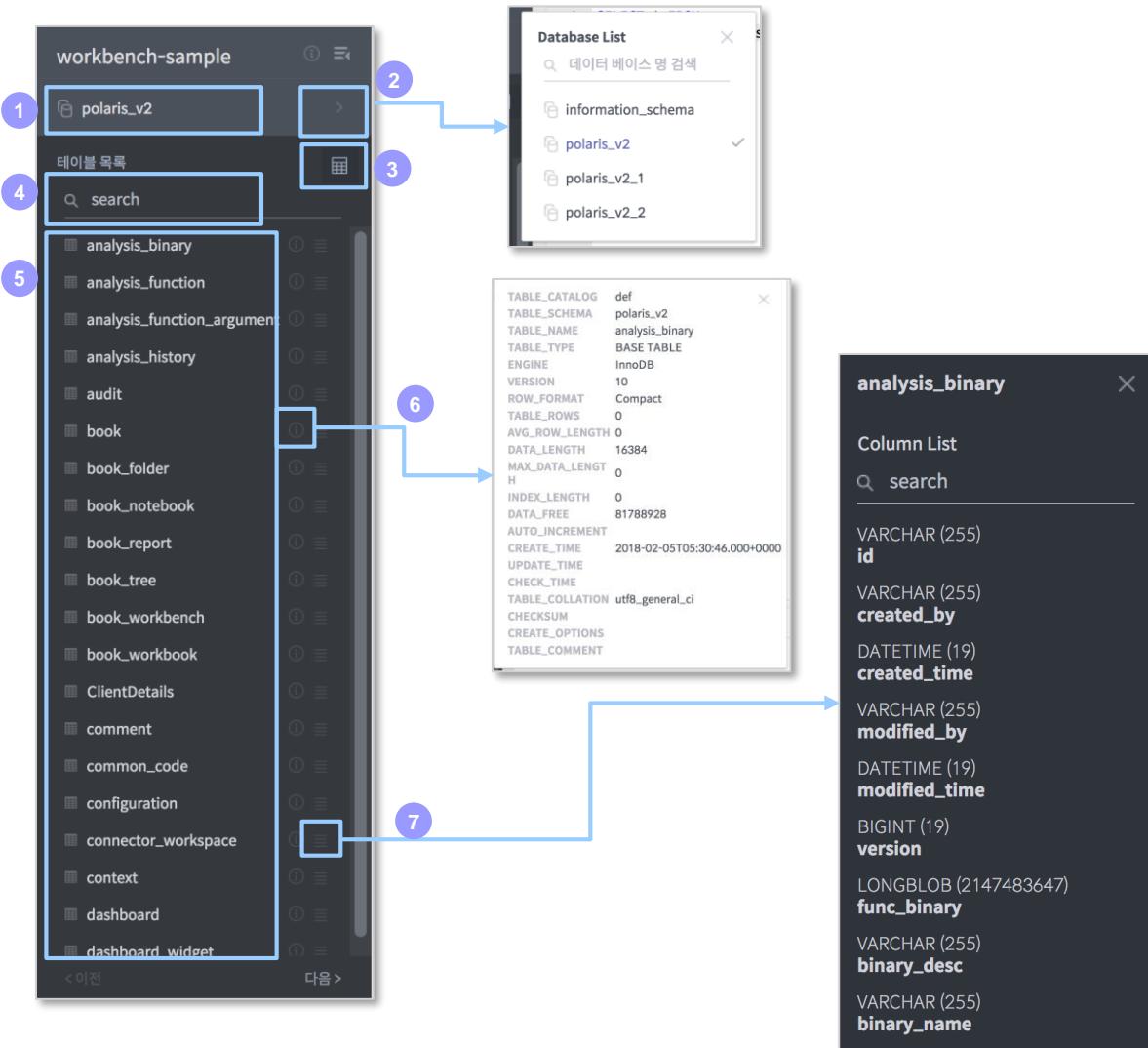
현재 작업하는 워크벤치에 관한 정보가 나타나는 영역입니다.



- ① 이름: 워크벤치의 이름입니다. 클릭하여 이름을 변경할 수 있습니다.
- ② 데이터 커넥션: 해당 워크벤치와 연결되어 있는 데이터 커넥션의 이름입니다. 아이콘을 클릭하면 자세한 정보가 나타납니다.
- ③ : 패널을 접고 펼치는 UI 버튼입니다

3.2 스키마 및 테이블 영역

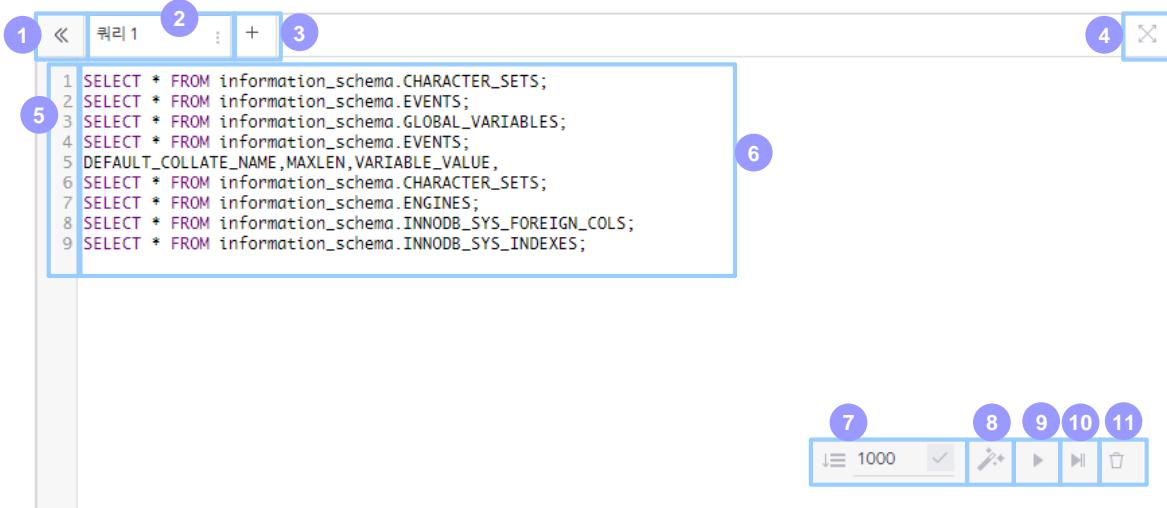
원하는 데이터베이스, 테이블, 컬럼을 손쉽게 쿼리 에디터에 입력할 수 있는 UI 기능입니다.



- 1** **데이터베이스 이름:** 현재 선택된 데이터베이스의 이름을 출력해줍니다. 해당 워크벤치에 등록된 데이터 커넥션의 첫 번째 테이블이 기본적으로 선택됩니다.
- 2** **데이터베이스 리스트:** 현재 선택된 데이터베이스를 바꿀 때 사용됩니다. 클릭하면 데이터 커넥션에 포함된 모든 데이터베이스가 조회되며, 그 중에 하나를 선택하면 해당 데이터베이스로 변경됩니다.
- 3** **스키마 정보:** 선택된 데이터베이스의 테이블 목록과 각 테이블에 속한 모든 컬럼 및 레코드 정보를 확인할 수 있습니다.
- 4** **search:** 선택된 데이터베이스에 등록된 테이블을 이름으로 검색합니다.
- 5** **테이블 이름:** 필요한 데이터를 담은 테이블을 선택하면, 오른쪽 쿼리 에디터에 해당 테이블에 대한 SELECT * FROM 쿼리가 자동으로 입력됩니다.
- 6** **테이블 정보:** 해당 테이블에 관한 기본적인 정보가 나타납니다.
- 7** **컬럼 목록:** 해당 테이블에 속한 모든 컬럼 이름과 각각의 데이터 타입이 나타납니다. 컬럼 이름을 클릭하면 쿼리 에디터에 자동으로 삽입됩니다.

3.3 쿼리 에디터 영역

쿼리를 작성하고 실행할 수 있는 에디터 화면입니다.



- 1 《: 탭이 너무 많을 경우 이전에 만들어져 있는 탭을 보여주는 역할을 합니다. 탭의 개수가 많지 않은 경우 현재 보여지고 있는 탭들의 리스트를 보여주고 탭 선택 시 해당 탭으로 이동 할 수 있습니다.
- 2 **탭**: 쿼리 관리를 위해 여러 개의 탭으로 나누어서 쿼리를 실행하거나 저장할 수 있습니다. ⋮ 버튼을 클릭 시 탭 제목의 수정과 탭 삭제가 가능합니다.
- 3 +: 이 버튼을 클릭하면 새로운 탭이 추가됩니다.
- 4 ✎: 이 버튼을 클릭하면 쿼리 에디터 영역을 전체화면으로 확장하거나 축소 할 수 있습니다.
- 5 **쿼리 행**: 쿼리 코드의 행 번호를 보여줍니다.
- 6 **에디터 화면**: 이 곳에 쿼리 문을 작성합니다. 멀티 쿼리의 실행 및 단일 쿼리의 실행이 가능합니다. ';'을 쿼리 문 끝날 때마다 삽입하면 쿼리를 나눠서 실행이 가능합니다. 자동완성 기능이 제공됩니다.
- 7 **최대 조회 행 수**: 쿼리 결과를 몇 줄까지 나타낼 것인지 설정합니다.
- 8 **SQL BEAUTIFIER**: 이 버튼을 클릭하면 작성된 쿼리 문이 쿼리 문법 표준에 맞게 정렬됩니다. 정렬하기를 원하는 쿼리를 선택한 후 버튼을 누르면 쿼리가 정리됩니다.
- 9 **ALL SQL EXECUTE**: 쿼리 문안에 들어있는 전체 쿼리를 실행합니다. (단축키: Ctrl + Enter)
- 10 **SELECTED SQL EXECUTE**: 쿼리 문 안에서 마우스가 위치된 특정 쿼리 또는 드래그하여 선택된 영역만 쿼리를 실행합니다. (단축키: Command + Enter)
- 11 **CLEAR SQL**: 쿼리 문을 모두 삭제 합니다.

3.4 쿼리 결과 영역

쿼리가 실행되면 그 결과가 쿼리 결과 탭에 나타납니다. 모든 쿼리 결과는 계속 누적되지만, 원하는 결과 탭을 자유롭게 삭제할 수 있습니다. 쿼리 결과는 텍스트 그리드 형태로 제공되며, 차트 미리보기, 데이터소스 저장, 다운로드 CSV, 온라인 엑셀 기능이 지원됩니다.

The screenshot shows the Metatron Workbench interface with the 'Results' tab selected. The results grid displays a table of data with columns: SEQ, lineitem.l_orderkey, lineitem.l_partkey, lineitem.l_suppkey, lineitem.l_linenumber, lineitem.l_quantity, and lineitem.l_extendedprice. There are 5 rows of data. Numbered callouts point to various UI elements: 1 points to the back arrow; 2 points to the tab header 'Query 1 - 결과1'; 3 points to the search bar; 4 points to the chart icon; 5 points to the download CSV icon; 6 points to the online Excel icon; 7 points to the clipboard icon; and 8 points to the refresh icon.

SEQ	lineitem.l_orderkey	lineitem.l_partkey	lineitem.l_suppkey	lineitem.l_linenumber	lineitem.l_quantity	lineitem.l_extendedprice
1	1	31037869	1537885	1	17.0	30690.27
2	1	13461816	1461817	2	36.0	63977.04
3	1	12739956	739957	3	8.0	15962.56
4	1	426299	926300	4	28.0	34307.56
5	1	4805327	305334	5	24.0	29569.92

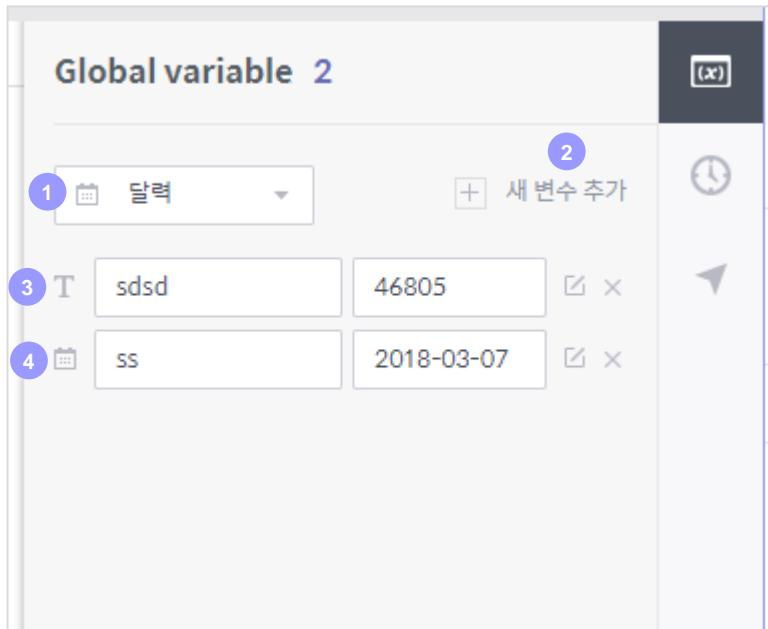
- 1 «: 탭이 너무 많아 한 화면에 모든 탭을 표시할 수 없을 경우, 이전에 만들어져 있는 탭을 보여주는 역할을 합니다. 탭의 개수가 많지 않은 경우에는 현재 보여지고 있는 탭들의 리스트를 보여주고 탭 선택 시 해당 탭으로 이동 할 수 있습니다.
- 2 탭: 결과 탭의 이름입니다. ✕ 버튼을 클릭 시 탭 삭제가 가능합니다.
- 3 컬럼 데이터 검색: 결과 내 컬럼 및 값을 검색할 수 있습니다.
- 4 차트 미리보기: 쿼리 결과를 이용하여 차트를 가상으로 그려볼 수 있습니다. 시각화를 위해 그려지는 것이고, 실제 워크스페이스 내용에 반영되지는 않습니다. (자세한 조작 방식은 '4부 워크북'에서 '5장 차트 생성·관리' 참조)
- 5 데이터 소스 저장: 쿼리 결과를 이용하여 워크스페이스 내 데이터 소스로 저장할 수 있습니다. 데이터 소스 생성 팝업이 나타나며, 데이터 커넥션 선택 및 테이블 선택 등의 과정은 워크벤치 결과 내용으로 대체됩니다. 따라서 스키마 정의 및 ingestion 주기 등의 과정이 곧바로 진행되게 됩니다. (자세한 절차는 '3부 운영관리'에서 2.1.3절 참조)
- 6 다운로드 CSV: 쿼리 결과를 로컬 파일(csv)로 다운로드가 가능합니다.
- 7 온라인 엑셀: 해당 버튼을 클릭하면, 온라인 엑셀에 결과값이 보여집니다. 온라인 엑셀은 Spread Sheet라는 HTML5로 구현된 온라인 엑셀 어플리케이션입니다. 현재 사용하고 있는 워크벤치가 아닌 새 창에 온라인 엑셀 어플리케이션이 담긴 브라우저가 열립니다.
- 8 데이터 내역: 쿼리 실행에 의해 출력된 데이터 내역입니다. 출력된 데이터는 클립보드에 복사하여 활용할 수도 있습니다.

3.5 추가 도구 영역

추가 도구 영역은 워크벤치 활용에 도움을 주는 유용한 기능들로 구성되어 있습니다.

3.5.1 전역변수(Global variable) 편집 기능

어떤 구문이 반복적으로 사용되고 그 내용을 계속 바꿔가면서 쿼리를 실행해야 할 경우, 그 구문을 '전역 변수'로 지정해서 사용하면 편리합니다.



- ① **변수 타입:** 전역 변수 타입으로는 달력과 텍스트가 제공합니다.
- ② **새 변수 추가:** 원하는 변수 종류를 선택한 뒤 새 변수 추가 버튼을 누릅니다. 쿼리 에디터 영역에 해당 전역변수가 추가됩니다.
- ③ **이름:** 변수 이름을 입력합니다.
- ④ **변수 값:** 달력은 날짜를 선택, 텍스트는 값을 입력하여 사용할 수 있습니다.

3.5.2 Query History 목록 조회

수행한 쿼리의 이력과 결과를 조회할 수 있으며 해당 쿼리를 다시 불러올 수 있습니다.

The screenshot shows the 'Query History' interface. At the top left is a search bar (1) with a magnifying glass icon. To its right is a refresh button (2). Below these are two small icons: a square with '(x)' and a clock. The main area displays a table of query history entries. The columns are labeled 'No.', 'Time', 'Action', 'Resp.', and 'Result'. The 'Action' column contains the actual SQL queries. The 'Result' column indicates if the query was successful ('Success') or failed ('Fail'). Some failed queries have a red circular icon with a question mark next to them. The table lists approximately 20 entries from April 9th and 10th, 2016, with response times ranging from 1 ms to 4 ms. At the bottom left of the table is a blue button labeled '+ 더보기' (4), which means 'View more'.

No.	Time	Action	Resp.	Result
44	04-09 15:55:00 04-09 15:55:...	SELECT * FROM information_sche ma.ENGINES	3 ms	Success
43	04-09 15:55:00 04-09 15:55:...	SELECT * FROM information_sche ma.CHARACTER_SETS	1 ms	Success
42	04-09 15:55:00 04-09 15:55:...	SELECT * FROM information_sche ma.CHARACTER_SETS	1 ms	Success
41	04-09 15:55:00 04-09 15:55:...	INSERT INTO information_schema SELECT * FROM information_sc...	2 ms	Fail ⓘ
40	04-09 15:55:00 04-09 15:55:...	SELECT * FROM information_sche ma.COLLATION_CHARACTER_...	2 ms	Success
39	04-09 15:55:00 04-09 15:55:...	.CHARACTER_SETS (CHARACTER _SET_SELECT * FROM informat...	2 ms	Fail ⓘ
38	04-09 15:55:00 04-09 15:55:...	NAME,DEFAULT_COLLATE_NAM E,DESCRIPTION,MAXLEN) value...	1 ms	Fail ⓘ
37	04-09 15:55:00 04-09 15:55:...	SELECT * FROM information_sche ma.GLOBAL_STATUS	3 ms	Success
36	04-09 15:55:00 04-09 15:55:...	SELECT * FROM information_sche ma.INDEX_STATISTICS	1 ms	Success
35	04-09 15:55:00 04-09 15:55:...	CHARACTER_SET_NAME,MAXLE N,CHARACTER_SET_NAME,Ho...	2 ms	Fail ⓘ
34	04-06 15:43:47 04-06 15:43:...	select * from information_schema. COLLATIONS	4 ms	Success
33	04-06 15:43:33 04-06 15:43:...	select * from information_schema. COLLATIONS	3 ms	Success
32	04-06 15:43:31 04-06 15:43:...	select * from information_schema. FILES	2 ms	Success
31	04-06 15:43:29 04-06 15:43:...	select * from information_schema. INDEX_STATISTICS	1 ms	Success
30	04-06 15:43:28 04-06 15:43:...	select * from information_schema. INNODB_BUFFER_PAGE	1 ms	Success

- ① **쿼리 검색:** 쿼리 이력에 남아있는 쿼리를 검색합니다.
- ② **전체 삭제:** 쿼리 이력에 남아있는 모든 쿼리 내역을 삭제합니다.
- ③ **쿼리 목록:** 쿼리 이력을 마지막으로 삭제한 이후에 수행된 쿼리들에 대한 정보를 실행 순서대로 보여줍니다. 나열된 쿼리 중 하나를 클릭하면 해당 쿼리 문이 쿼리 에디터에 복사됩니다.
- ④ **+ 더보기:** 이 버튼을 클릭하면 더 많은 쿼리 내역을 볼 수 있습니다.

3.5.3 워크벤치 네비게이션

다른 워크벤치로 이동하는 기능을 제공합니다. 이동하기를 원하는 워크벤치를 클릭하면 해당 워크벤치로 이동합니다.

Workbench Navigation (15)		
No.	Workbench name	수정일
35	경제지표	2018-04-09
34	111	2018-04-05
33	ddd	2018-04-05
32	test	2018-04-04
31	hive (2.3)	2018-04-03
30	work	2018-03-28
29	test-magenta	2018-03-27
28	NEW	2018-03-27
27	2	2018-03-23
26	테스트01	2018-03-23
25	ddd	2018-03-23
24	mysql test	2018-03-22
23	test-Magenta	2018-03-22
22	CrimeSQL	2018-03-20
21	nenv	2018-03-20

- ① **워크벤치 검색:** 워크스페이스에 저장된 워크벤치를 검색합니다.
- ② **워크벤치 목록:** 워크스페이스에 저장된 모든 워크벤치를 보여줍니다. 나열된 워크벤치 중 하나를 클릭하면 해당 워크벤치로 이동됩니다.