

1. Demo Instruction
2. AccuInsight+ Services
3. Demo Outline
4. Senario 1 : 회사별 일평균 좌석이용률
 - Dataset (좌석 예약 데이터, 좌석 메타 데이터)
5. Senario 2 : 좌석유형별 주 이용 나이
 - Dataset (좌석 예약 데이터, 좌석 메타 데이터)
6. Senario 3 : 회의실 규모별 이용현황
 - Dataset (회의실 예약 데이터, 회의실 메타 데이터)

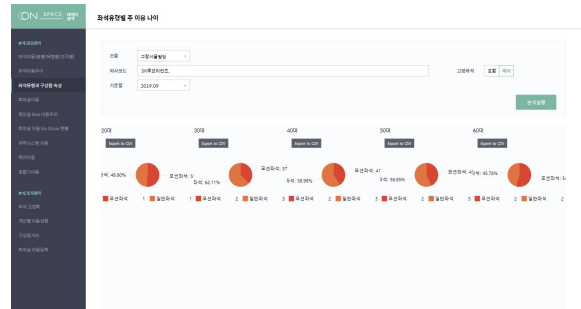
1. Demo Instruction

공유 오피스 분석 플랫폼 구축 프로젝트에 사용한 AccuInsight+ 서비스 Batch Pipeline, Data Insight를 공유오피스 분석 요건 45개 중 다음 4가지 요건을 통해 실습해본다.

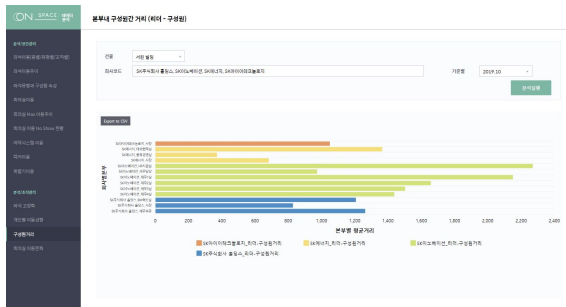
a. 회사별 일평균 좌석이용률



b. 좌석유형별 주 이용 나이



c. 본부내 리더와 구성원간 거리

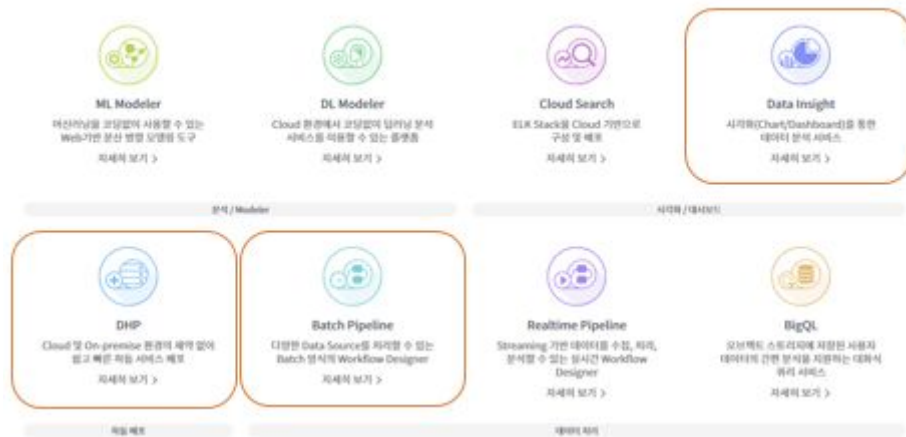


d. 회의실 규모별 이용현황

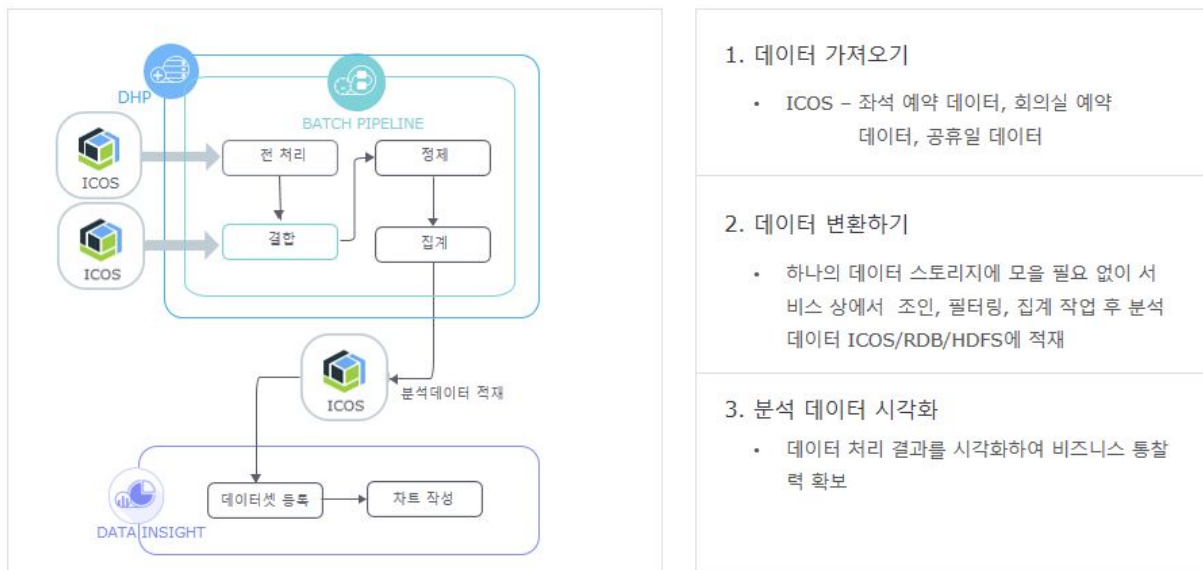


2. AccuInsight+ Services

빅데이터 처리에 필요한 모든 서비스 제공



3. Demo Outline



4. Senario 1 : 회사별 일평균 좌석이용률 데이터 준비

ICOS에 업로드된 데이터를 사용합니다.

스토리지 : IBMOSC1146611-6

버킷 : handson-bucket

a. Dataset 1 : 좌석 예약 데이터

column명	column 설명	Value Example
bldg	빌딩명	SKC 본사, 서린 빌딩, ...
desk_type	좌석 유형	일반좌석, 스탠딩좌석, ...
desk_id	각 좌석의 ID	3a27b3bf-d8de-11e8-975a-0a9726611f46
dt	예약 날짜	20190803

com	회사명	SK C&C, SK이노베이션, ...
age	나이대	20대,30대,...

b. Dataset 2 : 좌석 메타 데이터

column명	column 설명	Value Example
bldg	빌딩명	SKC 본사, 서린 빌딩, ...
desk_type	좌석 유형	일반좌석, 스탠딩좌석, ...
occupied	고정좌석 여부	N, Y
desk_id	각 좌석의 ID	3a27b3bf-d8de-11e8-975a-0a9726611f46
dt	날짜	20190803

c. 데이터 처리 (Batch Pipeline)

데이터파일 2개 조인하여 ETL작업 수행하기

ETL flow 구성내용 : ICOS에서 파일 불러오기 → 사용할 column만 선택 → 데이터 정제 → 통계
→ join key 생성 → 2개 파일 join → 새로운 column 생성 → 불필요한 column 삭제 → null값
제거→ column명 rename → ICOS에 저장

- 생성 메뉴로 이동

AccuInsight+ BATCH PIPELINE

ABOUT DASHBOARD WORKFLOW CONFIGURATION

생성
관리

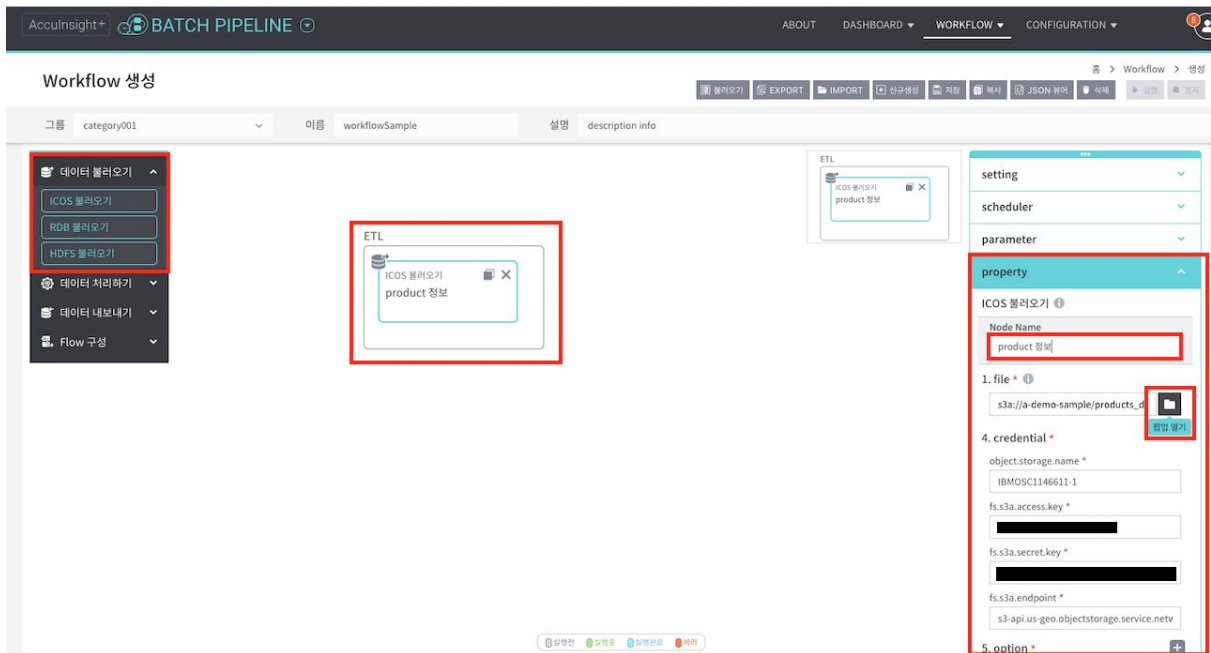
Batch Pipeline
다양한 데이터 소스를 처리할 수 있는 배치 방식의
Workflow Designer

지금 시작해 보세요 자세히 알아보기

복잡한 데이터의 배치 처리를 간단한 Workflow UI로 제공

고성능 ETL Batch Pipeline은 Hadoop MapReduce보다 5배 빠른 고유 Spark 코드를 생성 하여 고성능 처리를 제공합니다.
또한, 고유 Spark 코드를 생성하므로 추가 설치할 필요가 없고 불필요한 관리 비용과 처리 오버 헤드가 발생하지 않습니다. 기존 데이터 silo를 연결 하고
새로운 데이터 소스를 추가하면 데이터 분석에 새로운 기회를 부여 할 수 있습니다.

- 데이터 로딩



데이터 불러오기 그룹에서 ICOS 불러오기 노드를 캔버스 영역에 drag & drop하여 node 생성
우측 property 패널

- file : 불러오려는 ICOS 파일의 경로 지정
- option : 데이터에 header가 없는 경우 false로 변경
- schema : 자동 파일 열기

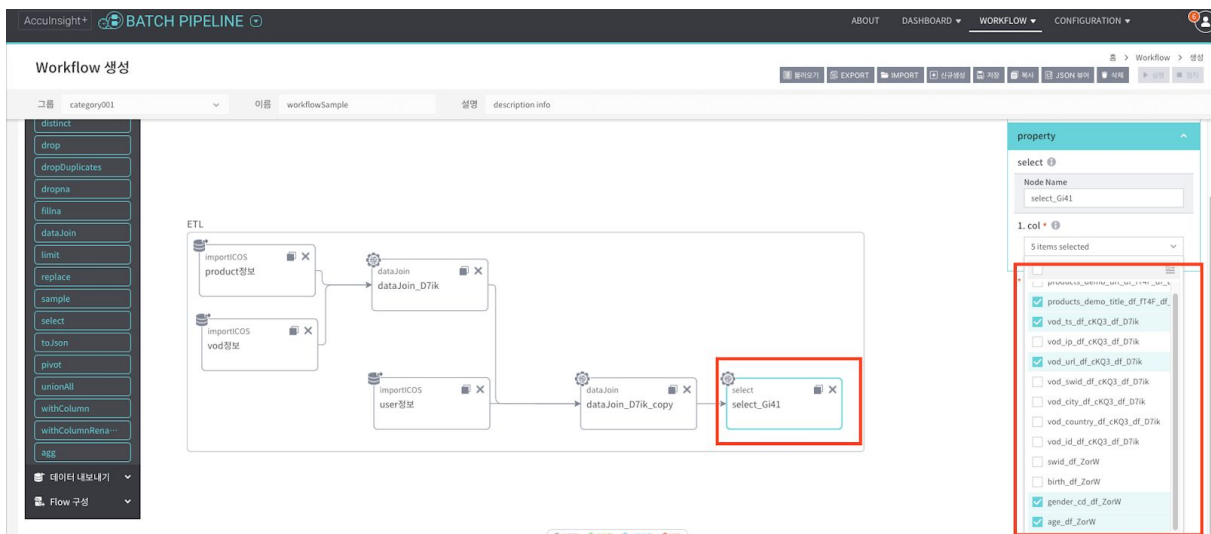
여기서 column명 지정

좌석 예약 데이터 : bldg, desk_type, desk_id, dt, com, age

동일한 방법으로 좌석 메타 데이터도 불러오기

좌석 메타 데이터 : bldg, desk_type, occupied, desk_id, dt

- 필요한 컬럼 선택



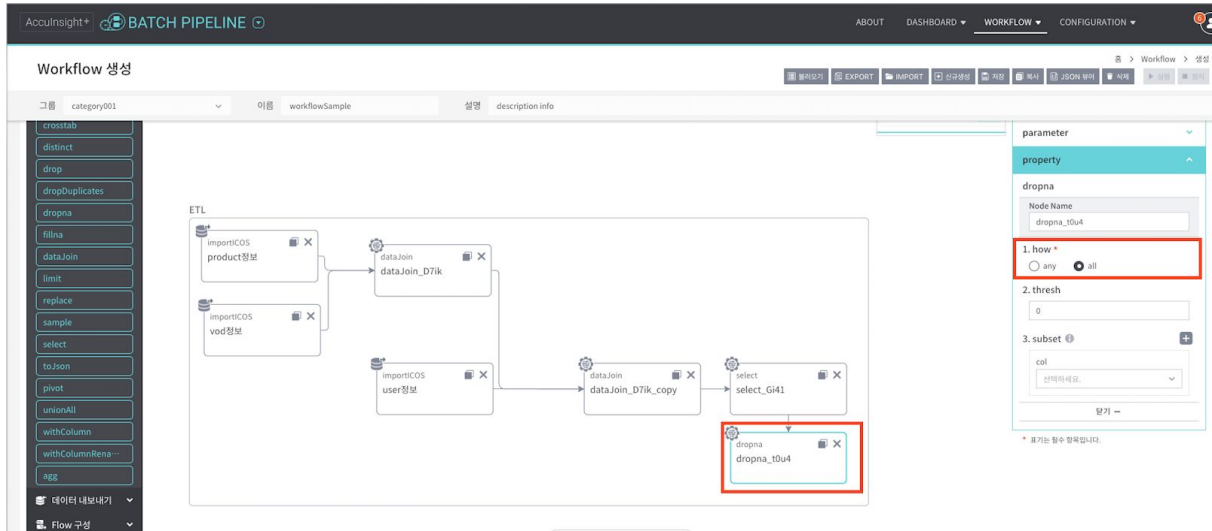
select 노드 drag & drop하여 생성

우측 property 패널

col checkbox에서 아래 5가지 column 체크

- 좌석 예약 데이터 : bldg, desk_id, dt, com
- 좌석 메타 데이터 : bldg, occupied, desk_id, dt

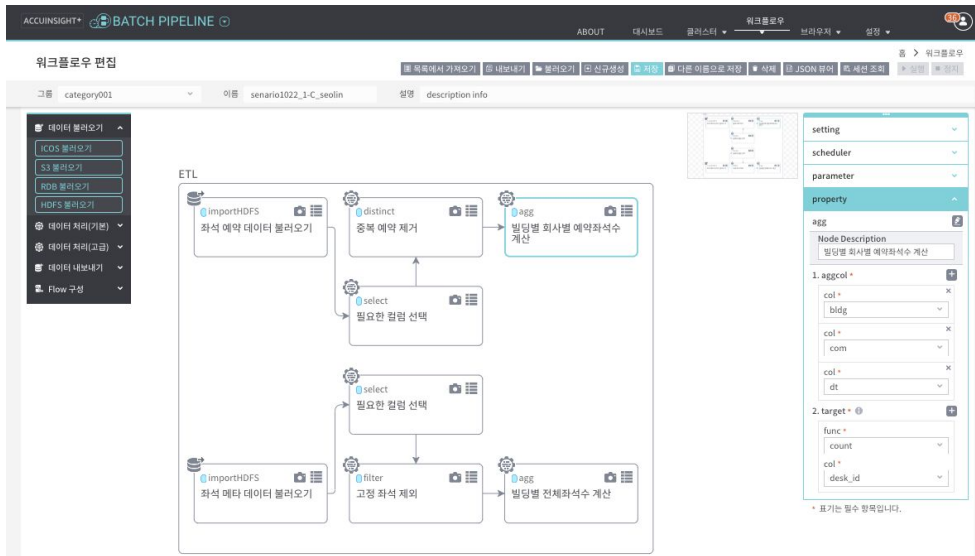
- 데이터 정제



distinct 노드 drag & drop하여 좌석 예약 데이터에서 중복 예약 제거
filter 노드 drag & drop하여 좌석 메타 데이터에서 고정 좌석 제외
우측 property 패널

- col : occupied
- filterOption : =:equal
- filterValue : 'N'

- 통계



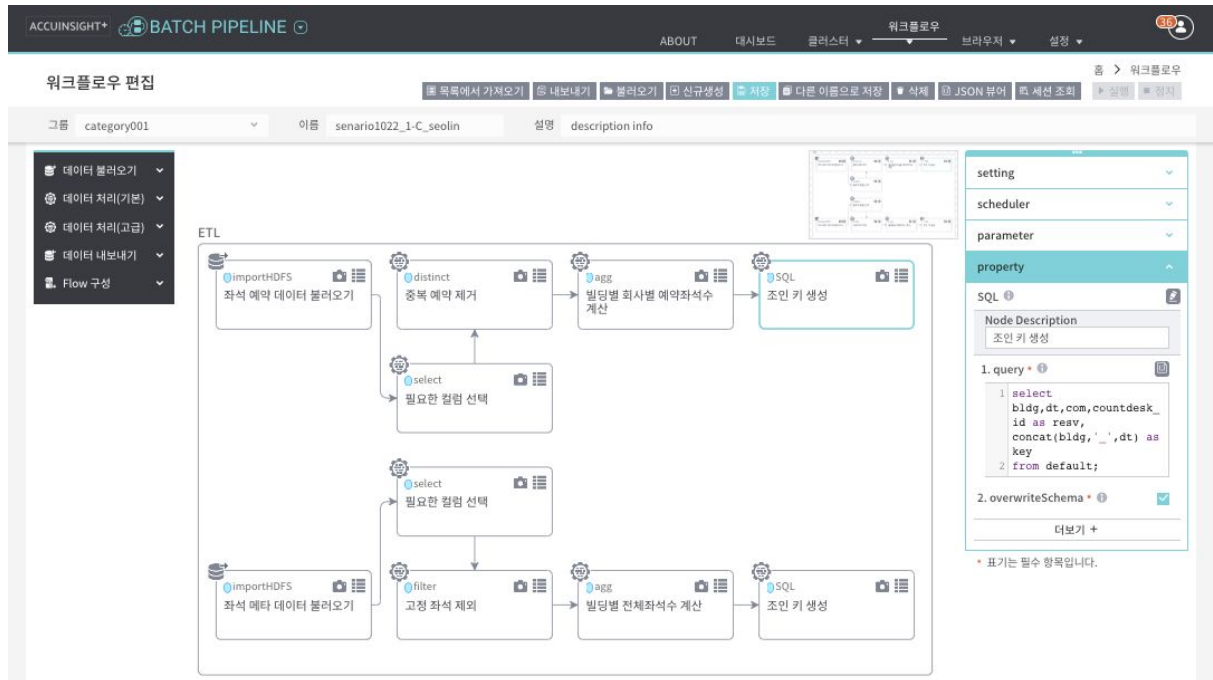
agg 노드 drag & drop하여 좌석 예약 데이터에서 빌딩별 회사별 예약좌석수 계산
우측 property 패널

- aggcol : bldg, com, dt
- target : func count, col desk_id

agg 노드 drag & drop하여 좌석 메타 데이터에서 빌딩별 전체좌석수 계산
우측 property 패널

- aggcol : bldg, dt
- target : func count, col desk_id

- join key 생성



SQL 노드 drag & drop하여 좌석 예약 데이터에서 join key 생성

우측 property 패널

- query

```
select bldg,dt,com,countdesk_id as resv, concat(bldg,'_',dt) as key
from default;
```

- overwriteSchema 체크 (SQL 결과로 데이터 변경)

SQL 노드 drag & drop하여 좌석 메타 데이터에서 join key 생성

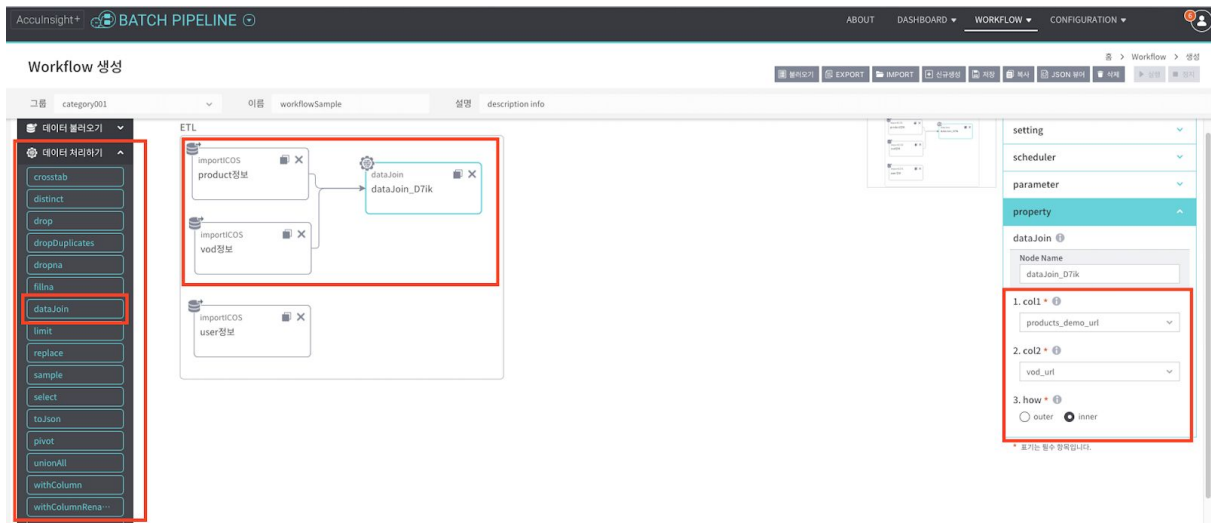
우측 property 패널

- query

```
select countdesk_id as total, concat(bldg,'_',dt) as key
from default;
```

- overwriteSchema 체크 (SQL 결과로 데이터 변경)

- 2개파일 조인



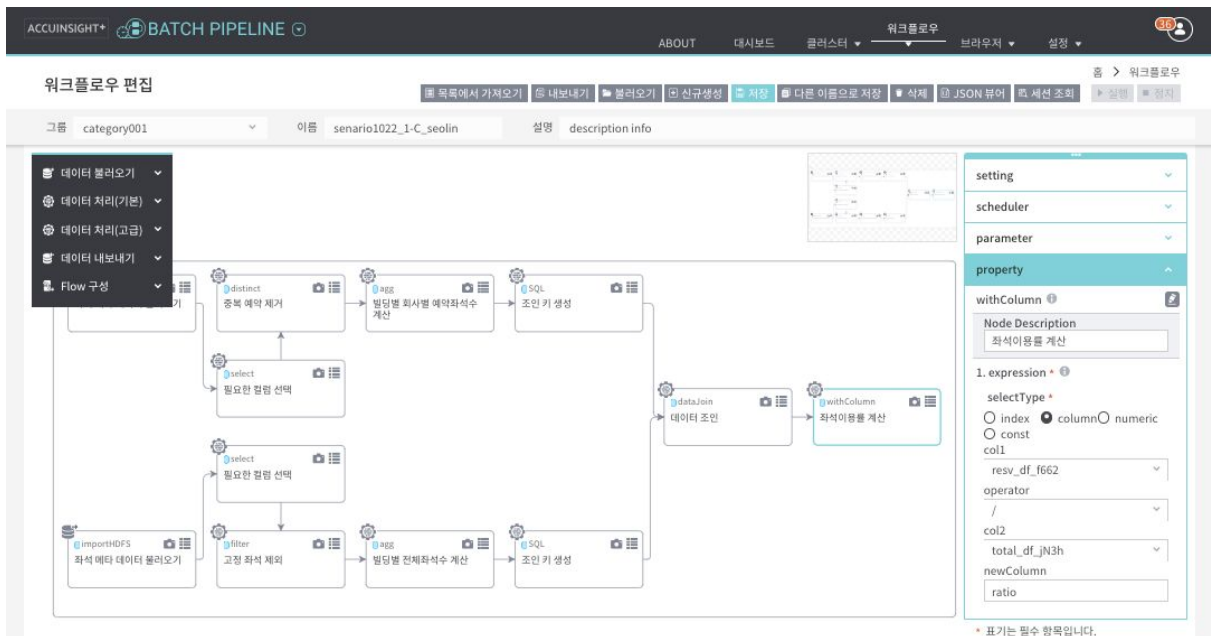
좌측의 데이터 처리하기 클릭

열린 패널에서 dataJoin 노드 우측 캔버스에 drag & drop하여 생성

우측 property패널

- col1 : key 선택
- col2 : key 선택
- how : right_outer 선택

- 새로운 column 생성



withColumn 노드 drag & drop하여 좌석이용률 계산

우측 property 패널

- selectType : column
- col1 : resv_dt_~
- operator : /
- col2 : total_dt_~
- newColumn : ratio

- 불필요한 column 삭제

The screenshot shows the ACCUINSIGHT+ BATCH PIPELINE interface. The workflow is titled 'scenario1022_1-C_seolin'. The 'drop' node is selected, and its properties are displayed on the right. The 'property' tab shows a list of columns to be dropped: total_df_jN3h, key_df_jN3h, ratio, bldg_df_f662, dt_df_f662, and com_df_f662. The 'Node Description' field contains '안쓰는 컬럼 삭제'.

drop 노드 drag & drop하여 불필요한 데이터 삭제

우측 property 패널

- total_df_~
- key_df_~
- resv_df_~
- key_df_~

- null값 제거

The screenshot shows the ACCUINSIGHT+ BATCH PIPELINE interface. The workflow is titled 'scenario1022_1-C_seolin'. The 'filter' node is selected, and its properties are displayed on the right. The 'property' tab shows a list of filters: col * (com_df_f662), filterOption * (!=: not equal), and filterValue * ('N'). The 'Node Description' field contains '회사가 NULL이면 삭제'.

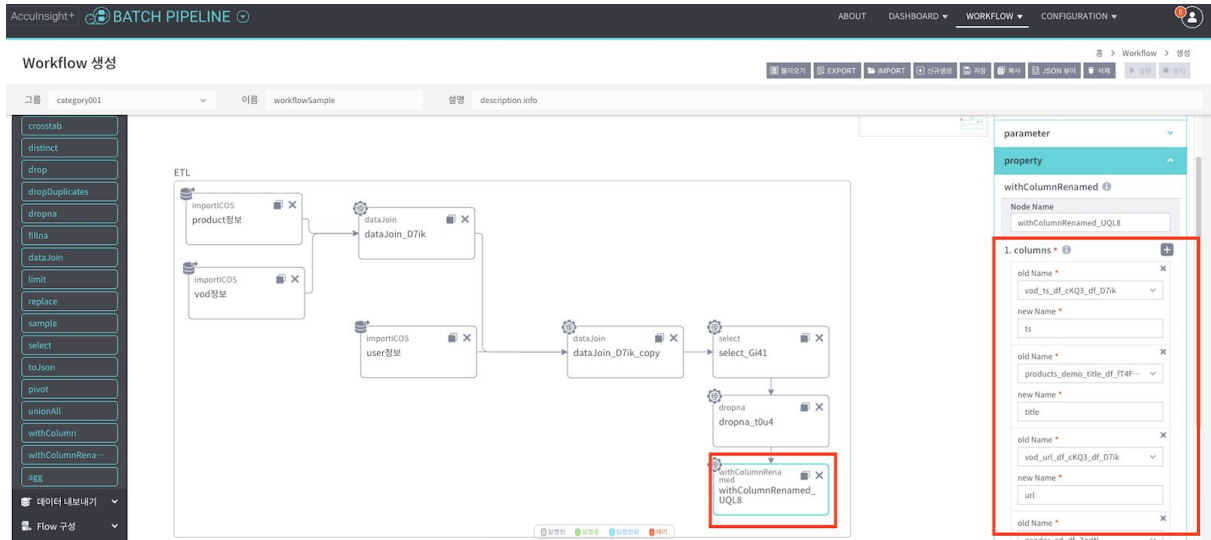
filter 노드 drag & drop하여 회사명이 'N' 이면 삭제

우측 property 패널

- col : com_df_~
- filterOption : !=:not equal

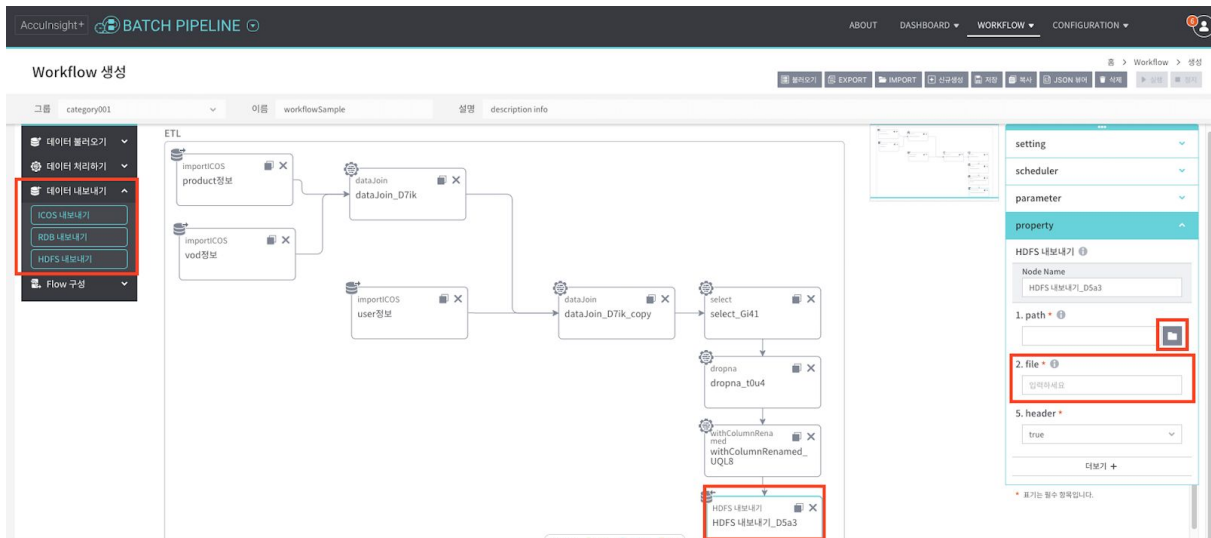
- filterValue : '\N'

- 컬럼명 변경



withColumnRenamed노드 drag&drop하여 생성
우측 property 패널
전체 컬럼 불러오기 아이콘을 통해 자동 파싱

- ICOS에 저장



좌측 데이터 내보내기 클릭
ICOS 내보내기 노드 drag & drop 하여 생성
withColumnRenamed 노드에서 ICOS 내보내기노드로 연결
우측 property 패널
- path의 browse 아이콘을 클릭하여 열리는 팝업에서 저장할 디렉토리 위치 선택 후 확인
클릭 (ex. /tmp)
- file에 생성할 디렉토리명 입력 (ex.modeling_data)

1. path * ⓘ

hdfs://10.42.175.152:8020//tmp

2. file * ⓘ

modeling_data

- 저장 및 실행

Accusight+ BATCH PIPELINE

ABOUT DASHBOARD WORKFLOW CONFIGURATION

Workflow 편집

category001 workflowSample 설명 description info

데이터 불러오기
데이터 처리하기
데이터 내보내기
Flow 구성

ETL

importCOS product정보
importCOS vod정보
dataJoin dataJoin_D7ik
dataJoin dataJoin_D7ik_copy
select select_G141
dropna dropna_10u4
withColumnRenamed withColumnRenamed_UQL8

setting

1. cluster *
stackid *
dlpe

2. notifyLevel * ⓘ
level * ERROR INFO

4. workflow *
name *
workflowSample

scheduler
parameter
property

더보기 +

실행

저장 클릭

실행 클릭

- 워크플로우 상태 확인

Accusight+ BATCH PIPELINE

ABOUT DASHBOARD WORKFLOW CONFIGURATION

Workflow 관리

검색 필터

실행 중지 삭제

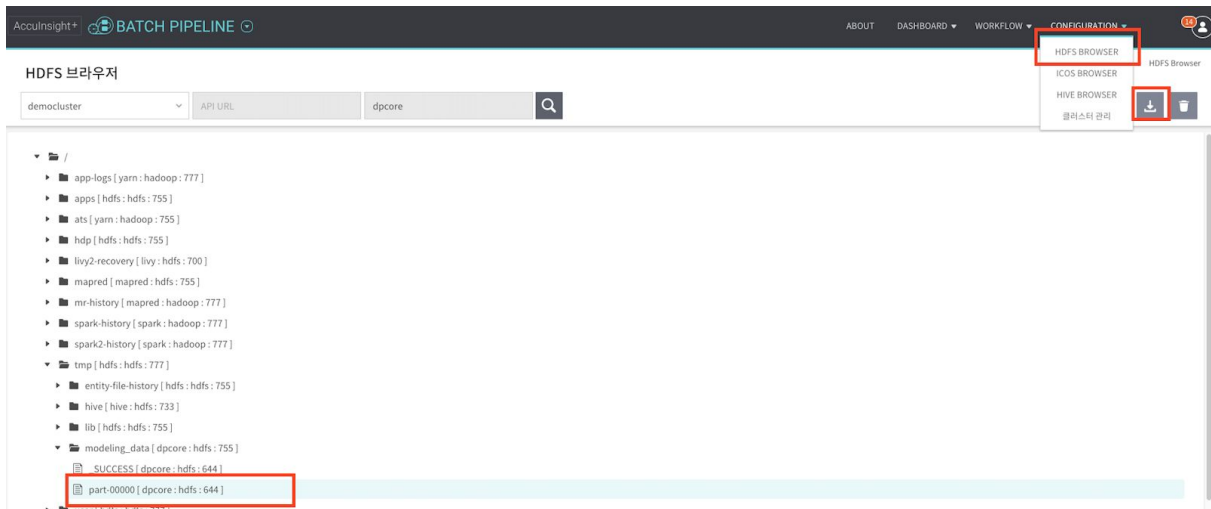
No.	그룹	이름	설명	생성된 시각	업데이트된 시각	상태	상세 정보
159	category001	workflowSample	description info	2018-11-09 15:32:20	2018-11-09 15:52:53	ACTIVE	Instance 목록

페이지당 행수 10

workflow 관리 화면으로 이동하여 상태 확인

이미 성공으로 종료되었거나 오류가 발생하여 종료된 경우 INACTIVE 상태
상세정보 컬럼의 Instance목록 아이콘 클릭

- 실행결과 확인



CONFIGURATION > ICOS BROWSER 메뉴로 이동

지정했던 ICOS directory위치로 이동하여 내보내기로 생성한 디렉토리 하단의

part-00000~파일 클릭

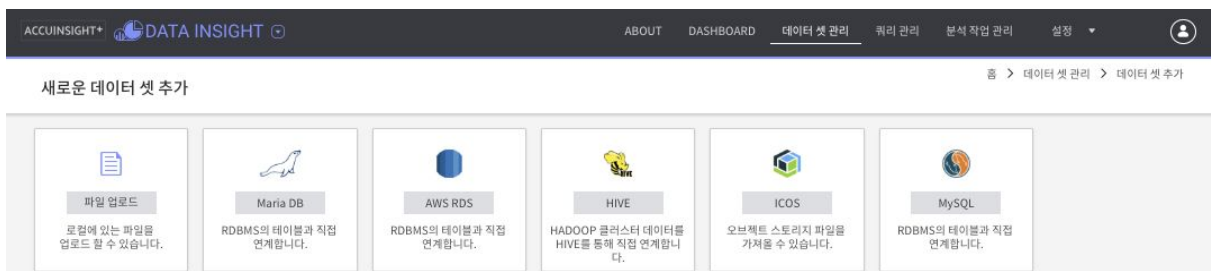
우측 상단의 다운로드 버튼 클릭하여 파일내용 확인

d. 데이터 시각화 (Data Insight)

데이터 준비

- 데이터 셋 추가

데이터 셋 관리 페이지에서 데이터 셋 추가를 클릭하면 추가 페이지로 이동합니다.



DataInsight 분석 및 시각화를 위해 MariaDB, ICOS, HIVE 등 다양한 종류의 데이터 셋을 추가할 수 있습니다.

지원하는 Data Source 타입

- ① 로컬 파일 : csv 타입 등 로컬에 있는 파일을 업로드
- ② MariaDB : MariaDB 테이블 연동하여 데이터 업로드
- ③ AWS RDS: AWS RDS 테이블 연동하여 데이터 업로드
- ④ HIVE : 하둡 클러스터의 데이터를 hive를 통하여 업로드
- ⑤ ICOS : IBM 오브젝트 스토리지의 파일을 업로드
- ⑥ MySQL : MySQL 테이블을 연동하여 데이터 업로드

데이터 셋 추가 : ICOS 선택

이전에 Batch Pipeline 실행 결과 파일(ICOS) 선택

스키마 편집

ACCUIINSIGHT+ DATA INSIGHT						
ABOUT DASHBOARD 데이터 셋 관리 쿼리 관리 분석 작업 관리 설정						
홈 > 데이터 셋 관리 > 데이터 셋 추가 > 스키마 편집						
< part-00000-82a70935-9787-44f2-8cc5-...						
필드 추가	빌딩명	좌석유형	날짜	회사명	나이대	좌석이용률
입력하세요.	String	String	Integer	String	String	Float
<input checked="" type="checkbox"/> 빌딩명	SKC 본사	일반좌석	20190802	SKC	60대	0.10526315789473684
<input checked="" type="checkbox"/> 좌석유형	SKC 본사	일반좌석	20190802	MCNS	30대	0.008771929824561403
<input checked="" type="checkbox"/> 날짜	SKC 본사	일반좌석	20190802	SKC HT&M	50대	0.03508771929824561
<input checked="" type="checkbox"/> 회사명	SKC 본사	일반좌석	20190802	SKC	20대	0.05263157894736842
<input checked="" type="checkbox"/> 나이대	SKC 본사	일반좌석	20190802	MCNS	60대	0.02631578947368421
<input checked="" type="checkbox"/> 좌석이용률	SKC 본사	일반좌석	20190802	SKC	40대	0.2807017543859649
	SKC 본사	일반좌석	20190802	SKC	30대	0.2631578947368421
	SKC 본사	일반좌석	20190802	MCNS	50대	0.03508771929824561
	SKC 본사	일반좌석	20190802	SKC	50대	0.21929824561403508
	SKC 본사	일반좌석	20190802	MCNS	40대	0.02631578947368421

데이터 셋 이름 변경

part-00000~ -> 회사별 일평균 좌석이용률

column명 변경

bldg -> 빌딩명, dt -> 날짜, com -> 회사명, ratio -> 좌석이용률

저장

차트 작성

- 분석 작업 관리

분석 작업 관리 페이지에서 분석 작업 생성을 클릭하면 추가 페이지로 이동합니다.

차트 작성할 데이터 셋 적용

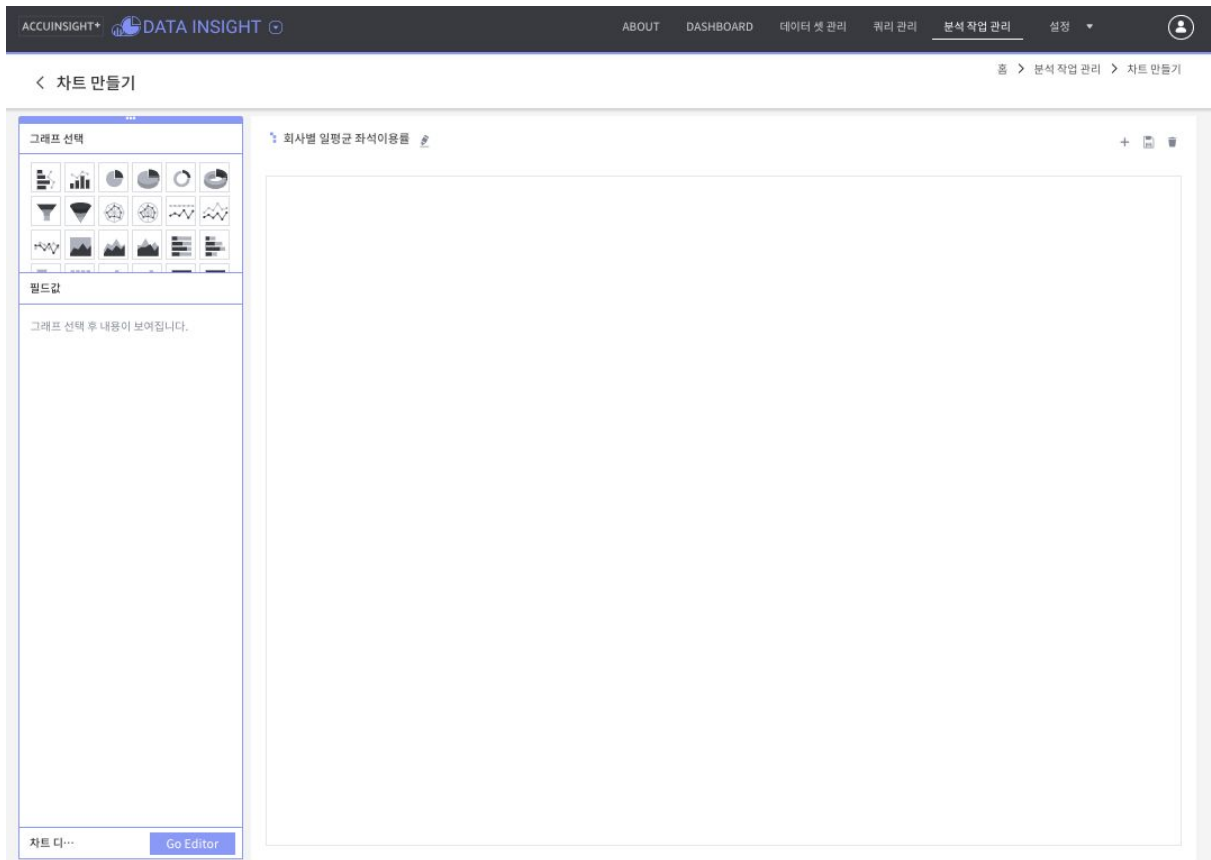
ACCUIINSIGHT+ DATA INSIGHT						
ABOUT DASHBOARD 데이터 셋 관리 쿼리 관리 분석 작업 관리 설정						
홈 > 분석 작업 관리 > 새로운 분석 작업						
<div> <div>검색 필터</div> <div> <div>새 쿼리로부터 생성</div> <div>새 데이터 소스로부터 생성</div> </div> </div>						
그룹	타입	데이터 셋 타입	이름	최종수정일	Action	
Default	데이터 셋		회사별 일평균 좌석이용률	2019-10-25	<div> <div>+</div> <div>삭제</div> </div>	
Default	데이터 셋		part-00000-45682442-d580-4cac-97c8-fe0472baee1f-c000.csv	2019-10-24	<div> <div>+</div> <div>삭제</div> </div>	
Default	쿼리		규모별 회의실 이용현황 쿼리	2019-10-23	<div> <div>+</div> <div>삭제</div> </div>	
Default	데이터 셋		part-00000-84cf0a49-f431-47f5-bb79-4da65cd7272b-c000.csv	2019-10-23	<div> <div>+</div> <div>삭제</div> </div>	
Default	데이터 셋		kmeans.result	2019-02-18	<div> <div>+</div> <div>삭제</div> </div>	

필터 추가

- 이름 : 날짜
- 날짜 = 20190820

분석 작업 차트 생성

차트 생성 버튼 클릭



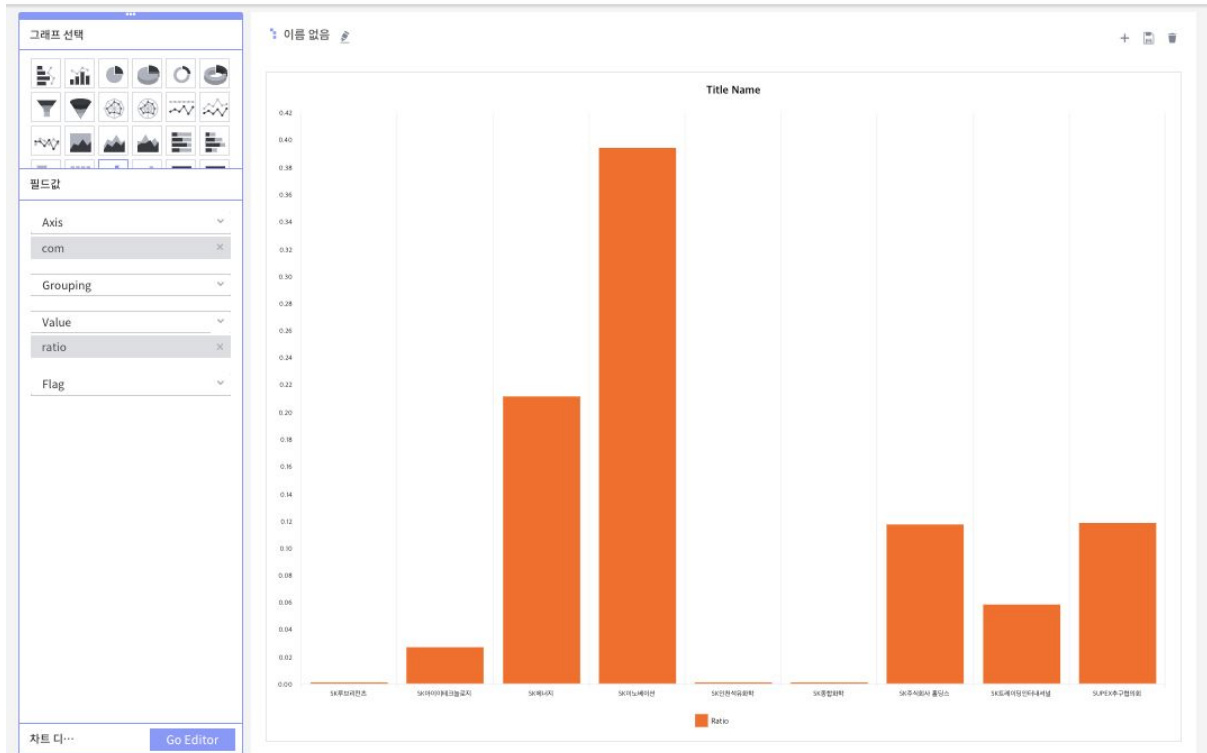
컬럼 스택 차트 선택

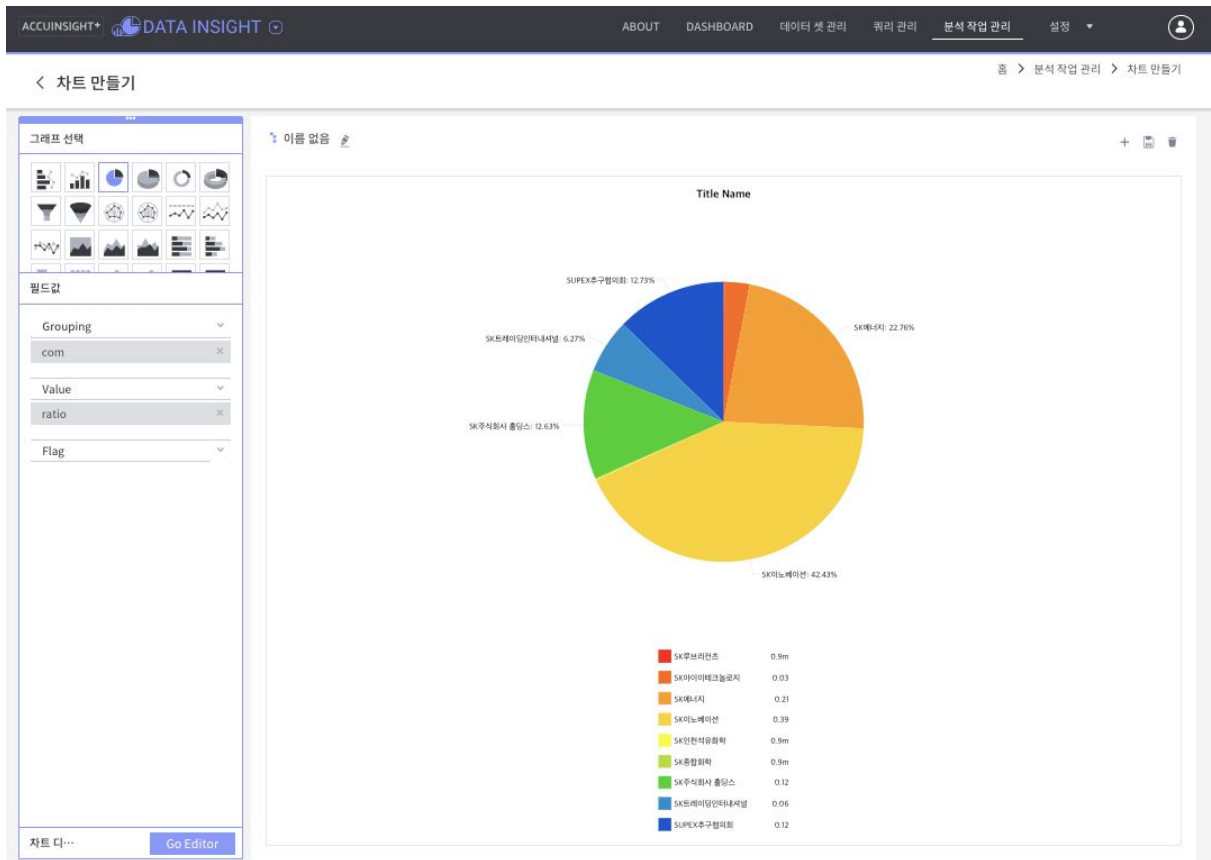
Axis : 회사명

Value : 좌석이용률

〈 차트 만들기

홈 > 분석작업 관리 > 차트 만들기





저장
대시보드 생성 버튼 클릭

DASHBOARD

대시보드는 작업했던 분석 결과들을 보관하고 확인할 수 있는 공간입니다.

분석 작업 만들기에서 대시보드 생성 버튼을 클릭하여 작업 결과물들을 대시보드에 보관할 수 있습니다

ACCUIINSIGHT+ DATA INSIGHT

ABOUT DASHBOARD 데이터셋 관리 쿼리 관리 분석 작업 관리 설정

홈 > DASHBOARD

대시보드 리스트

검색 필터

그룹	이름	차트	데이터셋	최종수정일	Action
Default	2019년 8월 20일의 회사별 좌석이용률	2	1	2019-10-25	
Default	tt	2	1	2019-10-24	
Default	1-C	2	1	2019-10-24	

<< < 1-3 / 3 > >> 페이지당 행수 10

대시보드 이미지로 내보내기
다운받은 이미지 확인

5. Senario 2 : 좌석유형별 주 이용 나이 데이터 준비

ICOS에 업로드된 데이터를 사용합니다.

스토리지 : IBMOSC1146611-6

버킷 : handson-bucket

a. Dataset 1 : 좌석 예약 데이터

column명	column 설명	Value Example
bldg	빌딩명	SKC 본사, 서린 빌딩, ...
desk_type	좌석 유형	일반좌석, 스탠딩좌석, ...
desk_id	각 좌석의 ID	3a27b3bf-d8de-11e8-975a-0a9726611f46
dt	예약 날짜	20190803
com	회사명	SK C&C, SK이노베이션, ...
age	나이대	20대,30대,...

b. Dataset 2 : 좌석 메타 데이터

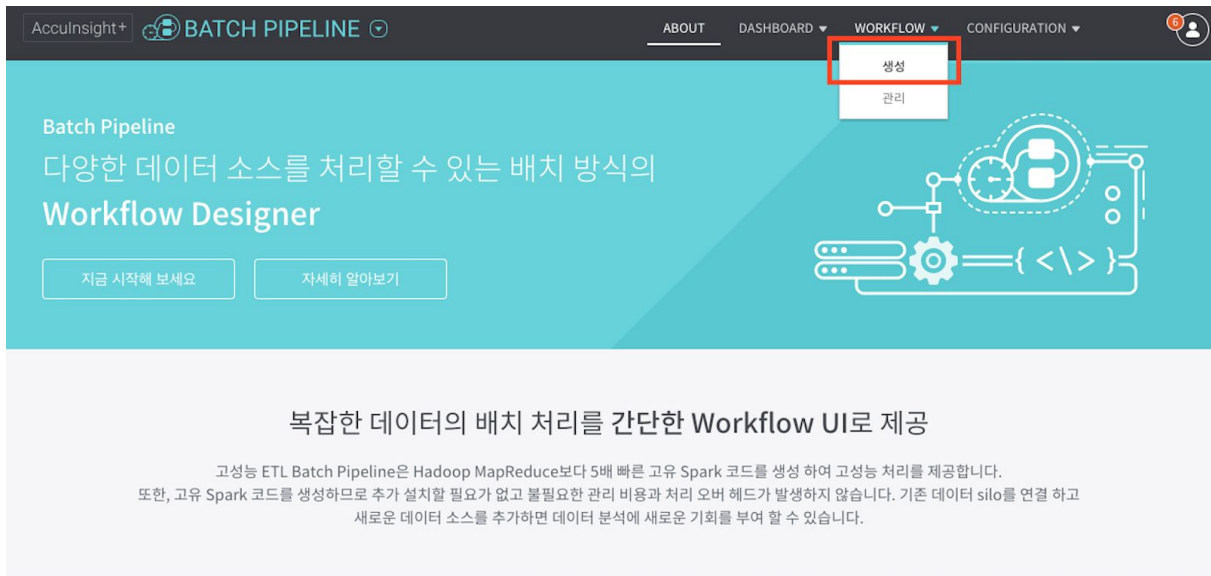
column명	column 설명	Value Example
bldg	빌딩명	SKC 본사, 서린 빌딩, ...
desk_type	좌석 유형	일반좌석, 스탠딩좌석, ...
occupied	고정좌석 여부	N, Y
desk_id	각 좌석의 ID	3a27b3bf-d8de-11e8-975a-0a9726611f46
dt	날짜	20190803

c. 데이터 처리 (Batch Pipeline)

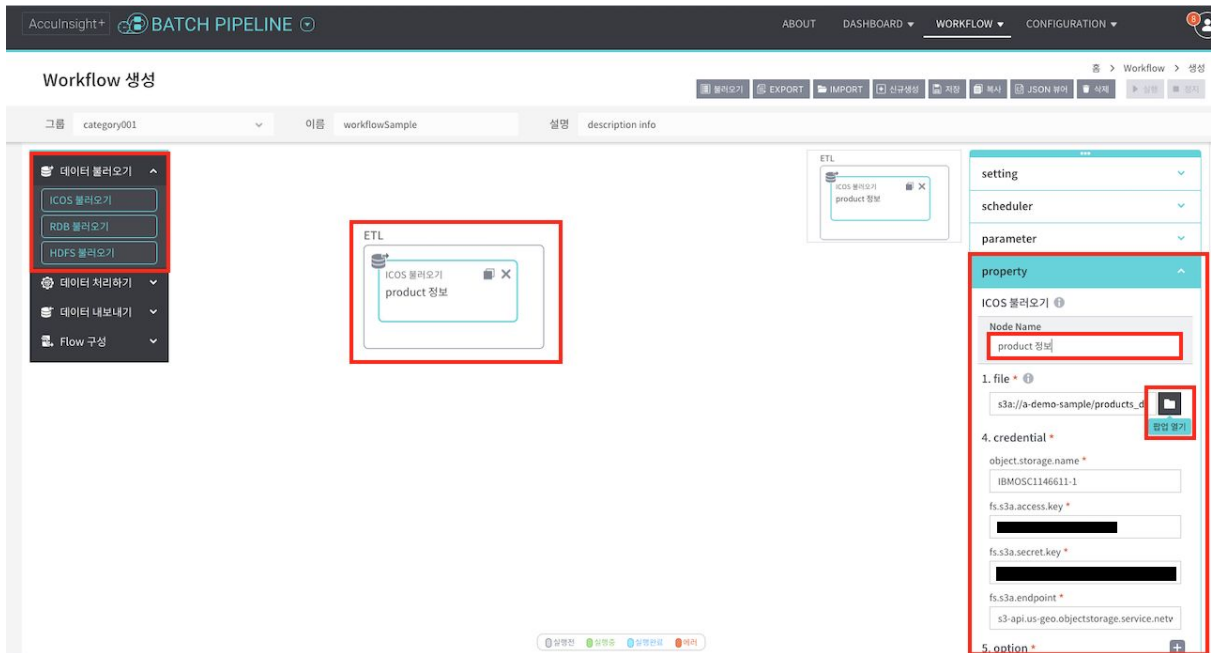
데이터파일 2개 조인하여 ETL작업 수행하기

ETL flow 구성내용 : ICOS에서 파일 불러오기 → 데이터 정제 → 통계 → join key 생성 → 2개 파일 join → 통계 → 불필요한 column 삭제 → column명 rename → 통계 및 정제를 통한 새로운 파일 생성 → 기존 파일과 통합 → ICOS에 저장

- 생성 메뉴로 이동



- 데이터 로딩



데이터 불러오기 그룹에서 ICOS 불러오기 노드를 캔버스 영역에 drag & drop하여 node 생성
우측 property 패널

- file : 불러오려는 ICOS 파일의 경로 지정
- option : 데이터에 header가 없는 경우 false로 변경
- schema : 자동 파일 열기

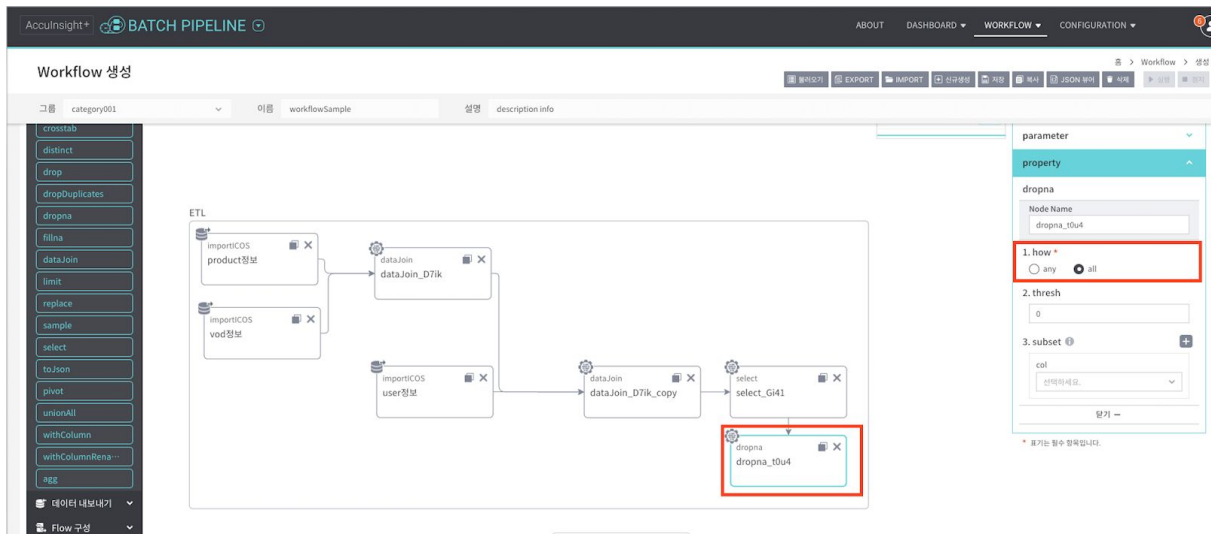
여기서 column명 지정

좌석 예약 데이터 : bldg, desk_type, desk_id, dt, com, age

동일한 방법으로 좌석 메타 데이터도 불러오기

좌석 메타 데이터 : bldg, desk_type, occupied, desk_id, dt

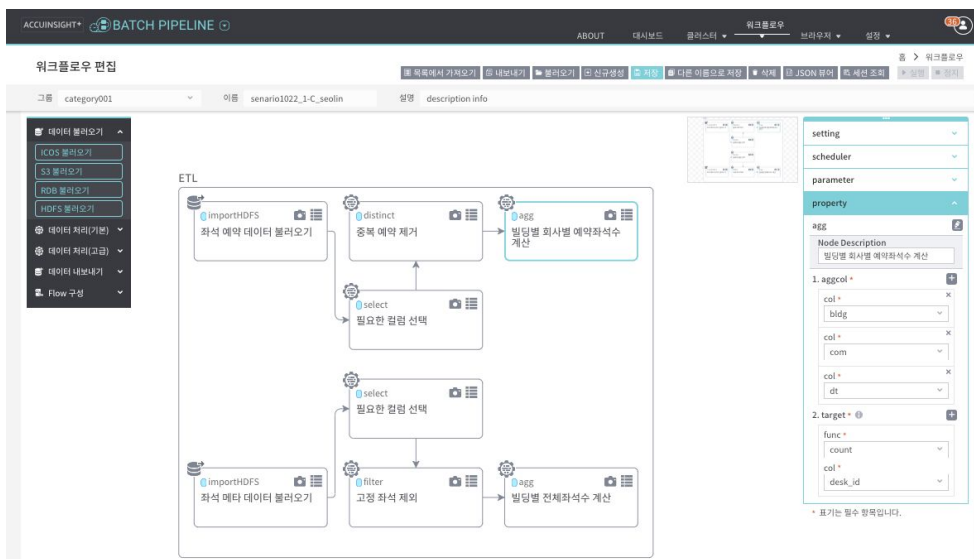
- 데이터 정제



distinct 노드 drag & drop하여 좌석 예약 데이터에서 중복 예약 제거
 filter 노드 drag & drop하여 좌석 메타 데이터에서 고정 좌석 제외
 우측 property 패널

- col : occupied
- filterOption : =:equal
- filterValue : 'N'

- 통계



agg 노드 drag & drop하여 좌석 예약 데이터에서 빌딩별 좌석유형별 회사별 나이대별
 예약좌석수 계산

우측 property 패널

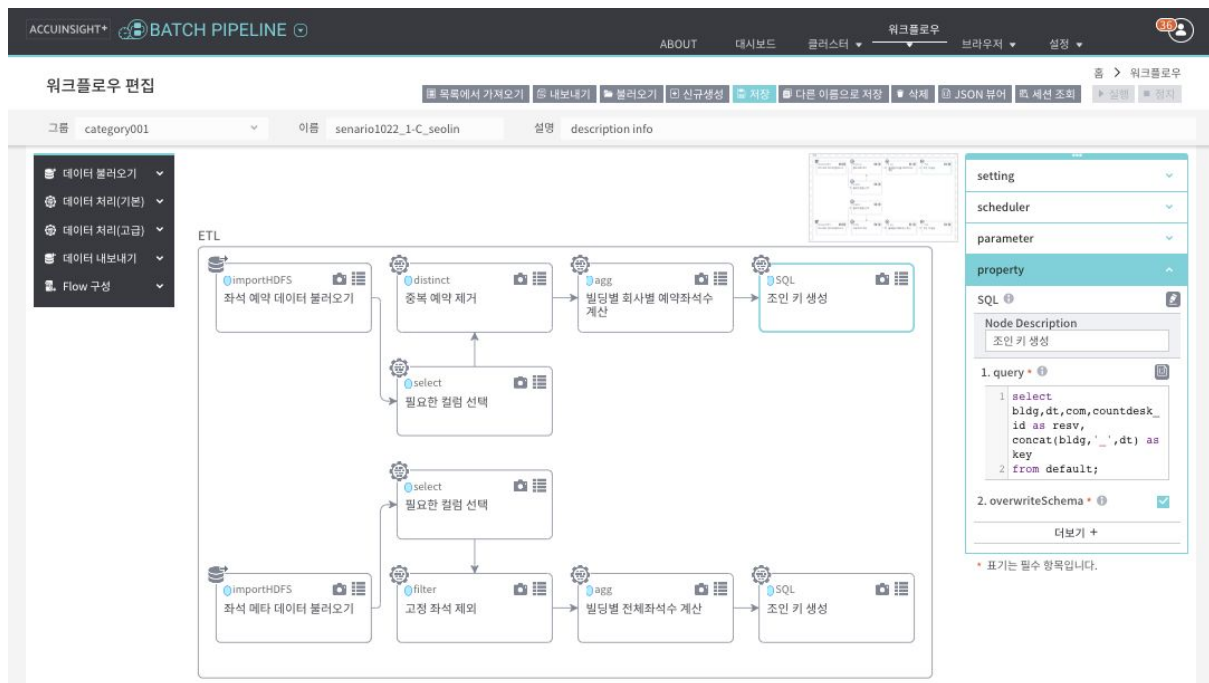
- aggcol : bldg, desk_type, dt, com, age
- target : func count, col desk_id

agg 노드 drag & drop하여 좌석 메타 데이터에서 빌딩별 좌석유형별 전체좌석수 계산

우측 property 패널

- aggcol : bldg, desk_type, dt
- target : func count, col desk_id

- join key 생성



SQL 노드 drag & drop하여 좌석 예약 데이터에서 join key 생성
우측 property 패널

- query

```
select bldg,desk_type,dt,com,age,countdesk_id as resv, concat(bldg,'_',desk_type,'_',dt)
as key
from default;
```

- overwriteSchema 체크 (SQL 결과로 데이터 변경)

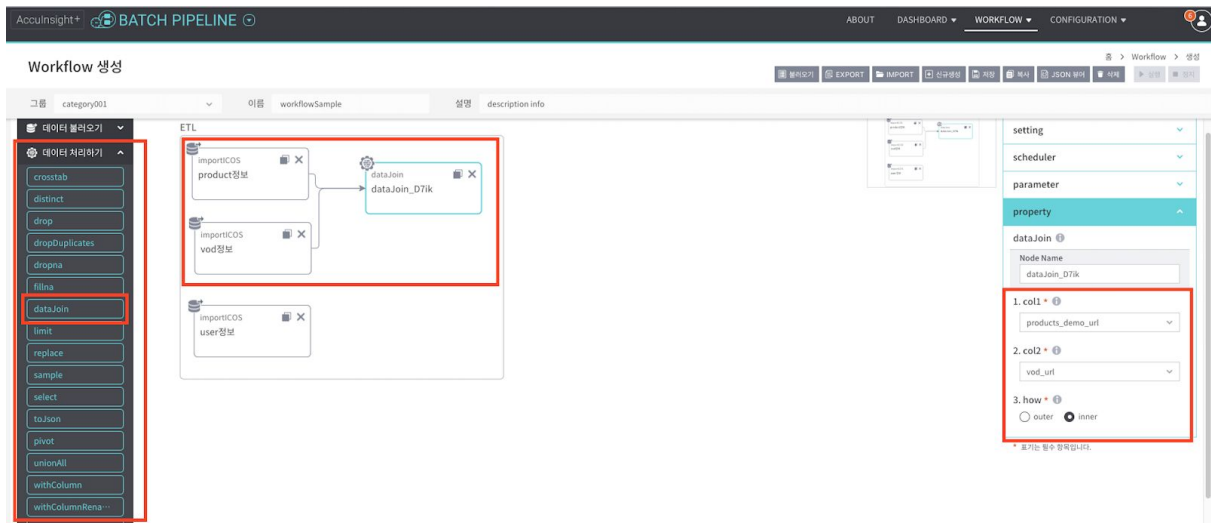
SQL 노드 drag & drop하여 좌석 메타 데이터에서 join key 생성
우측 property 패널

- query

```
select countdesk_id as total, concat(bldg,'_',desk_type,'_',dt) as key
from default;
```

- overwriteSchema 체크 (SQL 결과로 데이터 변경)

- 2개파일 조인



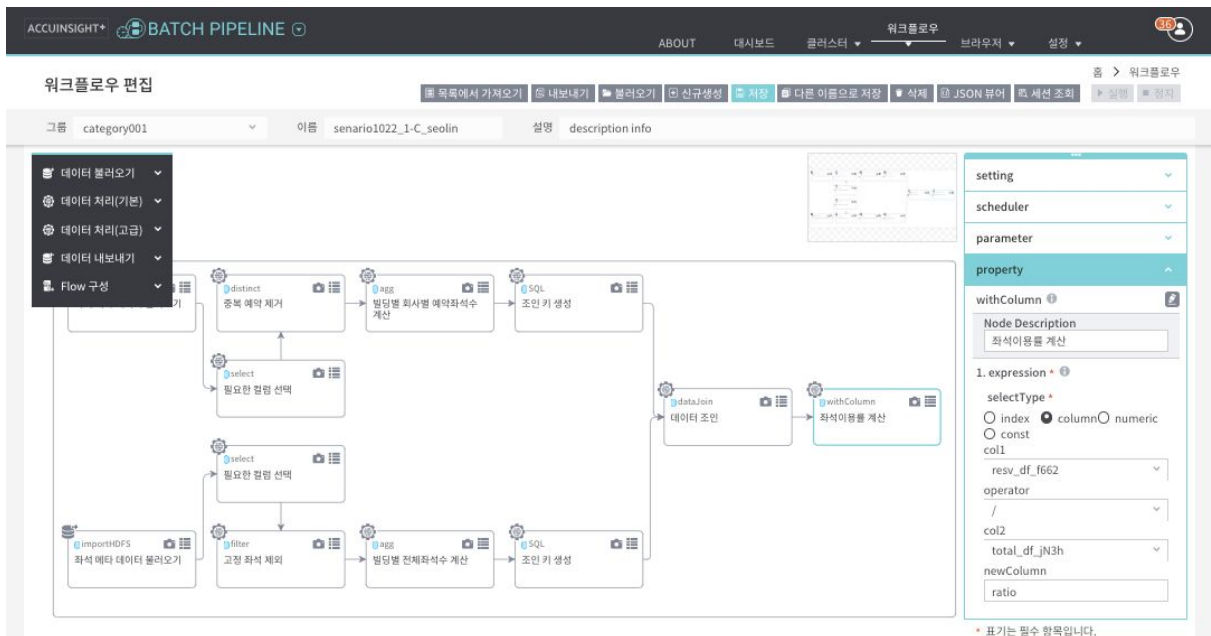
좌측의 데이터 처리하기 클릭

열린 패널에서 dataJoin 노드 우측 캔버스에 drag & drop하여 생성

우측 property패널

- col1 : key 선택
- col2 : key 선택
- how : right_outer 선택

- 새로운 column 생성



withColumn 노드 drag & drop하여 좌석이용률 계산

우측 property 패널

- selectType : column
- col1 : resv_dt_~
- operator : /
- col2 : total_dt_~
- newColumn : ratio

- 불필요한 column 삭제

The screenshot shows the ACCUINSIGHT+ BATCH PIPELINE interface. The workflow is titled 'scenario1022_1-C_seolin'. The 'drop' node is highlighted, and its properties are shown on the right. The properties include a list of columns to be dropped: total_df_jN3h, key_df_jN3h, ratio, bldg_df_f662, dt_df_f662, and com_df_f662. The 'key_df_jN3h' column is selected for dropping.

drop 노드 drag & drop하여 불필요한 데이터 삭제

우측 property 패널

- total_df_~
- key_df_~
- resv_df_~
- key_df_~

- null값 제거

The screenshot shows the ACCUINSIGHT+ BATCH PIPELINE interface. The workflow is titled 'scenario1022_1-C_seolin'. The 'filter' node is highlighted, and its properties are shown on the right. The properties include a list of filters: col * (com_df_f662), filterOption * (!=: not equal), and filterValue * ('N'). The 'filter' node is used to remove NULL values from the data.

filter 노드 drag & drop하여 회사명이 'N' 이거나 나이가 'N'이면 삭제

우측 property 패널

- col : com_df_~
- filterOption : !=:not equal

- filterValue : '\N'
- 필터 추가 & filterOperator는 OR로 설정
- col : age_df_~
- filterOption : !=:not equal
- filterValue : '\N'

- 컬럼명 변경

withColumnRenamed노드 drag&drop하여 생성
우측 property 패널
전체 컬럼 불러오기 아이콘을 통해 자동 파싱

- ICOS에 저장

좌측 데이터 내보내기 클릭
ICOS 내보내기 노드 drag & drop 하여 생성
withColumnRenamed 노드에서 ICOS 내보내기노드로 연결
우측 property 패널
- path의 browse 아이콘을 클릭하여 열리는 팝업에서 저장할 디렉토리 위치 선택 후 확인
클릭 (ex. /tmp)

- file에 생성할 디렉토리명 입력 (ex.modeling_data)

1. path * ?

hdfs://10.42.175.152:8020//tmp

2. file * ?

modeling_data

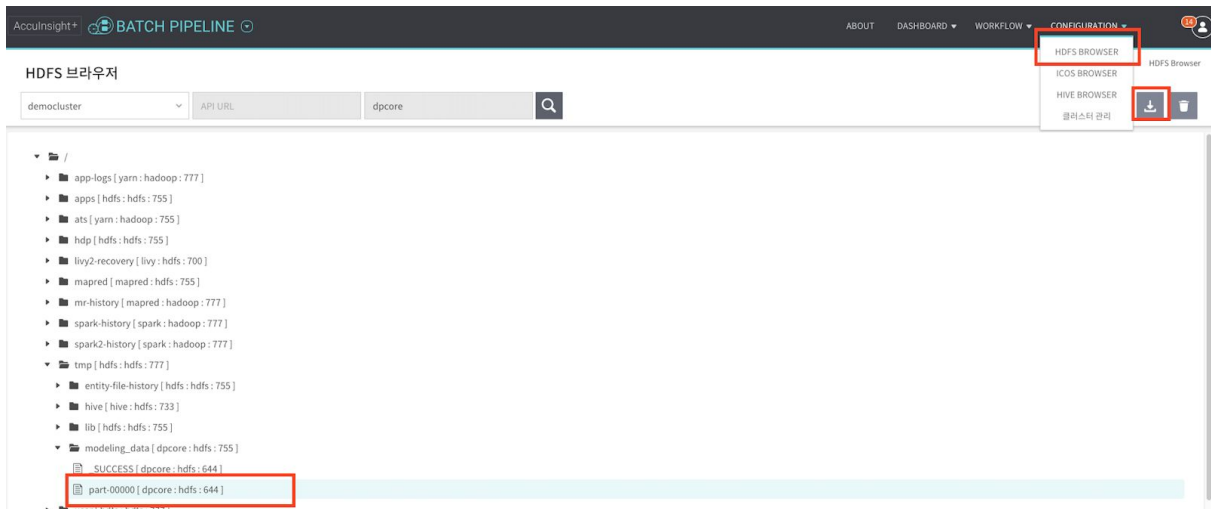
- 저장 및 실행

저장 클릭
실행 클릭

- 워크플로우 상태 확인

workflow 관리 화면으로 이동하여 상태 확인
이미 성공으로 종료되었거나 오류가 발생하여 종료된 경우 INACTIVE 상태
상세정보 컬럼의 Instance목록 아이콘 클릭

- 실행결과 확인



CONFIGURATION > ICOS BROWSER 메뉴로 이동

지정했던 ICOS directory위치로 이동하여 내보내기로 생성한 디렉토리 하단의

part-00000~파일 클릭

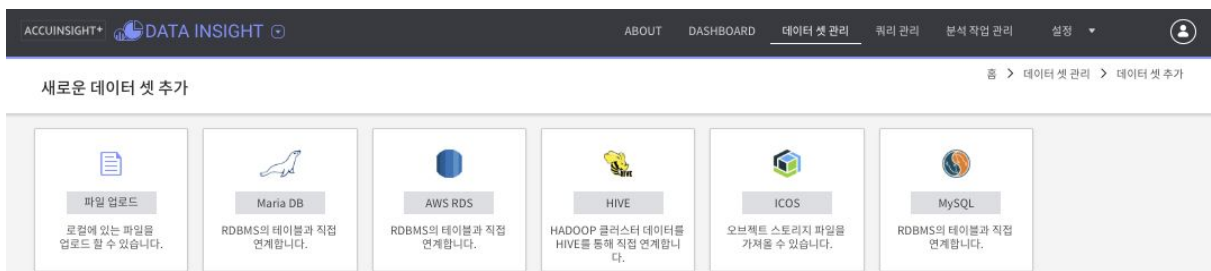
우측 상단의 다운로드 버튼 클릭하여 파일내용 확인

d. 데이터 시각화 (Data Insight)

데이터 준비

- 데이터 셋 추가

데이터 셋 관리 페이지에서 데이터 셋 추가를 클릭하면 추가 페이지로 이동합니다.



DataInsight 분석 및 시각화를 위해 MariaDB, ICOS, HIVE 등 다양한 종류의 데이터 셋을 추가할 수 있습니다.

지원하는 Data Source 타입

- ① 로컬 파일 : csv 타입 등 로컬에 있는 파일을 업로드
- ② MariaDB : MariaDB 테이블 연동하여 데이터 업로드
- ③ AWS RDS: AWS RDS 테이블 연동하여 데이터 업로드
- ④ HIVE : 하둡 클러스터의 데이터를 hive를 통하여 업로드
- ⑤ ICOS : IBM 오브젝트 스토리지의 파일을 업로드
- ⑥ MySQL : MySQL 테이블을 연동하여 데이터 업로드

데이터 셋 추가 : ICOS 선택

이전에 Batch Pipeline 실행 결과 파일(ICOS) 선택

스키마 편집

ACCUIINSIGHT+ DATA INSIGHT						
ABOUT DASHBOARD 데이터 셋 관리 쿼리 관리 분석 작업 관리 설정						
홈 > 데이터 셋 관리 > 데이터 셋 추가 > 스키마 편집						
part-00000-82a70935-9787-44f2-8cc5-...						
필드 추가	빌딩명	좌석유형	날짜	회사명	나이대	좌석이용률
입력하세요.	String	String	Integer	String	String	Float
<input checked="" type="checkbox"/> 빌딩명	SKC 본사	일반좌석	20190802	SKC	60대	0.10526315789473684
<input checked="" type="checkbox"/> 좌석유형	SKC 본사	일반좌석	20190802	MCNS	30대	0.008771929824561403
<input checked="" type="checkbox"/> 날짜	SKC 본사	일반좌석	20190802	SKC HT&M	50대	0.03508771929824561
<input checked="" type="checkbox"/> 회사명	SKC 본사	일반좌석	20190802	SKC	20대	0.05263157894736842
<input checked="" type="checkbox"/> 나이대	SKC 본사	일반좌석	20190802	MCNS	60대	0.02631578947368421
<input checked="" type="checkbox"/> 좌석이용률	SKC 본사	일반좌석	20190802	SKC	40대	0.2807017543859649
	SKC 본사	일반좌석	20190802	SKC	30대	0.2631578947368421
	SKC 본사	일반좌석	20190802	MCNS	50대	0.03508771929824561
	SKC 본사	일반좌석	20190802	SKC	50대	0.21929824561403508
	SKC 본사	일반좌석	20190802	MCNS	40대	0.02631578947368421

데이터 셋 이름 변경

part-00000~ -> 회사별 일평균 좌석이용률

column명 변경

bldg -> 빌딩명, desk_type -> 좌석유형, dt -> 날짜, com -> 회사명, age -> 나이대, ratio -> 좌석이용률

저장


차트 작성

- 분석 작업 관리

분석 작업 관리 페이지에서 **분석 작업 생성**을 클릭하면 추가 페이지로 이동합니다.

차트 작성할 데이터 셋 적용

ACCUIINSIGHT+

 DATA INSIGHT

ABOUT


DASHBOARD

데이터 셋 관리

쿼리 관리

분석 작업 관리

설정













홈 > 분석 작업 관리 > 새로운 분석 작업

검색 필터

새 쿼리로부터 생성

새 데이터 소스로부터 생성



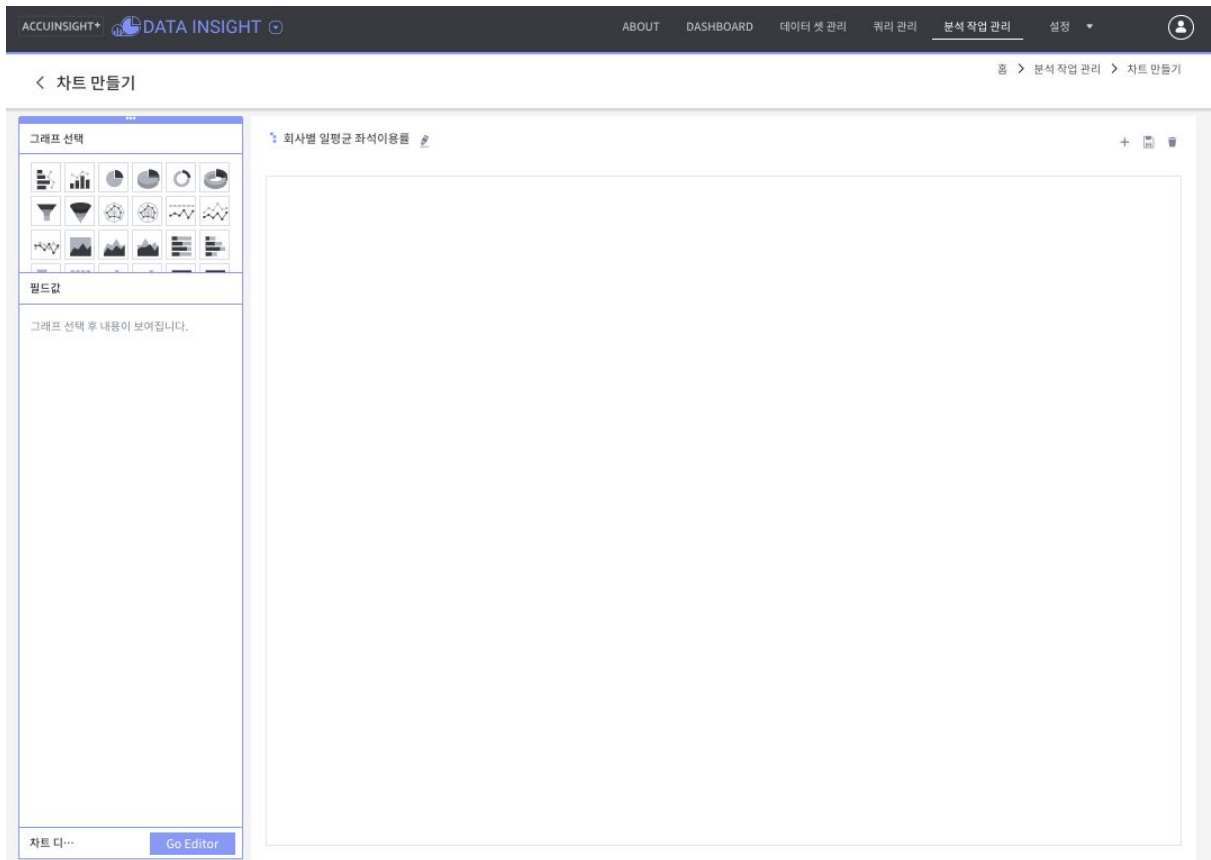
그룹	타입	데이터 셋 타입	이름	최종수정일	Action
Default	데이터 셋		좌석유형별 주 이용 나이	2019-10-25	<div><div></div><div>적용</div></div>
Default	데이터 셋		회사별 일평균 좌석이용률	2019-10-25	<div><div></div><div>적용</div></div>
Default	쿼리	-	규모별 회의실 이용현황 쿼리	2019-10-23	<div><div></div><div>적용</div></div>
Default	데이터 셋		part-00000-84cf0a49-f431-47f5-bb79-4da65cd7272b-c000.csv	2019-10-23	<div><div></div><div>적용</div></div>
Default	데이터 셋		kmeans result	2019-02-18	<div><div></div><div>적용</div></div>

필터 추가

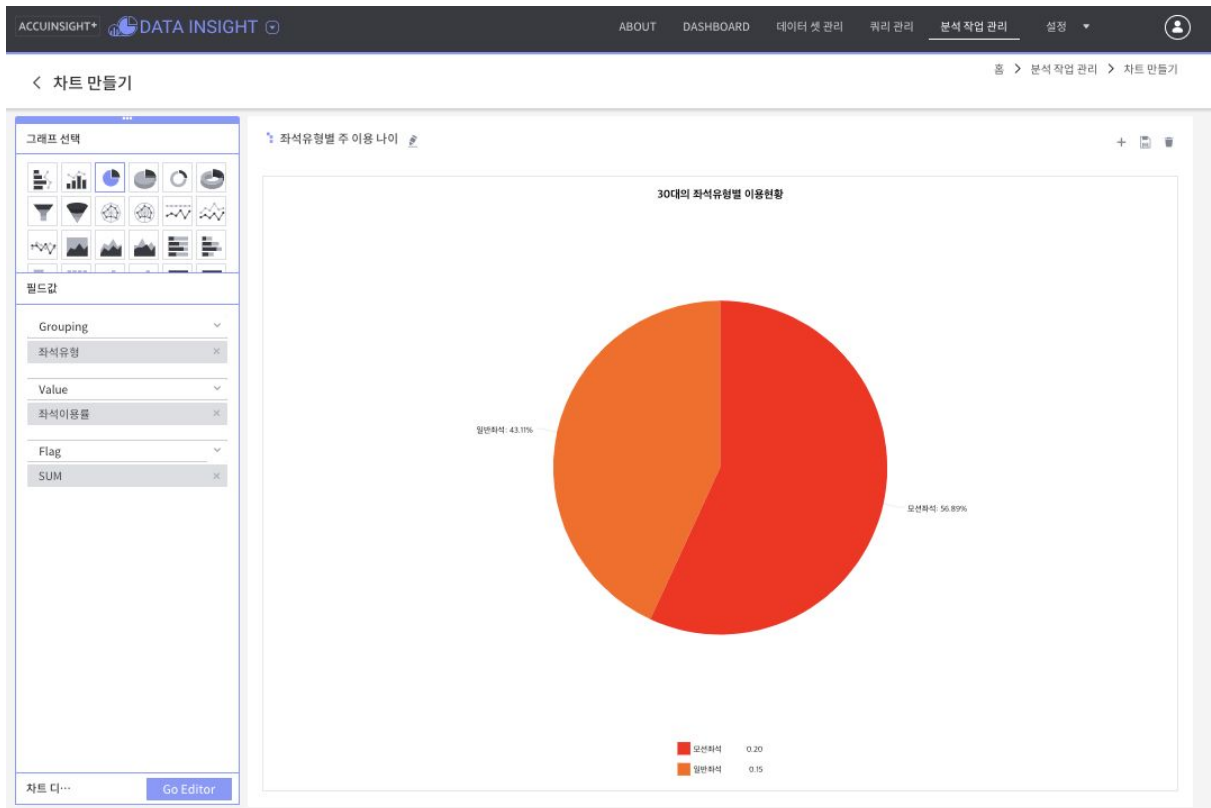
- 빌딩명 = 서린 빌딩
- 회의 = SK트레이딩인터내셔널 or SK이노베이션 or SK에너지 or SKC (카테고리 내에서 필터 추가)
- 날짜 = 20190820
- 나이대 = 30대

분석 작업 차트 생성

차트 생성 버튼 클릭



파이 차트 선택
Grouping : 좌석유형
Value : 좌석이용률
Flag : SUM



저장
차트 생성 버튼 클릭

Go Editor

소수점 2째 자리까지 : Number formatting에서 Precision을 2로 변경

애니메이션 삭제 : Miscellaneous에서 Start duration을 0으로 변경

Label 위치 변경 : Legend에서 Align 콤보박스 선택

Label의 Column 개수 증가 : Legend에서 Max columns 개수 지정

타이틀 설정 : Title에서 Text에 '30대의 좌석유형별 이용현황' 입력

6. Senario 3 : 회의실 규모별 이용현황

데이터 준비

ICOS에 업로드된 데이터를 사용합니다.

스토리지 : IBMOSC1146611-6

버킷 : handson-bucket

a. Dataset 1 : 회의실 예약 데이터

column명	column 설명	Value Example
bldg	빌딩명	SKC 본사, 서린 빌딩, ...
room_scale	회의실 규모	2, 4, 5, ...
room_resv_id	회의실 예약 ID	3a27b3bf-d8de-11e8-975a-0a9726611f46
time	회의실 사용 시점 (30분 단위)	09:30:00, 11:00:00, ...

dt	예약 날짜	20190803
----	-------	----------

b. Dataset 2 : 회의실 메타 데이터

column명	column 설명	Value Example
bldg	빌딩명	SKC 본사, 서린 빌딩, ...
room_scale	회의실 규모	2, 4, 5, ...
room_id	각 회의실의 ID	3a27b3bf-d8de-11e8-975a-0a9726611f46
room_type	회의실 유형	일반회의실, 프로젝트룸, ...
dt	날짜	20190803

c. 데이터 처리 (Batch Pipeline)

데이터파일 2개 조인하여 ETL작업 수행하기

ETL flow 구성내용 : ICOS에서 파일 불러오기 → 데이터 정제 → 통계 → join key 생성 → 2개 파일 join → 통계 → 불필요한 column 삭제 → column명 rename → 통계 및 정제를 통한 새로운 파일 생성 → 기존 파일과 통합 → ICOS에 저장

- 생성 메뉴로 이동



AccuInsight+ BATCH PIPELINE

ABOUT DASHBOARD WORKFLOW CONFIGURATION

생성
관리

Batch Pipeline

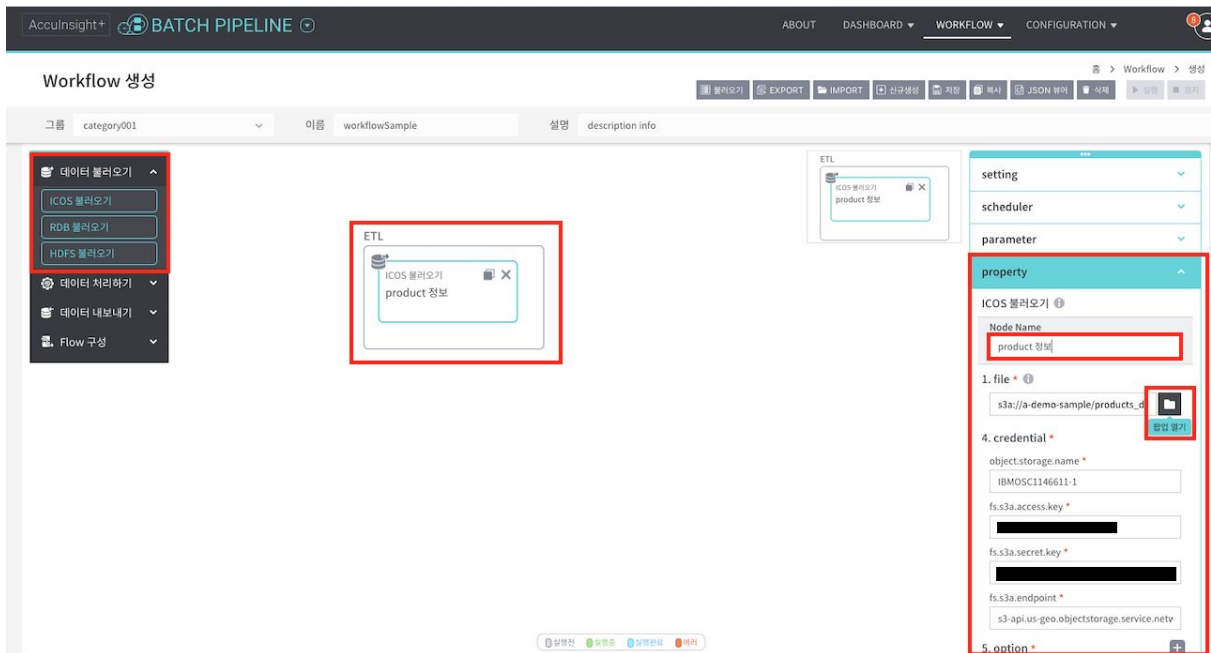
다양한 데이터 소스를 처리할 수 있는 배치 방식의 Workflow Designer

지금 시작해 보세요 자세히 알아보기

복잡한 데이터의 배치 처리를 간단한 Workflow UI로 제공

고성능 ETL Batch Pipeline은 Hadoop MapReduce보다 5배 빠른 고유 Spark 코드를 생성 하여 고성능 처리를 제공합니다.
또한, 고유 Spark 코드를 생성하므로 추가 설치할 필요가 없고 불필요한 관리 비용과 처리 오버 헤드가 발생하지 않습니다. 기존 데이터 silo를 연결 하고 새로운 데이터 소스를 추가하면 데이터 분석에 새로운 기회를 부여 할 수 있습니다.

- 데이터 로딩



데이터 불러오기 그룹에서 ICOS 불러오기 노드를 캔버스 영역에 drag & drop하여 node 생성
우측 property 패널

- file : 불러오려는 ICOS 파일의 경로 지정
- option : 데이터에 header가 없는 경우 false로 변경
- schema : 자동 파일 열기

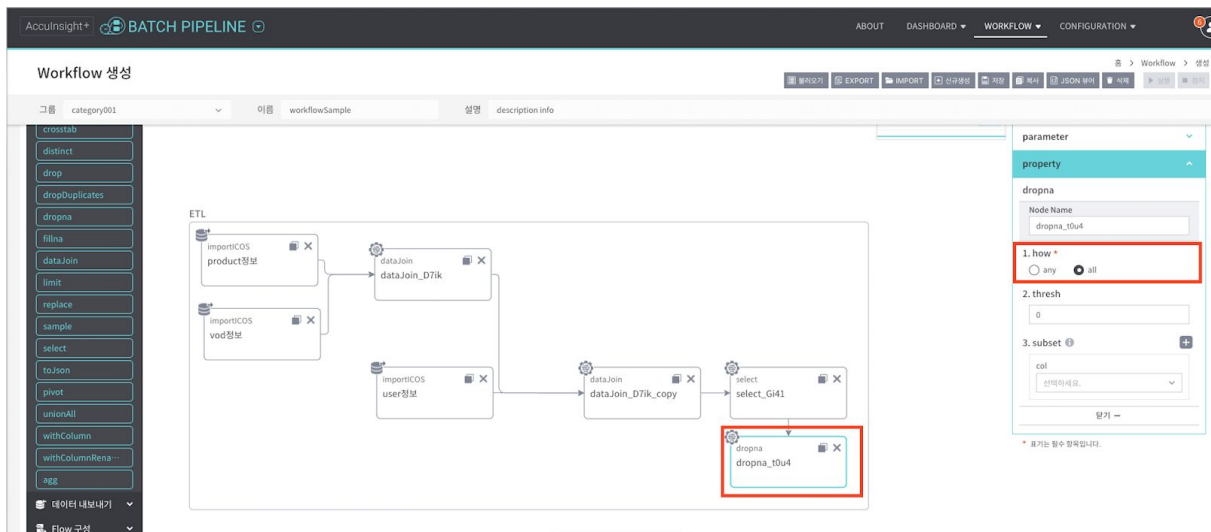
여기서 column명 지정

회의실 예약 데이터 : bldg, room_scale, room_resv_id, time, dt

동일한 방법으로 회의실 메타 데이터도 불러오기

회의실 메타 데이터 : bldg, room_scale, room_id, room_type, dt

- 데이터 정제



distinct 노드 drag & drop하여 회의실 예약 데이터에서 중복 예약 제거

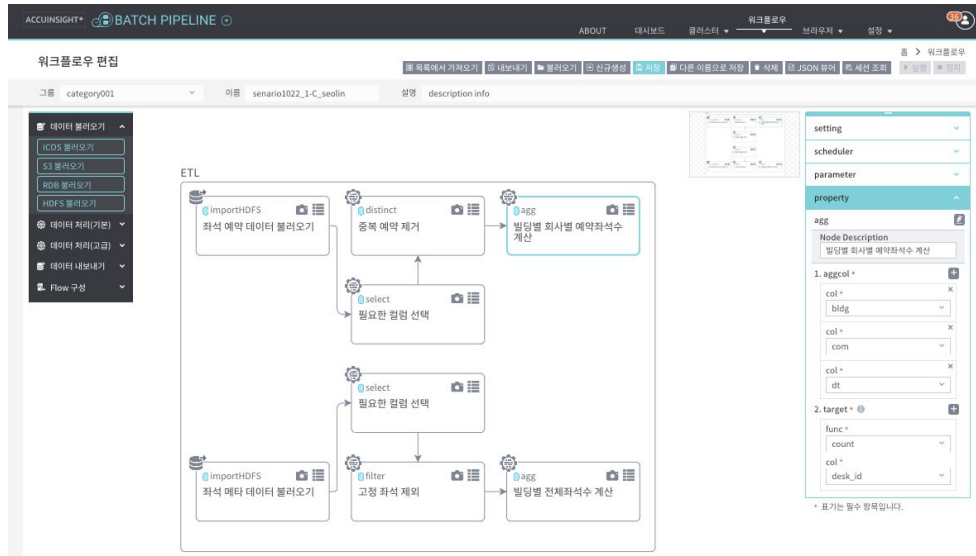
filter 노드 drag & drop하여 회의실 메타 데이터에서 일반회의실만 필터링

우측 property 패널

- col : room_type
- filterOption : =:equal

- filterValue : '일반회의실'

- 통계



agg 노드 drag & drop하여 회의실 예약 데이터에서 빌딩별 규모별 시간대별 예약회의실수 계산

우측 property 패널

- aggcol : bldg, room_scale, dt, time

- target : func count, col room_resv_id

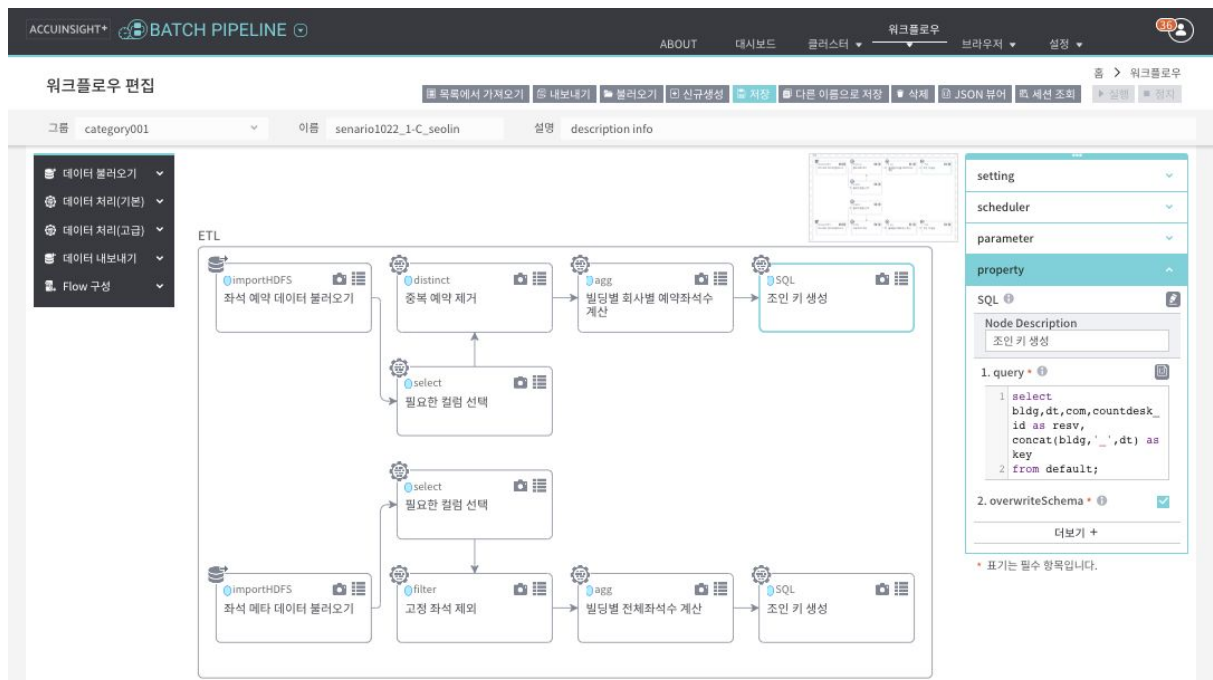
agg 노드 drag & drop하여 회의실 메타 데이터에서 빌딩별 규모별 전체회의실수 계산

우측 property 패널

- aggcol : bldg, room_scale, dt

- target : func count, col room_id

- join key 생성



SQL 노드 drag & drop하여 회의실 예약 데이터에서 join key 생성
우측 property 패널

- query

```
select bldg,room_scale,dt,time,countroom_resv_id as  
resv,concat(bldg,'_',room_scale,'_',dt) as key  
from default;
```

- overwriteSchema 체크 (SQL 결과로 데이터 변경)

SQL 노드 drag & drop하여 회의실 메타 데이터에서 join key 생성
우측 property 패널

- query

```
select countroom_id as total,concat(bldg,'_',room_scale,'_',dt) as key  
from default;
```

- overwriteSchema 체크 (SQL 결과로 데이터 변경)

- 2개파일 조인

The screenshot displays the Accusight+ BATCH PIPELINE interface. On the left, a sidebar titled '데이터 처리하기' (Data Processing) lists various SQL operations, with 'dataJoin' highlighted. The main workspace shows a workflow diagram with three input nodes ('importCOS product정보', 'importCOS vod정보', 'importCOS user정보') feeding into a 'dataJoin' node. The right-hand 'setting' panel is open to the 'property' tab for the 'dataJoin' node, showing configuration for 'col1' (products_demo_url), 'col2' (vod_url), and 'how' (inner join).

좌측의 데이터 처리하기 클릭

열린 패널에서 dataJoin 노드 우측 캔버스에 drag & drop하여 생성
우측 property패널

- col1 : key 선택

- col2 : key 선택

- how : right_outer 선택

- 새로운 column 생성

ACCUSIGHT+ BATCH PIPELINE

ABOUT 대시보드 클러스터 워크플로우 브라우저 설정

워크플로우 편집

그룹 category001 이름 scenario1022_1-C_seolin 설명 description info

데이터 불러오기
데이터 처리(기본)
데이터 처리(고급)
데이터 내보내기
Flow 구성

distinct 중복 예약 제거
select 필요한 컬럼 선택
importHDFS 작성 메타 데이터 불러오기
filter 고정 작성 제외
HBS 범당별 회사별 예약작성수 계산
SQL 조인 키 생성
dataJoin 데이터 조인
withColumn 작성이용률 계산

setting
scheduler
parameter
property

withColumn
Node Description
작성이용률 계산

1. expression
selectType
index
column
numeric
col1
resv_df_f662
operator
/
col2
total_df_jN3h
newColumn
ratio

* 표기는 필수 항목입니다.

withColumn 노드 drag & drop하여 회의실이용률 계산

우측 property 패널

- selectType : column
- col1 : resv_dt_~
- operator : /
- col2 : total_dt_~
- newColumn : ratio

- 불필요한 column 삭제

ACCUSIGHT+ BATCH PIPELINE

ABOUT 대시보드 클러스터 워크플로우 브라우저 설정

워크플로우 편집

그룹 category001 이름 scenario1022_1-C_seolin 설명 description info

데이터 불러오기
데이터 처리(기본)
데이터 처리(고급)
데이터 내보내기
Flow 구성

ETL

importHDFS 작성 예약 데이터 불러오기
distinct 중복 예약 제거
select 필요한 컬럼 선택
importHDFS 작성 메타 데이터 불러오기
filter 고정 작성 제외
HBS 범당별 회사별 예약작성수 계산
SQL 조인 키 생성
dataJoin 데이터 조인
withColumn 작성이용률 계산
drop 안쓰는 컬럼 삭제

setting
scheduler
parameter
property

drop
Node Description
안쓰는 컬럼 삭제

1. col
선택하세요.
total_df_jN3h
key_df_jN3h
ratio
bidg_df_f662
dt_f662
com_df_f662

* 표기는 필수 항목입니다.

drop 노드 drag & drop하여 불필요한 데이터 삭제

우측 property 패널

- total_df_~

- key_df_~
- resv_df_~
- key_df_~

- 컬럼명 변경

Workflow 생성

parameter

property

withColumnRenamed

Node Name

withColumnRenamed_UQL8

1. columns *

old Name *

vod_ts_df_ckQ3_df_D7ik

new Name *

ts

old Name *

products_demo_title_df_FT4F...

new Name *

title

old Name *

vod_url_df_ckQ3_df_D7ik

new Name *

url

old Name *

withColumnRenamed노드 drag&drop하여 생성
우측 property 패널
전체 컬럼 불러오기 아이콘을 통해 자동 파싱

- 통계

워크플로우 편집

setting

scheduler

parameter

property

agg

Node Description

일당별 회사별 예약좌석수 계산

1. aggcol *

col *

bldg

col *

com

col *

dt

2. target *

func *

count

col *

desk_id

agg 노드 drag & drop하여 회의실 전체에 대한 이용률 계산
우측 property 패널

- aggcol : bldg, dt, time
- target : func avg, col ratio

- 새로운 column 생성

ACCUSIGHT+ BATCH PIPELINE

ABOUT 대시보드 클러스터 워크플로우 브라우저 설정

워크플로우 편집

그룹 category001 이름 scenario1022_1-C_seolin 설명 description info

데이터 불러오기
데이터 처리(기본)
데이터 처리(고급)
데이터 내보내기
Flow 구성

distinct 중복 예약 제거
select 필요한 컬럼 선택
importHDFS 좌석 메타 데이터 불러오기
filter 고정 좌석 제외
groupBy 범당별 회차별 예약좌석수 계산
groupBy 범당별 회차별 예약좌석수 계산
SQL 조인 키 생성
dataJoin 데이터 조인
withColumn 좌석이용률 계산

setting
scheduler
parameter
property

withColumn
Node Description
좌석이용률 계산

1. expression
selectType
index
const
column
numeric
col1
resv_df_f662
operator
/
col2
total_df_jN3h
newColumn
ratio

* 표기는 필수 항목입니다.

withColumn 노드 drag & drop하여 회의실 규모가 '전체'인 컬럼 추가
우측 property 패널

- selectType : const
- newColumn : room_scale
- constantValue : all
- valueType : String

- 기존 파일과 통합하기 위해 컬럼 순서 변경

ACCUSIGHT+ BATCH PIPELINE

ABOUT 대시보드 클러스터 워크플로우 브라우저 설정

워크플로우 편집

그룹 category001 이름 scenario1022_4-B 설명 description info

데이터 불러오기
데이터 처리(기본)
데이터 처리(고급)
데이터 내보내기
Flow 구성

distinct 중복 예약 제거
groupBy 범당별 회차별 예약좌석수 계산
SQL 조인 키 생성
dataJoin 데이터 조인
withColumn 회의실이용률 계산
drop 안쓰는 컬럼 삭제
withColumnRenamed 컬럼명 변경
importHDFS 회의실 메타 데이터 불러오기
filter 일반 회의실만 필터링
groupBy 범당별 규모별 전체회의실수 계산
groupBy 범당별 규모별 전체회의실수 계산
SQL 조인 키 생성
withColumn 규모가 '전체'인 컬럼 추가
SQL 컬럼 순서 변경

setting
scheduler
parameter
property

SQL
Node Description
컬럼 순서 변경

1. query
select bldg,
room_scale, dt, time,
avgratio as ratio
from default;

2. overwriteSchema
더보기 +

* 표기는 필수 항목입니다.

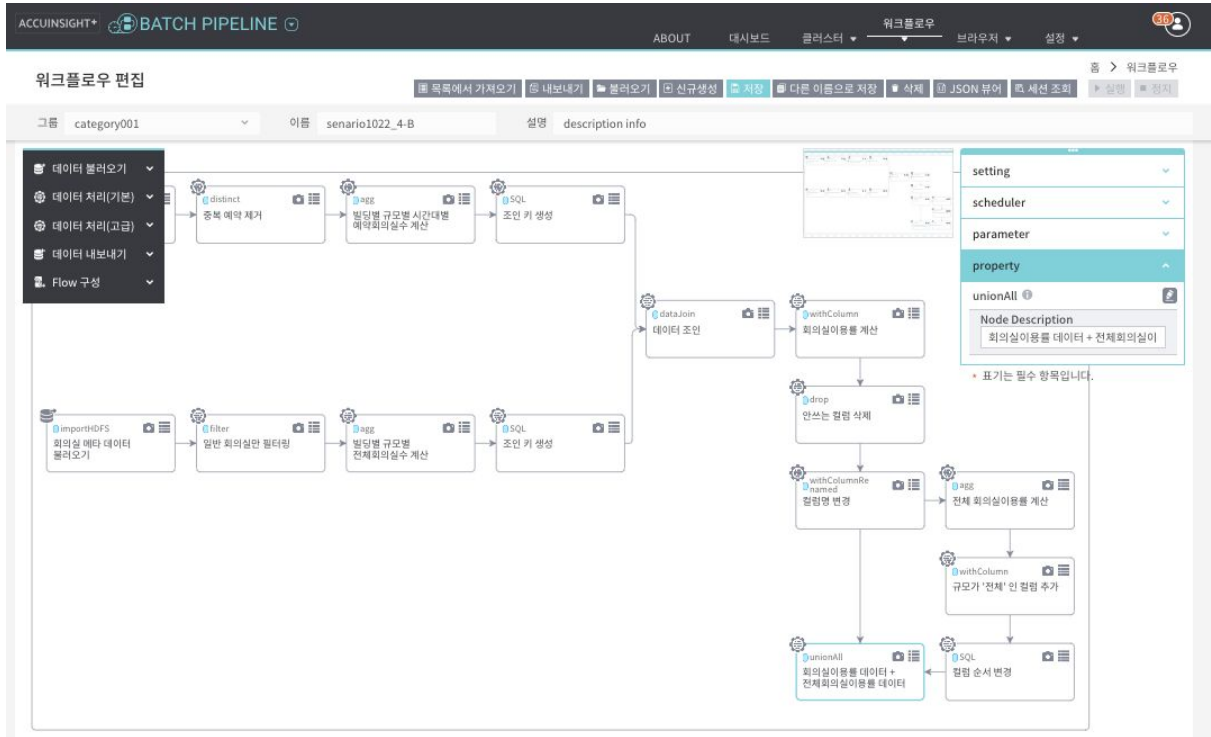
SQL 노드 drag & drop하여 컬럼 순서 변경
우측 property 패널

- query

```
select bldg, room_scale, dt, time, avgratio as ratio
from default;
```

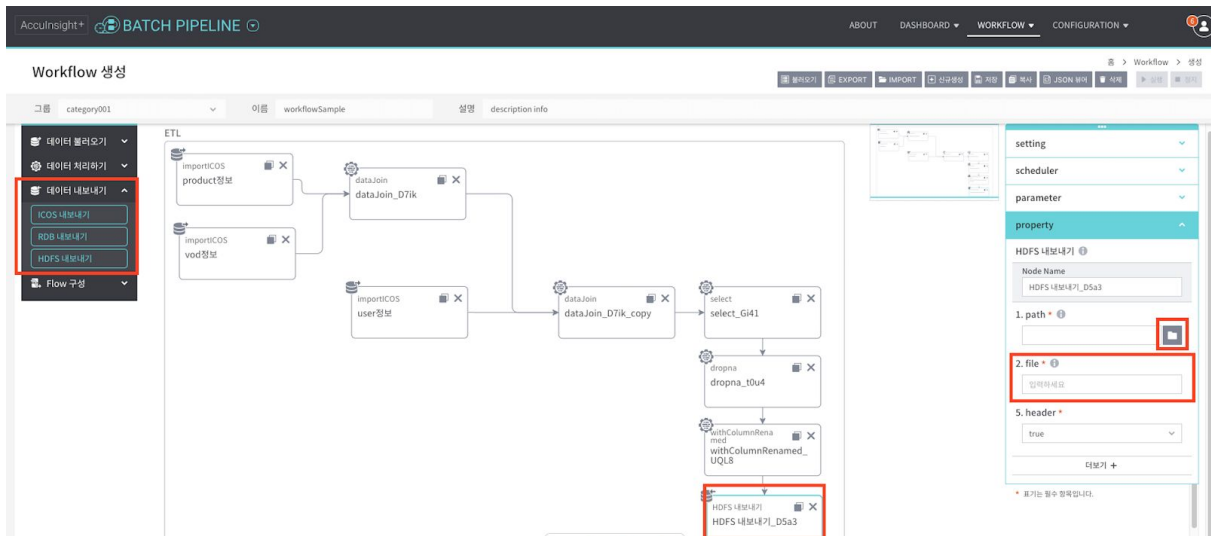
- overwriteSchema 체크 (SQL 결과로 데이터 변경)

- 기존 파일과 통합



unionAll 노드 drag & drop하여 데이터 통합

- ICOS에 저장



좌측 데이터 내보내기 클릭

ICOS 내보내기 노드 drag & drop 하여 생성

withColumnRenamed 노드에서 ICOS 내보내기노드로 연결

우측 property 패널

- path의 browse 아이콘을 클릭하여 열리는 팝업에서 저장할 디렉토리 위치 선택 후 확인 클릭 (ex. /tmp)
- file에 생성할 디렉토리명 입력 (ex.modeling_data)

1. path * ?

2. file * ?

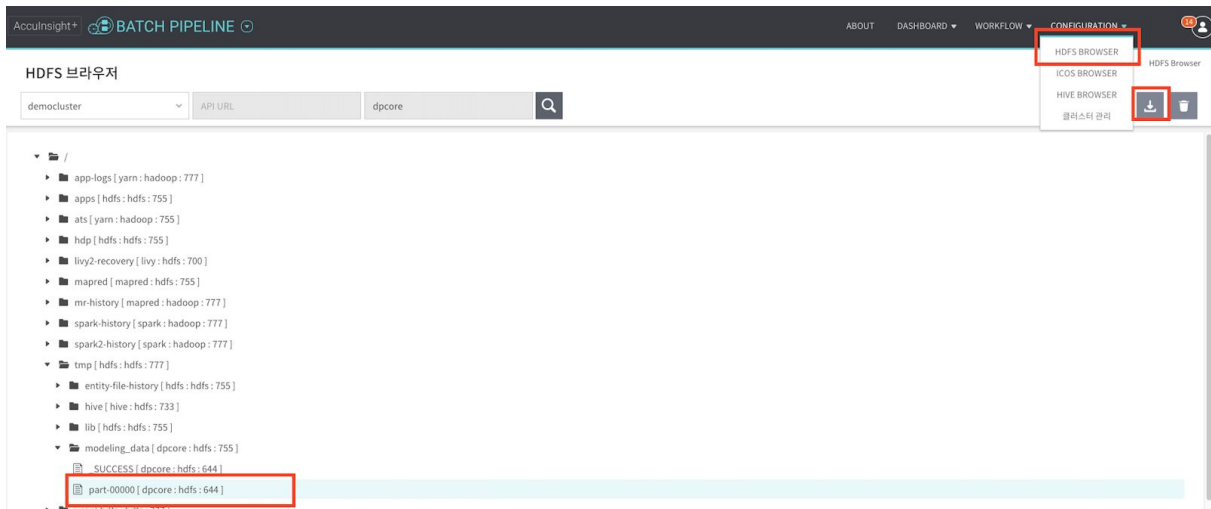
- 저장 및 실행

저장 클릭
실행 클릭

- 워크플로우 상태 확인

workflow 관리 화면으로 이동하여 상태 확인
이미 성공으로 종료되었거나 오류가 발생하여 종료된 경우 INACTIVE 상태
상세정보 컬럼의 Instance목록 아이콘 클릭

- 실행결과 확인



CONFIGURATION > ICOS BROWSER 메뉴로 이동

지정했던 ICOS directory위치로 이동하여 내보내기로 생성한 디렉토리 하단의

part-00000~파일 클릭

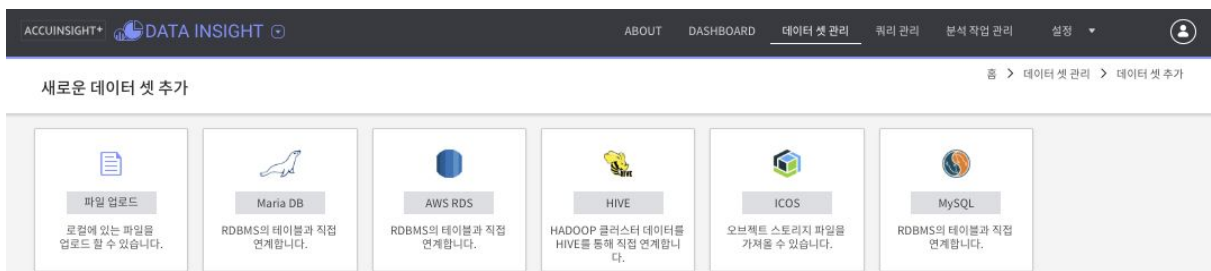
우측 상단의 다운로드 버튼 클릭하여 파일내용 확인

d. 데이터 시각화 (Data Insight)

데이터 준비

- 데이터 셋 추가

데이터 셋 관리 페이지에서 데이터 셋 추가를 클릭하면 추가 페이지로 이동합니다.



DataInsight 분석 및 시각화를 위해 MariaDB, ICOS, HIVE 등 다양한 종류의 데이터 셋을 추가할 수 있습니다.

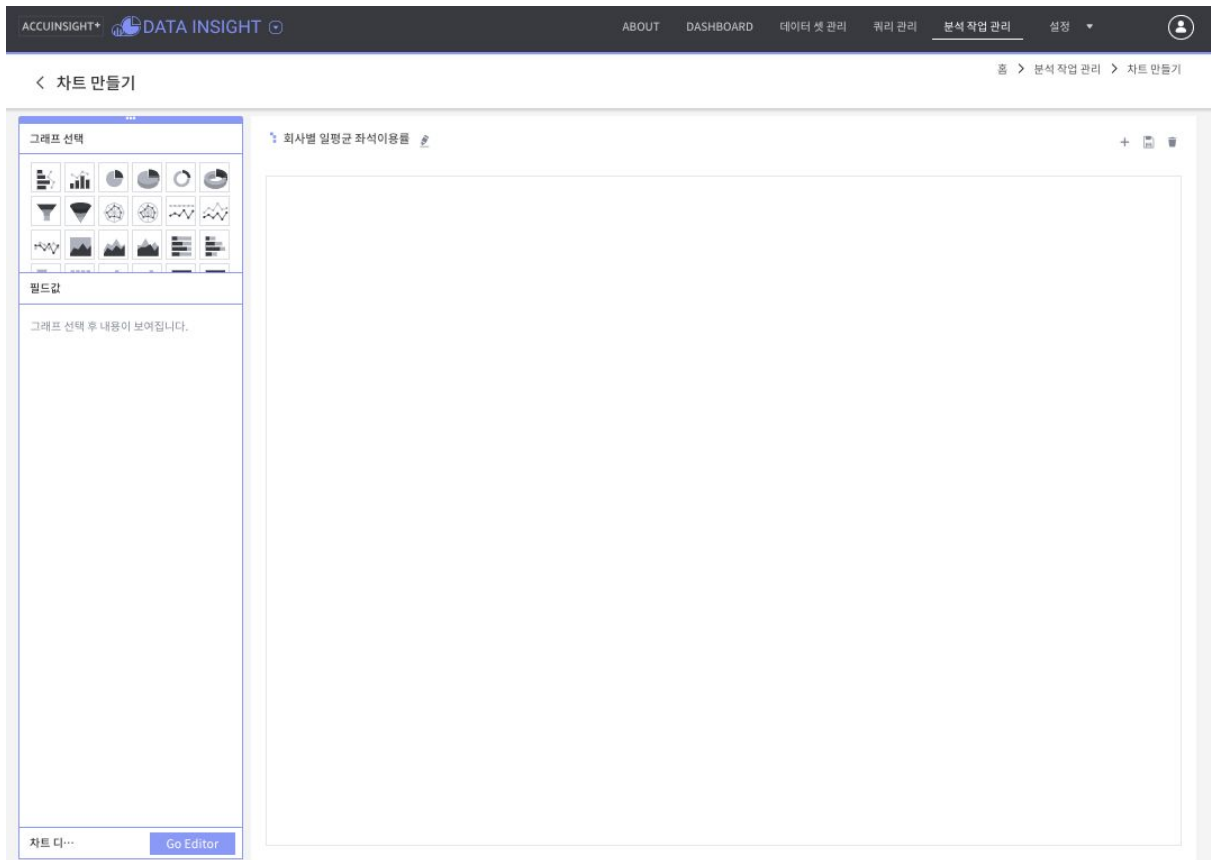
지원하는 Data Source 타입

- ① 로컬 파일 : csv 타입 등 로컬에 있는 파일을 업로드
- ② MariaDB : MariaDB 테이블 연동하여 데이터 업로드
- ③ AWS RDS: AWS RDS 테이블 연동하여 데이터 업로드
- ④ HIVE : 하둡 클러스터의 데이터를 hive를 통하여 업로드
- ⑤ ICOS : IBM 오브젝트 스토리지의 파일을 업로드
- ⑥ MySQL : MySQL 테이블을 연동하여 데이터 업로드

데이터 셋 추가 : ICOS 선택

이전에 Batch Pipeline 실행 결과 파일(ICOS) 선택

스키마 편집



라인 차트 선택

Axis : 시점

Grouping : 회의실규모

Value : 회의실이용률

Flag : SUM



저장

차트 생성 버튼 클릭

Go Editor

소수점 2째 자리까지 : Number formatting에서 Precision을 2로 변경

애니메이션 삭제 : Miscellaneous에서 Start duration을 0으로 변경

Label 위치 변경 : Legend에서 Align 콤보박스 선택

Label의 Column 개수 증가 : Legend에서 Max columns 개수 지정

타이틀 설정 : Title에서 Text에 '회의실 규모별 이용현황' 입력