

Data Analytics

# Data Wrangler

## 쉽고 빠르게 데이터를 탐색하고 원하는 형태로 재구성해주는 서비스

Data Wrangler는 수집된 Raw 데이터를 사용자에게 익숙한 엑셀 형태의 시각적인 데이터로 변환하여 제공하며, 이를 통해 데이터를 정리하고 분석하는 시간을 줄일 수 있습니다. 데이터 탐색부터 변환까지 데이터 처리 과정은 모두 시각적으로 확인 가능하며, 데이터 프로파일 정보, Join 유형 및 조건, 데이터 원천 등을 클릭만으로 간편하게 확인할 수 있습니다.

### 시각적 데이터 준비

웹 기반으로 편리하게 Self Service로 데이터 전처리가 가능합니다. 데이터를 탐색하고 원하는 데이터를 로드하여, 시각적으로 확인하면서 데이터의 보정, 변환 등을 처리하는 데이터 처리 함수를 제공합니다.

### 데이터 프로파일링 기반 탐색

데이터 프로파일링을 통해 데이터 분포 현황, 데이터 유효성 검사, 통계 정보 등을 확인할 수 있습니다. 선택한 데이터 셋의 전체적인 현황을 한 눈에 파악하여 데이터 정제 대상 및 처리 유형을 결정할 수 있습니다. 컬럼별 데이터 분포와 Invalid data를 확인하여 데이터 보정을 제안하며, 최대값, 최소값, 평균값, 최빈값 등을 확인할 수 있습니다.

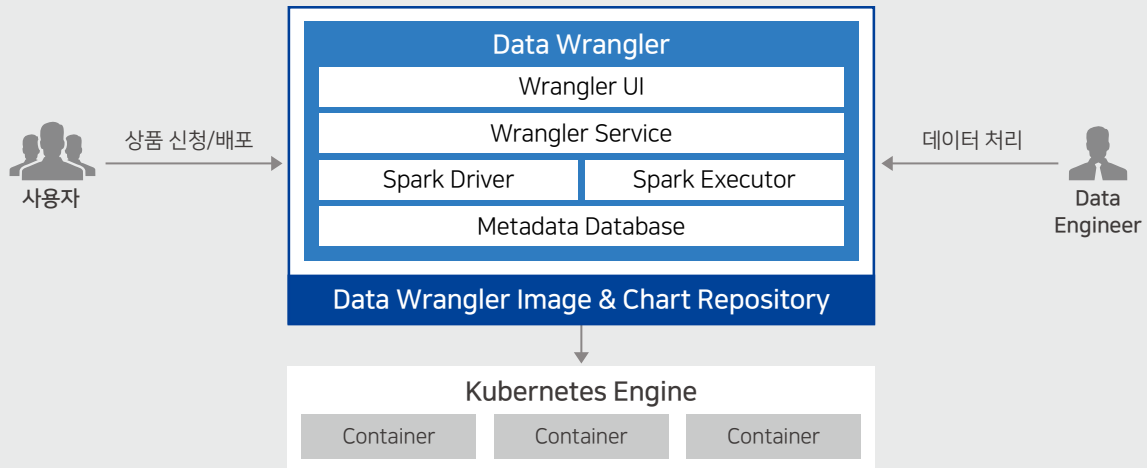
### 작업 내용 Recipe 관리

데이터가 변환된 히스토리를 저장, 관리하여 변환된 스크립트를 확인하거나 로딩된 데이터, 컬럼 변환 내역 관리 등 작업 내용에 대한 커뮤니케이션 기능을 제공합니다.

### 데이터 원천 확인

서로 다른 출처의 데이터를 병합 처리할 수 있으며, 변환된 데이터에 대한 출처와 적용된 Recipe간의 관계는 Lineage Diagram을 통해 간편하게 확인할 수 있습니다.

## 서비스 구성도



## 주요 기능




- 간편한 설치
  - Data Wrangler 신청 시 Kubernetes Engine 까지 한번에 신청 가능
  - Kubernetes Engine의 자원은 Data Wrangler에서 신청한 자원 규모 이상으로 설정할 수 있어 사용자 실수에 의한 에러 방지 가능
- 다양한 데이터 연계 기능
  - 연결된 데이터 소스의 Schema 정보 사용 (Hive Schema, RDB Schema)
  - SQL을 이용한 데이터 로드
  - Local File 기능을 이용한 대상 데이터 업로드
- 다양한 데이터 분석 함수 제공
  - Group 함수 : count, sum, avg, min, max, first, last, countDistinct, sumDistinct, collect\_list, collect\_set 등
  - Window 함수 : lag, lead, rank, dense\_rank, row\_number 등
  - 데이터 전처리에 필요한 함수, Math 함수 뿐만 아니라 내장된 다양한 Scalar 함수 사용
- Job 관리 및 모니터링
  - 전처리 작업의 결과물인 Recipe를 전체 데이터에 적용하는 Job을 관리하고 실행 현황을 모니터링
  - Job 상태별 조회, Name 조회
  - Job 목록, 상태, 실행시간 등의 상세 현황 모니터링

## 요금 기준

- 과금
  - Data Wrangler가 사용하는 Kubernetes Engine의 Pod들의 CPU 사용시간
  - Kubernetes Engine, Worker Node(VM), 스토리지 사용 요금은 별도 부과

FOR MORE INFORMATION

**SAMSUNG SDS**

 [www.samsungsds.com](http://www.samsungsds.com) / [cloud.samsungsds.com](http://cloud.samsungsds.com)  
 [contact.sds@samsung.com](mailto:contact.sds@samsung.com) / [scp\\_sales@samsung.com](mailto:scp_sales@samsung.com)  
 [youtube.com/samsungsds](https://youtube.com/samsungsds)

