

Data Analytics

Data Ops



Builds Workflow and Automate Job Execution for Data Processing Operations

Data Ops is an Apache Airflow-based workflow orchestration service that creates data processing workflow and automates job scheduling. It can be used independently in a Kubernetes Engine cluster environment of Samsung Cloud Platform or along with other application SW.

Easy Installation

A web-based console allows for easy installation of Data Ops in a standard Kubernetes cluster environment and Apache Airflow and management modules are installed automatically. In addition, the Data Ops dashboard provides comprehensive monitoring of a web server and a scheduler.

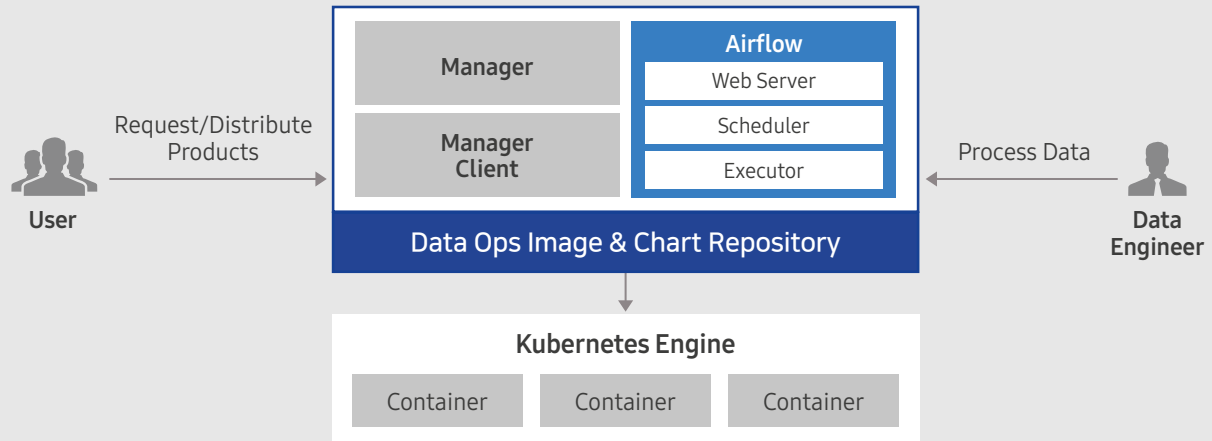
Configuration of Dynamic Pipeline

A pipeline is configured for data processing using Python code. Since tasks are created dynamically by working with a date-time format for scheduling data operations, users can configure the desired workflow type and scheduling.

Easy Workflow Management

DAG (Direct Acyclic Graph) configuration is visualized and managed using a web-based UI, making it easier to understand parallel processes of data flow. Managing timeout of an individual task, number of retries, and priorities also become easy.

Service Architecture



Key Features




- **Easy installation**
 - Install open source Airflow in a container environment
- **GUI-based easy management**
 - Easily manage Airflow settings in a container environment
 - Distribute Airflow plug-in
 - Status monitoring of Airflow services
- **Write/Schedule workflow**
 - Easy scalability thanks to python-based workflow
 - Task performed automatically by a scheduler
 - Manage resource by DAG task
 - Reprocess data processing error/failure
- **Airflow components**
 - Web server : Support visualization of DAG components and status, manage Airflow configurations
 - Scheduler : Orchestrate various DAG and tasks, support DAG reservation/execution
 - Executor : Provide KubernetesExecutor, a Kubernetes-based dynamic executor
 - Metadata DB : Store metadata on DAG, execution and user, role and connection, and other Airflow components

Pricing

- **Billing**
 - The amount of CPU time used by the Kubernetes Engine pod of Data Ops

FOR MORE INFORMATION

SAMSUNG SDS

 www.samsungsds.com / cloud.samsungsds.com
 contact.sds@samsung.com / scp_sales@samsung.com
 youtube.com/samsungsds

