Data Analytics

Data Flow

Extracts, Transforms and Moves Data from Various Sources and Automates Data Processing Flow

Data Flow is data processing flow tool that extracts large data from various data sources. It enables the visual establishment of how stream/batch data are converted/transmitted and also provides open source Apache NiFi services. It can be used independently in a Kubernetes Engine cluster environment of Samsung Cloud Platform or along with other application SW.

Easy Installation

The installation of Data Flow in a standard Kubernetes cluster environment is easy using a web-based console. By auto-configuring the architecture needed for open source Apache NiFi-based scalable clustering, ZooKeeper, registry and management modules are installed automatically. Using Data Flow, configuration files and NiFi template needed for service connection can be configured as well as deployed.
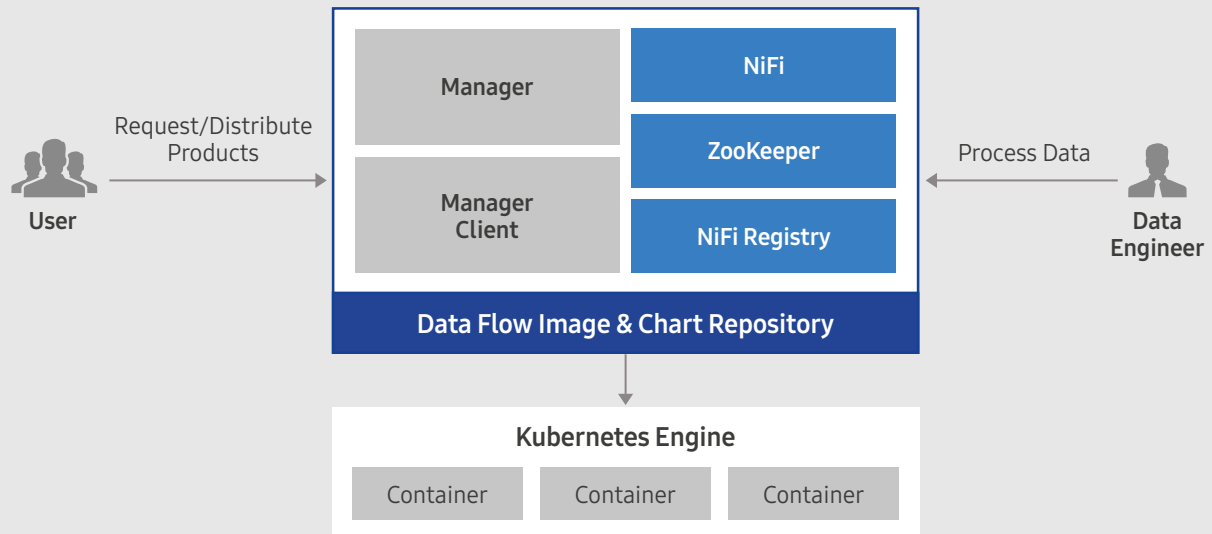
Easier Data Flow Management

You can easily create GUI-based data stream/batch processing and processing flow according to user environment. Creating the flowchart of stream/batch data between systems is easy as it is GUI-based. This also makes the extraction, moving, and processing of data more efficient.

NiFi Template Gallery

Using Data Flow, reference NiFi templates are shared and deployed. Data Flow provides a gallery of frequently-used NiFi templates and users can share their data processing jobs.

Service Architecture

```
                  Request/Distribute          ┌─────────────────────────────────────┐
                       Products                │  Manager         NiFi               │        Process Data
   [User]  ─────────────────────────────►      │                  ZooKeeper          │   ◄───────────────────  [Data
                                                │  Manager                            │                          Engineer]
                                                │  Client          NiFi Registry      │
                                                ├─────────────────────────────────────┤
                                                │   Data Flow Image & Chart Repository │
                                                └─────────────────────┬───────────────┘
                                                                      │
                                                ┌─────────────────────▼───────────────┐
                                                │         Kubernetes Engine           │
                                                │   Container  Container  Container    │
                                                └─────────────────────────────────────┘
```

Key Features

- **Easy installation**
  - Automatically install container-based open source NiFi scalable cluster environment using a web console
  - Pre-installed components : NiFi, ZooKeeper, NiFi Registry

- **Data processing**
  - Data sources available for connection : HDFS, HIVE, HBASE, Solr, Cassandra, MongoDB, Elasticsearch, Kafka, RabbitMQ, Syslog, HTTPS, SFTP, etc
  - Connect with data at a real-time, track data flow, and prevent data loss
  - Processing large amounts of data and parallel tasks

- **GUI-based easy management**
  - GUI-based deployment/management of NiFi settings and Nar files in a cluster environment
  - Share/deploy NiFi templates

Pricing

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