Union Platform Architecture

https://docs.union.ai/byoc/union-overview
https://docs.union.ai/byoc/platform-architecture

Service Platform

- The typical deployment has the customer use a multi-tenant control plane hosted by Union, where all the administration and management functionality is provided

- The customer's actual data and computation takes place in the execution plane that is hosted inside the customer's perimeter but where the customer grants the required permissions for the administration and management of the data plane.

- The data plane can be implemented in AWS / GCP / Azure and various customer owned VPC/perimeter(*)
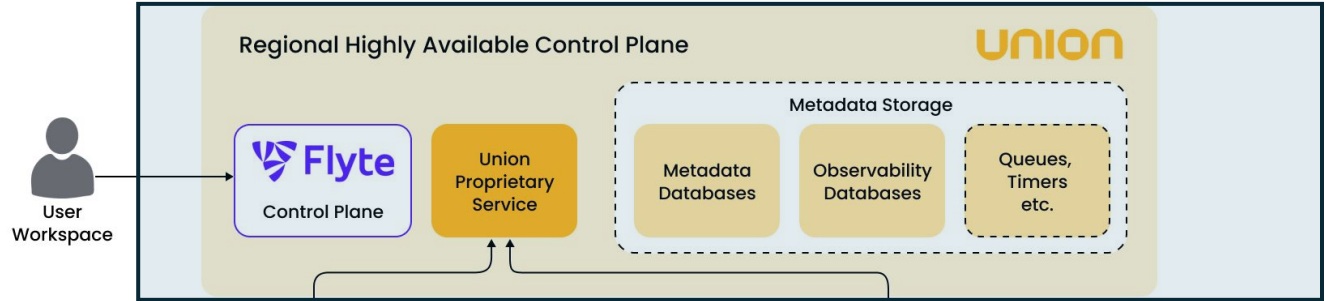   - * has specific requirements
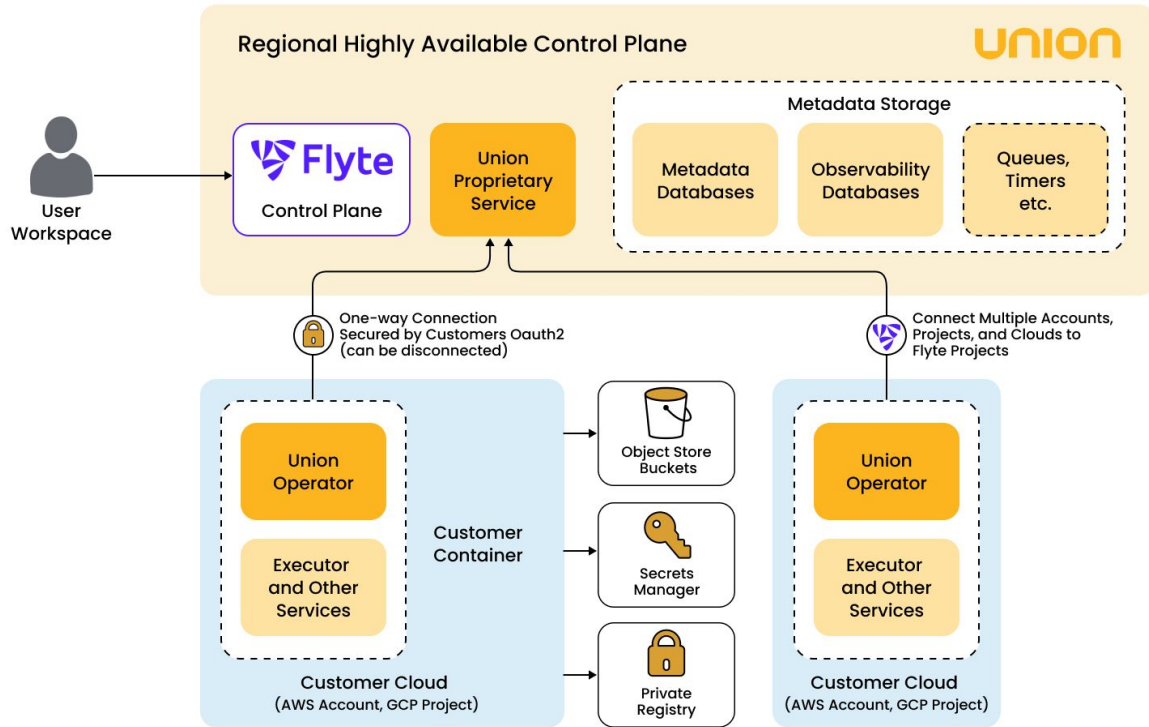
Architecture

- The platform is intentionally split into two planes, each which can be placed in its own VPC:
  - Control Plane:
    - User interface – Authentication, Authorization, Observation and Management
    - Control – placement of execution workloads and performs cluster control and management
  - Data Plane:
    - Includes a resident Union Operator which executes all tasks and workflows in clusters and spawns to node types:
      - Default Nodes – provide operator, autoscaling, monitoring, and other infrastructure maintenance services
      - Worker Nodes – execute all workloads under full user control, and scale down to the configured minimum when not in use (usually zero)
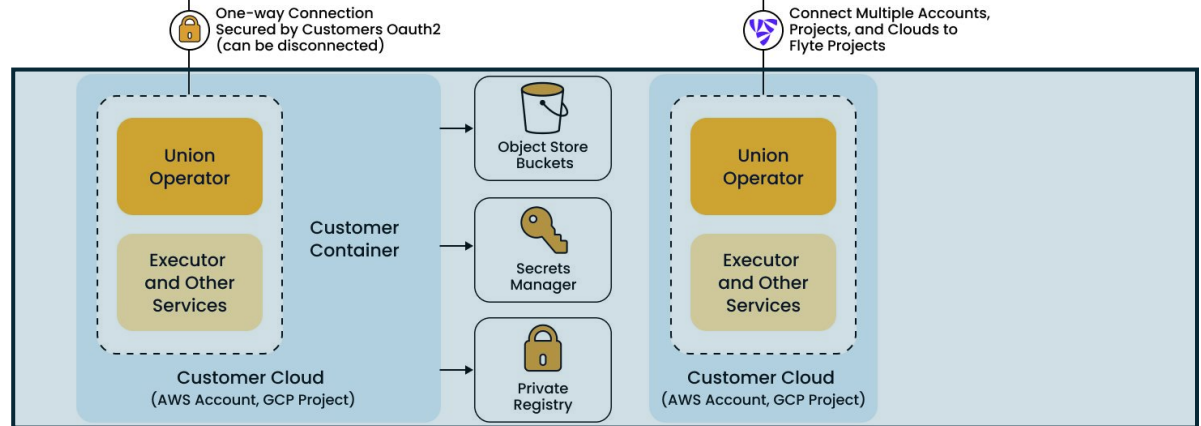
Control Plane

Multi Tenant SaaS

Data Plane

Customer's Private Cloud

Regional Highly Available Control Plane

UNION

User Workspace

Flyte Control Plane

Union Proprietary Service

Metadata Storage

Metadata Databases

Observability Databases

Queues, Timers etc.

One-way Connection Secured by Customers Oauth2 (can be disconnected)

Connect Multiple Accounts, Projects, and Clouds to Flyte Projects

Customer Container

Union Operator

Executor and Other Services

Object Store Buckets

Secrets Manager

Private Registry

Customer Cloud (AWS Account, GCP Project)

Union Operator

Executor and Other Services

Customer Cloud (AWS Account, GCP Project)

Tenancy & Security

- Multi Tenant SaaS Service:
    - Provides a control plane in US and EU regions (east/west) in each. The standard control plane is multi tenanted, managing customer data planes which are single tenanted and resident in the customer's own cloud/VPC/Physical
    - Is SOC-2 type 2 certified
    - The Operator service performs its functions with only the minimum set of required permissions:
        - DOES NOT – provide direct access to data or secrets
        - ONLY allows the control plane to spin up and down clusters and enables support engineers to access system level logs and apply changes requested by customers

- Private Single Tenant version: provides a customized control plane implemented in customer's own infrastructure

Data types and privacy

- Union uses two types of data to function:
  - In the control plane it stores **Registry** data used to manage workflow execution eg.:
    - Names of the workflows, tasks, launch plans and artifacts
    - Workflow and task Input/Output types
    - Execution status, start, end and duration times for workflows and tasks
    - Version information
    - Artifact definitions/descriptions as entered by customer users
  - **DOES NOT** include workflow or task code or data processed by these

Data types and privacy

- In an object store in the customer's data plane storing Execution Data that is comprised of:
  - Event Data
  - Workflow inputs and outputs
  - Task input/output which is further split in two categories:
    - **RAW** data only ever read and possibly temporarily cached in the control plane if needed
      - Files and Directories
      - Dataframes
      - Models
      - Python-pickled types
    - **Literal** data only ever passed by value and NOT reference, may be stored in the Union control plane
      - Primitive execution inputs (int, string… etc.)
      - JSON-serializable dataclasses
- Strict data privacy practices will always ensure no private information is passed in literal form between tasks