

salesforce

Airflow at Salesforce

Building a fully managed workflow orchestration system

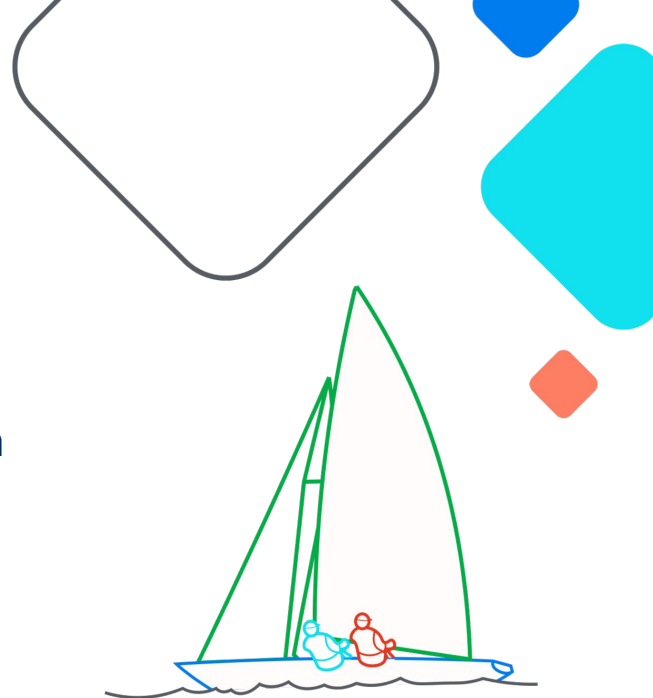
Varun Srinivas & Raj Ramalingam



Airflow at Salesforce

Building a “fully managed” workflow orchestration system

Raj Ramalingam, Varun Srinivas



 **Airflow Summit**
Let's flow together

September 19-21, 2023,
Toronto, Canada

Forward Looking Statements

This presentation contains forward-looking statements about, among other things, trend analyses and future events, future financial performance, anticipated growth, industry prospects, environmental, social and governance goals, and the anticipated benefits of acquired companies. The achievement or success of the matters covered by such forward-looking statements involves risks, uncertainties and assumptions. If any such risks or uncertainties materialize or if any of the assumptions prove incorrect, Salesforce's results could differ materially from the results expressed or implied by these forward-looking statements. The risks and uncertainties referred to above include those factors discussed in Salesforce's reports filed from time to time with the Securities and Exchange Commission, including, but not limited to: impact of, and actions we may take in response to, the COVID-19 pandemic, related public health measures and resulting economic downturn and market volatility; our ability to maintain security levels and service performance meeting the expectations of our customers, and the resources and costs required to avoid unanticipated downtime and prevent, detect and remediate performance degradation and security breaches; the expenses associated with our data centers and third-party infrastructure providers; our ability to secure additional data center capacity; our reliance on third-party hardware, software and platform providers; the effect of evolving domestic and foreign government regulations, including those related to the provision of services on the Internet, those related to accessing the Internet, and those addressing data privacy, cross-border data transfers and import and export controls; current and potential litigation involving us or our industry, including litigation involving acquired entities such as Tableau Software, Inc. and Slack Technologies, Inc., and the resolution or settlement thereof; regulatory developments and regulatory investigations involving us or affecting our industry; our ability to successfully introduce new services and product features, including any efforts to expand our services; the success of our strategy of acquiring or making investments in complementary businesses, joint ventures, services, technologies and intellectual property rights; our ability to complete, on a timely basis or at all, announced transactions; our ability to realize the benefits from acquisitions, strategic partnerships, joint ventures and investments, including our July 2021 acquisition of Slack Technologies, Inc., and successfully integrate acquired businesses and technologies; our ability to compete in the markets in which we participate; the success of our business strategy and our plan to build our business, including our strategy to be a leading provider of enterprise cloud computing applications and platforms; our ability to execute our business plans; our ability to continue to grow unearned revenue and remaining performance obligation; the pace of change and innovation in enterprise cloud computing services; the seasonal nature of our sales cycles; our ability to limit customer attrition and costs related to those efforts; the success of our international expansion strategy; the demands on our personnel and infrastructure resulting from significant growth in our customer base and operations, including as a result of acquisitions; our ability to preserve our workplace culture, including as a result of our decisions regarding our current and future office environments or work-from-home policies; our dependency on the development and maintenance of the infrastructure of the Internet; our real estate and office facilities strategy and related costs and uncertainties; fluctuations in, and our ability to predict, our operating results and cash flows; the variability in our results arising from the accounting for term license revenue products; the performance and fair value of our investments in complementary businesses through our strategic investment portfolio; the impact of future gains or losses from our strategic investment portfolio, including gains or losses from overall market conditions that may affect the publicly traded companies within our strategic investment portfolio; our ability to protect our intellectual property rights; our ability to develop our brands; the impact of foreign currency exchange rate and interest rate fluctuations on our results; the valuation of our deferred tax assets and the release of related valuation allowances; the potential availability of additional tax assets in the future; the impact of new accounting pronouncements and tax laws; uncertainties affecting our ability to estimate our tax rate; uncertainties regarding our tax obligations in connection with potential jurisdictional transfers of intellectual property, including the tax rate, the timing of the transfer and the value of such transferred intellectual property; uncertainties regarding the effect of general economic and market conditions; the impact of geopolitical events; uncertainties regarding the impact of expensing stock options and other equity awards; the sufficiency of our capital resources; the ability to execute our Share Repurchase Program; our ability to comply with our debt covenants and lease obligations; the impact of climate change, natural disasters and actual or threatened public health emergencies; and our ability to achieve our aspirations, goals and projections related to our environmental, social and governance initiatives.

Outline



Understanding the Landscape

Key Architectural Components

Putting it all together
Operationalizing our Architecture

Conclusion & Takeaways



Understanding the Landscape

A stylized, flat-design illustration of a landscape. In the foreground, there are rolling green hills with a winding orange path. The hills are dotted with small yellow flowers and have several rounded, bushy trees with brown trunks and green foliage. In the middle ground, a dense line of dark green coniferous trees stretches across the horizon. In the background, a large, white, snow-capped mountain range is visible under a light blue sky with a few white, fluffy clouds.

Why Salesforce Managed Airflow?

Trust is our #1 priority



Infrastructure As Code



Immutable Deployments



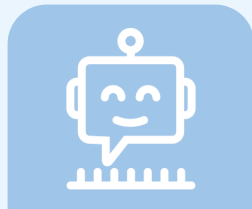
Compliance



Service Ownership



Cost to Serve



Elastic & Event-Driven Ops



Zero Trust Architecture



Service Resilience





Petabyte Scale Data Lakes

Orchestration is Imperative

*One-stop turnkey destination for all of Salesforce's
Internal orchestration needs*





Multi-tenant Data Platform - Simplified View




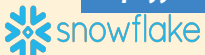

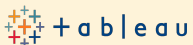
 Data Engineers  Data Scientists  ML Ops  Engineers

Superset UI/API Tableau Jupyter


   

  Airflow Kubernetes Runtime

mTLS Kerberos

  snowflake  trino  + a b l e a u

Data Processing

 einstein

ML

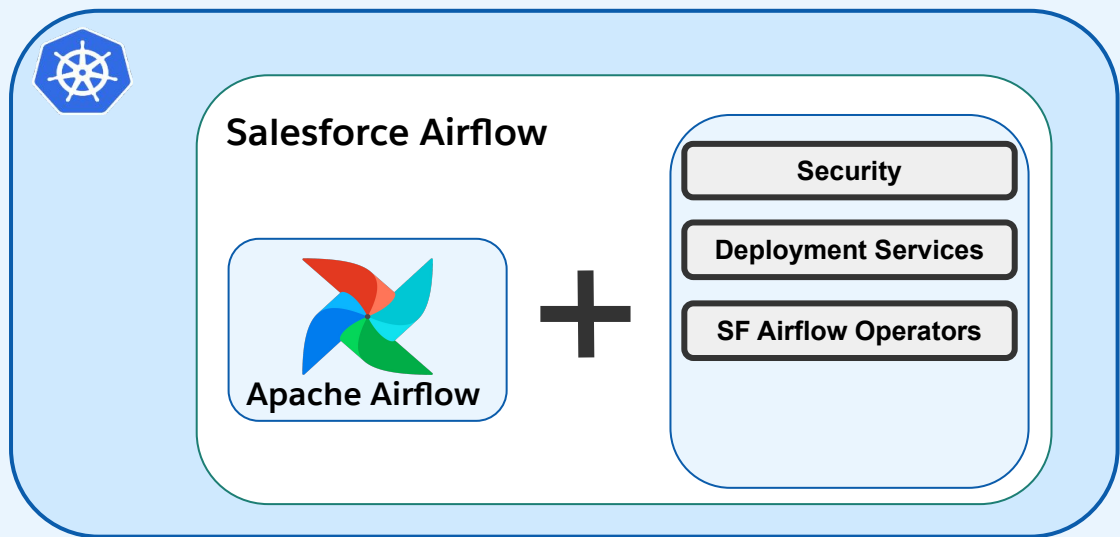
Use Cases / Tools

- ETL/ELT
- ML
- Data analytics

Data Processing Engines

- Multi-tenant
- Different AuthN/AuthZ

Salesforce Airflow



Hyperforce

Build & Delivery Monitoring & Metrics Network Services Security Services

Trust / Security

- SF Operated PKI
- Secrets Management
- Service Account / Team

Deployment Services

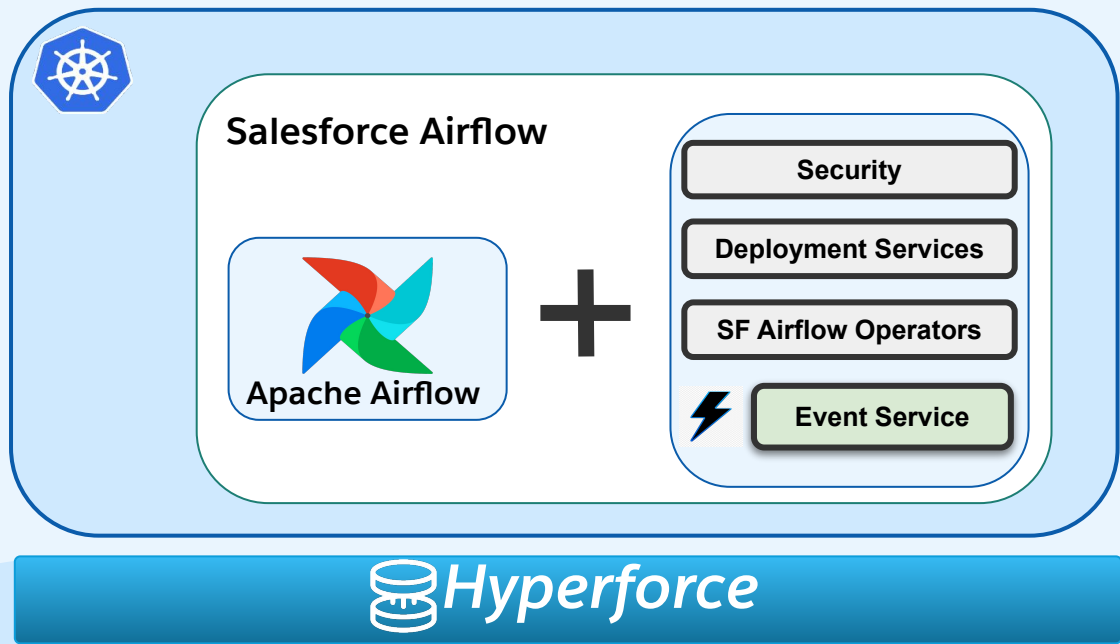
- DAGs, PySpark
- Dependency management

SF Airflow Operators

- Light Abstraction over core Operators
- Distribute uniformly (Sandbox/Prod)



Salesforce Airflow



Trust / Security

- SF Operated PKI
- Secrets Management
- Service Account / Team

Deployment Services

- DAGs, PySpark
- Dependency management

Event Service

- Add on Service - Async / Remote Execution

Salesforce Airflow - Goals

- Security/Trust
- Provide Productive Environment
- “Easy” to use - No configuration
- “Easy” to onboard



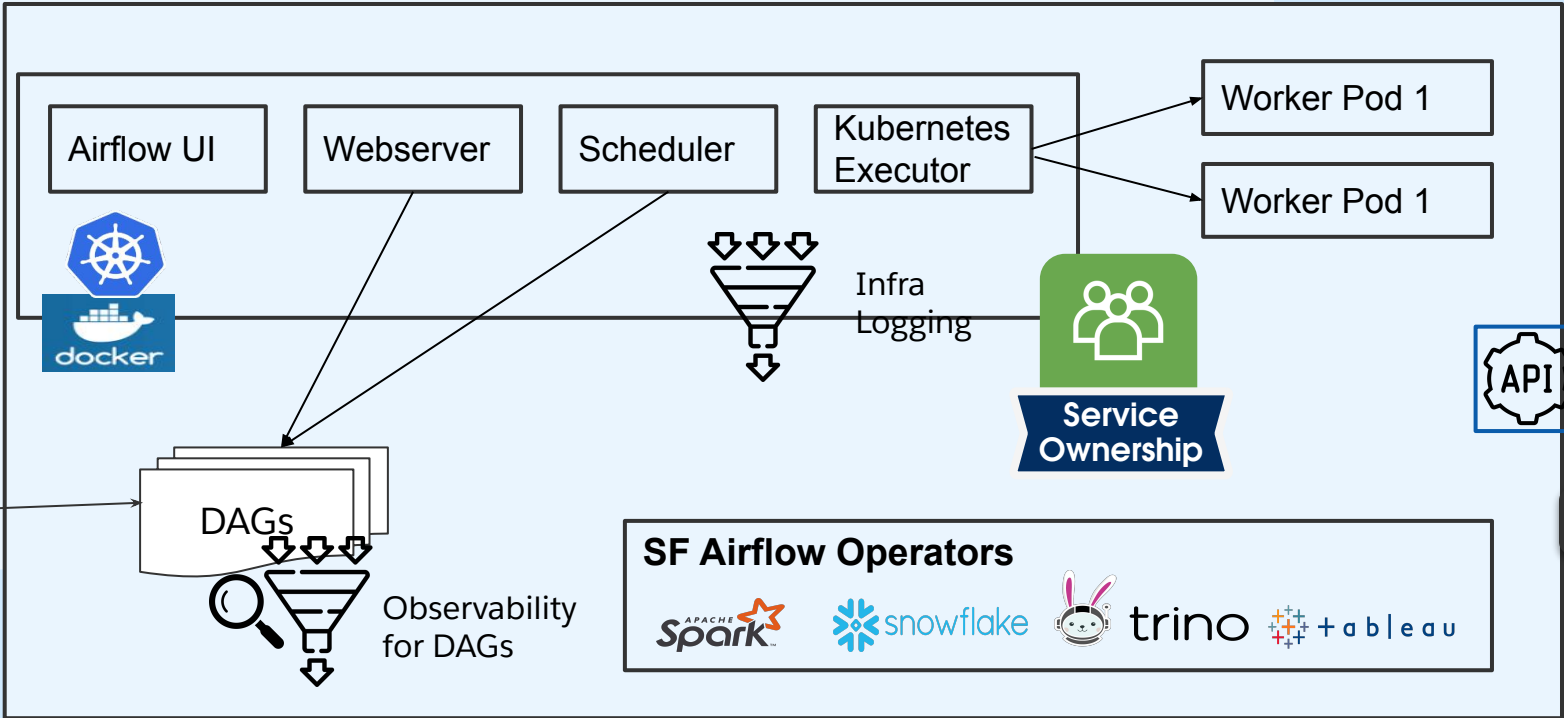
Salesforce Airflow - Key Decisions



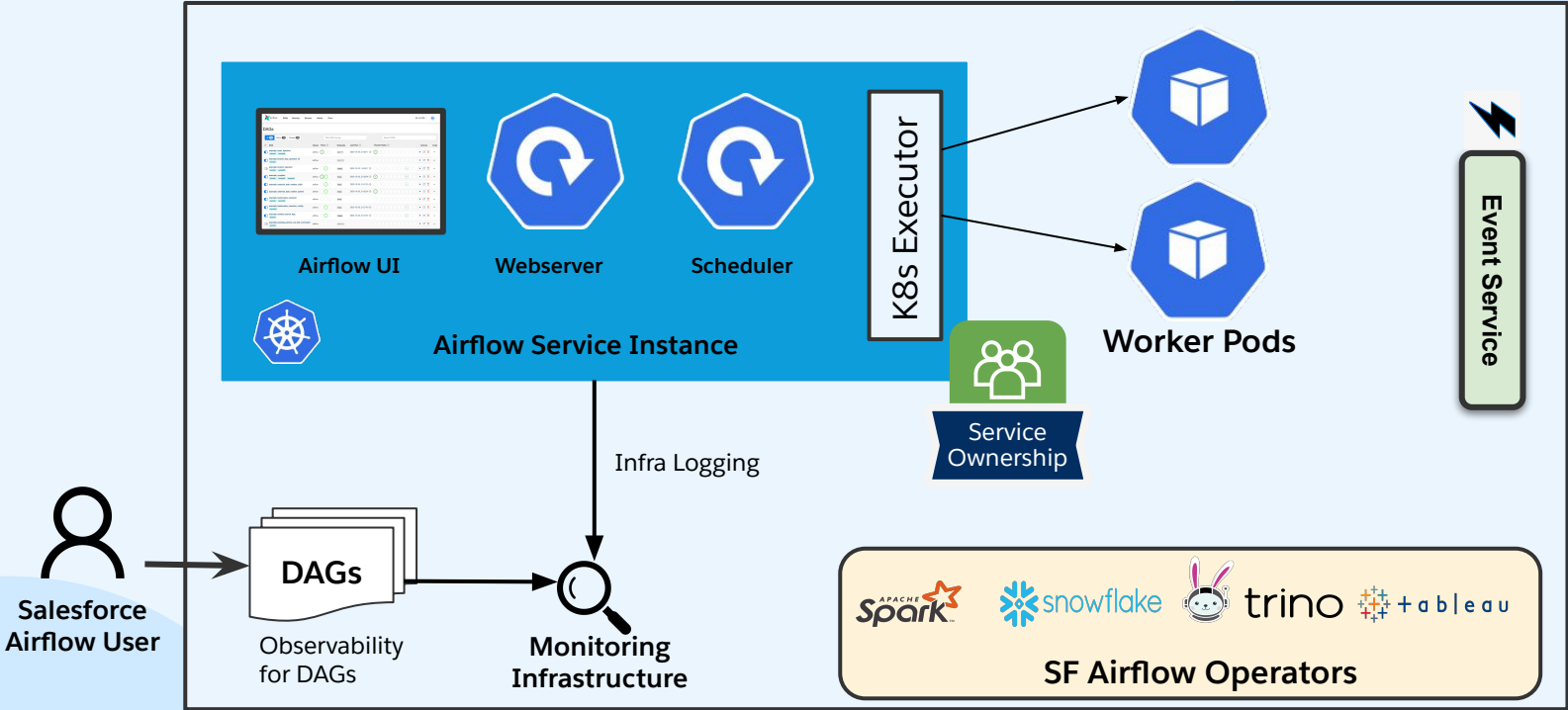
- **K8s Executors**
- **Infrastructure abstractions**
 - Standardized base images
 - Logging / Secrets
- **SLO (Service Level Ownership)**
 - SF Airflow Level
 - DAG Level
- **SF Airflow Operators**
 - Livy, Trino, Snowflake, K8 Pod Op, Event Service, Observability Plugin...
 - Easier Distribution through deployment services
- **Guard Rails - Deployment Services.**
 - DAGs/Manage Artifacts / Dependencies
- **Event Service**
 - To Avoid short/long polling for data, job completion
 - Remote Execution



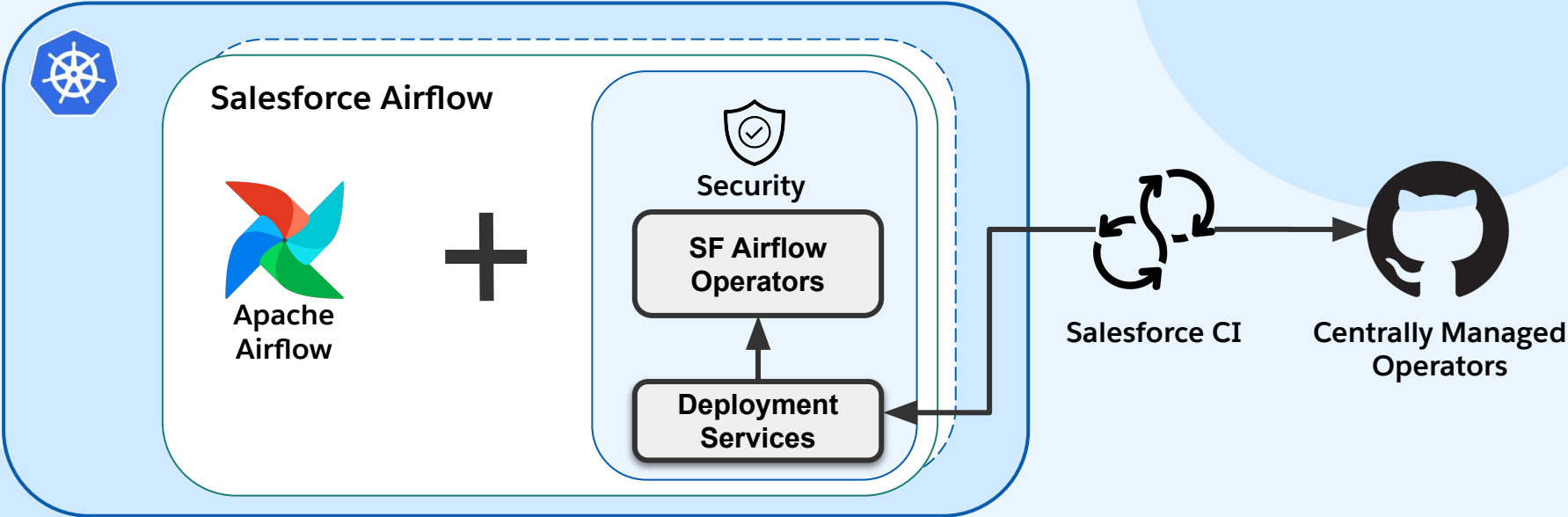
Salesforce Airflow - Decisions, Capabilities



Salesforce Airflow - Decisions, Capabilities



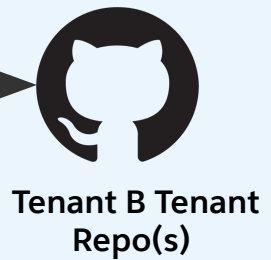
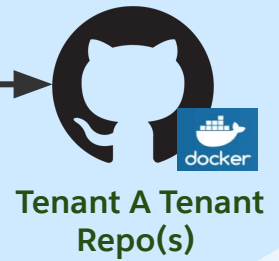
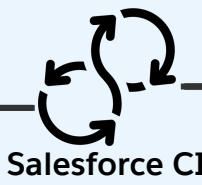
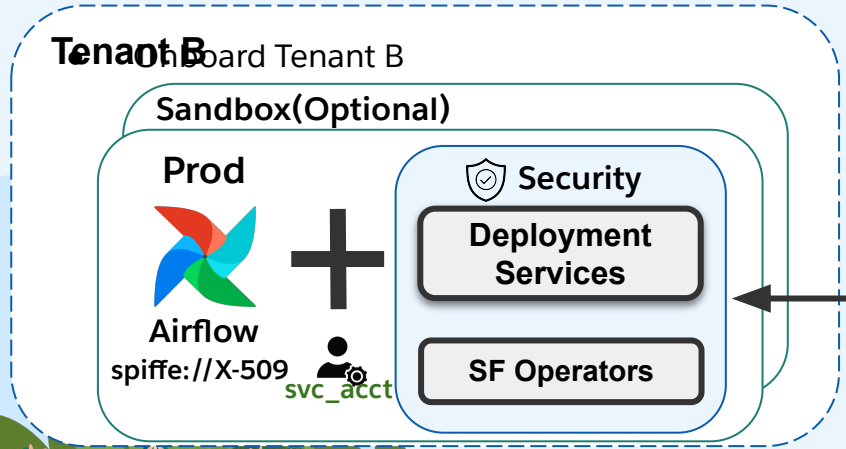
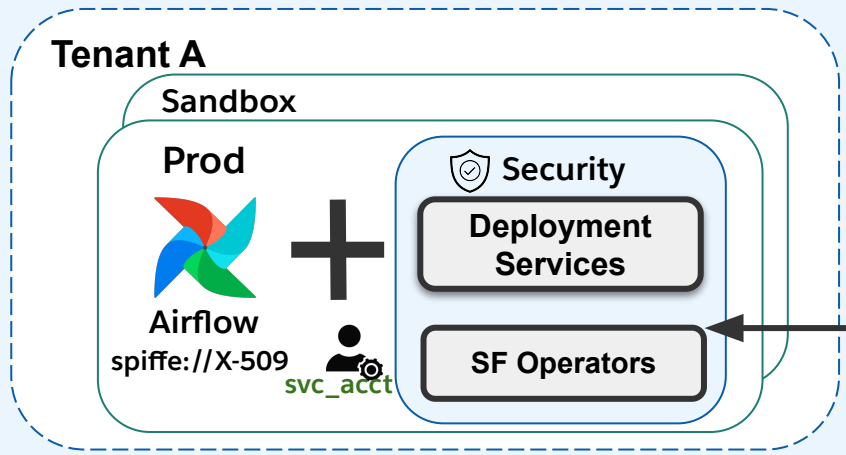
Salesforce Airflow - Simplified Blueprint Prod/Sandbox



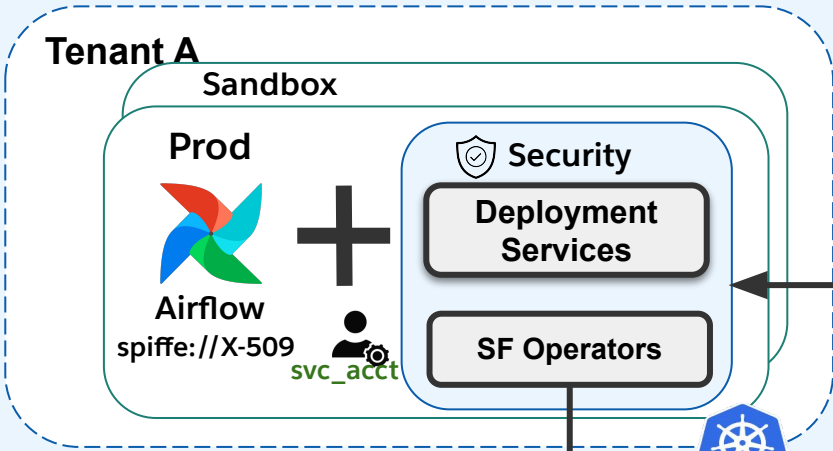
Salesforce Airflow - Tenant View



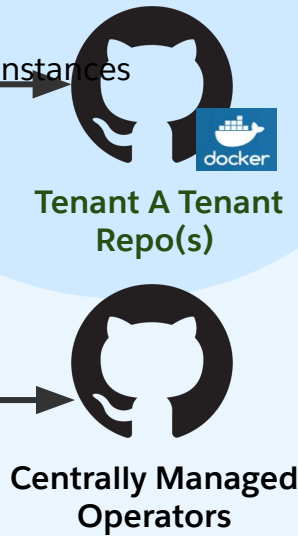
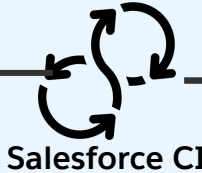
- Testing DAGs can be challenging
- High fidelity between Sandbox and Prod



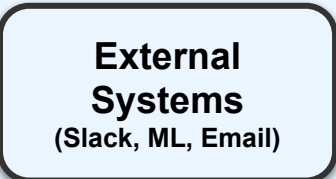
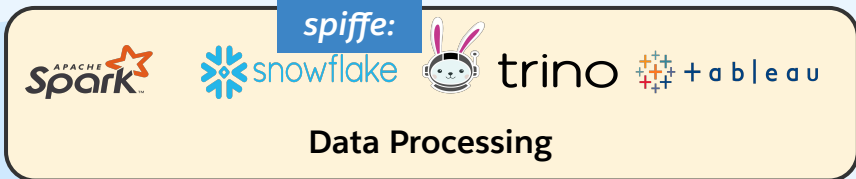
Salesforce Airflow - Tenant View



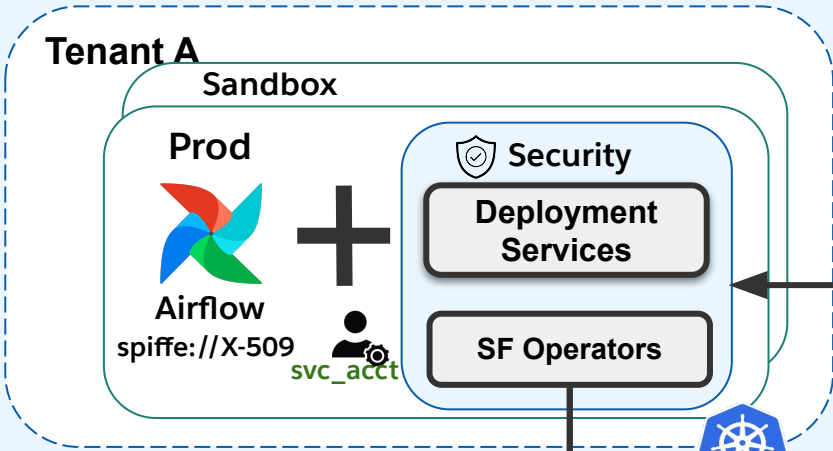
- Preconfigured Productive Environment.
- Testing DAGs in Prod can be challenging
- High fidelity between Sandbox and Prod instances



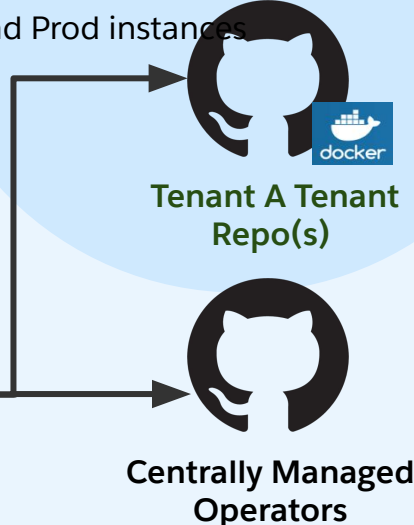
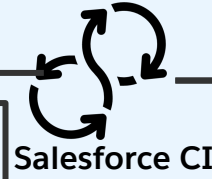
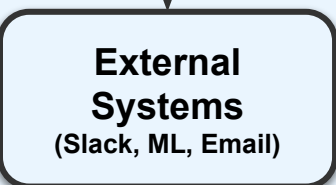
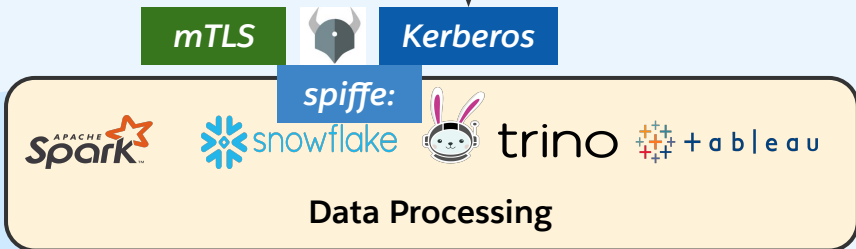
mTLS Kerberos



Salesforce Airflow - Tenant View



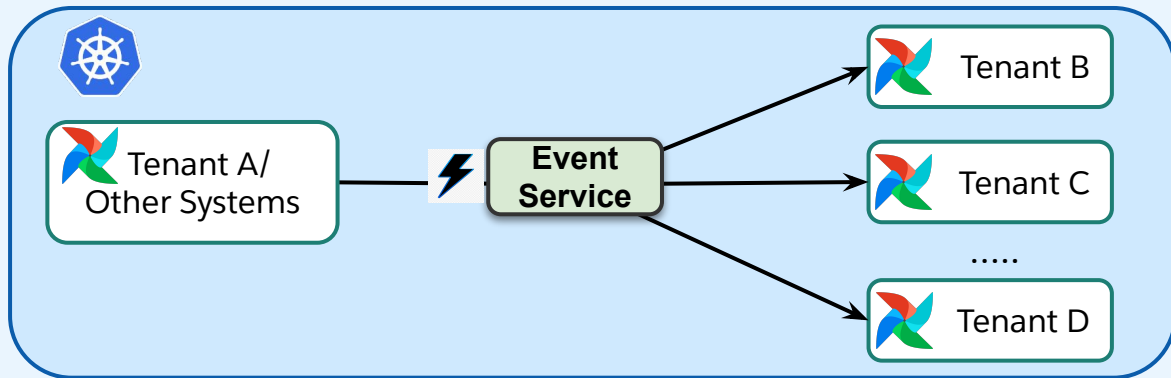
- Preconfigured Productive Prod Environment.
- Testing Complex DAGs in Prod can be challenging
- High fidelity between Sandbox and Prod instances
- DAG that depend on other DAGs



Event Service - Use Cases



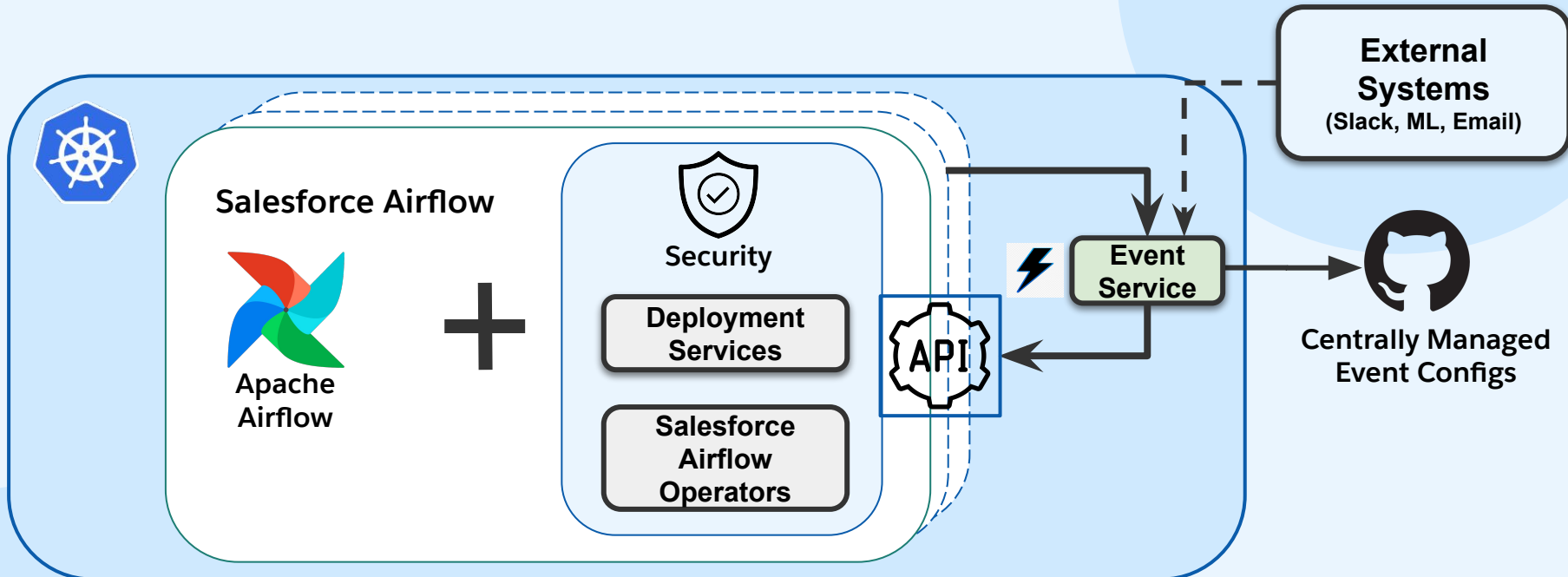
Event Based Dag Triggers across Airflow Tenant Instances.



Use cases

- Job Dependent
- Data Dependent
- Airflow Ops- Fleet wide

Airflow (*Services*) + Event (*Service*) View



Services Interaction at Run-time - Secure Access



Service authentication and authorization at every step

Zero Trust Architecture

There is no implicit trust between services

Salesforce-Operated PKI Service

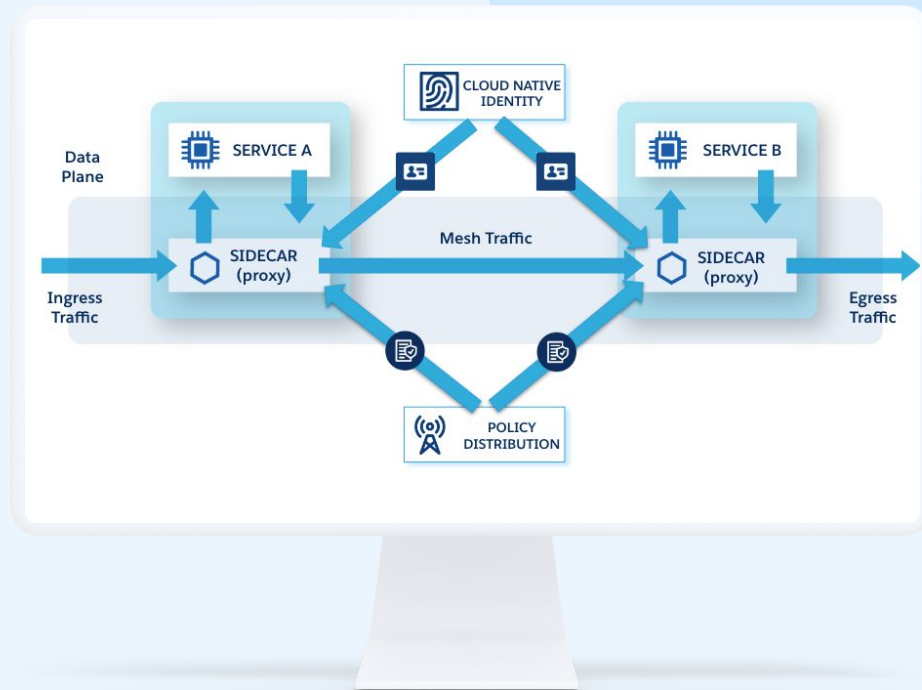
Authentication using digital certificates

Open Policy Agent

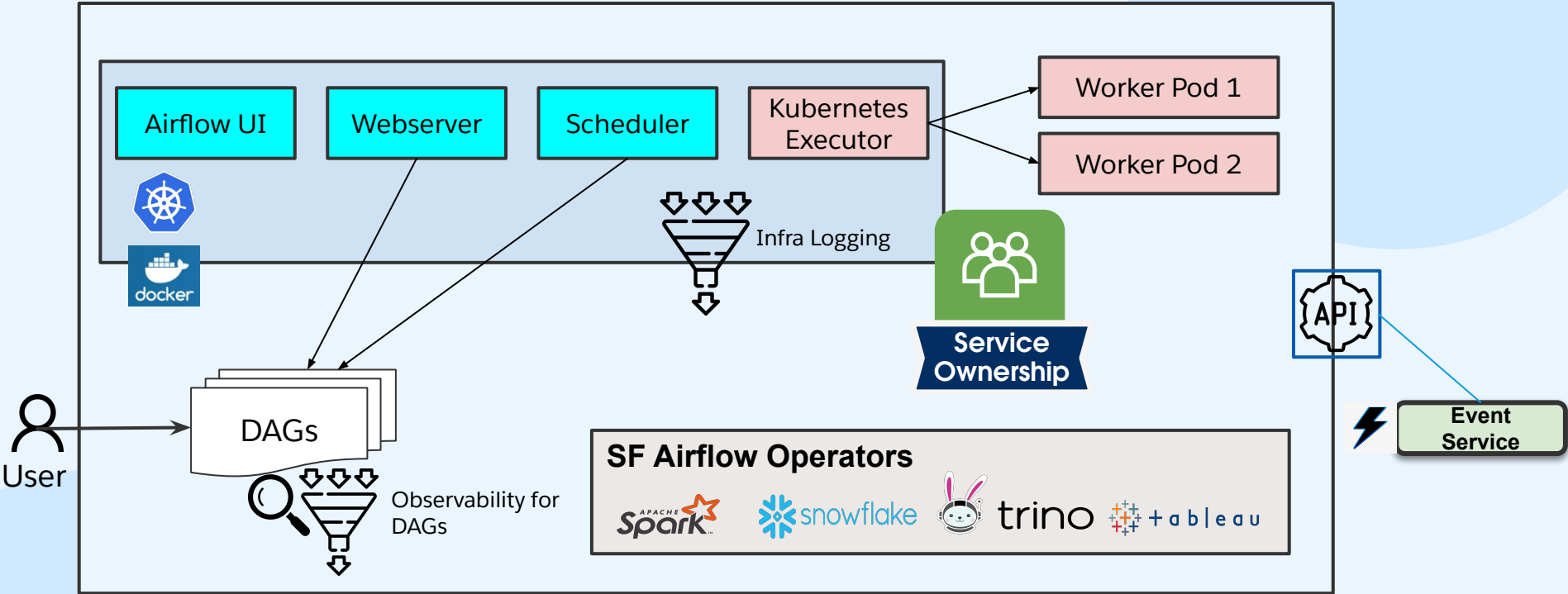
Dynamic and granular authorization

Least Privilege Access Policy

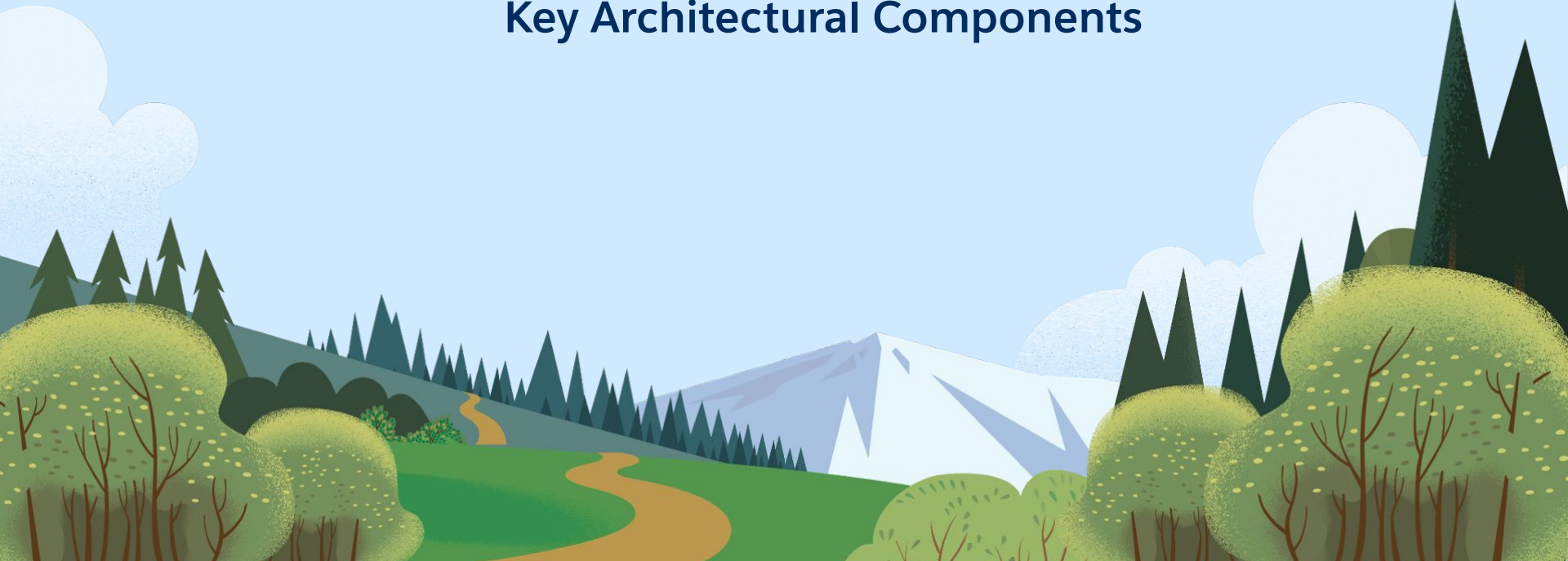
Requestors receive no more than they need



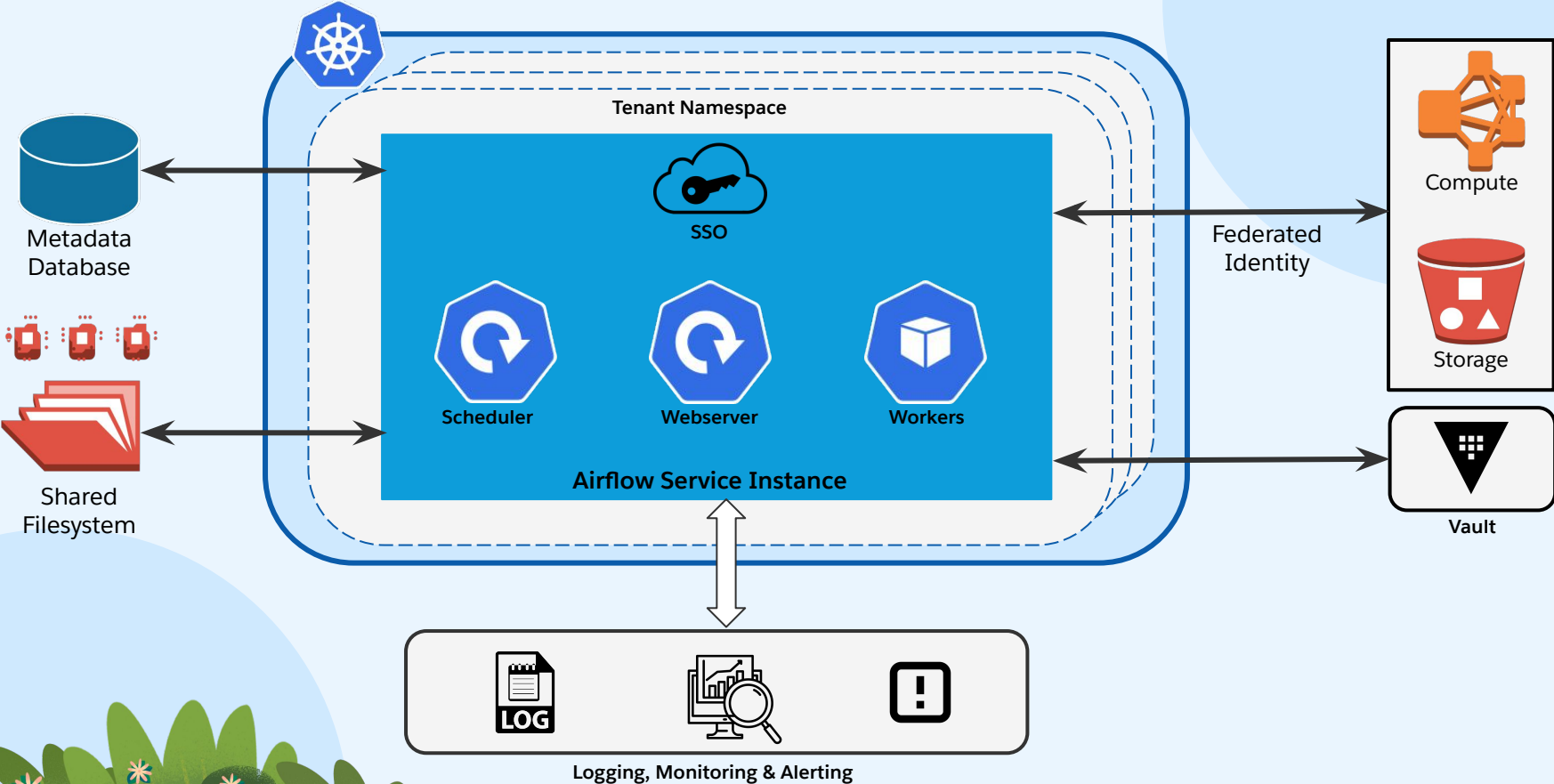
SF Airflow Summary - Decisions, Capabilities



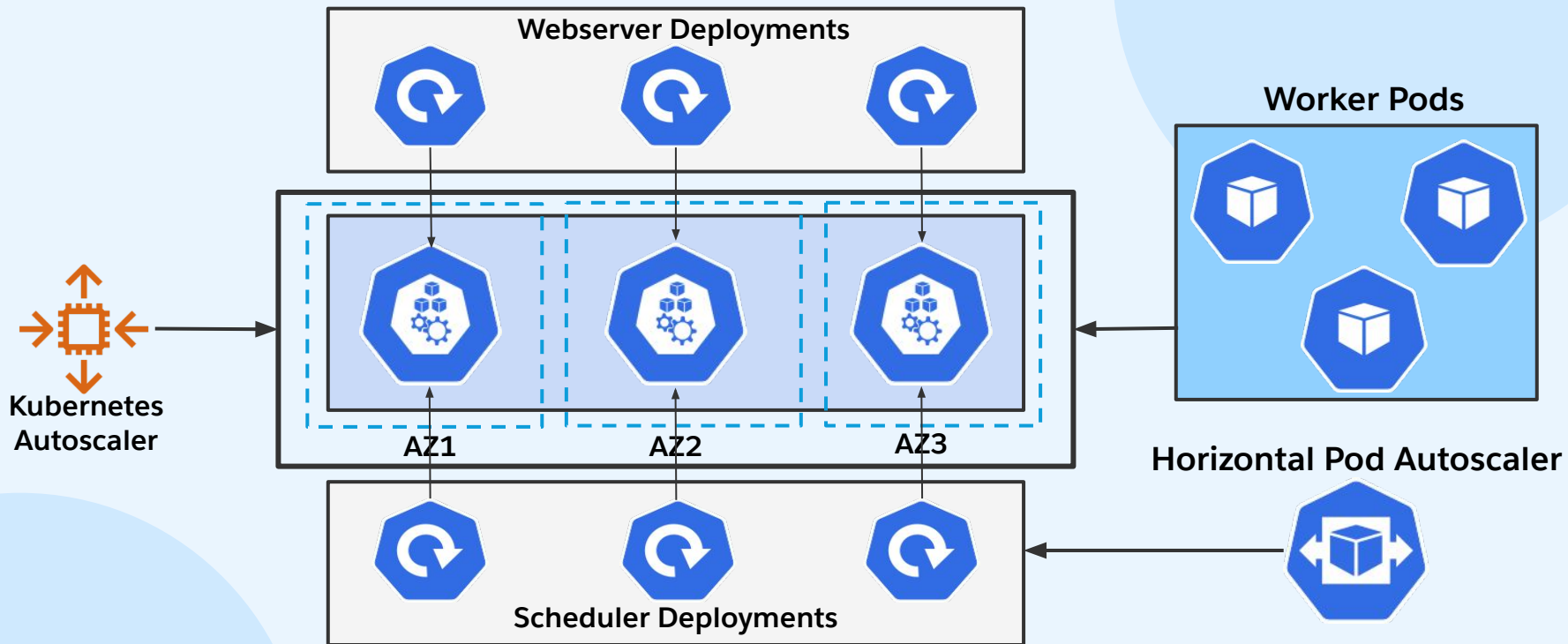
Key Architectural Components



High Level Architecture



Availability, Scaling & Resiliency



Logging, Monitoring and Alerting



Cloud Native
Monitoring
Infrastructure



Salesforce Airflow
Infrastructure



Log/Metric
Ingestion
REST



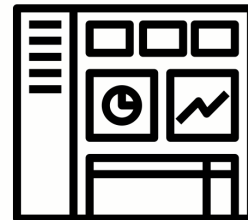
Funnel



Salesforce Monitoring Infrastructure



Kafka
Aggregations



Dashboarding
Visualization
Analytics



Alerting

Putting it all together: Operationalizing our Architecture

A stylized landscape illustration. In the foreground, there are rolling green hills with a winding orange path. The hills are dotted with small yellow flowers and have some brown, leafless tree branches. In the middle ground, there is a dense forest of dark green coniferous trees. In the background, there are white, snow-capped mountains under a light blue sky with a few white, fluffy clouds.

User Journey



Develop DAGs

Source DAGs

Source
Dependencies

Execute
Workflows



Dev Airflow Instance



DAGs & Configuration



Airflow Kubernetes Runtime



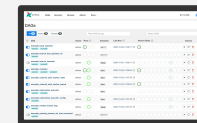
Airflow Kubernetes Runtime



Event Service



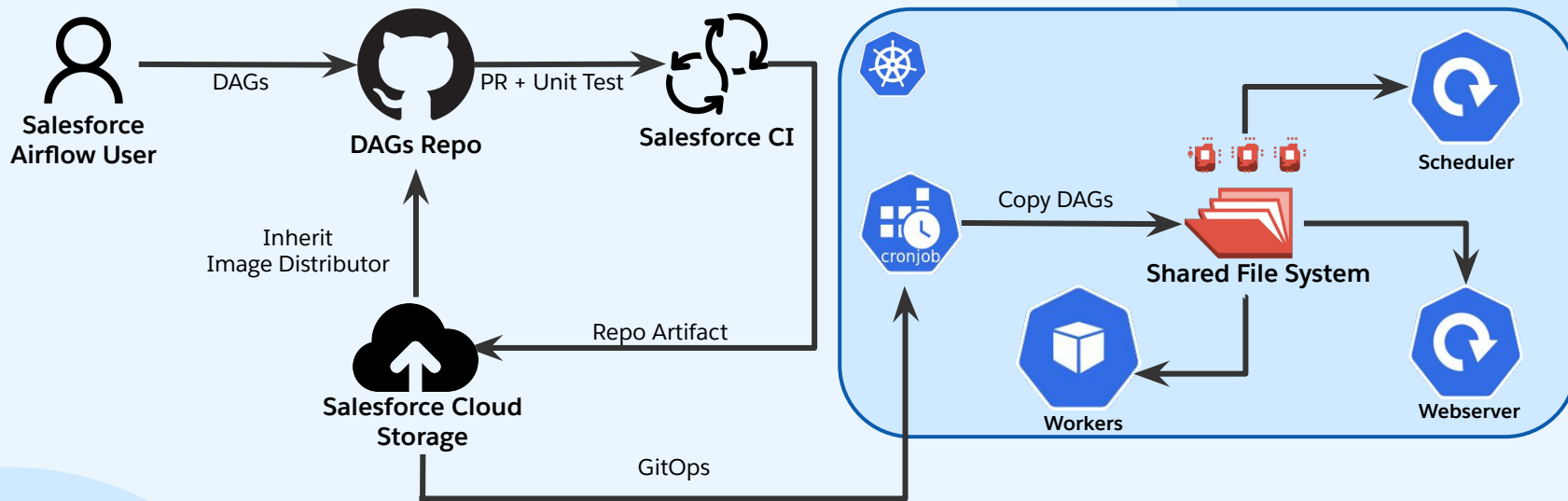
Cron



Ad Hoc

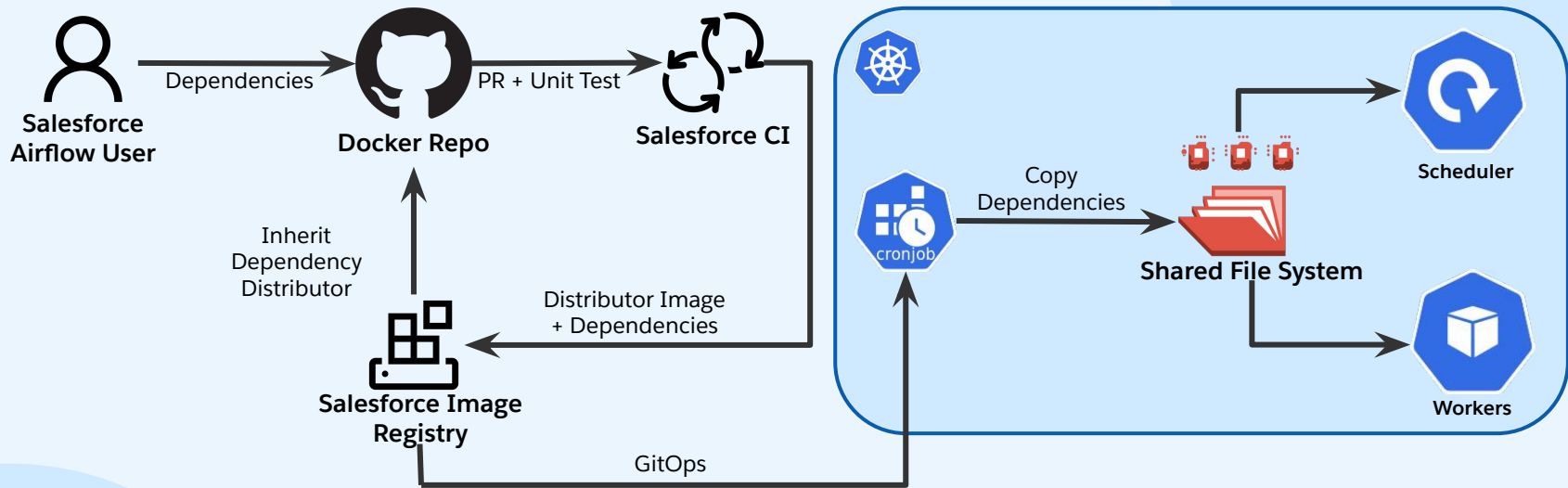
User Journey

Sourcing DAGs



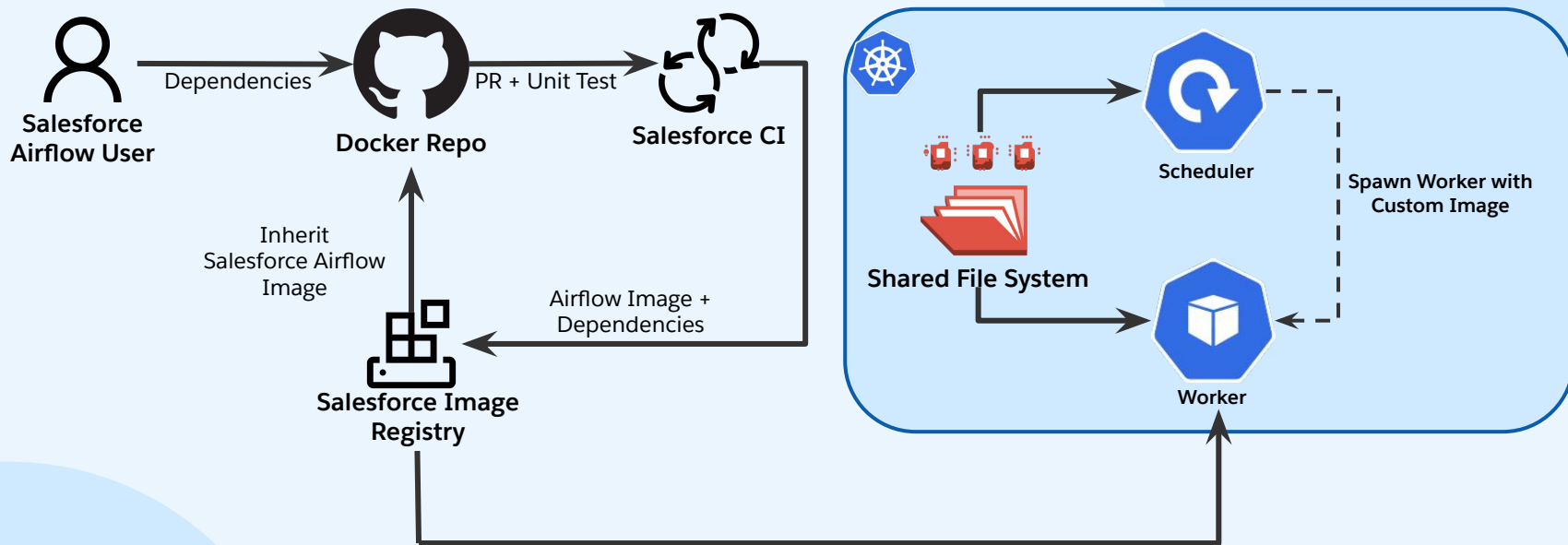
User Journey

Sourcing Dependencies for Kubernetes Executor



User Journey

Sourcing Dependencies for Kubernetes Pod Operator



Conclusion & Takeaways



Airflow “Easy” at Scale

Unburden Customer Operations, Focus on Business Excellence

Unified Developer Experience



Hyperforce

Securely Connected Platforms



TRUST

Monitoring, Alerting & Support



Salesforce Compliant



Salesforce Airflow

Flexible,
Scalable and
Automated
Orchestration





Thank you

