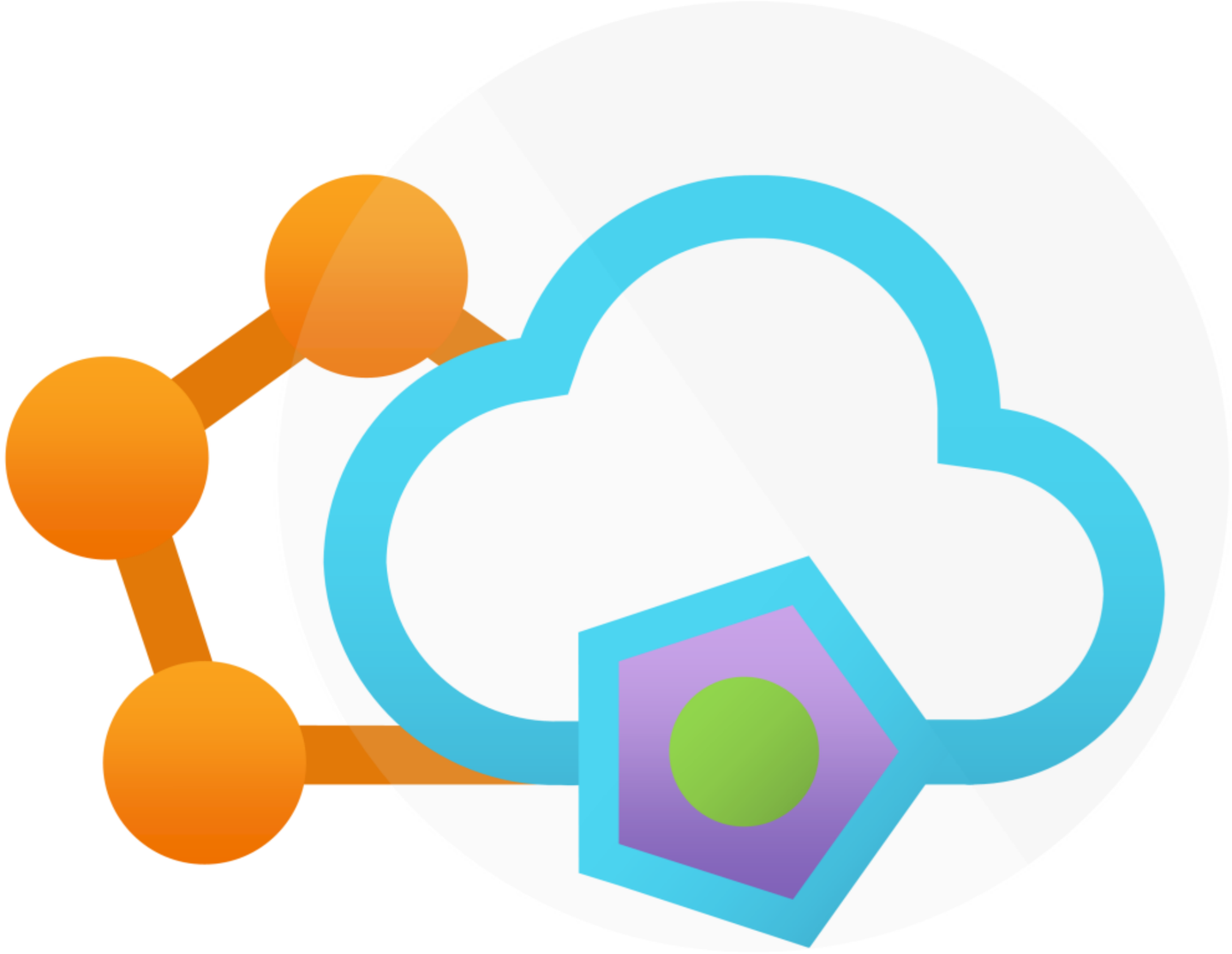
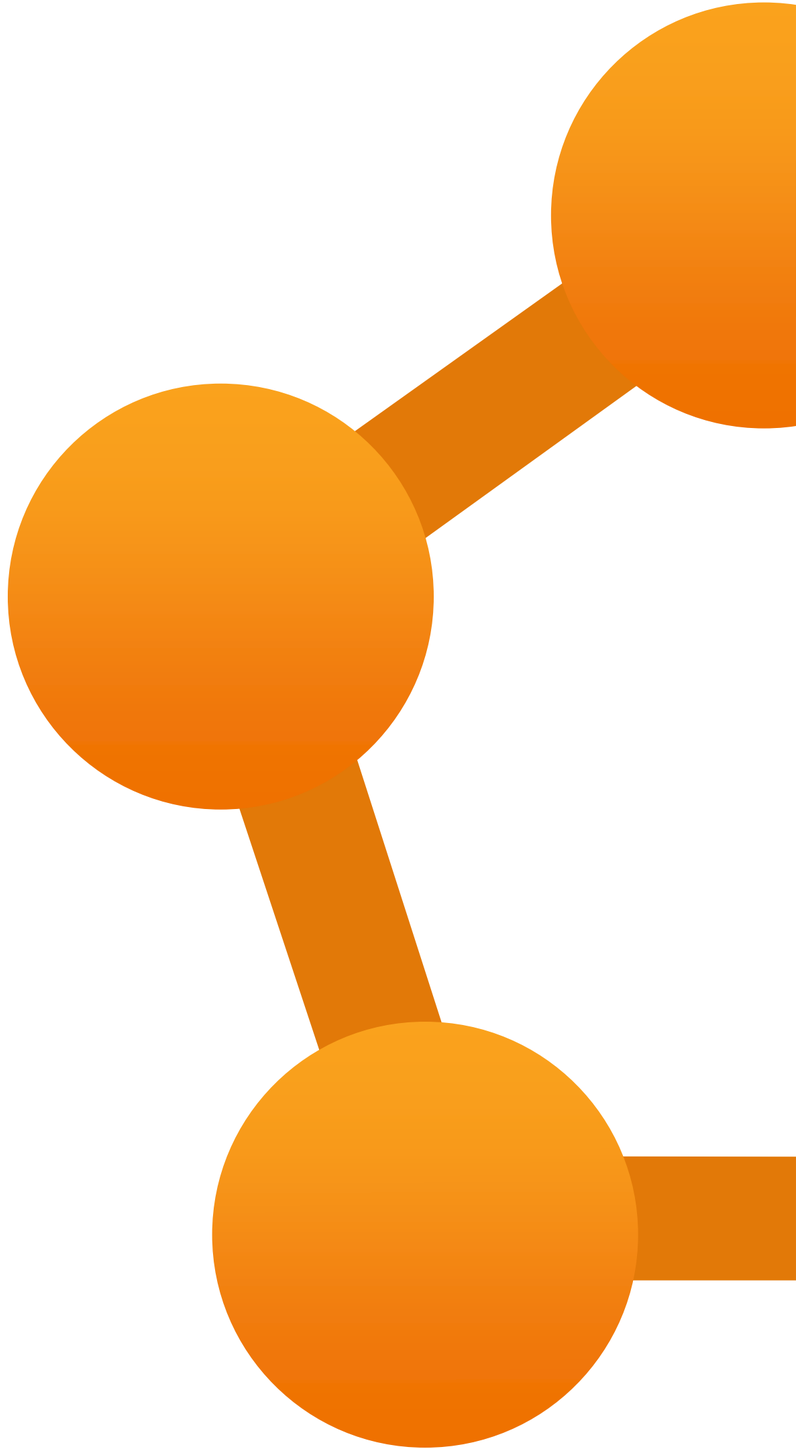


CF.Cumulus ™

An open-source cloud data platform product to accelerate time to insights.

Cloud Formations

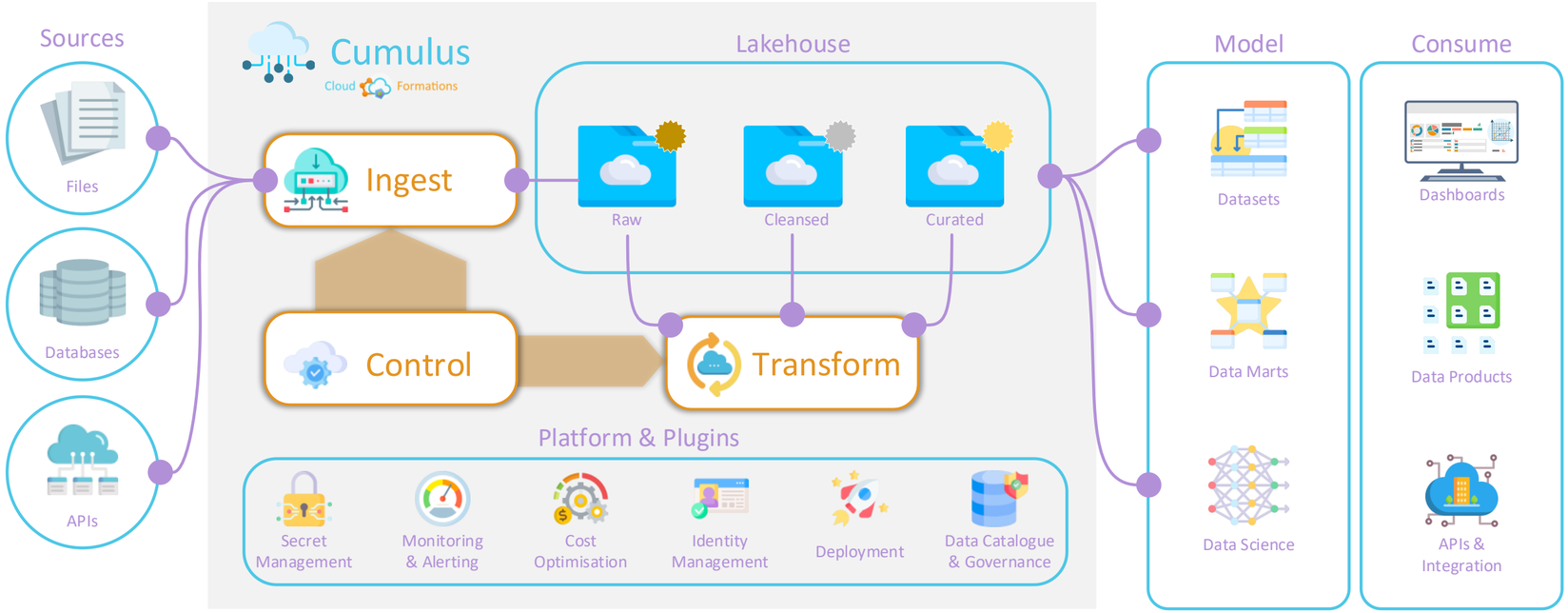
What is CF.Cumulus



Cloud Formations

CF.Cumulus Framework

- **Open Source*:**
  Zero cost, unlimited access to code
- **Scalable:**
  Horizontal & vertical scaling for Big Data executions
- **Cost-Efficient:**
  Faster time-to-insight reduces delivery costs
- **Extensible:**
  Supports complex & unique data handling
- **Collaborative:**
  Brings development & operations together
- **Rapid Development:**
  Accelerates data solution implementation
- **Proven Framework:**
  Adopted and used with 100+ repository forks
- **Transparent & Trustworthy:**
  Tried & tested patterns using mature technologies



Underlying Technology: Spark, DELTA LAKE™

Works With:

Control

- Metadata-driven workflow orchestration
- Managed dependency chains
- Concurrent batch processing
- Multi-threaded execution
- Resilient exception handling
- Enterprise grade reliability

Ingest

- Metadata-driven data ingestion
- Adaptable to all scenarios
- Extensible data sources & formats
- Flexible operating modes:
  - Deploy-time code gen
  - Run-time code gen†

Transform

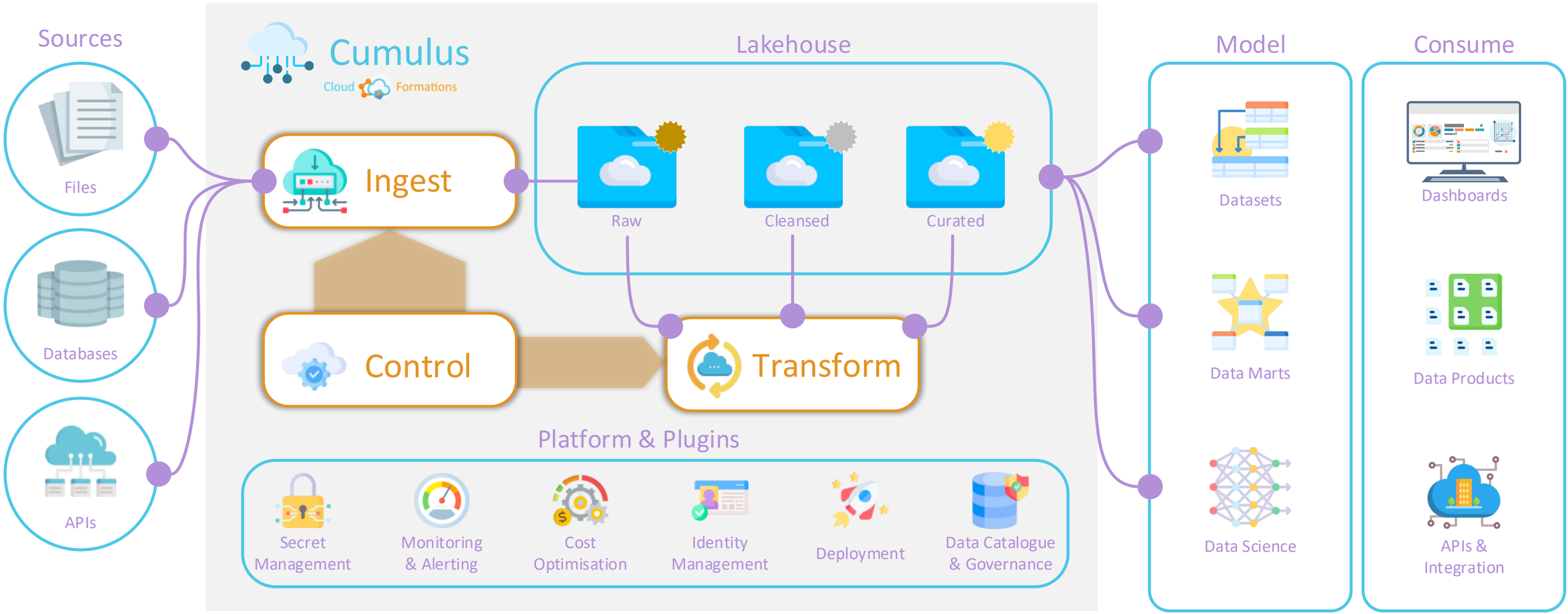
- Metadata-driven data transformation
- Rapid engineering of common transformations
- Scalable data processing
- Baked in support for dimensional models
- Supports AI & Machine Learning model inference
- Complete flexibility through custom notebooks

Cloud Formations - Logical Architecture

* In accordance with MIT licencing.
† Coming soon.

© 2024 Cloud Formations Ltd

CF.Cumulus Framework



Underlying Technology:

Works With:

Cloud Formations - Logical Architecture

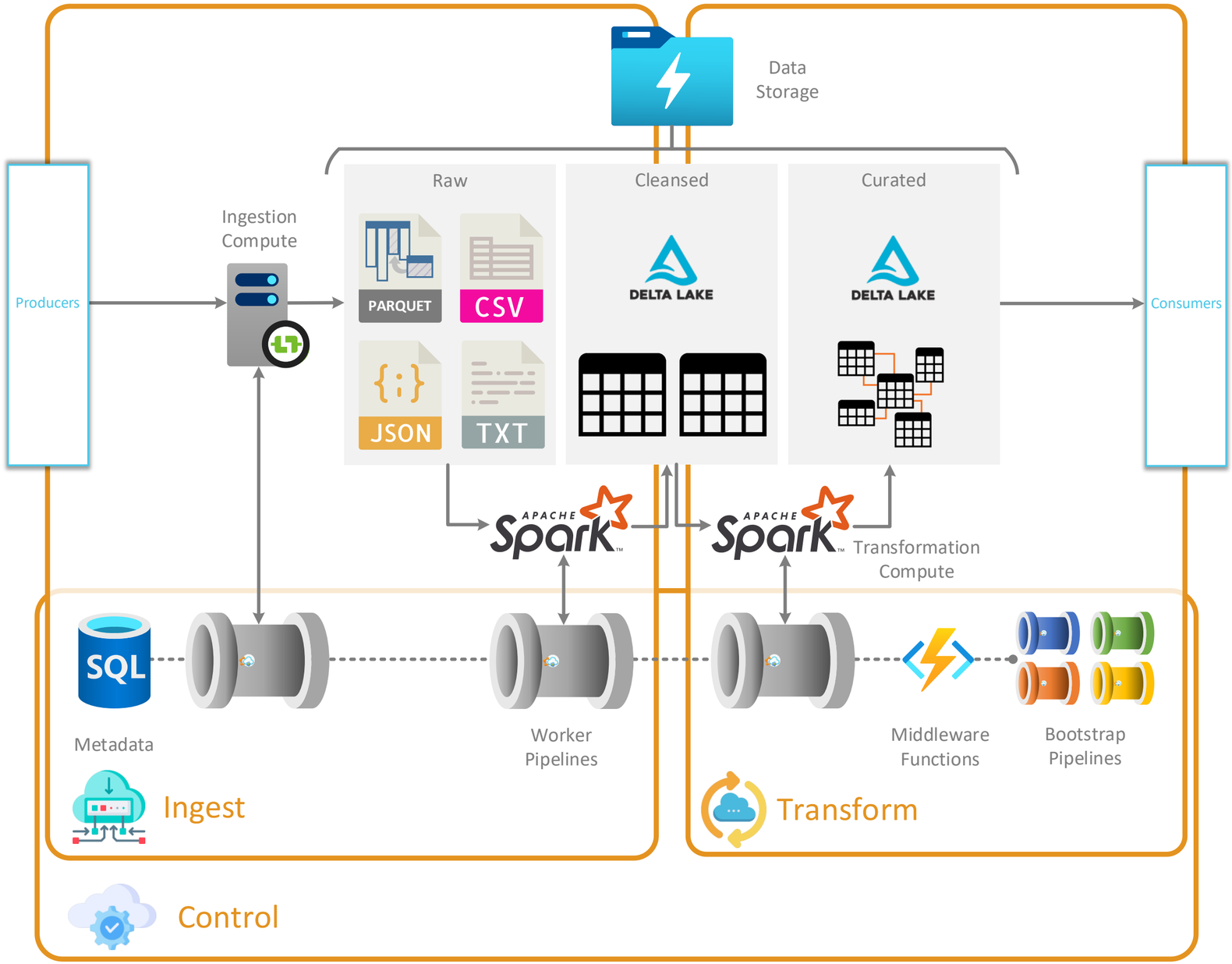
© 2024 Cloud Formations Ltd

CF.Cumulus Framework

An Azure product agnostic view of compute, storage and orchestration resources as used by CF.Cumulus.

A primary design focus for CF.Cumulus is to align with open-standards and open technologies. Offering a flexible, plug and play approach to the core components.

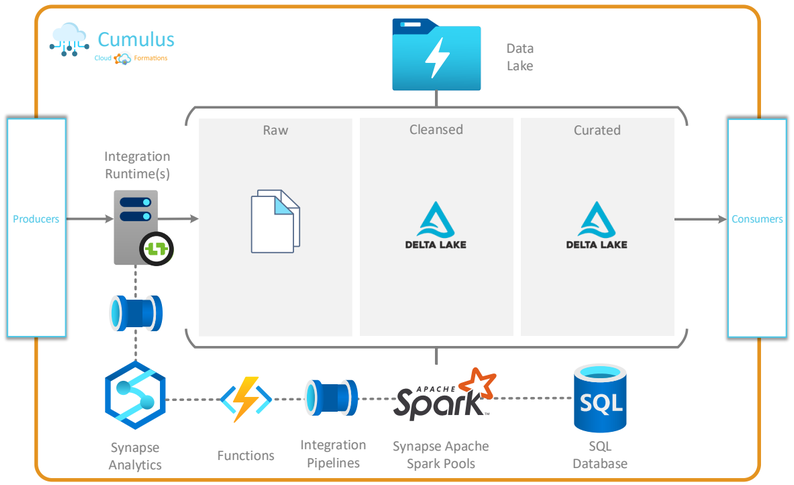
In this context, using Apache Spark compute and Delta Lake structures for the storage layer.



© 2024 Cloud Formations Ltd

Flexible Technology Choices

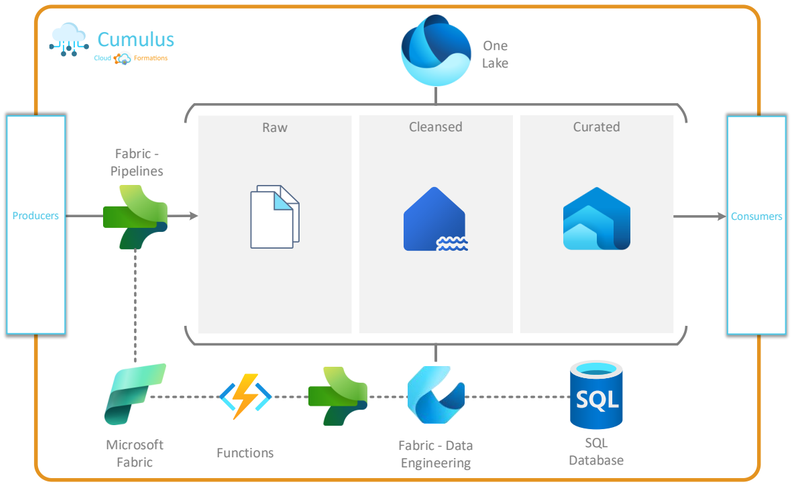
Building on the generic visuals above, but now applying a pure Microsoft product perspective to the logical delivery of Cumulus. The following diagrams could apply to a solution depending on the choice of tooling for Orchestration, Compute and Storage with each offering different advantages and coupling of resources.



Using Azure Synapse Analytics

This implementation represents a partly decoupled architecture, in terms of cloud resources used to delivery Cumulus.

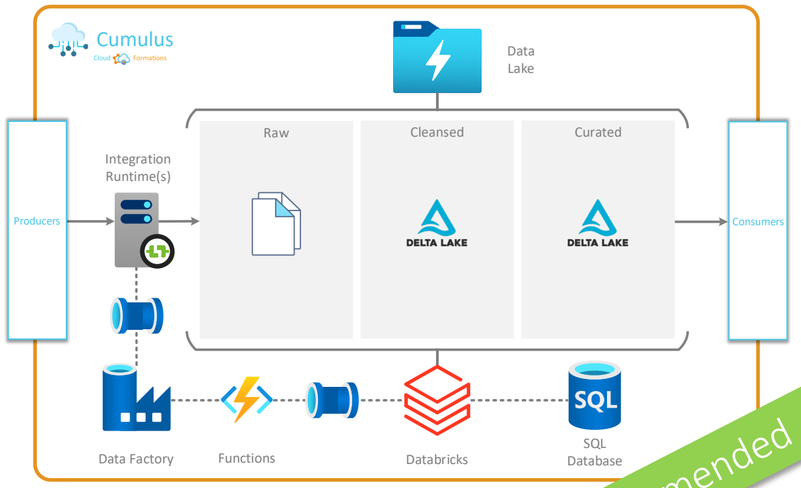
- **Orchestration** and **Compute** delivered using Azure Synapse Analytics.
- **Storage** delivered using Azure Data Lake.



Using Microsoft Fabric

This implementation represents a coupled architecture, in terms of cloud resources used to delivery Cumulus.

- **Orchestration, Compute** and **Storage** delivered using Microsoft Fabric experiences.



Using Azure Data Factory & Azure Databricks

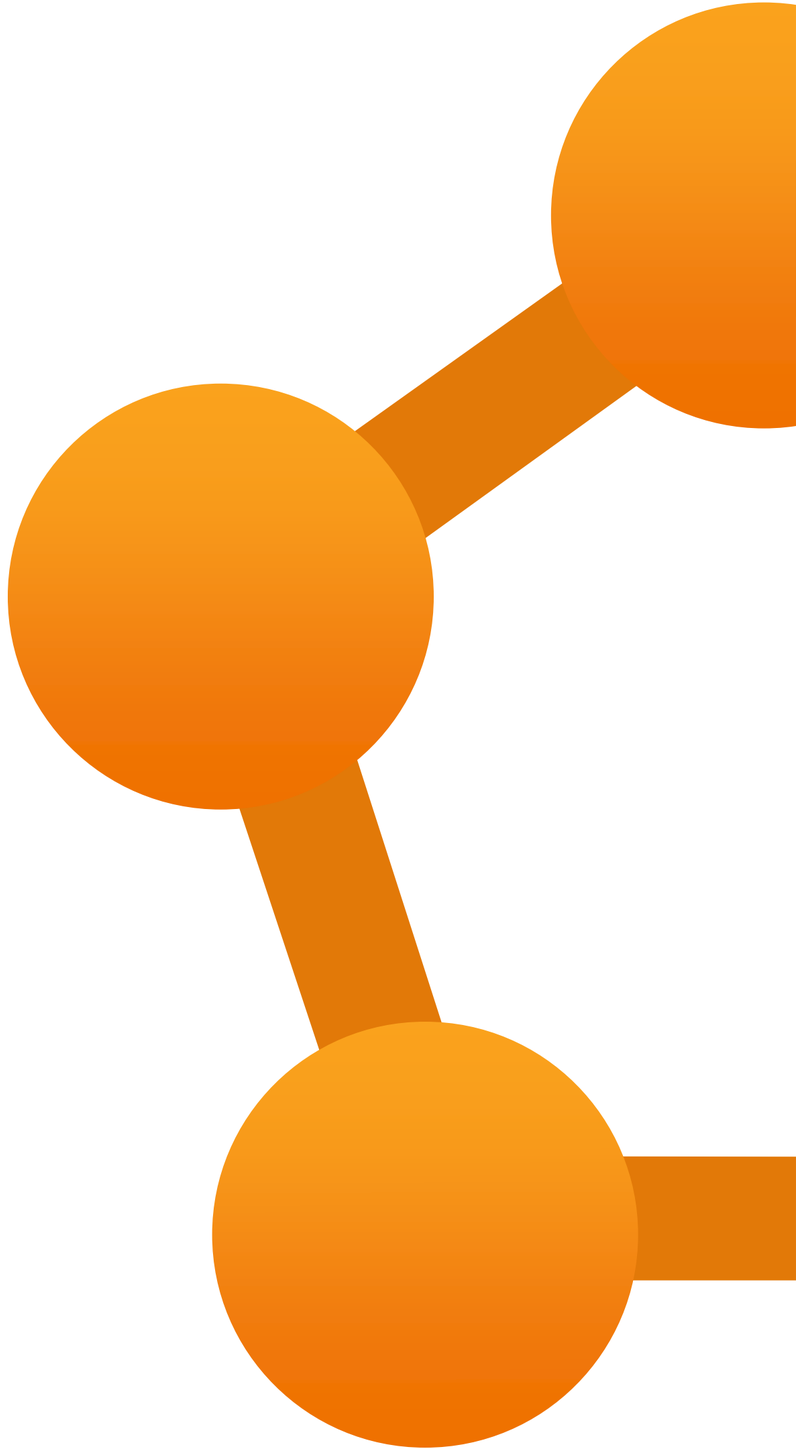
Recommended

This implementation represents a fully decoupled architecture, in terms of cloud resources used to delivery Cumulus.

- **Orchestration** delivered using Azure Data Factory.
- **Compute** delivered using Azure Databricks.
- **Storage** delivered using Azure Data Lake.

© 2024 Cloud Formations Ltd

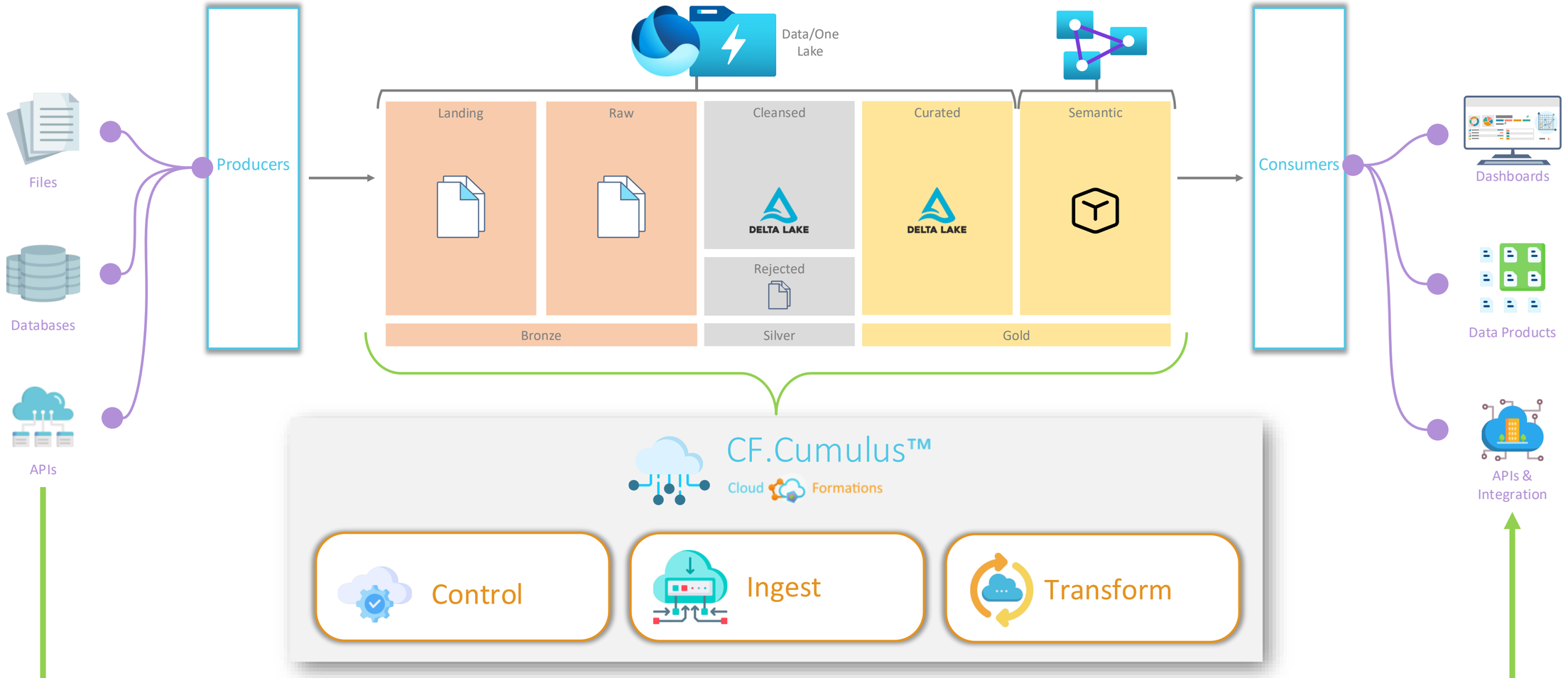
Why CF.Cumulus



Cloud Formations

Lakehouse Implementations - Simplified Through Metadata

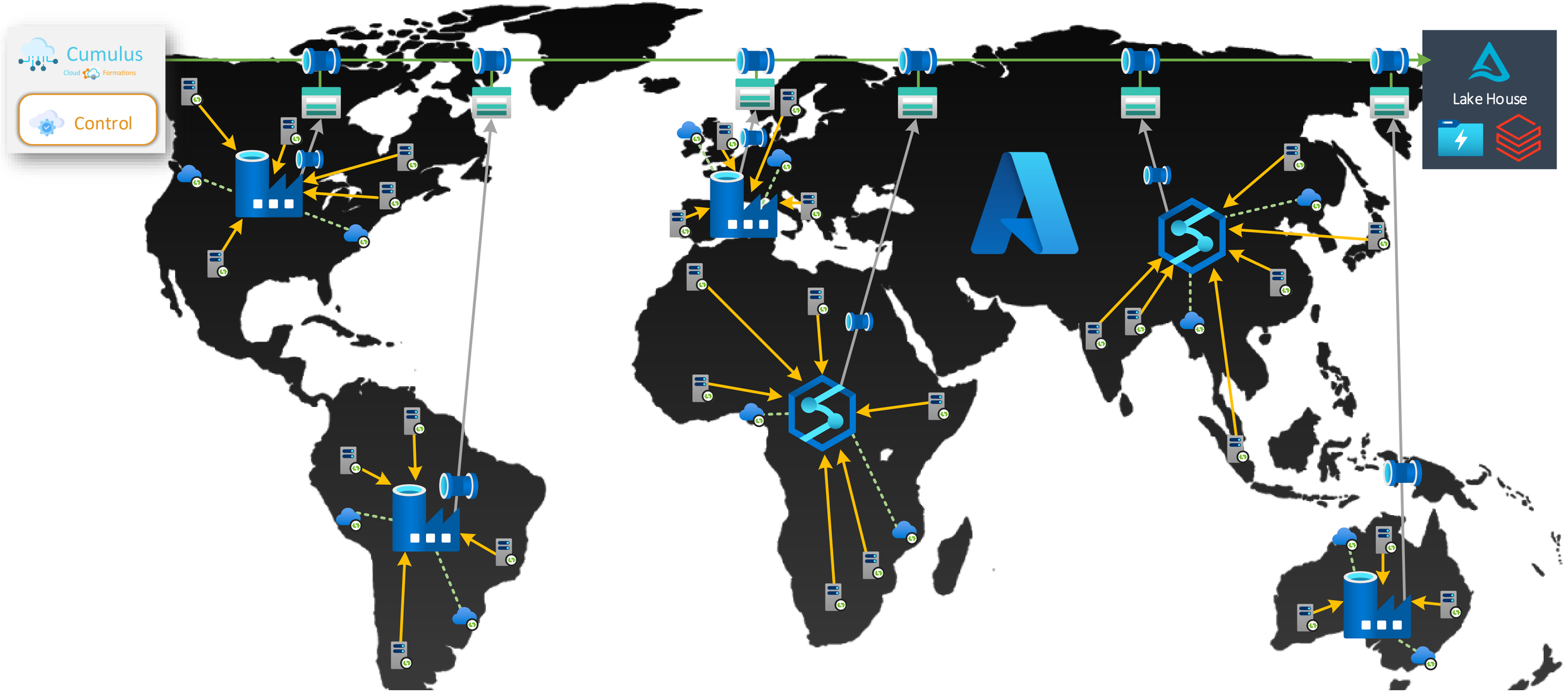
New source datasets delivered from producers to consumers with ease using only a few lines of metadata and open cloud native tooling.



© 2024 Cloud Formations Ltd. Confidential

Lakehouse Deployments - Simplified Through Metadata

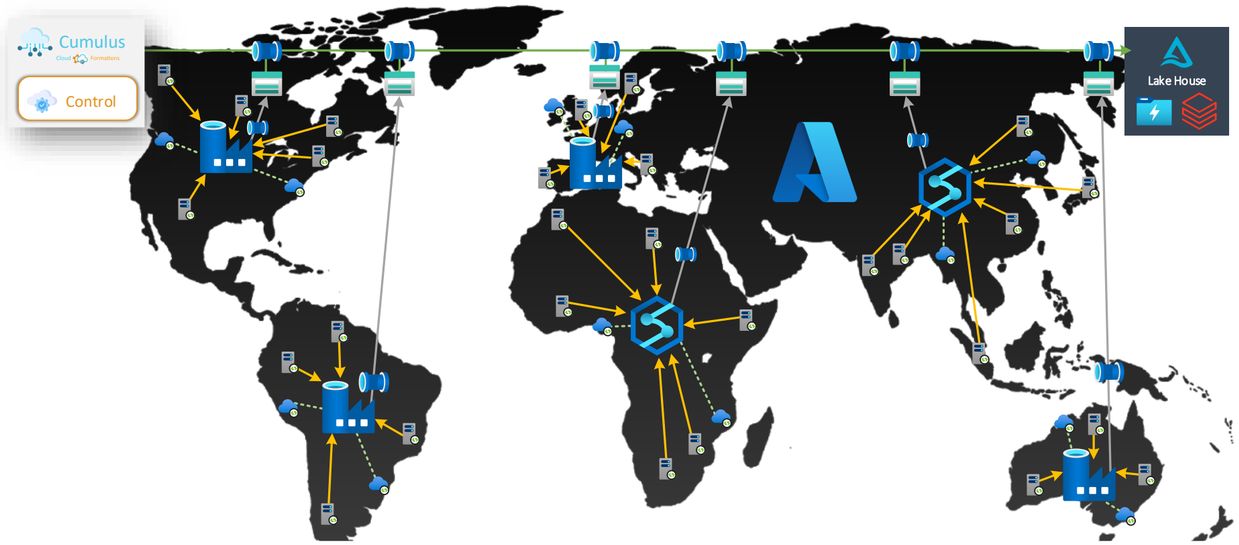
Multinational companies used CF.Cumulus to deliver cloud native orchestration at scale across 6 Azure Regions and 25 data source locations.

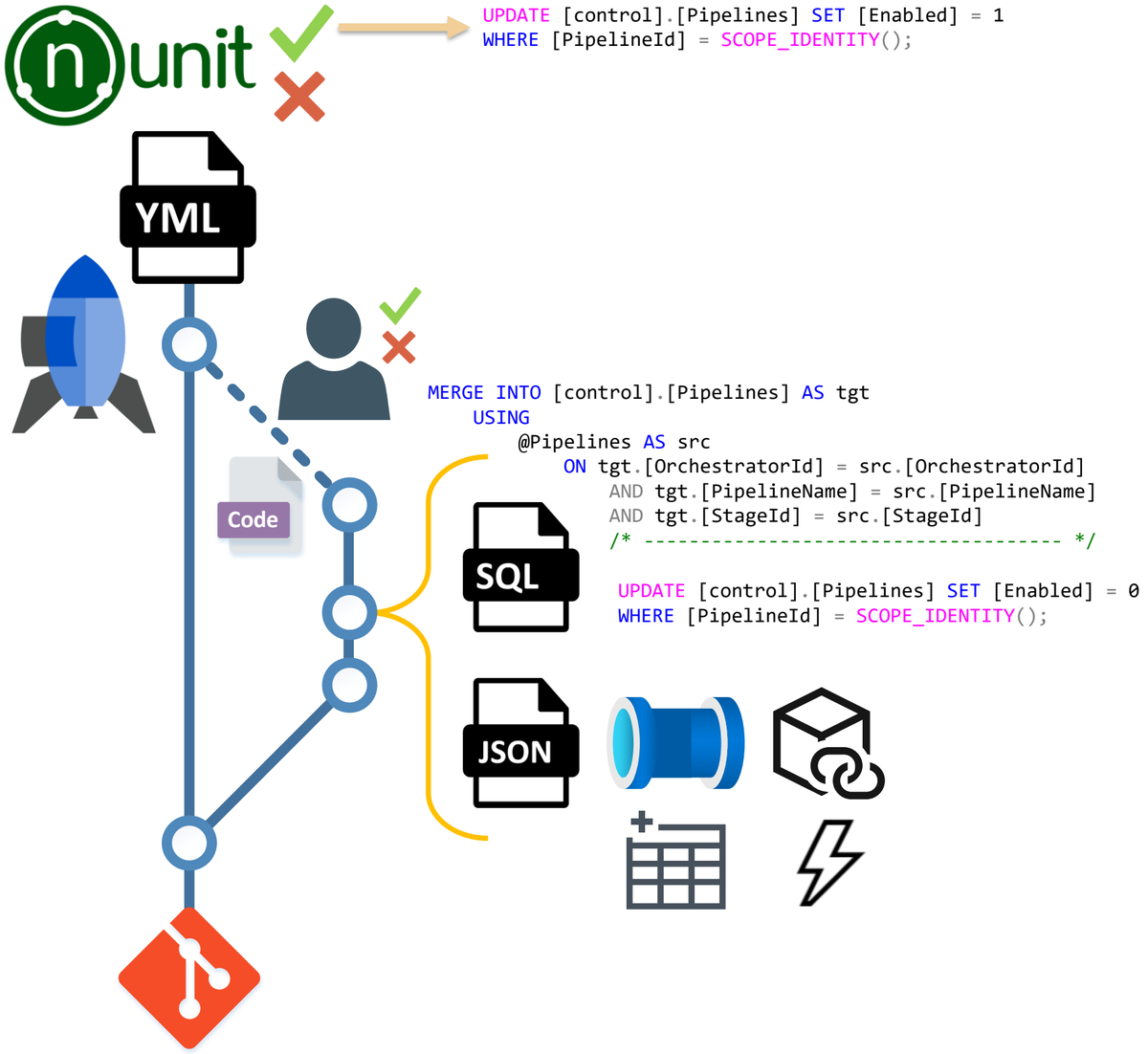


© 2024 Cloud Formations Ltd. Confidential

Lakehouse Deployments - Simplified Through Metadata

Multinational companies used CF.Cumulus to deliver cloud native orchestration at scale across 6 Azure Regions and 25 data source locations.

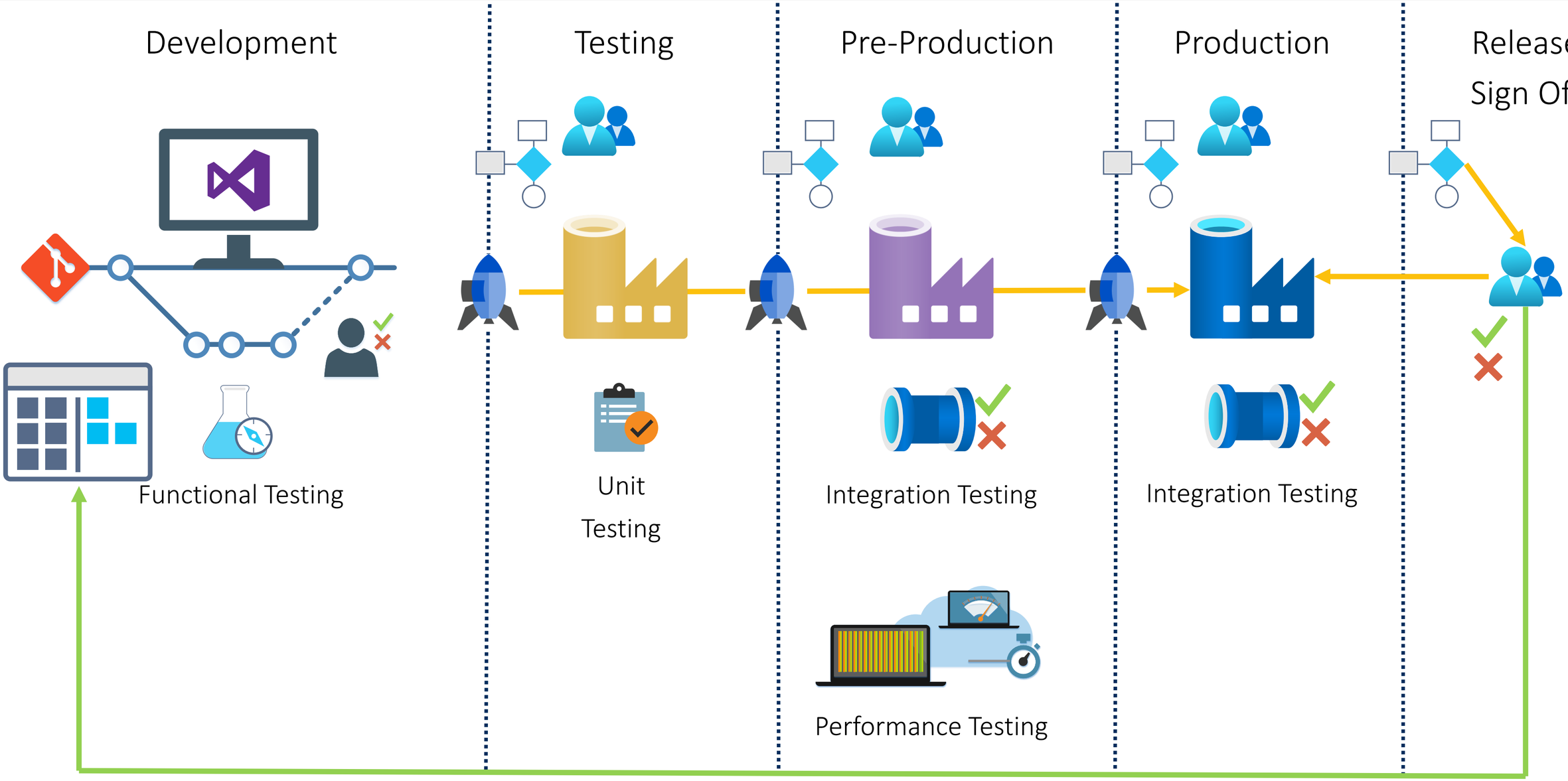




Cloud Formations

© 2024 Cloud Formations Ltd. Confidential

Lakehouse Deployments - Simplified Through Metadata



© 2024 Cloud Formations Ltd. Confidential

CF.Cumulus Indicative Run Costs

Assuming the use of an Azure UK Region priced for GBP, with all CF.Cumulus compute processes running for 4 hours every day of the year, with locally redundant storage, the following Azure run costs **per month** can be assumed:

SQL Database
250GB of Metadata
£12.08

Data Factory
Azure IR Usage
£25.84

Functions
2 Node Dedicated App Service Plan
£44.30

Databricks
3 Node Cluster with 4 vCPUs & 16GB RAM
£152.92

Blob Storage
1TB of Log Data
£18.45

Data Lake
1TB Analytics Data
£38.82

Key Vault
50,000 Requests
£0.14

Monitoring
500MB Log & App Insights Data per Day
£45.61

£338.16*

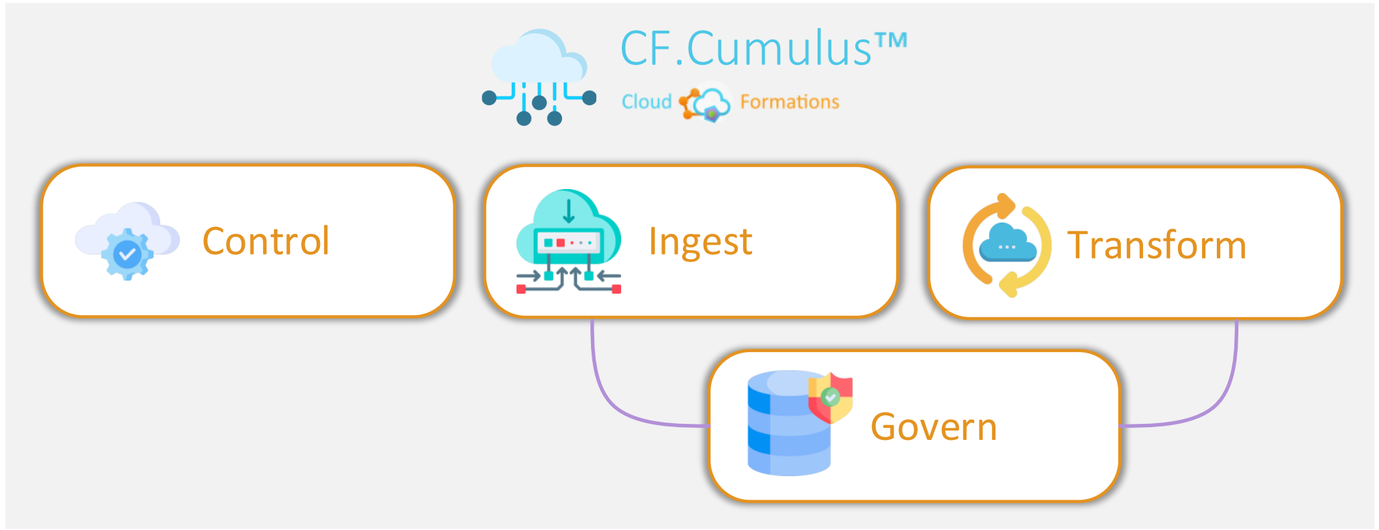
*With every customer implementation data volumes, table complexity and serving queries (to dashboards) may differ. All affecting the above run costs.

Azure Price Calculator workings available for review and adjustment: https://azure.com/e/30f8aeedfd564ba189d33f278845bb89

Cloud Formations - Product Run Costs

© 2024 Cloud Formations Ltd

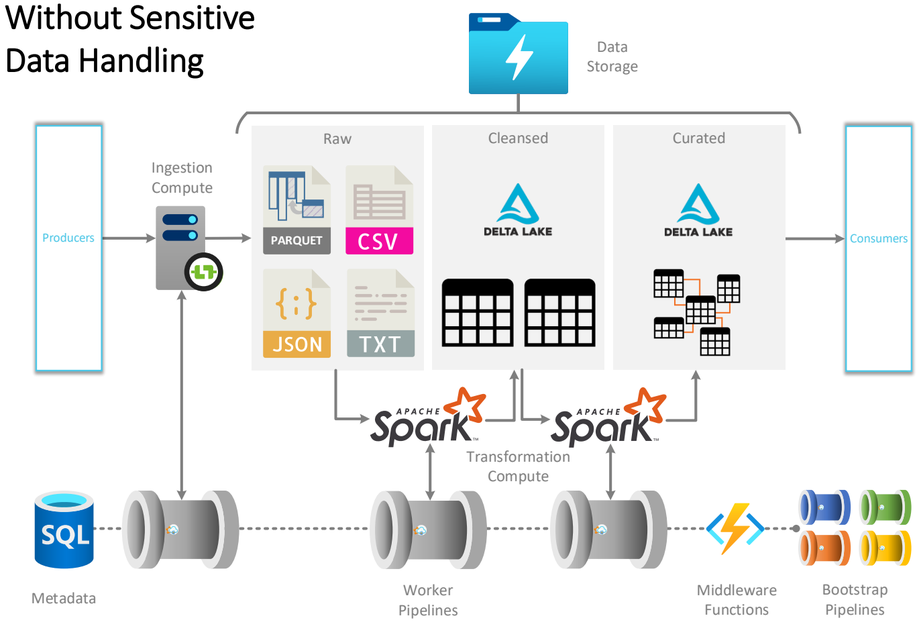
CF.Cumulus – Govern Plugin



Govern

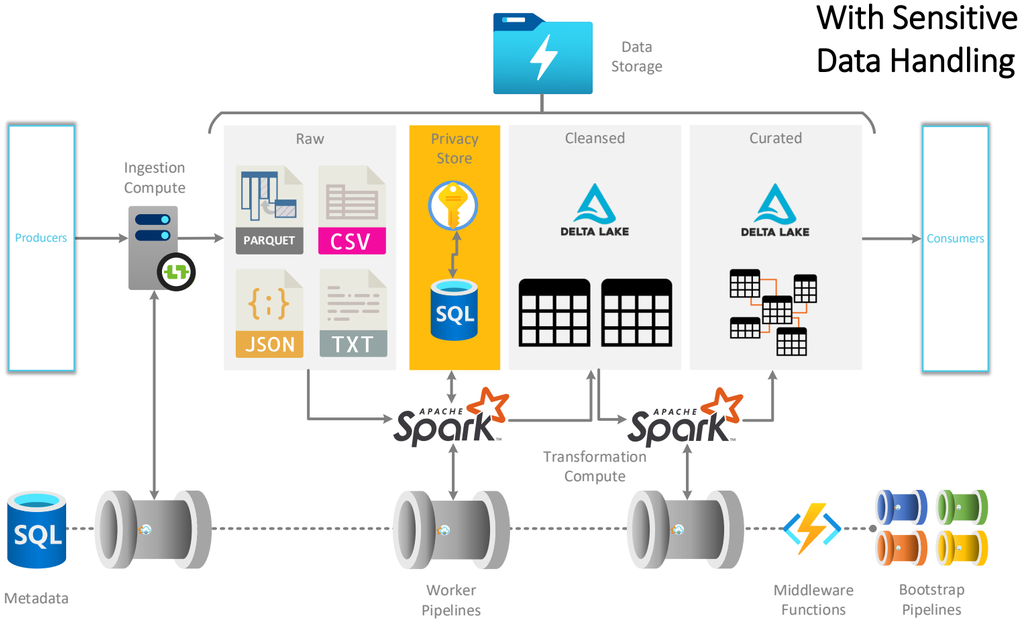
- Metadata-driven sensitive data handling.
- Apply ingestion data policies to support GDPR use cases.
- Custom value encryption and decryption.
- Custom value type handling for all sensitive values.
- Downstream reporting content/aggregations maintained.
- Isolation of sensitive data into a dedicated Privacy Store.
- Implemented data retention policies across Raw data.
- Synchronous and asynchronous, performant value handling.
- Integration with CF.Cumulus.Ingest and Control.
- API value retrieval for trusted Workers.

Without Sensitive Data Handling



A logical view of the CF.Cumulus product architect with and without sensitive data handling implemented as a plugin for the Ingest and Control components. Relational values diverted to a dedicated Privacy Store with custom encryption and full GDPR support.

With Sensitive Data Handling



Cloud Formations - Logical Architecture Overview

© 2024 Cloud Formations Ltd

CF.Cumulus.Support

Our enduring support gives you delivery confidence when using CF.Cumulus from the product creators, with level 3 support and guidance.

CF.Cumulus User Training

- Deep Dive CF.Cumulus Training Workshop
- Licenced Training Content for Internal Reuse

Teaching your engineering teams 'how to fish' with the open-source CF.Cumulus product. Onboarding new datasets with simple metadata changes. Trained on your own CF.Cumulus instance if available.

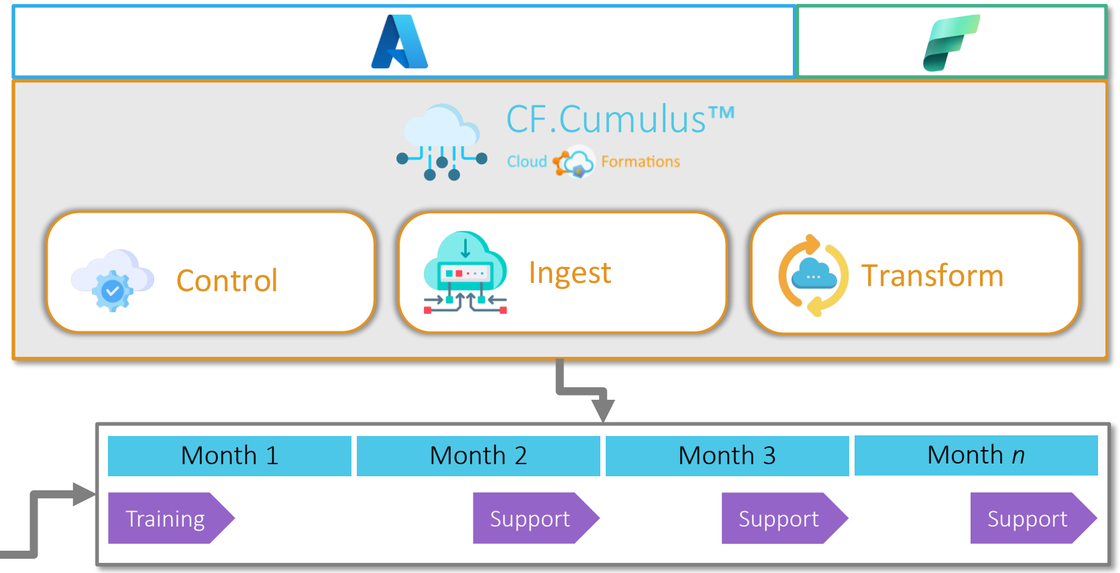
1 Day Training

CF.Cumulus Product Support

- Knowledge Base Access
- Product Roadmap Collaboration
- Feature Technical Guidance
- Early Feature Access & Upgrades
- Level 3 Service Desk
- Product Defect Resolution

On hand technical advice and support for the CF.Cumulus product available ad-hoc with access to our complete knowledge base, product roadmap and upgrade support as we continue to innovate and build out new product features (based on your business priorities).

Ongoing Innovation & Support

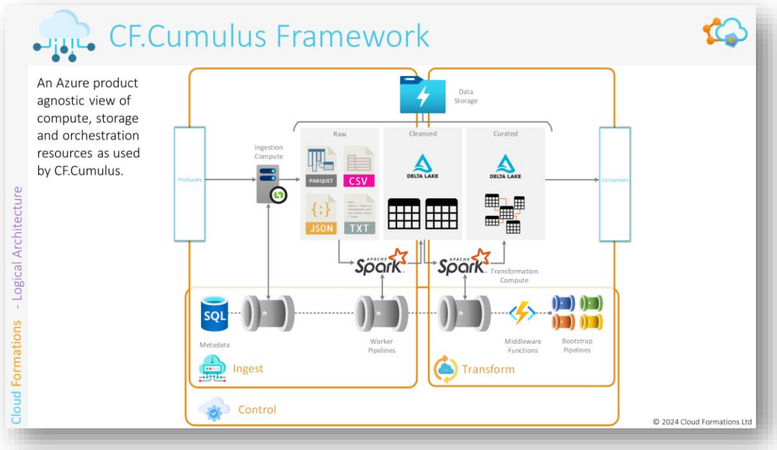


Seamless Support, Every Step of the Way

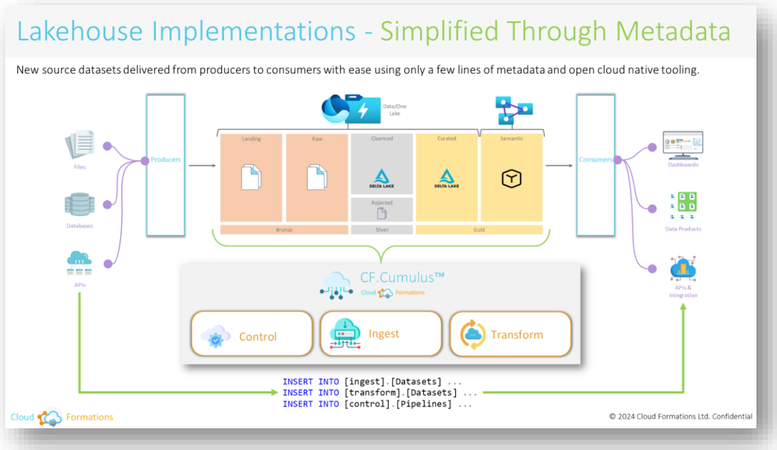
Cloud Formations - Product Wrapper

© 2024 Cloud Formations Ltd

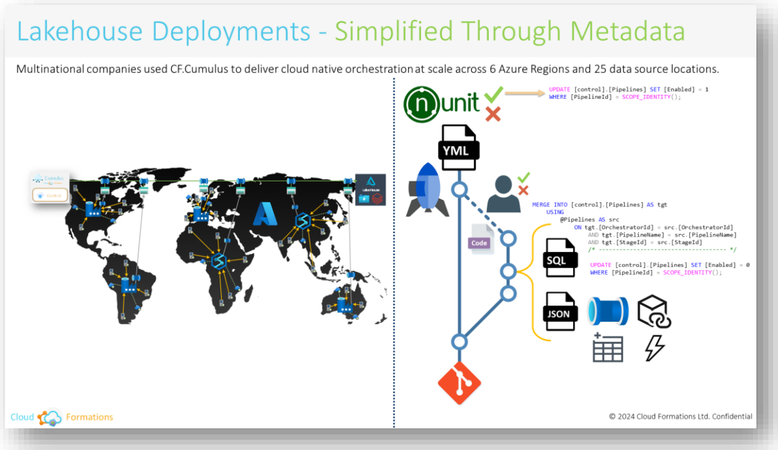
In Summary



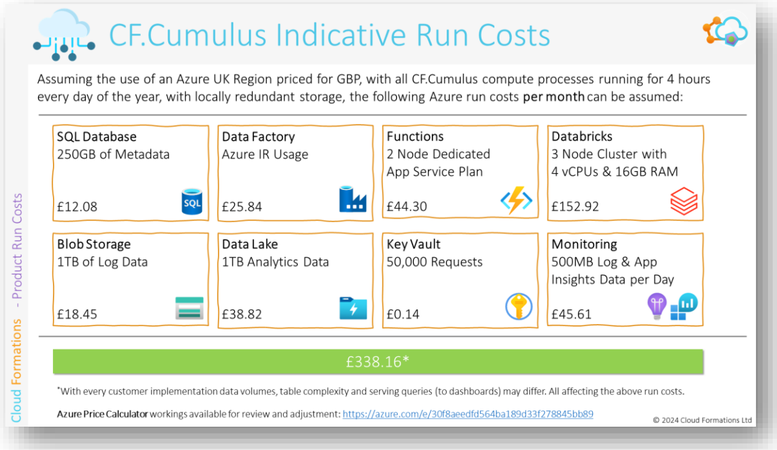
1. Open-source and open-standards.



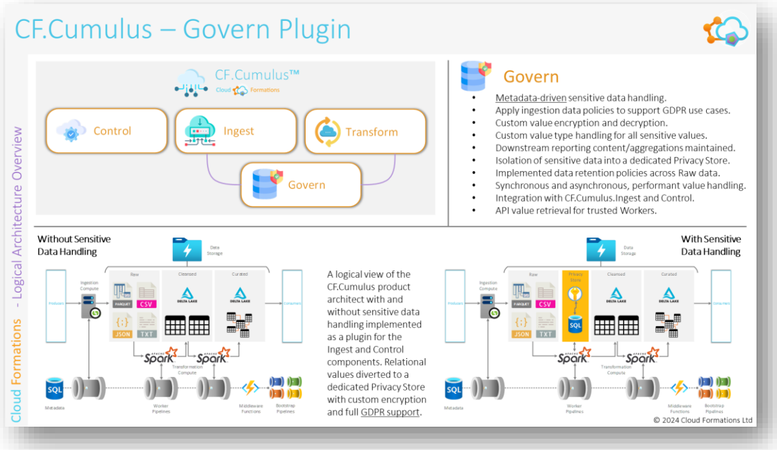
2. Rapid dataset delivery.



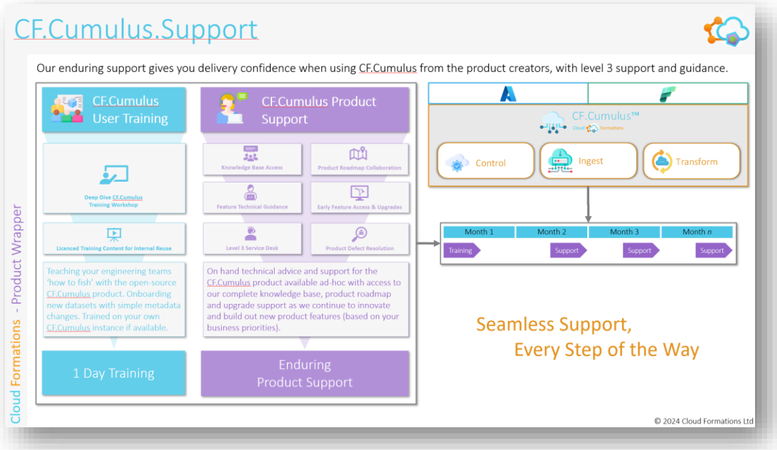
3. Deployment control.



4. Low running costs.



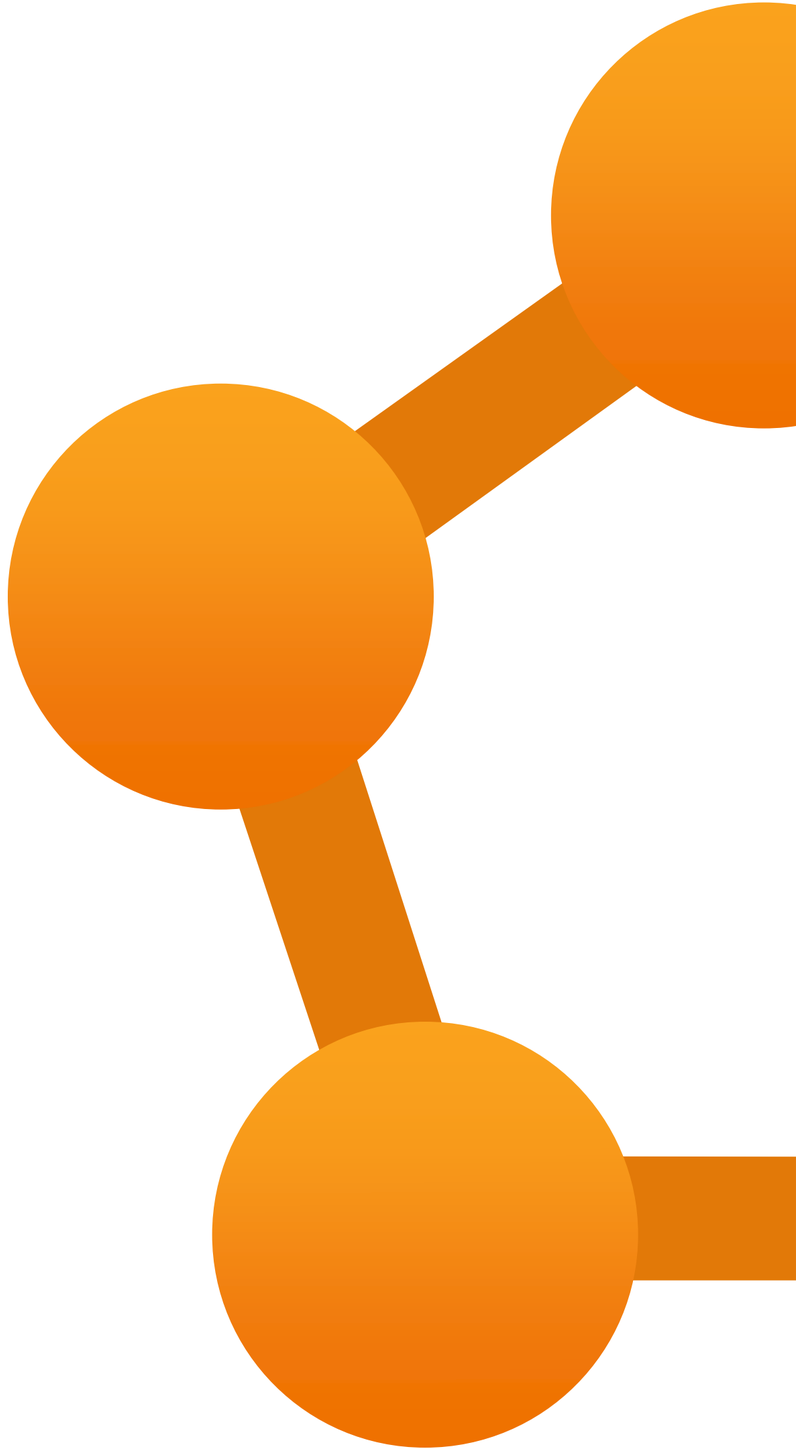
5. Plugins and extensibility.



6. Comprehensive support.

© 2024 Cloud Formations Ltd. Confidential

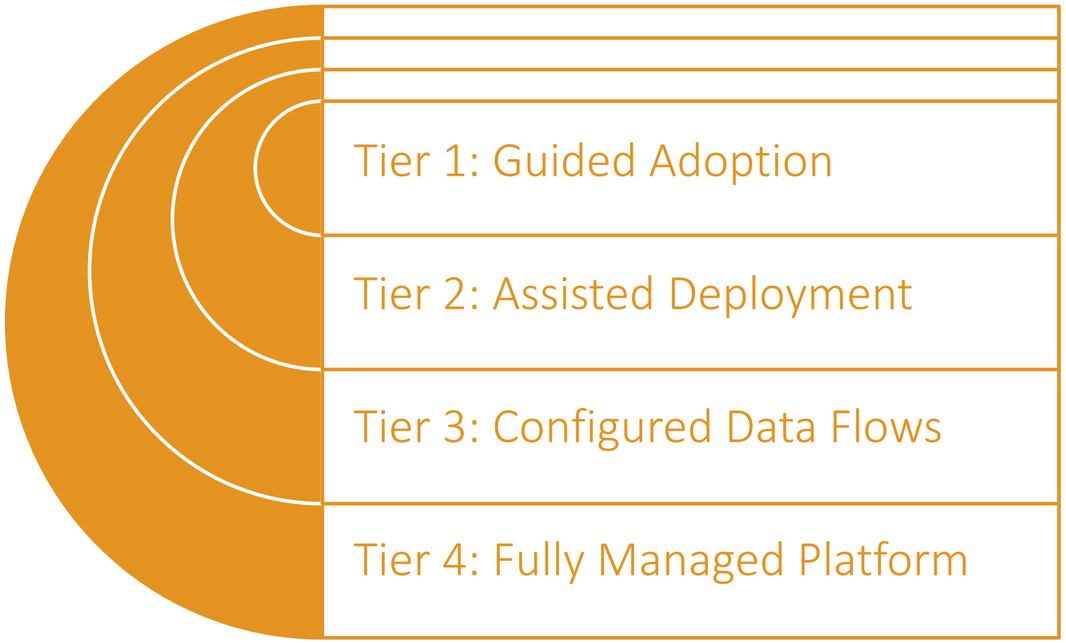
Engagement Options



Cloud Formations

CF.Cumulus: Flexible Tiers of Engagement

Unlock the full potential of the **CF.Cumulus** at your own pace. Whether you're looking to deploy and manage it yourself, or prefer a fully managed service, we have the right level of support for you.



Cumulus is now available as an open-source framework, allowing businesses to leverage cutting-edge data orchestration technology without license fees. However, the real value comes from partnering with us to ensure a successful deployment. Our decades of experience in building scalable, resilient data platforms will ensure that you extract the maximum value from your data.

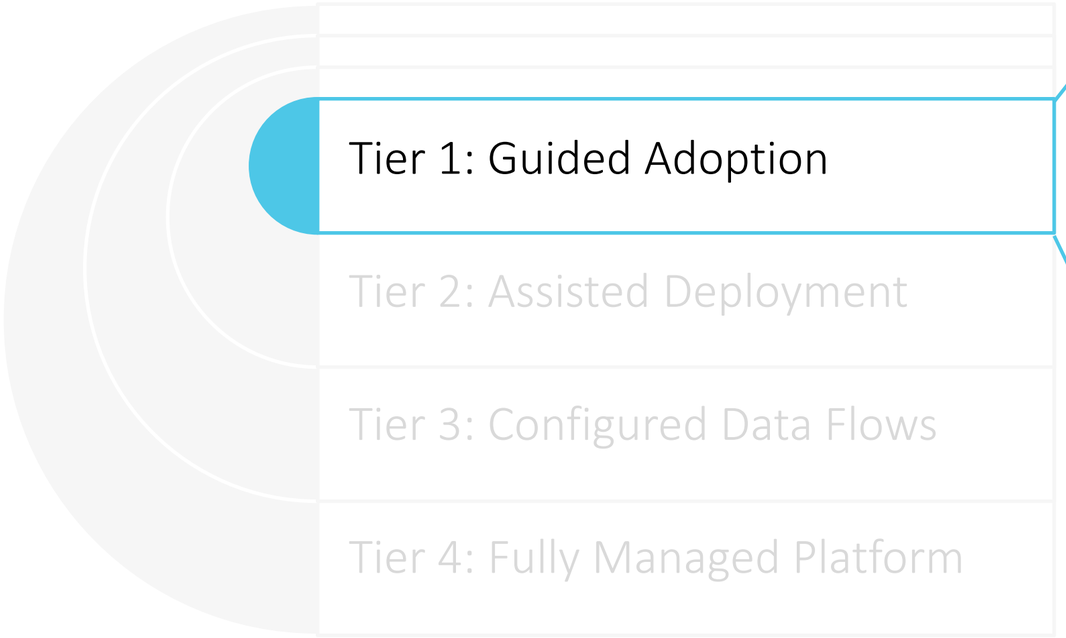
Open-Source Flexibility, Professional Expertise

Cloud Formations

© 2024 Cloud Formations Ltd. Confidential

CF.Cumulus: Flexible Tiers of Engagement

Unlock the full potential of the **CF.Cumulus** at your own pace. Whether you're looking to deploy and manage it yourself, or prefer a fully managed service, we have the right level of support for you.



Tier 1: Guided Adoption – Training & Product Support

Take full control with expert guidance

We equip your team with the knowledge and tools to confidently deploy, configure, and use the framework. After training, our team remains available to support you through your initial implementation phase, ensuring a smooth transition to self-sufficiency.

What You Get

- Hands-on training covering deployment, configuration, and usage
- On-demand support during early implementation stages
- Easy transition into ongoing product support with continued guidance and access to our L3/L4 service desk for Cumulus-specific issues

Who is this for?

For teams that want to take ownership of the Cumulus platform but need expert training and a support safety net during the initial stages.

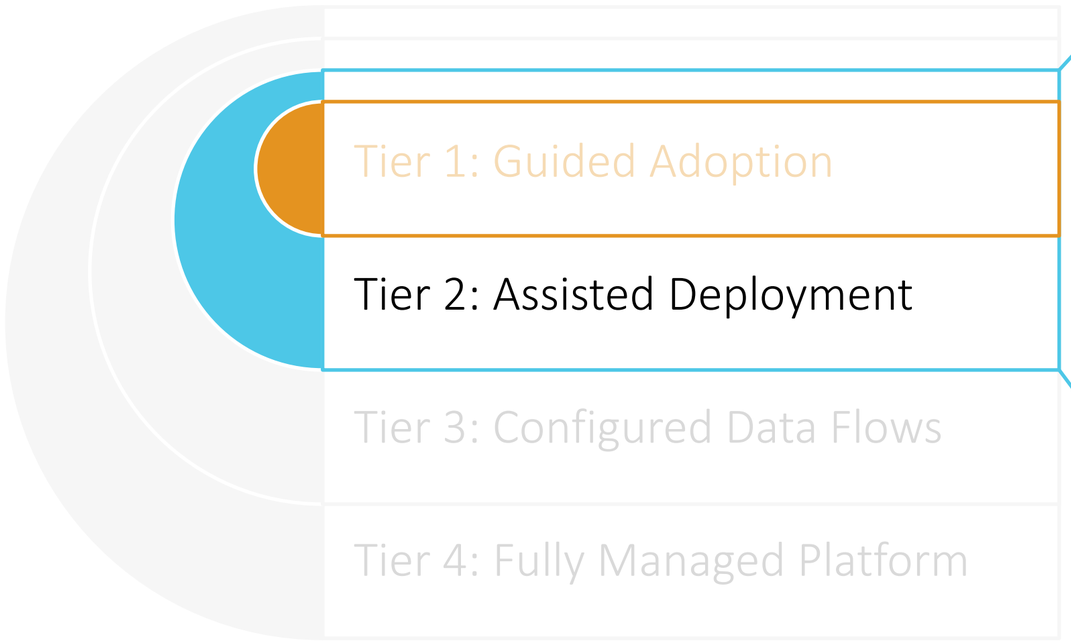
Open-Source Flexibility, Professional Expertise

Cloud Formations

© 2024 Cloud Formations Ltd. Confidential

CF.Cumulus: Flexible Tiers of Engagement

Unlock the full potential of the **CF.Cumulus** at your own pace. Whether you're looking to deploy and manage it yourself, or prefer a fully managed service, we have the right level of support for you.



Tier 2: Assisted Deployment – Design & Deployment

Handle complex environments with confidence.

For organizations with more intricate security, networking, or infrastructure requirements, we go beyond training by helping you design and deploy the framework in your environment. You'll get the best of both worlds—our expertise to handle complexity, plus the flexibility to manage the platform post-deployment.

What You Get

- Everything from Tier 1
- Custom architecture design to accommodate security and networking needs
- Initial deployment in your environment with hands-on assistance

Who is this for?

For businesses with complex infrastructure that want expert help with deployment but still wish to manage the platform internally after setup.

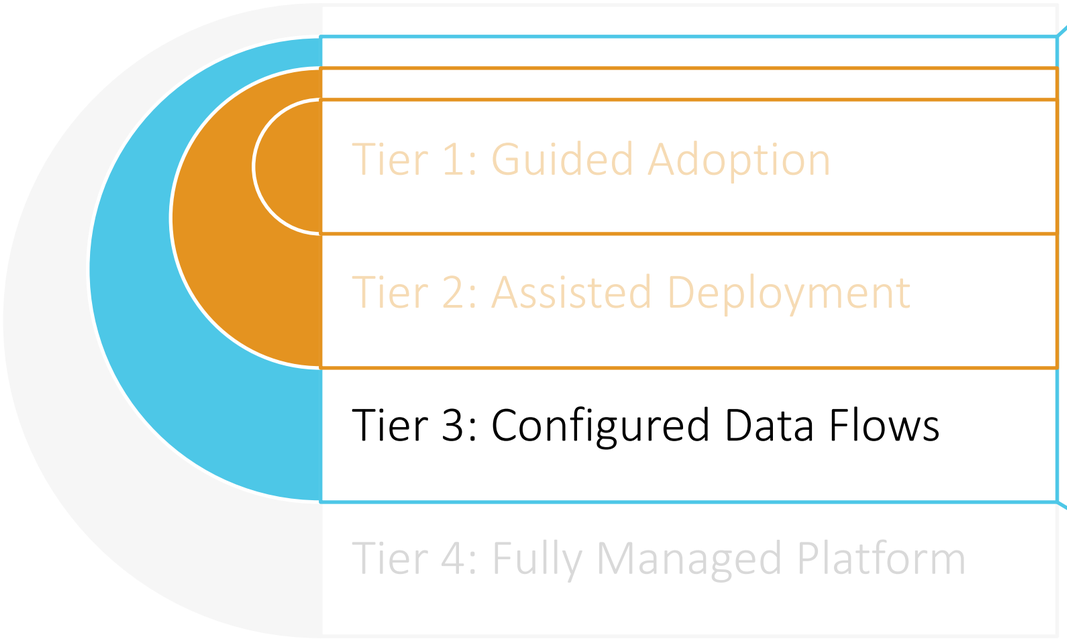
Open-Source Flexibility, Professional Expertise

Cloud Formations

© 2024 Cloud Formations Ltd. Confidential

CF.Cumulus: Flexible Tiers of Engagement

Unlock the full potential of the **CF.Cumulus** at your own pace. Whether you're looking to deploy and manage it yourself, or prefer a fully managed service, we have the right level of support for you.



Tier 3: Configured Data Flows – Initial Configuration

Jumpstart your data workflows.

In addition to deployment, we take the complexity out of configuring data flows and transformations. We'll handle the setup of key workflows, ensuring your platform is ready to generate insights from day one. This tier is perfect for businesses that need a fully operational data solution from the outset.

What You Get

- Everything from Tier 2
- Full configuration of data ingestion and transformations based on your data sources and business needs
- Scaled pricing depending on the number of data sources and transformations

Who is this for?

For businesses that need to get their data flowing quickly but still want to manage day-to-day operations.

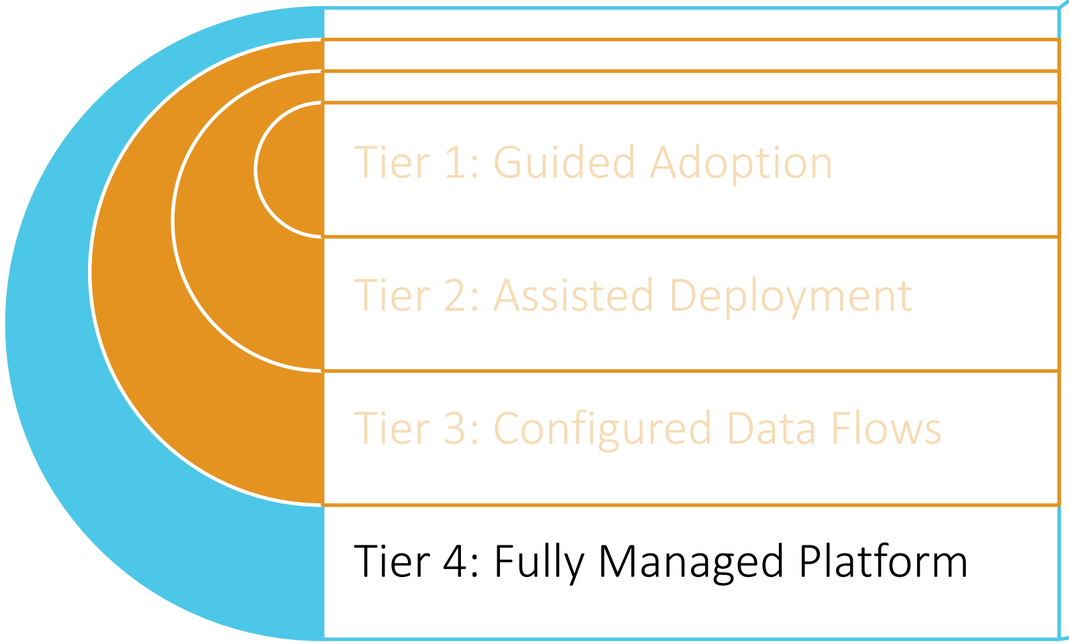
Open-Source Flexibility, Professional Expertise

Cloud Formations

© 2024 Cloud Formations Ltd. Confidential

CF.Cumulus: Flexible Tiers of Engagement

Unlock the full potential of the **CF.Cumulus** at your own pace. Whether you're looking to deploy and manage it yourself, or prefer a fully managed service, we have the right level of support for you.



Tier 4: Fully Managed Platform – End-to-End Support

Leave the heavy lifting to us.

This fully managed solution offers complete peace of mind. We handle everything from deployment to ongoing platform management and support. Your team can focus on deriving insights, while we ensure your data platform operates smoothly.

What You Get

- Everything from Tier 3
- Full platform management including support for both the CF.Cumulus framework and your data workflows.
- Ongoing monitoring and optimization to ensure peak performance.

Who is this for?

For businesses looking for a turnkey solution with ongoing management and support to ensure a reliable, scalable data platform without internal management overhead.

Open-Source Flexibility, Professional Expertise

Cloud Formations

© 2024 Cloud Formations Ltd. Confidential

CF.Cumulus: Flexible Tiers of Engagement

Tier 1: Guided Adoption – Training & Product Support

Take full control with expert guidance

We equip your team with the knowledge and tools to confidently deploy, configure, and use the framework. After training, our team remains available to support you through your initial implementation phase, ensuring a smooth transition to self-sufficiency.

What You Get

- Hands-on training covering deployment, configuration, and usage
- On-demand support during early implementation stages
- Easy transition into ongoing product support with continued guidance and access to our L3/L4 service desk for Cumulus-specific issues

Who is this for?

For teams that want to take ownership of the Cumulus platform but need expert training and a support safety net during the initial stages.

Tier 2: Assisted Deployment – Design & Deployment

Handle complex environments with confidence.

For organizations with more intricate security, networking, or infrastructure requirements, we go beyond training by helping you design and deploy the Cumulus framework in your environment. You'll get the best of both worlds—our expertise to handle complexity, plus the flexibility to manage the platform post-deployment.

What You Get

- Everything from Tier 1
- Custom architecture design to accommodate security and networking needs
- Initial deployment in your environment with hands-on assistance

Who is this for?

For businesses with complex infrastructure that want expert help with deployment but still wish to manage the platform internally after setup.

Tier 3: Configured Data Flows – Initial Configuration

Jumpstart your data workflows.

In addition to deployment, we take the complexity out of configuring data flows and transformations. We'll handle the setup of key workflows, ensuring your platform is ready to generate insights from day one. This tier is perfect for businesses that need a fully operational data solution from the outset.

What You Get

- Everything from Tier 2
- Full configuration of data ingestion and transformations based on your data sources and business needs
- Scaled pricing depending on the number of data sources and transformations

Who is this for?

For businesses that need to get their data flowing quickly but still want to manage day-to-day operations.

Tier 4: Fully Managed Platform – End-to-End Support

Leave the heavy lifting to us.

This fully managed solution offers complete peace of mind. We handle everything from deployment to ongoing platform management and support. Your team can focus on deriving insights, while we ensure your data platform operates smoothly.

What You Get

- Everything from Tier 3
- Full platform management including support for both the CF.Cumulus framework and your data workflows.
- Ongoing monitoring and optimization to ensure peak performance.

Who is this for?

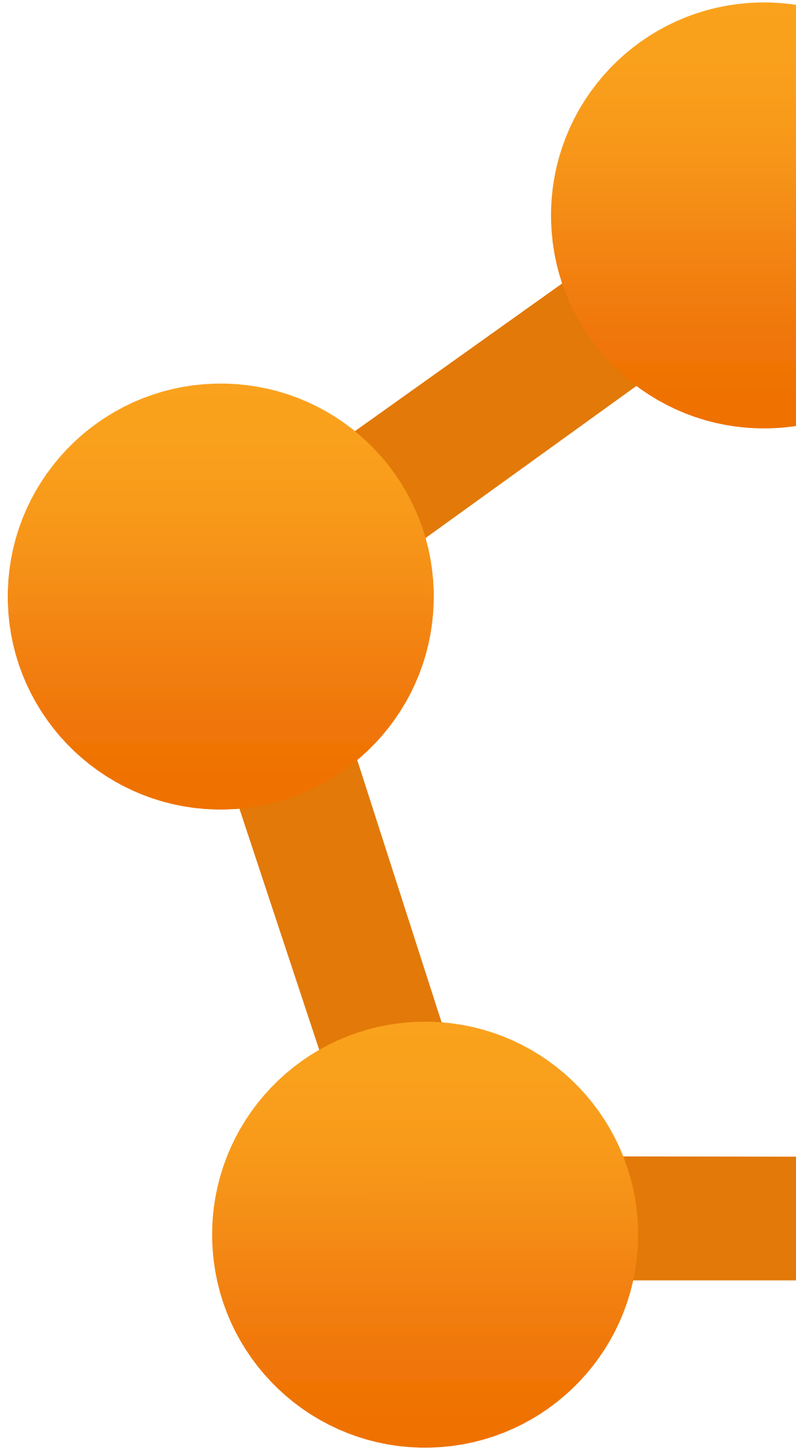
For businesses looking for a turnkey solution with ongoing management and support to ensure a reliable, scalable data platform without internal management overhead.

Open-Source Flexibility, Professional Expertise

Cloud Formations

© 2024 Cloud Formations Ltd. Confidential

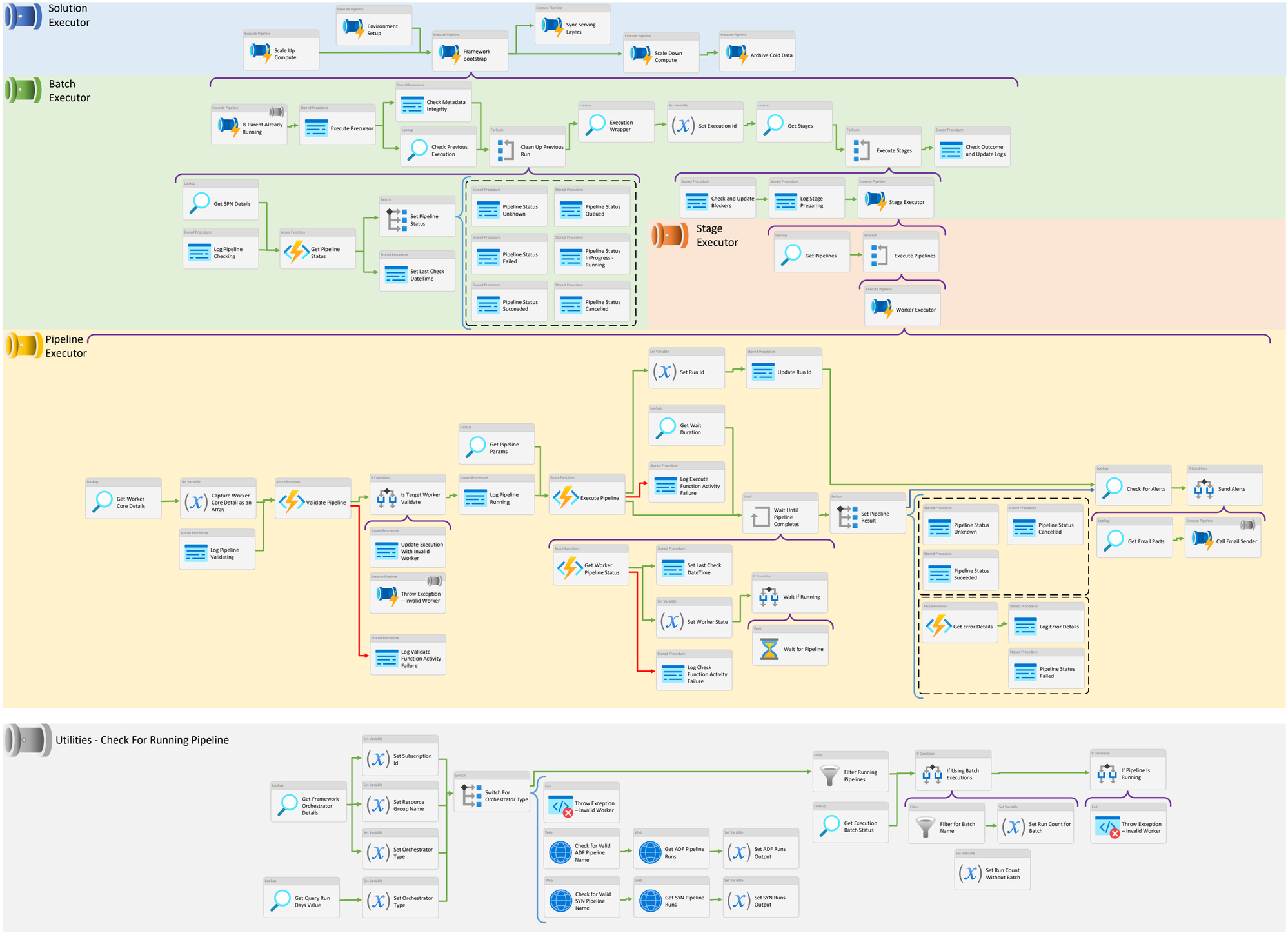
Product Internals



Cloud Formations

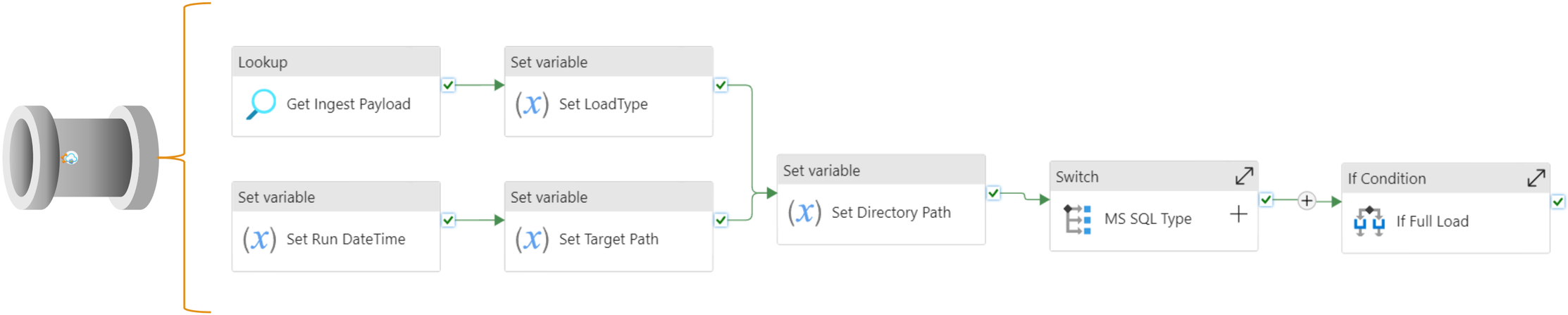
CF.Control Pipeline Internals

A view on the activity structure within the CF.Cumulus control pipelines, executing each of the Ingest and Transform workers.



Cloud Formations

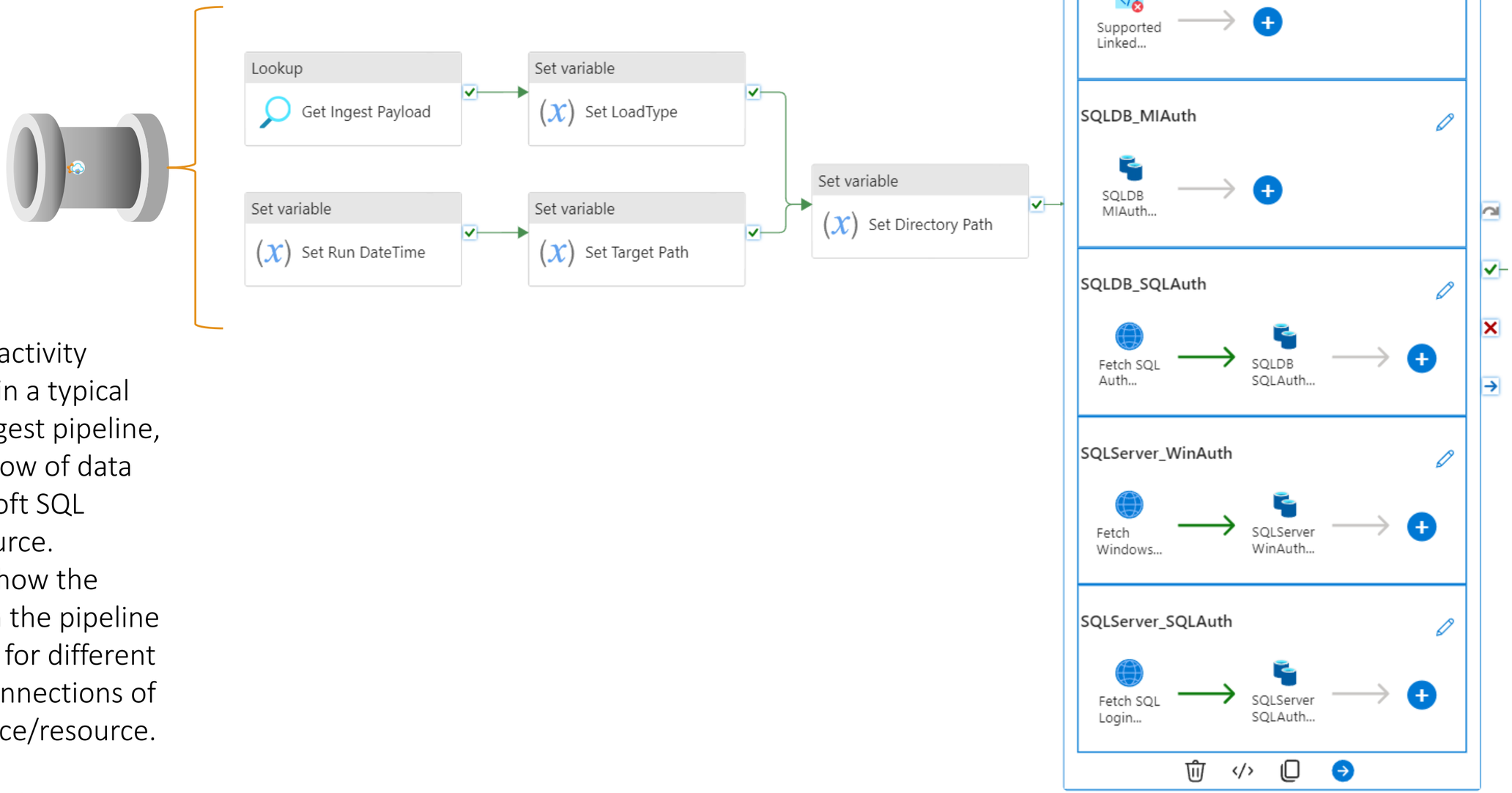
CF.Ingest Pipeline Internals



A view on the activity structure within a typical CF.Cumulus ingest pipeline, handling the flow of data from a Microsoft SQL based data source.

© 2024 Cloud Formations Ltd. Confidential

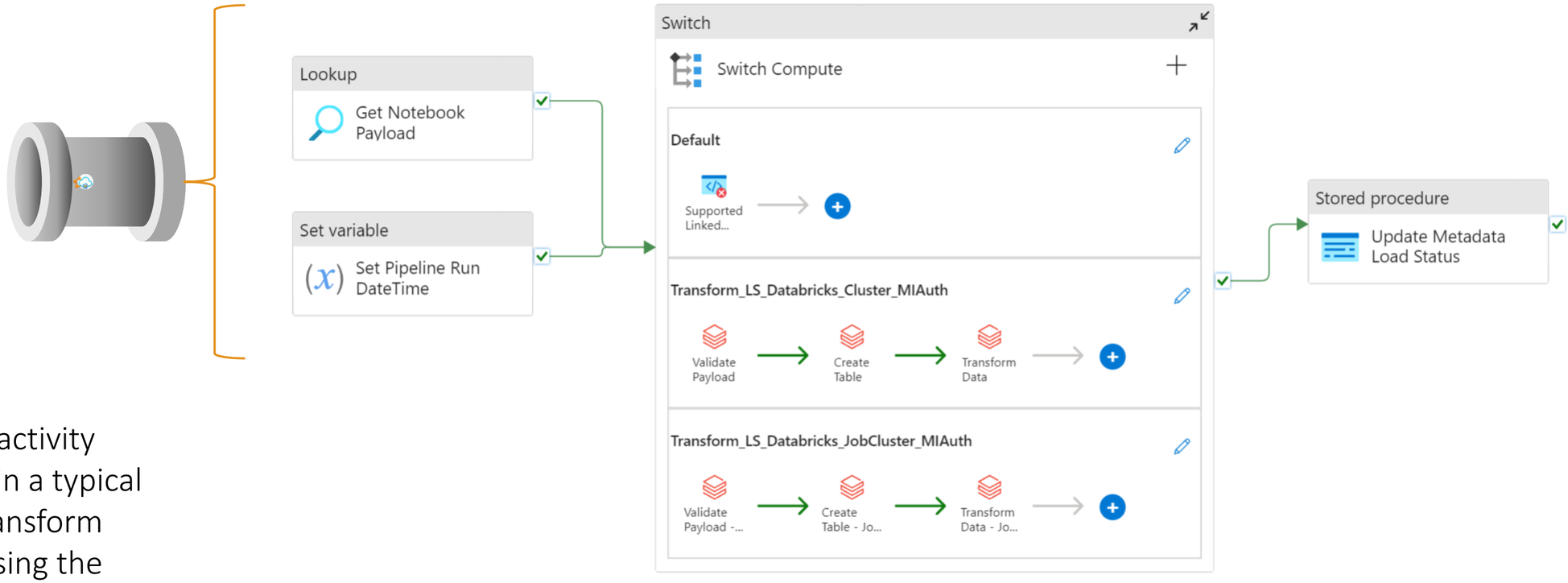
CF.Ingest Pipeline Internals



A view on the activity structure within a typical CF.Cumulus ingest pipeline, handling the flow of data from a Microsoft SQL based data source. Expanded to show the variations with the pipeline Switch activity for different version and connections of the SQL instance/resource.

© 2024 Cloud Formations Ltd. Confidential

CF.Transform Pipeline Internals

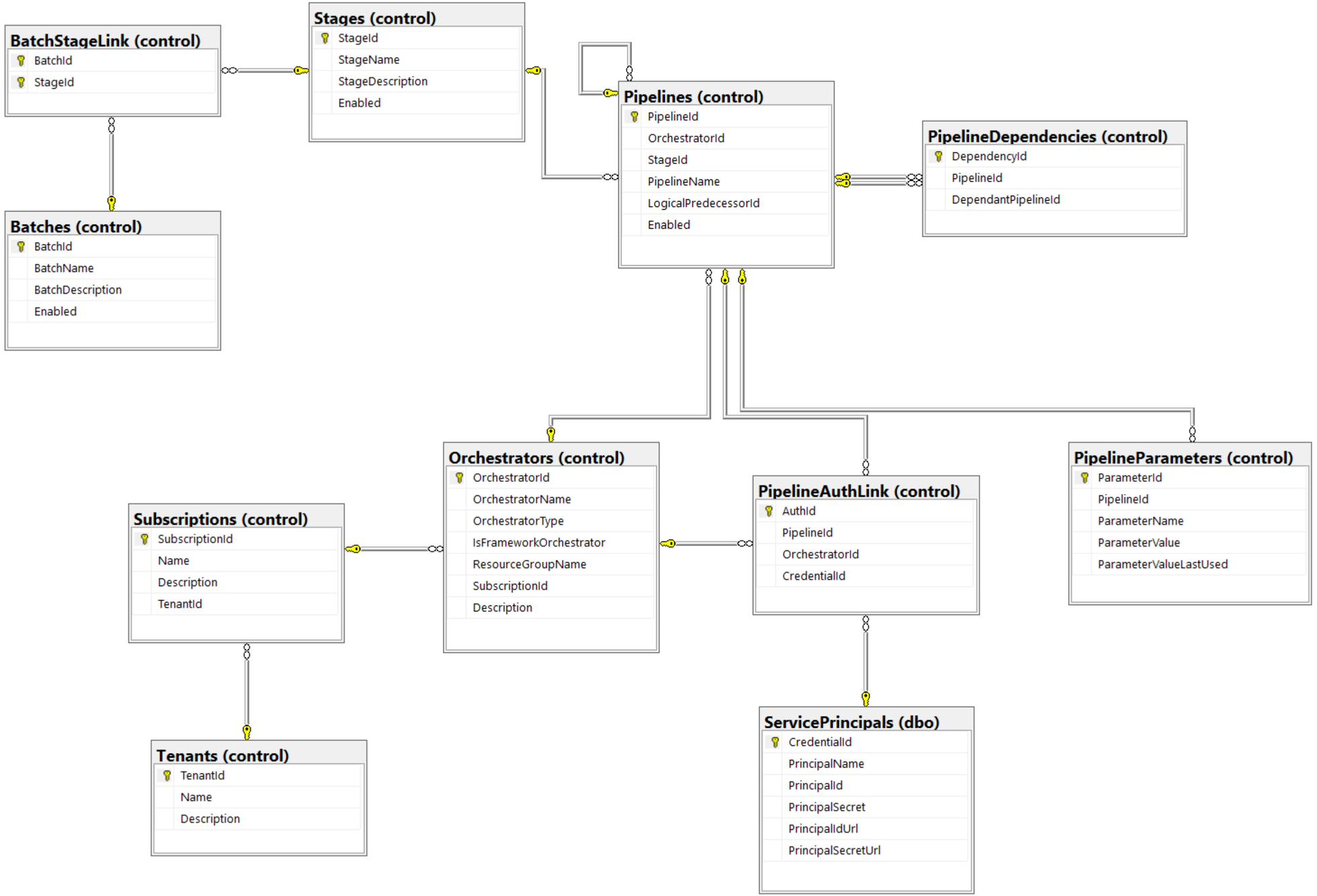


A view on the activity structure within a typical CF.Cumulus transform pipeline, exposing the Switch activity calling different variations of Notebook, depending on cluster sizing and authentication requirements.

Cloud Formations

© 2024 Cloud Formations Ltd. Confidential

CF.Control Database Internals

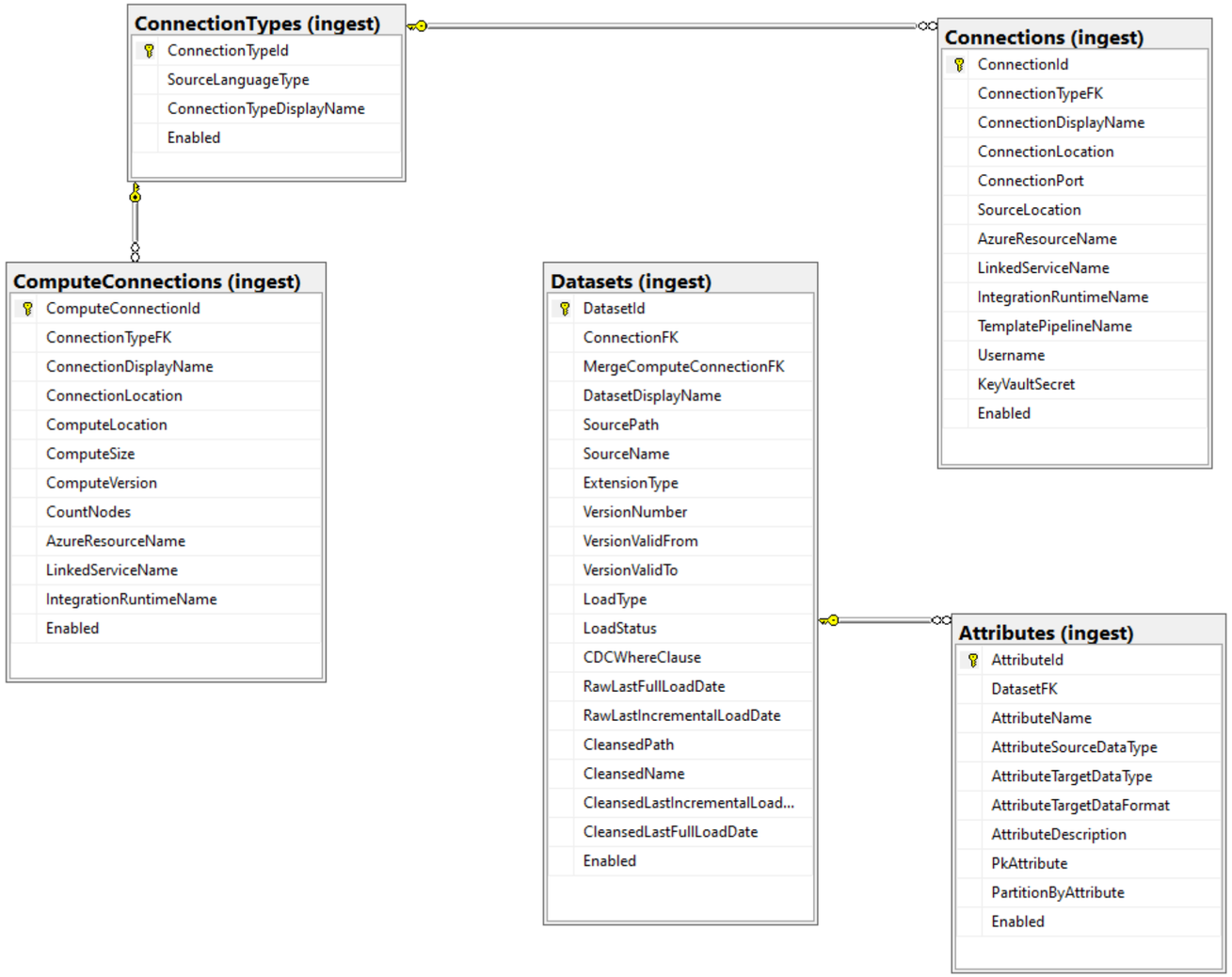


A view on the entities and relationships with CF.Cumulus metadata database. As presented for use by the Control pipelines.

Cloud Formations

© 2024 Cloud Formations Ltd. Confidential

CF.Ingest Database Internals

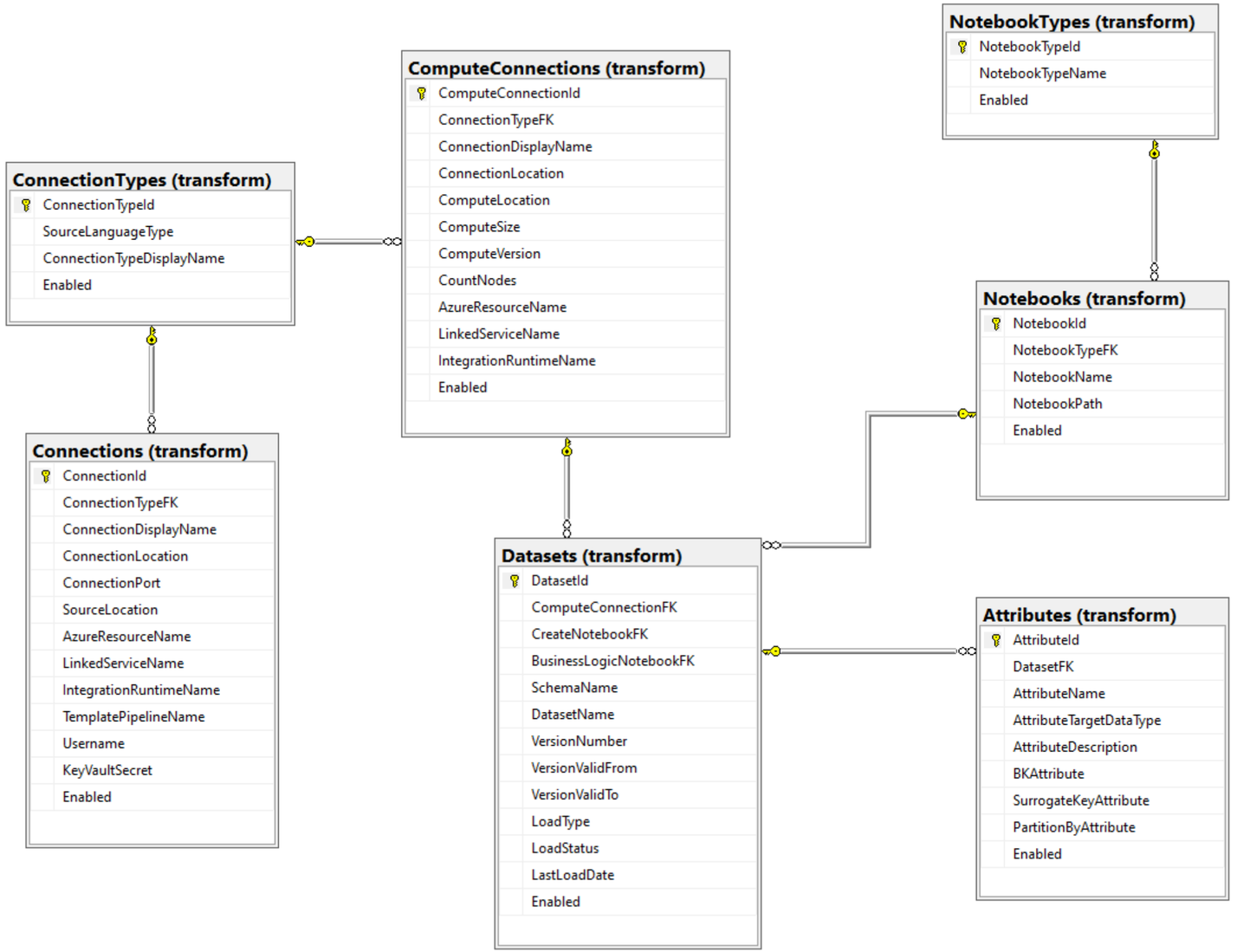


A view on the entities and relationships with CF.Cumulus metadata database. As presented for use by the Ingest pipelines.

Cloud Formations

© 2024 Cloud Formations Ltd. Confidential

CF.Transform Database Internals



A view on the entities and relationships with CF.Cumulus metadata database. As presented for use by the Transform pipelines.

Cloud Formations

© 2024 Cloud Formations Ltd. Confidential

Contact Us

https://cloudformations.org
contactus@cloudformations.org
In/CloudFormations
@CloudFormsLtd
CloudFormationsLtd

bit.ly/cf-meet

© 2024 Cloud Formations Ltd. Confidential