Cumulus Framework

Cloud Formations - Logical Architecture
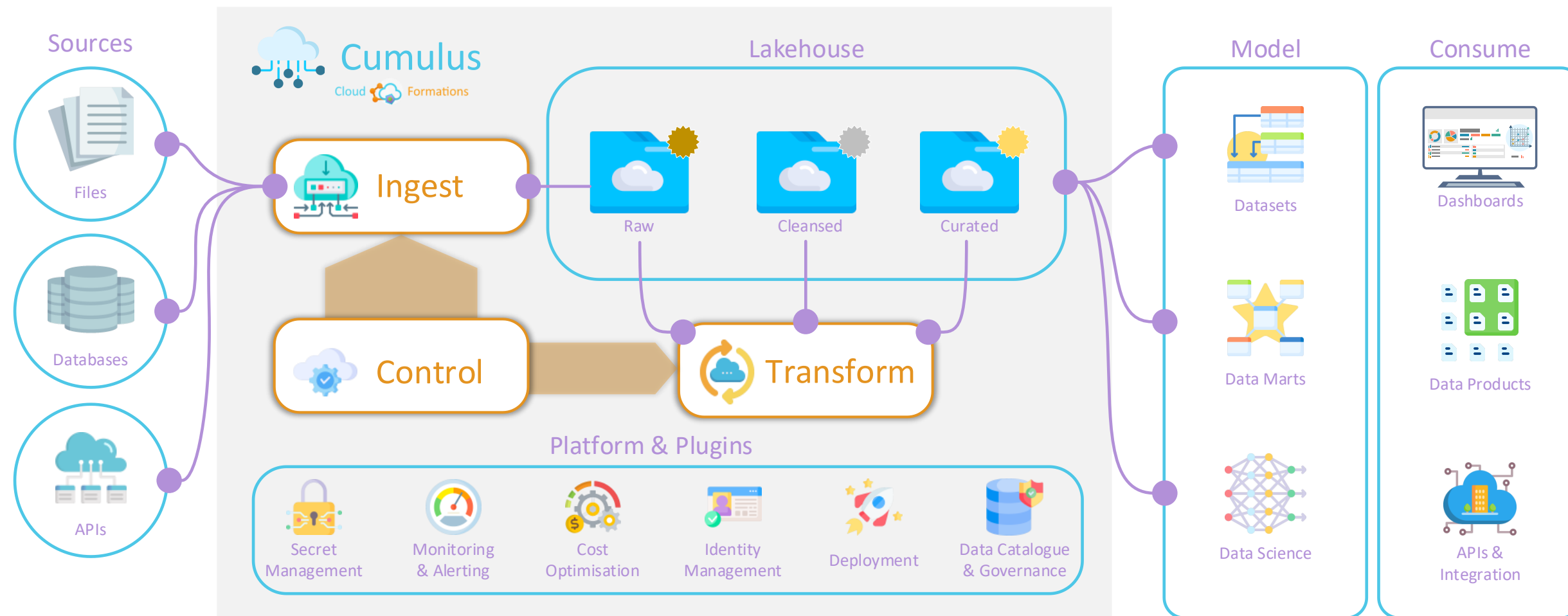
- **Open Source***:
  Zero cost, unlimited access to code
- **Scalable:**
  Horizontal & vertical scaling for Big Data executions
- **Cost-Efficient:**
  Faster time-to-insight reduces delivery costs
- **Extensible:**
  Supports complex & unique data handling
- **Collaborative:**
  Brings development & 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

Diagram

Sources: Files, Databases, APIs

Cumulus — Ingest, Control, Transform

Lakehouse: Raw, Cleansed, Curated

Platform & Plugins: Secret Management, Monitoring & Alerting, Cost Optimisation, Identity Management, Deployment, Data Catalogue & Governance

Model: Datasets, Data Marts, Data Science

Consume: Dashboards, Data Products, APIs & Integration

Underlying Technology: Spark, DELTA LAKE

Works With:

Control

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

Ingest

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

Transform

- <u>Metadata-driven</u> 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

* In accordance with MIT licencing.
† Coming soon.