Data Lake & Warehouse Provisioning

Automate provisioning & orchestration of your data storage and presentation

Overview

Data Lakes & warehouses now function as the heart of business data storage and presentation. Applications & analytics consumption of this data drives business operations, provides the capability to extract value, and ensure business teams can make timely decisions with high levels of accuracy. The quality of the data presentation is paramount to ensuring high value extraction for your business goals. Provisioning and orchestration has traditionally been a manual task and fraught with error and poor orchestration. The solution presented here brings full automation of provisioning and orchestration eliminating errors and providing a high level of consistency to data presentation and consumption.

CI/CD for Data Lakes & Warehouses

DataOps extends the CI/CD pipeline concept to your data resource provisioning process. The CI/CD pipeline brings a consistent and structured approach to the traditionally manual process of provisioning and operating a data lake or warehouse. Similar to application development, you can now utilize pipeline stages to enforce quality & security standards, automate testing & feedback, and automate the deployment process as well.

How It Works

Data teams codify a 'stack' defining data lake or warehouse properties including ingestion sources, storage requirements, preparation, conditioning, transforms, models and presentation or serving settings. Access and security controls are also defined. This template is stored in version control where data teams can collaborate and iterate on configurations. As an example, a new ingestion source can be added by simply adding the parameters to the template. Upon new commits the pipeline is triggered and executes a series of stages to provision or modify the required infrastructure. The pipeline then executes a series of validation tests and provides rapid feedback for data teams. Additional stages of the pipeline can enforce quality and security standards by conducting static analysis of Infrastructure code. Finally, upon successful completion of testing, the pipeline automatically deploys to production.

1. Data Operations teams define a new Data Lake or Warehouse configuration stack including ingestion, storage, transform model attributes as well as access controls and security. Configuration is codified in a YML template.
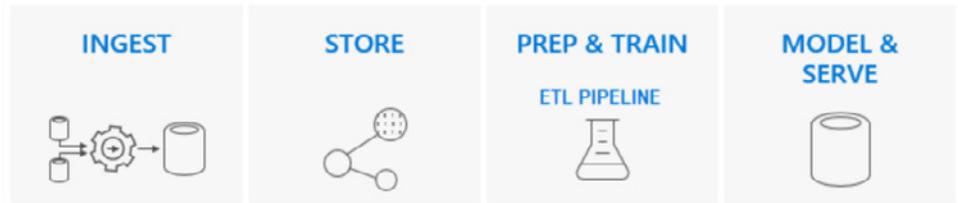
- OR -

2. Data Operations iterate on an existing template to add or update new sources of ingestion, storage changes, or new presentation models. Updates are version controlled.

3. The provisioning pipeline is triggered by commits to the template. It conducts validation checks and full integration checks in a 'clean' environment. Upon passage of tests it promotes to the production environment.

INGEST   **STORE**   **PREP & TRAIN**   **MODEL & SERVE**

ETL PIPELINE

Testing Environment

Production Environment

- ✔ Validate tables/functions follow conventions
- ✔ Check dependencies
- ✔ Validate query syntax
- ✔ Unit tests for complex views
- ✔ Integration testing
- ✔ Security controls

End Result

DataOps brings the concept of Continuous Integration and Continuous Deployment (CI/CD) to your data lake & warehouse provisioning process. Automating provisioning ensures:

- Consistent deployment of infrastructure and configuration of data presentation, access, and consumption models
- Fast deployment of new data ingestion sources and presentation modes
- Enforced access and security controls
- Enforced data lifecycle controls including PII and localization

By eliminating manual processes, an automated provisioning pipeline brings consistency to your data lake & warehouse configuration. This ensures high quality ingestion, transform and presentation of your data sets. Shortening the cycle to add new ingestions sources, transforms, and presentation models brings the capability to develop new insights and extract value from your data at high velocity.

ABOUT MPHASIS STELLIGENT

With over a decade of experience, Mphasis Stelligent is a Premier Amazon Web Service (AWS) Consulting Partner, AWS Public Sector Partner, and holds competencies in DevOps, Security and Financial Services. It has a demonstrated track record in assisting enterprise customers benefit from AWS' continuous innovation. Mphasis Stelligent brings in-depth expertise in DevOps, DevSecOps, and Data/MLOps automation services to enable security-conscious enterprises to focus on developing business-critical software. It uniquely brings a data-driven approach to assess and steamline DevOps maturity and apply proven 'deep automation' techniques to codify and accelerate complex enterprise migration programs for apps and data that is aligned with the AWS Prescriptive Migration Framework. Learn more at www.stelligent.com

aws partner network

Premier

Consulting Partner

DevOps Competency

Financial Services Competency

Public Sector Partner

For more information, contact us at: info@stelligent.com

11710 Plaza America Drive
Suite 2000
Reston, VA 20190-4743
Tel.: +1 888 924 4539

Mphasis
stelligent™

The Next Applied