Top Reasons to Choose Pachyderm

1. Cost-Effective Scalability

◆ Deliver reliable results optimizing resource utilization and maximizing developer efficiency.

◆ Optimize resource utilization with completely automated data-driven pipelines.

◆ Deduplication of data and code saves infrastructure costs.

2. Reproducibility

◆ Ensure reproducibility and compliance

◆ Immutable data lineage and data versioning.

◆ Familiar git-like structure of commits, branches, & repos.

3. Flexibility

◆ Leverage your infrastructure investments.

◆ Run on your existing cloud or on-premises infrastructure.

◆ Run again any type, size, or scale of data in both batch or real-time pipelines.

> ❝Pachyderm is on its way to becoming the next big data infrastructure company. ❞

NAGRAJ KASHYAP
CORPORATE VICE PRESIDENT
MICROSOFT

Microsoft

Select Reference Customers

adarga

AGBIOME®

BCG

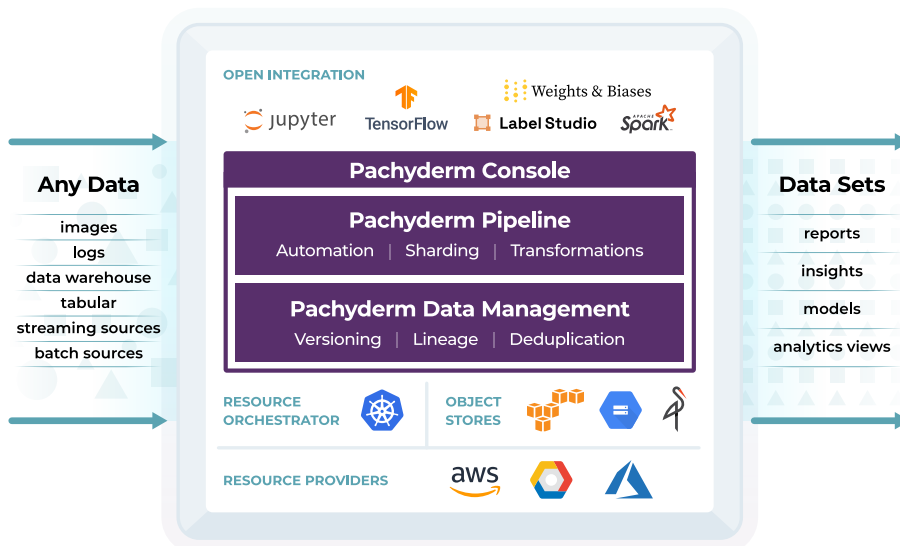Digital Reasoning

Epona

Fraunhofer

generalfusion

LIVEPERSON

LogMeIn

rtl

SEER AI

Any Data
images
logs
data warehouse
tabular
streaming sources
batch sources

OPEN INTEGRATION
Jupyter | TensorFlow | Label Studio | Weights & Biases | Spark

Pachyderm Console

Pachyderm Pipeline
Automation | Sharding | Transformations

Pachyderm Data Management
Versioning | Lineage | Deduplication

RESOURCE ORCHESTRATOR

OBJECT STORES

RESOURCE PROVIDERS
aws

Data Sets
reports
insights
models
analytics views

Pachyderm Products

Pachyderm is available in two major ediitons - Community and Enterprise. Community Edition is available on GitHub and licensed under a developer-friendly open license. The Enterprise Edition is available directly from the company under a commercial license. Both editions enable the building of a robust data-pipelining solution with data versioning and lineage.

Community Edition

Pachyderm Community Edition includes the core platform with data-driven pipelines, versioning, and immutable data lineage. The Community Edition is designed for small teams who prefer to build and support their own software. There are limits on the number of pipelines and parallel workers supported.

Enterprise Edition

Pachyderm Enterprise Edition is designed for organizations and teams that require more advanced administrative features and no limits with world-class support. The Enterprise Edition expands on the features of Community Edition, with pluggable authentication and role-based access controls.

Both editions run across all major cloud providers and on-premises installations.

FEATURES	Community	Enterprise
Console	✔	✔
Notebook Support	✔	✔
Immutable Data Lineage	✔	✔
Data Version Control	✔	✔
Deduplication	✔	✔
Data-Driven Pipelines	16	Unlimited
Parallel Processing (Parallel Workers)	8	Unlimited
Role Based Access Controls (RBAC)		✔
Pluggable Auth – Login with your IdP		✔
Enterprise Support		✔

Key Features

CONSOLE

Console is a complete web UI for visualizing running pipelines and exploring your data.

- Map out the overall structure and flow of all pipelines.

- View repositories, commit histories, and preview data directly in your browser.

- Follow job statuses, pipeline processes, and execution history.

NOTEBOOK

JupyterLab Mount Extension that selectively maps the contents of data repositories right into your Jupyter environment.

- Ideal for Data Scientists to explore and analyze data.

- Run and test pipeline code against versioned data.

- Create reliable, shareable development environments.

Contact Pachyderm

To learn more about the Pachyderm data-driven pipelines, contact us:

info@pachyderm.com
888-338-9597
www.pachyderm.com