

# Top Reasons to Choose Pachyderm

## 1. Automated Data Versioning

### Automated and Performant Way to Keep Track of All Data Changes

Pachyderm utilizes a Git-like structure that enables effective team collaboration through commits, branches and rollbacks. Its powerful content-based deduplication reduces the cost of storing and accessing large data sets, and file-based versioning also provides a complete audit trail for all data and artifacts across pipeline stages including intermediate results. All data is stored as native objects (not metadata pointers) so that versioning is automated and guaranteed.

## 2. Data-Driven Pipelines

### Containerized Pipelines Speed Data Processing while Lowering Compute Costs

Pachyderm's Kubernetes native approach supports any library or language, and autoscales with parallel processing of data, eliminating the need to write any additional code. Automated pipelines execute whenever new data is committed, and incremental processing saves compute by only processing differences and automatically skipping duplicate data.

## 3. Immutable Data Lineage

### Provides an Immutable Record for All Activities and Assets in the ML lifecycle

Pachyderm's immutable data lineage tracks every version of your code, models, and data, while maintaining reproducibility of data and code for compliance. Pachyderm helps manage relationships between historical data states. Pachyderm's Global IDs make it easy for teams to track any result all the way back to its raw input, including all analysis, parameters, code, and intermediate results.

“Pachyderm is on its way to becoming the next big data infrastructure company.”

**NAGRAJ KASHYAP**  
CORPORATE VICE PRESIDENT  
MICROSOFT



### Select Reference Customers

adarga

AGBIOME

BCG

Digital Reasoning

Epona

Fraunhofer

generalfusion

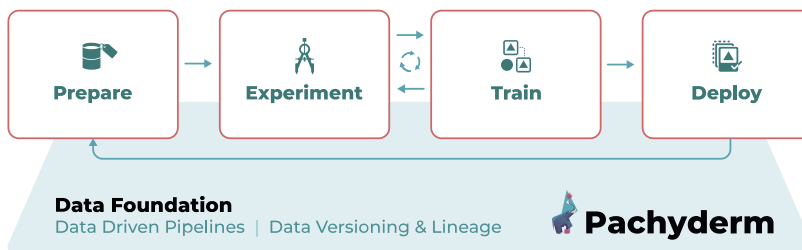
LIVEPERSON

LogMeIn

RTL

## The Leader in Data Versioning and Pipelines

Pachyderm is the data foundation that powers your machine learning lifecycle



## 4. Console

### Provides an Intuitive Visualization of Your DAG and Aids in Reproducibility

Pachyderm enables you to see the overall structure and flow of all your pipelines, while easing pipeline and workflow design, and facilitating collaboration across teams on shared DAGs. Pachyderm helps you drill into pipelines and job details for easy debugging.

## 5. Notebooks

### Pachyderm's JupyterLab Mount Extension provides a point-and-click interface to Pachyderm versioned data

Pachyderm accelerates experimentation with easy and intuitive access to versioned data. Mount any Pachyderm data repository locally for convenient access. It works with versioned data like it's on your own file system — no Pachyderm knowledge required. Explore data with Pachyderm's built in file browser and collaborate across teams with a single source of truth for your data.

## 6. Enterprise Administration

### Robust Set of Tools for Deploying and Administering Pachyderm at Scale

Pachyderm's support for Helm 3 provides robust and standards-based deployment on any public or private cloud. Enterprise Server provides easy centralized licensing and administration of all Pachyderm clusters / workspaces. Pachyderm's pluggable authentication enables you to use any identity provider. Pachyderm's Role Based Access Control (RBAC), allows for fine grained control over access to clusters and data.

## About Pachyderm

Pachyderm is the leader in data versioning and pipelines for MLOps. We provide the data foundation that allows data science teams to automate and scale their machine learning lifecycle while guaranteeing reproducibility.

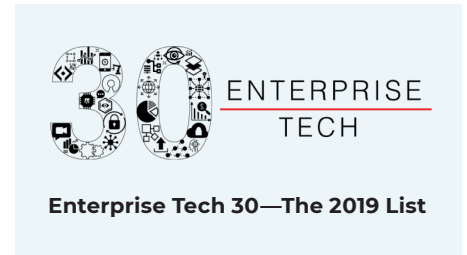
With over \$40 million in three rounds of funding from leading investors like Benchmark, Microsoft M12, Y Combinator, and others, Pachyderm, Inc. offers a commercial Pachyderm Enterprise Edition and an open source Pachyderm Community Edition.

Pachyderm helps customers get their ML and AI projects to market faster, lower data processing and storage costs, and supports strict data governance requirements.

## Contact Pachyderm

To learn more about Pachyderm's machine learning solutions, contact us:

[info@pachyderm.com](mailto:info@pachyderm.com) • 888-338-9597 • [www.pachyderm.com](http://www.pachyderm.com)



“The difference was an order of magnitude faster... if it took 10 hours on the old system then it would only take an hour with Pachyderm.”

GEORGE BONEV, PHD  
MACHINE LEARNING ENGINEER  
LIVEPERSON

