

# Top Reasons to Choose Pachyderm

**Pachyderm**

**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.

**The Leader in Data Versioning and Pipelines**  
Pachyderm is the data foundation that powers your machine learning lifecycle

**Data Foundation**  
Data Driven Pipelines | Data Versioning & Lineage

**Microsoft**  
NAGRAJ KASHYAP  
CORPORATE VICE PRESIDENT  
MICROSOFT

**Select Reference Customers**

- adarga
- ACBIOME
- BCG
- Digital Reasoning
- Epona
- Fraunhofer
- generalfusion
- LIVEPERSON
- LogMeIn
- rTL

© Pachyderm Inc. www.pachyderm.com

Pachyderm is the leader in data versioning and pipelines for MLOps. With a strong foundation of automation and lineage, data science teams can automate the machine learning lifecycle and process data at scale.

With investment from Benchmark, Microsoft M12, and others, Pachyderm offers a full-featured Enterprise Edition with powerful parallelization and management console, and the open source Community Edition.

Pachyderm helps customers get their ML and AI projects to market faster with containerized pipelines and immutable data lineage. Get the executive summary on the top reasons customers choose Pachyderm.