About Pachyderm Enterprise

Despite the many amazing applications of machine learning and artificial intelligence, many attempts at doing “data science” are anything but scientific. Anyone that's tried knows that machine learning workflows are awkward to deploy, hard to maintain, and often cause friction with engineering/IT teams. Even the best implementations out there often lack critical capabilities such as reproducibility, data provenance, and data governance.

Pachyderm lets you deploy and manage multi-stage, language-agnostic data pipelines while maintaining complete reproducibility and provenance. Machine learning pipelines built with Pachyderm make data and results more accessible to your organization without sacrificing functionality or security. Trusted by thousands of data-scientists and engineers, Pachyderm is the ideal platform for anyone looking to make machine learning successful in their organization.

Data Versioning

At last, data gets distributed version control the same way code does. Track the state of your data over time, backtest models on historical data, share data with teammates, and revert to previous states of data, all with one click.

Containerized Data Pipelines

Pachyderm utilizes software containers as the main element of data processing, data scientists and data engineers can use and combine any tools they need for a certain set of analyses or modeling.

Data Provenance

Pachyderm is the only data science platform that provides the ability to track any result all the way back to its raw input data, including all analysis, code, and intermediate results.