

# Containerized Data Science For The Enterprise

Organizations have gone to great lengths and spent millions to protect their data. However, these bastion efforts to preserve the integrity of and to isolate data from being compromised have created internal fortresses that halt machine learning (ML) efforts. Consequently, IT organizations are stuck in a familiar conundrum: protect or progress. As a result, any potential insights in the data to be trapped behind miles of red tape.

Pachyderm provides a third option: empower data scientists and ML engineers to use data as needed in a way that doesn't sacrifice integrity or security. Currently, data is being treated as a single source where the quality degrades with each new user. Pachyderm builds trackable, version-controlled, data pipelines enabling data scientists and ML engineers to leverage data on demand. Essentially, it's git for data.

Data is packaged, deployed, and isolated in Docker containers arranged by Kubernetes. These technologies allow data to inherit the modern production practices like automated deployments, infrastructure independence, scalability, rollbacks, and more. Best of all, IT organizations don't have to refactor, stored differently, or change data to use Pachyderm. Data scientists and ML engineers can cheaply pull data from virtually any source, structured or unstructured.

## Why Pachyderm



Reproducibility



Data Provenance



Collaboration



Incrementality

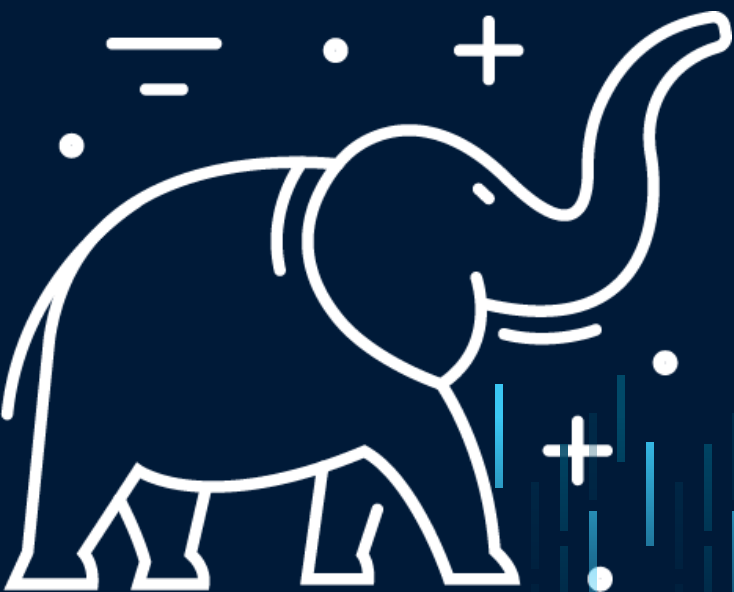


Data Science  
Autonomy



Infrastructure  
Agnostic

The time has come for every organization looking to utilize the potential of ML to come to grips with its impact and challenges. Organizations need to build roadmaps that enable rather than hinder. Pachyderm is the only enterprise-grade solution that eliminates obstacles that translates directly into business results.



Find out more at  
**pachyderm.io**

pachyderm