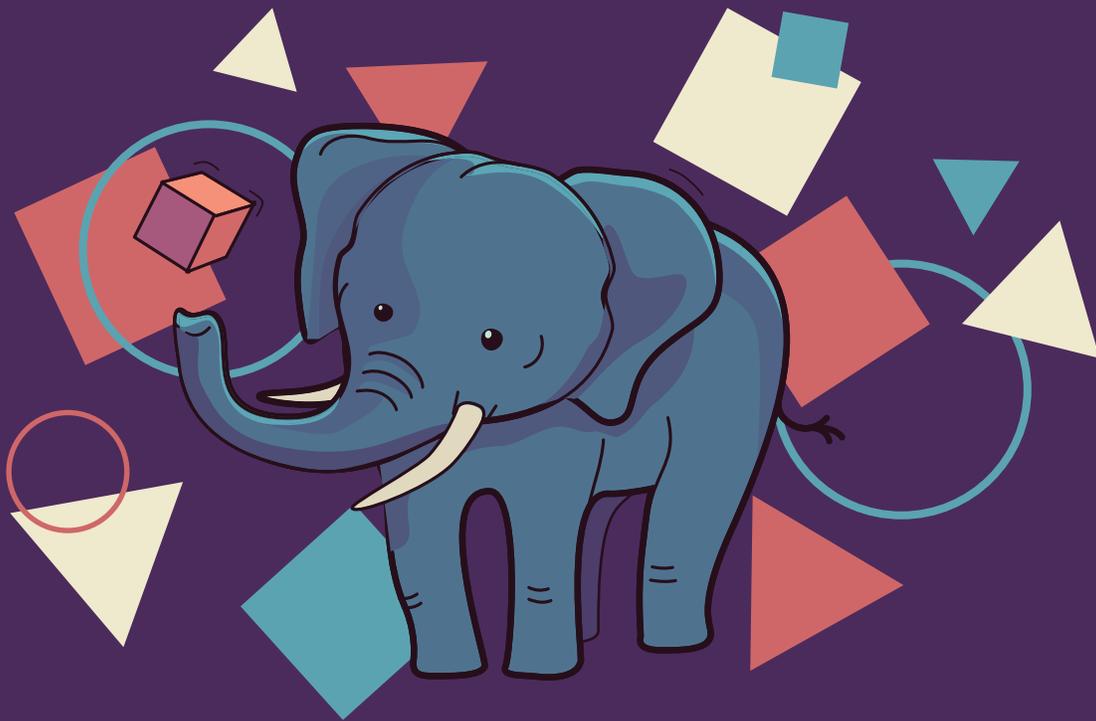# Pachyderm Customer Case Study:
LogMeIn

LogMeIn, Inc. creates innovative, cloud-based SaaS services for Unified Communication & Collaboration, Identity and Access Management, and Customer Engagement and Support. Founded in 2003 and based in Boston, LogMeIn began by building web-based software that gave IT administrators access to remote desktops. They've been a catalyst for reimagining how and where people work, delivering security in a bring-your-own device world and providing customer support and engagement for the digital-first generation.

Since its IPO in 2009, LogMeIn has steadily diversified its offerings to include collaboration and remote meetings with GoToMeeting, as well as everything from remote support for smartphones, tablets, and computers (with LogMeIn Rescue) to password and access management (with LastPass). Their Bold360 system delivers AI-powered engagement to customers, contact centers and employees, bringing together the best of human and chatbot support under a single platform.

"Pachyderm reduced our data processing time from
7 weeks, to just 7 hours"

– Eyal Heldenberg, Voice AI Product Manager LogMeIn AI Center of Excellence

https://www.pachyderm.com/case-studies/logmein/

The Challenge

At the LogMeIn AI Center of Excellence in Israel, the company's team deals with a lot of text, audio, and video that needs to get quickly processed and labeled for its data scientists to go to work delivering machine learning capabilities across their product lines. "Our job at the AI hub is to bring the best-in-class ML models of, in our case, Speech Recognition and NLP," said Eyal Heldenberg, Voice AI Product Manager at the LogMeIn AI Center of Excellence. "It became clearer that the ML cycle was not only training but also included lots of data preparation steps and iterations, and we were changing preparation logic quite often! That lack of parallelization and scale really hurt our ability to get datasets to our researchers so they could get to the real work of testing, training and building models for our products."

"For example, one of our steps is a heavy processing of audio for sort of specific cleaning," said Moshe Abramovitch, LogMeIn Data Science Engineer. "To process only one iteration of all our training data would sum up to seven weeks on the biggest compute machine AWS has to offer — and this is only one step. That means lots of unproductive time for the research team."

"We had started to look for a parallel compute solution that would be friendly with our technology stack and knowledge — Dockers and Kubernetes, Abramovitch continued. "We just wanted things to work without becoming experts in data pipelines."
That's where Pachyderm came into the picture.

Why LogMeIn Chose Pachyderm

Speed and Parallelization

"Pachyderm's parallelism helps us run the transformer at scale. Basically, there is no limit of how many datum transformers we can run at once because as Pachyderm runs on Kubernetes, we can scale up to however much we want," said Abramovitch.

LogMeIn did a small POC at first and realized that instead of taking seven to eight weeks to transform their data, Pachyderm crunched that time down to an amazing seven to ten hours. LogMeIn's research and business teams immediately saw the impact of Pachyderm's speed and scale. "Our models are more accurate, and they are getting to production and to the customer's hands much faster," said Heldenberg. "Once you remove time-wasting, building block-like data preparation, the whole chain is affected by that. If we can go from weeks to hours processing data, it greatly affects everyone. This way we can focus on the fun stuff: the research, manipulating the models and making greater models and better models."

With Pachyderm, LogMeIn has scaled their pipelines tremendously because it can do so much of the work in parallel, without the team having to rewrite its software to take advantage of that parallelization. Pachyderm does the scaling and chunking for them.

"The largest pipeline that we ever ran is around 2,000 or 3,000 containers for a single pipeline," said Abramovitch. "It's something like 15 nodes, and each node has 96 CPUs."

Flexibility

Pachyderm also delivers tremendous flexibility because it's agnostic to the tools data scientists need to get their work done right. LogMeIn uses different ML frameworks like TensorFlow and PyTorch, and also utilizes in-house and open-source toolkits like Kaldi. The LogMeIn team wrote its own pre-processing tools to adjust it to the different frameworks.

"You need to go fast," said Heldenberg. "You need to work with your existing tools, your existing languages, your existing dependencies. You want to invest as little as possible in learning, right? You just need stuff to be processed. And since Pachyderm utilizes really flexible tools like Docker and Kubernetes, it's very democratizing."

Instead of thinking about building a monstrous infrastructure that takes months and months to do, LogMeIn was up and running with Pachyderm in a few days and delivering real impact on the business in weeks, as they reworked their pre-processing to take advantage of its capabilities.

Buisness Impact

When other teams and companies are running into data processing challenges, Heldenberg has some simple advice for them: "First of all, I would recommend they evaluate Pachyderm. I already recommend it to my friends."

"Not everyone on the AI research team understands what Pachyderm does, they just know it's fast and delivers what they need, when they need it," he observed. "That's a good thing because it lets the data science team focus on what it does best — doing research and training models — instead of focusing on the infrastructure. "Everyone knows that Pachyderm is the processing framework, and it will just go fast."

"The fact that we're able to prepare our data so fast helped them to run a lot of training. Prior to using Pachyderm, we thought we'd never be able to execute those training sessions so fast. But because the data preparation process became so short, the research team was able to deliver much faster and create a lot of new models because of it." When LogMeIn researchers come to them now, the AI Center of Excellence team knows what to say: "We'll just do it in Pachyderm."