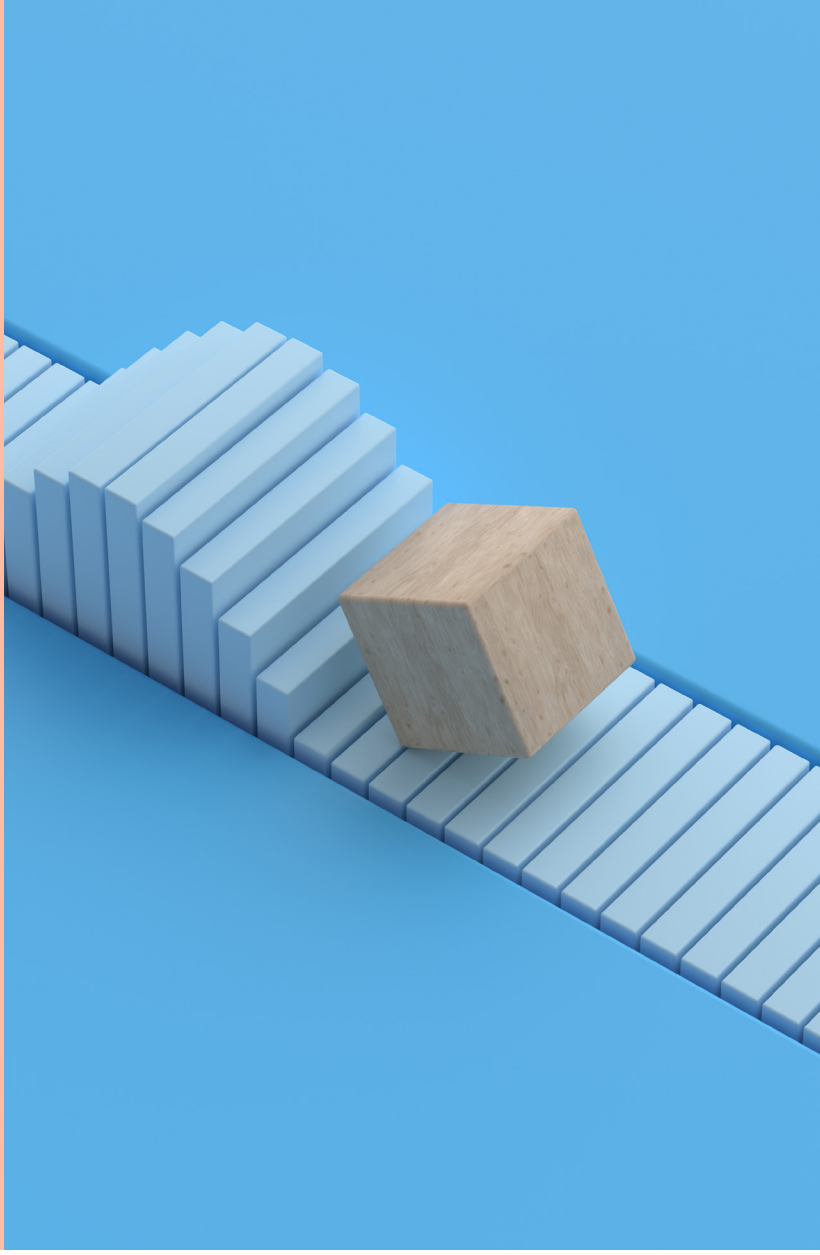
SOLUTION BRIEF

NetApp DataOps Toolkit

Python library for simplified data management



NetApp

The challenge

Artificial intelligence (AI) and machine learning (ML) are changing the game by helping organizations around the world to accelerate innovation and gain a competitive advantage. As organizations increase their use of AI and ML, they face challenges in data management, deployment complexity, and data availability. Many frameworks and toolkits attempt to improve data scalability and ease of data deployment, but most fail to address the crucial challenge of data management and availability. Many also feature proprietary data platforms that lack proven, enterprise-class reliability.

The solution

The NetApp® DataOps Toolkit is a Python library that makes it easy for developers, data scientists, and data engineers to perform numerous data management tasks. These tasks include provisioning a new data volume or development workspace, cloning a data volume or development workspace almost instantaneously, and creating a NetApp Snapshot™ copy of a data volume or development workspace for traceability and baselining. This Python library can function as either a command-line utility or a library of functions that can be imported into any Python program or Jupyter Notebook.

The DataOps Toolkit supports Linux and macOS hosts. The toolkit must be used in conjunction with a NetApp data storage system or service. It simplifies various data management tasks that are executed by the data storage system or service. To facilitate this simplification, the toolkit communicates with the data storage system or service through an API.

Key benefits

Simplify data management

- Quickly and easily provision, clone, and create a Snapshot copy of a data volume.
- Focus on the science instead of on IT.

Streamline AI workflows

- Simplify AI workloads and data science workspace provisioning.
- Drive your AI and ML projects forward faster.

Get the most out of your data

- Trust that your data is always available whenever and wherever it's needed, no matter where it resides.
- Simplify building your DevOps and MLOps pipelines.

The DataOps Toolkit comes in two versions: one for Kubernetes-based environments, and one for traditional virtualized or bare-metal environments. Both versions are compatible with NetApp AFF A-Series all-flash arrays, NetApp FAS arrays, NetApp Cloud Volumes ONTAP® software, and NetApp ONTAP Select software. The Kubernetes version also supports NetApp Cloud Volumes Service, Azure NetApp Files, and NetApp EF-Series all-flash arrays.

The DataOps Toolkit enhances the NetApp AI Control Plane, a full-stack AI data and experiment management reference architecture, by making it easier to invoke data management tasks. A data scientist working within a Jupyter Notebook that was provisioned using the NetApp AI Control Plane can use the DataOps Toolkit to implement a data management task in one simple line of Python code. Likewise, a data engineer can easily invoke a DataOps Toolkit operation as a step within an Apache Airflow or Kubeflow Pipeline automated workflow.

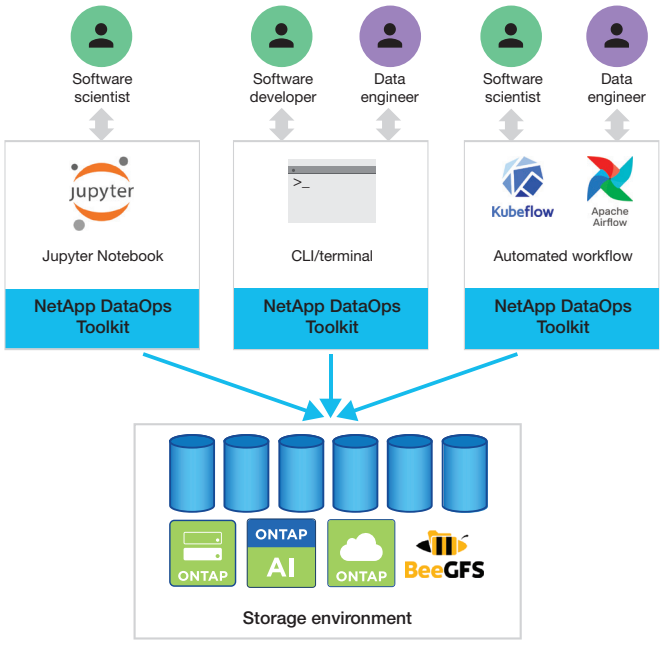


Figure 1: NetApp DataOps Toolkit, simplifying AI data management.

About NetApp

In a world full of generalists, NetApp is a specialist. We're focused on one thing, helping your business get the most out of your data. NetApp brings the enterprise-grade data services you rely on into the cloud, and the simple flexibility of cloud into the data center. Our industry-leading solutions work across diverse customer environments and the world's biggest public clouds.

As a cloud-led, data-centric software company, only NetApp can help build your unique data fabric, simplify and connect your cloud, and securely deliver the right data, services, and applications to the right people—anytime, anywhere. www.netapp.com

NetApp