NSDF-Cloud: Enabling Ad-Hoc Compute Clusters Across Academic and Commercial Clouds

Jakob Luettgau*, Paula Olaya*, Naweiluo Zhou*, Giorgio Scorzelli§,
Valerio Pascucci§, Michela Taufer
*University of Tennessee Knoxville §University of Utah

Motivation

- Computing resources are increasingly made available to researchers through cloud-like interfaces or APIs.
- The diversity of APIs creates unnecessary burden, especially for researchers new to the cloud.
- Even for experts, using the diverse APIs effectively can result in wasting countless person-hours.

Lessons Learned from Observations

- Cloud accessibility and ease-of-use are central concerns that all providers address through different means. Providers often offer new comers friendly web dashboards both to monitor and control resources.
- But most providers introduce their own vocabulary for different services. As such, a common challenge is to identify equivalent services when moving from one cloud platform to another.
- Most providers offer command-line interface (CLI) tools or API access in one form or another, but **no standard for Identity and Access Management (IAM) has yet emerged**, requiring special procedures for each provider.
- **Providers have their own unique sequence of steps** such as setting up security groups or networks **when VMs are launched**.
- Gathering the **public IPs** and injecting **SSH access credentials require significant customization for each provider.**
- Academic clouds provide high-level abstractions that are transparent by offering pathways to access the underlying open-source infrastructure, in case users request it.
- Commercial clouds offer their own stack and often hide details of the underlying infrastructure.

NSDF-Cloud: Toolkit Definition and Core Functions

NSDF-Cloud is an unified toolkit that enables users to create, state list, and delete multiple VMs in ad-hoc clusters across academic and commercial clouds through:

- Parallel creation/deletion of many VMs by using command-line tools;
- Automatic generation of Ansible inventory files; and
- Integration of credentials for multiple providers via configuration file.

NSDF-Cloud APIs

The NSDF-Cloud's unified APIs, both Python and CLI tools, consist of:

```
nsdf-cloud <account> create nodes <prefix> --num <N>
nsdf-cloud <account> get nodes <prefix>
nsdf-cloud <account> delete nodes <prefix>
```

and support credential management of multiple accounts across many providers.

Supported Cloud Providers and their Characteristics

Provider	*Type*	*Credentials*	*Regions*	*Stack*	*Custom Images*
AWS	Commercial	Token+Secret	Yes (Int.)	Custom	Yes
Chameleon	Academic	Token	Yes (US)	CHI on OpenStack	Yes*
CloudLab	Academic	Certificate	Yes (US)	Custom	Yes
Vultr	Commercial	Token+IP-Whitelist	Yes (Int.)	Custom	Yes
JetStream	Academic	Token	Yes (US)	Atmosphere on OpenStack	Yes*

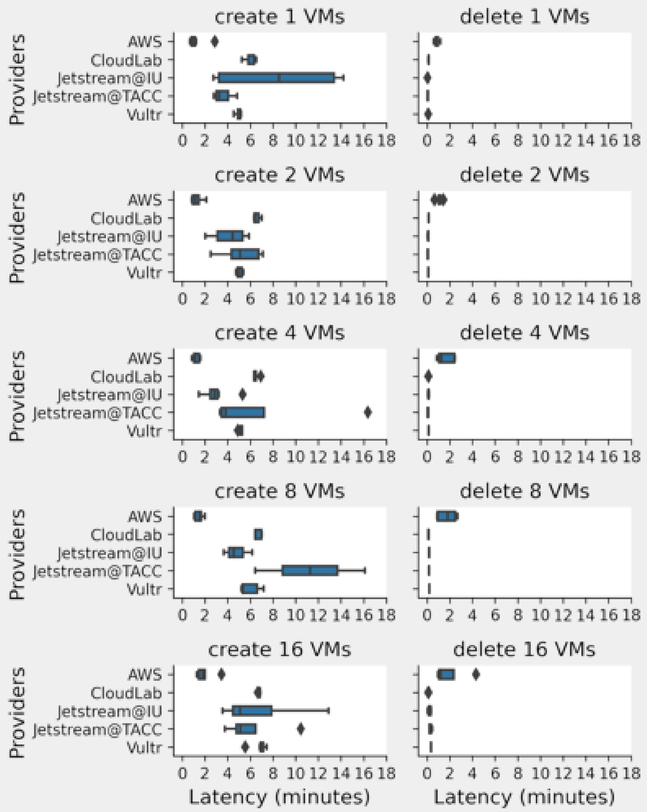
* Provider accepts user provided images but they will be public.

Try out NSDF-Cloud tools on Github or from nationalsciencedatafabric.org

NSDF-Cloud Latency

We measure the NSDF-Cloud latency for:

- Create ad-hoc clusters of up to 16 VMs and SSH them → task performed in < 15 minutes.
- Delete a set of up to 16 VMs in a cluster once an experiment is over → task performed in < 5 minute.



Take-away Message

NSDF-Cloud facilitates users at any entry level in the deployment of the cloud → one single API can generate a cluster of many VMs across multiple providers.

Acknowledgments:
This material is based upon work supported by the National Science Foundation under Grant No. 2138811.This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Funds for experiments on AWS were generously provided by Cloudbank to the NSDF project.