



EOSDIS

NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

Leveraging the Cloud for HDF¹ Software Testing

2019 ESIP Summer Meeting

Larry Knox

EED-2 Software Test Engineer

lrknox@hdfgroup.org

This work was supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C.
This document does not contain technology or Technical Data controlled under either the U.S. International Traffic
in Arms Regulations or the U.S. Export Administration Regulations.

Current in-house HDF test platforms

- IBM ppc64 and SunOS 5.11 sparc servers that allow us to test HDF software on big-endian machines. Big-endian format is used in storing NPP² HDF5 files, and in HDF4 and NetCDF-3³ file formats.
- Several servers running a variety of Windows and CentOS Linux VM⁴s.
- A collection of assorted re-purposed laptops and desktops running MacOS, Windows or CentOS Linux.

Current in-house HDF Daily Testing

- Test system consists of:
 - Shell scripts on Unix (mostly Linux) systems
 - BuildBot master and workers running on Windows and Linux VMs
- We test 3 HDF5 versions, HDF4 and a dozen applications that use these libraries, with more than 750 test configurations.
- Each HDF5 version has about 200 test configurations; running all configurations to test a code revision may take 16 hours over 2 days.

Why leverage the Cloud for HDF software testing?

- Cloud resources are used in conjunction with in-house testing platforms to provide more test coverage and provide more testing time.
- Users report problems with systems and distros we do not have installed in-house.
- Cost of testing in the cloud may be more economical than procuring additional test servers and maintaining them on premises.

More reasons to test in the Cloud

- Testing for issues on multiple distros
- Immediate availability
- No Overhead – pay only when in use
- No system maintenance
- Flexibility
- Consistent environment (for performance testing)

Linux Distributions Tested on AWS⁵

- Amazon Web Services (AWS) Spot instances are being used for testing HDF5 on these Linux Distributions: Amazon Linux, Centos, Debian, Fedora, SUSE and Ubuntu
- BuildBot launches spot instances
- AWS spot instances to run HDF5 build and regression tests in ~30 minutes are available at ~\$.02 per hour

AWS Cost Management

- Spot instances
 - 75% – 85% discount from On-Demand price in us-east-2 (Ohio).
 - Charges for Linux spot instances are based on usage (number of seconds).
 - Prefer instance type suitable for testing that has price history both low and stable. Spot instances are subject to termination; stability will minimize chances of termination.

Test results available on CDash

- BuildBot test results:
<https://cdash-internal.hdfgroup.org/>
- External test results:
<https://cdash.hdfgroup.org/>
- We intend to move all HDF daily tests to BuildBot and Cdash.

Parallel Performance Testing

- Purpose: Track and quantify performance improvements due to code change
- Limit effects often seen on busy systems
- Flexible cluster size
- H5cluster tool for HDF5 installs OrangeFS on NVMe⁶ or SSD⁷ and launches cluster with spot instances

Customized for HDF5 by Steven Varga,
<http://vargaconsulting.ca/>

Acronyms

1. HDF - Hierarchical Data Format
2. NPP - National Polar-orbiting Partnership
3. netCDF-3 – Network Common Data Format version 3
4. VM - Virtual Machine
5. AWS – Amazon Web Services
6. NVMe – non-volatile memory express
7. SSD – solid-state drive

This work was supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C.

Raytheon

in partnership with

