HUF 2024

Monday, September 9, 2024 - Thursday, September 12, 2024

SCC

Book of Abstracts

Contents

20

Restful SSM

Author: Fabi Adams[None]

2024 HUF Presentations by IBM

‼duration 1h

Timeslot:

11

IN2P3 Site Status

Author: Pierre-Emmanuel BRINETTE[1]

[1] *IN2P3 / CNRS*

Corresponding Author: pbrinette@cc.in2p3.fr

Updates on HPSS at IN2P3 Computing Center

- Infrastructure and HPSS upgrade

- ARM architecture support

Timeslot:

15 min

37

HPSS S3 Scalability With Rubin LFA S3 Store Use Case

SLAC National Accelerator Laboratory
Technology and Inovation Department
Scientific Computing Systems

With a growing demand on HPSS S3 support from SLAC science user's community, we eagerly started testing HPSS S3 interface since the pre-GA release in July 2023. From the initial fragile and immature version to today's more robust and resilient state, we worked directly with HPSS S3 developer's team to troubleshoot and triage many challenging issues faced with the scalability, data IO performance and the large and deeply nested data structure handling for very small files from Rubin LFA ceph S3 store use case. In this presentation we'll tell our stories in the journey of bring HPSS S3's capability to a next level.

13

Staging ~2 Million Files from Tape for a User

Author: Geoff Cleary[1]

[1] *LLNL*

Corresponding Author: gcleary@llnl.gov

Imagine you turn on your work laptop or arrive at the office and find this message from the customer service team: "we have a user that is trying to retrieve *many* files from the HPSS archive. At the current retrieval rate, we estimate it will take 6 months for the user to retrieve all the files in the dataset. Can you help?"

What do you do? How do you proceed? What features does HPSS offer to help with this situation?

I'll answer those questions and more as we examine LLNL's approach to retrieving nearly 2 million files across 10's of tape volumes with tools like quaid, SQLite, and RabbitMQ (along with a bit of custom Python code).

Timeslot:

15 min

34

Have I right-sized my disk cache?

Corresponding Author: francisdequenne@lbl.gov

Abstract: For most sites, the HPSS disk cache is a critical component of the HPSS configuration, helping boosting performances of storing and retrieving data from the archive. However, it may be a bit of a black art to assess how big the disk cache should be, especially in environments that have grown over the tears. This talk will present a couple of tools that have been developed at NERSC, that allow us to assess the effectiveness of an existing cache, and give some insight on the impact of increasing or decreasing that disk cache size.

10

JAXA Site Report

Authors: Naoyuki FUJITA[1]; Kazushige SATO[2]

[1] *Japan Aerospace Exploration Agency(JAXA)*

[2] *Kyndryl Japan*

Corresponding Author: fujita.naoyuki@jaxa.jp

The recent operation status of JAXA HPSS "J-SPACE", the plans and issues for its replacement in 2025, and monitoring functionality will be reported.

Timeslot:

15 min

7

Implementing a Virtualized HPSS Deployment for Testing and Development

Author: Forrest Greenwood[1]

[1] *HPSS Subsriber*

Corresponding Author: fgreenwo@iu.edu

As part of our efforts to upgrade our site to HPSS 10.3, Indiana University recently began development of a virtualized HPSS environment that we can use to quickly iterate on testing and development initiatives without tying up limited bare-metal hardware resources. This virtual-machine environment is patterned on the implementation created by IBM's HPSS Support team for use at the recent HPSS 10.3 Training held in May 2024.

Topics to be discussed include provisioning the VM, installing and configuring a virtual tape library using the mhVTL open-source software package, installing and configuring both DB2 and HPSS, and possibly a sampling of the sorts of issues we intend to test using this environment.

If technical affordances permit, this presentation could potentially include a live demonstration of the VM running on an external flash drive. Otherwise, we would be happy to present using the traditional static PowerPoint.

Timeslot:

30 min

5

NeRSC Site Report

Author: Owen James[1]

[1] *NERSC/LBNL*

Corresponding Author: o1james@lbl.gov

NERSC Site Report HUF24
Karlsruhe Institute of Technology Campus Nord from 09-12 September 2024

Abstract

Topics to discuss

- NERSC Stats - PBs, etc

- Upgrade HPSS 7.4.3 to 9.3 New RHEL8 Core Servers. New FS7300 metadata arrays. Update existing movers to RHEL8. Updated PAM auth module to work with NERSC Auth.

- Install 4th TS4500 tape library 16 Frame 1188+ slots. testing 10.0.1 firmware with SSL for REST over Ethernet. TS1160 drives JE media, while we figure out TS1170/JF. Total Theoretical capacity 950PB on JE, 2.37EB on JF media.

- Issues deploying TS1170 in our air cooled environment

- Monitoring update (brief, specific talk to follow)

- REST over ethernet testing

Timeslot:

30 min

25

HPSS monitoring at KIT

Author: Preslav Konstantinov[1]

[1] *KIT*

Preslav

Timeslot:

30 min

8

Transitioning HPSS Monitoring from Nagios to VictoriaMetrics

Author: Basil Lalli[1]

[1] *NERSC - LBNL*

Corresponding Author: bdlalli@lbl.gov

Last year at NERSC we retired our long-standing nagios-based HPSS monitoring deployment in favor of VictoriaMetrics, Loki and Alertmanager. We would like to share our experience and lessons learned on the way.

- Motivation for making this transition
    - Limitations of Nagios-style monitoring
    - How does VictoriaMetrics address these?
- General overview of our monitoring deployment
    - 3rd party exporters
    - Custom exporters/"plugins"
- Demonstration of some of the dashboards we use and alerts we generate.
- Future areas of improvement
    - Standardizing our HPSS-specific data collection
    - Service discovery

Timeslot:

15 min

24

KIT's Site report

Author: Dorin Lobontu[None]

Dorin

Timeslot:

30 min

4

Managing Data Throughout Its Lifecycle: Lessons Learned and Future Directions

Authors: Charles McClary[1]; Forrest Greenwood[2]

[1] *HPSS Subscriber*

[2] *HPSS Subsriber*

Corresponding Authors: cmcclary@iu.edu, fgreenwo@iu.edu

Abstract: Data lifecycle management poses significant challenges, particularly in academic and research environments where data accumulation is rapid and perpetual. This presentation delves into the complexities surrounding data retention and abandonment, highlighting the prevalent issues of data hoarding and the lack of structured deletion policies. Specifically, it addresses the dilemma wherein users, especially researchers, find little incentive to delete data, leading to a cluttered and often inaccessible data landscape. Furthermore, the departure of users from institutions like Indiana University (IU) exacerbates the problem, as data may be left behind with no clear ownership or accessibility.

Indiana University is tackling these issues gradually. We'll discuss our efforts to address data management and abandonment through:

New usage constraints: Instituting new quotas with tiered growth guidelines.

Simplified Archiving and Movement: Providing user-friendly tools to archive and migrate data to appropriate storage tiers.

Data Management Education: Empowering researchers with best practices for data stewardship.

Insuring allocation value: Requiring annual renewal of desired resources.

The "Digital Will" Concept: Developing a system where departing users can designate data inheritors and define deletion policies.

By examining the successes and pitfalls of these initiatives, this presentation provides valuable insights into effective data lifecycle management strategies. It underscores the importance of fostering a culture of responsible data stewardship while leveraging technological innovations to facilitate seamless data management throughout its lifecycle.

Timeslot:

45 min

16

Upcoming HPSS Features (Restricted)

Author: Michael Meseke [None]

2024 HUF Presentations by IBM

‼ duration: 1,5h

Timeslot:

15

HPSS Release Roadmap (Restricted)

Author: Michael Meseke[None]

Timeslot:

45 min

9

MPCDF Site Report

Author: Manuel Panea[1]

[1] *Max Planck Computing and Data Facility*

We will present our activities with HPSS since the last HUF, including our upgrade to HPSS 10.3.

Timeslot:

30 min

35

SSC Site Report

SSC Site Report will be a recap of previous HUF presentations and focus on :

- Solution overview
- Review of components throughout upgrades
- HPNLS (High Performance Nearline Storage) architecture
- HPSS and RHEL Upgrade
- HPSS monitoring

- User tools and environment

30

Introduction to GridKa Tour

Timeslot:

33

Exploring storage technologies for HPSS disk caches

At KIT we operate HPSS as a tape system for the GridKa WLCG Tier-1 and for the Baden-Württemberg Data Archive service. Performance limitations of the HPSS disk cache systems led us to explore new technology options for the disk cache, based on classic storage systems with SSDs, and storage servers with local NVMe devices. We will present details on the different possible solutions, including benchmarks.

18

Burning Issues (Restricted)

Author: Jonathan Procknow[None]

2024 HUF Presentations by IBM

Timeslot:

45 min

17

Support Update

Author: Jonathan Procknow[None]

Timeslot:

30 min

36

Closing HUF 2024

Corresponding Author: doris.ressmann@kit.edu

26

Introduction to HUF 2024

Corresponding Author: doris.ressmann@kit.edu

32

Spectra Logic

23

Welcome

Author: Achim Streit[1]

[1] *KIT-SCC*

Welcome

Timeslot:

15 min

6

Testing HPSS S3 Interface at MPCDF

Author: Elena Summer[1]

[1] *Max Planck Computing and Data Facility (MPCDF)*

Corresponding Author: elena.summer@mpcdf.mpg.de

Starting from version 10.3, HPSS has an S3 interface. We at MPCDF have installed it on our test system to try it out in several usage scenarios including cloud sync - using Ceph Cloud Sync module as well as rclone, generating presigned URLs, and just using different S3 clients. Among our test actions, we are trying out put, get, remove S3 objects as well as getting S3 objects from tape. This talk covers test scenarios, their setup and results, encountered issues and their fixes.

Timeslot:

15 min

21

HPSS Object Storage Class Deep Dive

Author: Greg Thorsness[None]

2024 HUF Presentations by IBM

‼duration 1h

Timeslot:

19

Generative AI and HPSS

Author: Greg Thorsness[None]

2024 HUF Presentations by IBM
‼ duration 1h

Timeslot:

3

HPSS Deployment and Support (Restricted)

Authors: Thuy Tran[1]; Aaron Watson[1]

[1] *IBM*

Corresponding Authors: thuytran@us.ibm.com, watsonaa@us.ibm.com

New deployment(s), completed upgrades, site statistics.

Timeslot:

30 min

14

HPSS Core Servers on Commodity Hardware or: How We Learned to Love Databases on ZFS

Author: Herb Wartens[None]

Corresponding Author: wartens2@llnl.gov

At LLNL we have been using commodity hardware more and more to serve our parallel filesystems and archival storage clusters. We wanted to explore how to use this same hardware for our HPSS Core Server systems. In order to make the system as reliable as possible, ZFS emerged as the underlying filesystem we wanted to utilize for its reliability and other advanced features. How would traditional databases perform on top of ZFS? Could we design a production-worthy system using this hardware?

Timeslot:

45 min

40

BOF Client Interfaces

22

GHI Backup and Restore

Author: Jonathan Procknow[None]

2024 HUF Presentations by IBM

!!duration 1h

Timeslot:

27

Registration

28

GridKa Tour 1

Timeslot:

29

GridKa Tour 2

Timeslot:

31

ZKM Tour

38

BOF Monitoring

39

Group Foto