Best Practices in Managing Tens of Petabytes of Data

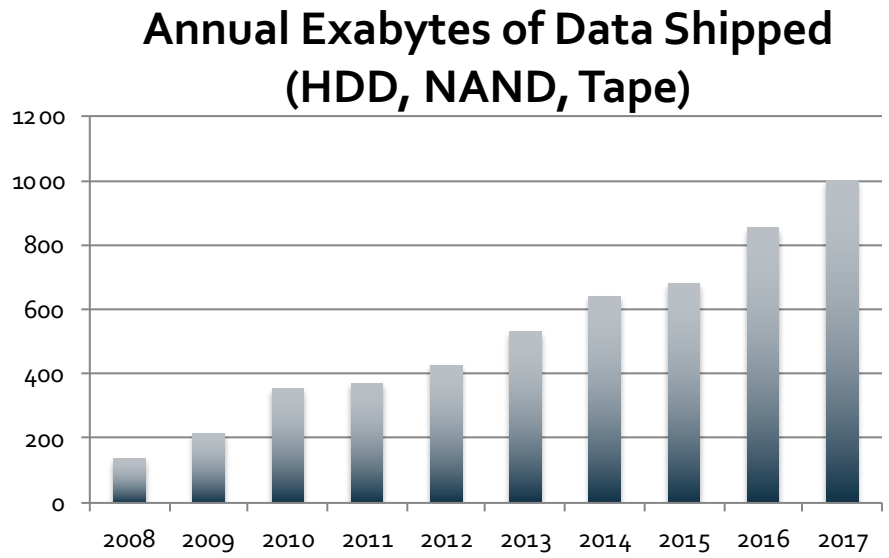
Bill Anderson
National Center for Atmospheric Research
HPCSYSPROS 2018

Acknowledgements

- Pam Hill
- Irfan Elahi
- Marc Genty
- Chris Hoffman
- Zach Mance
- Joey Mendoza

Introduction

- Challenges in managing large amounts of data:
  - Data growth continues



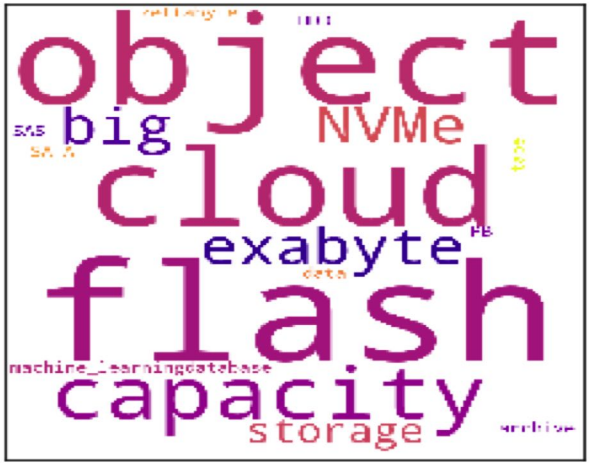
* Data Source: Fontana, R., Decad, G. 2018. A Ten Year Storage Landscape LTO Tape Media, HDD, NAND, *34th International Conference on Mass Storage Systems and Technologies (May 2018).*

Introduction

- Challenges in managing large amounts of data:
  - Stakes are high
    - Security break-ins can result in loss of data (Code Spaces went out of business)
    - Firmware/software bugs
    - Administrator error

Introduction

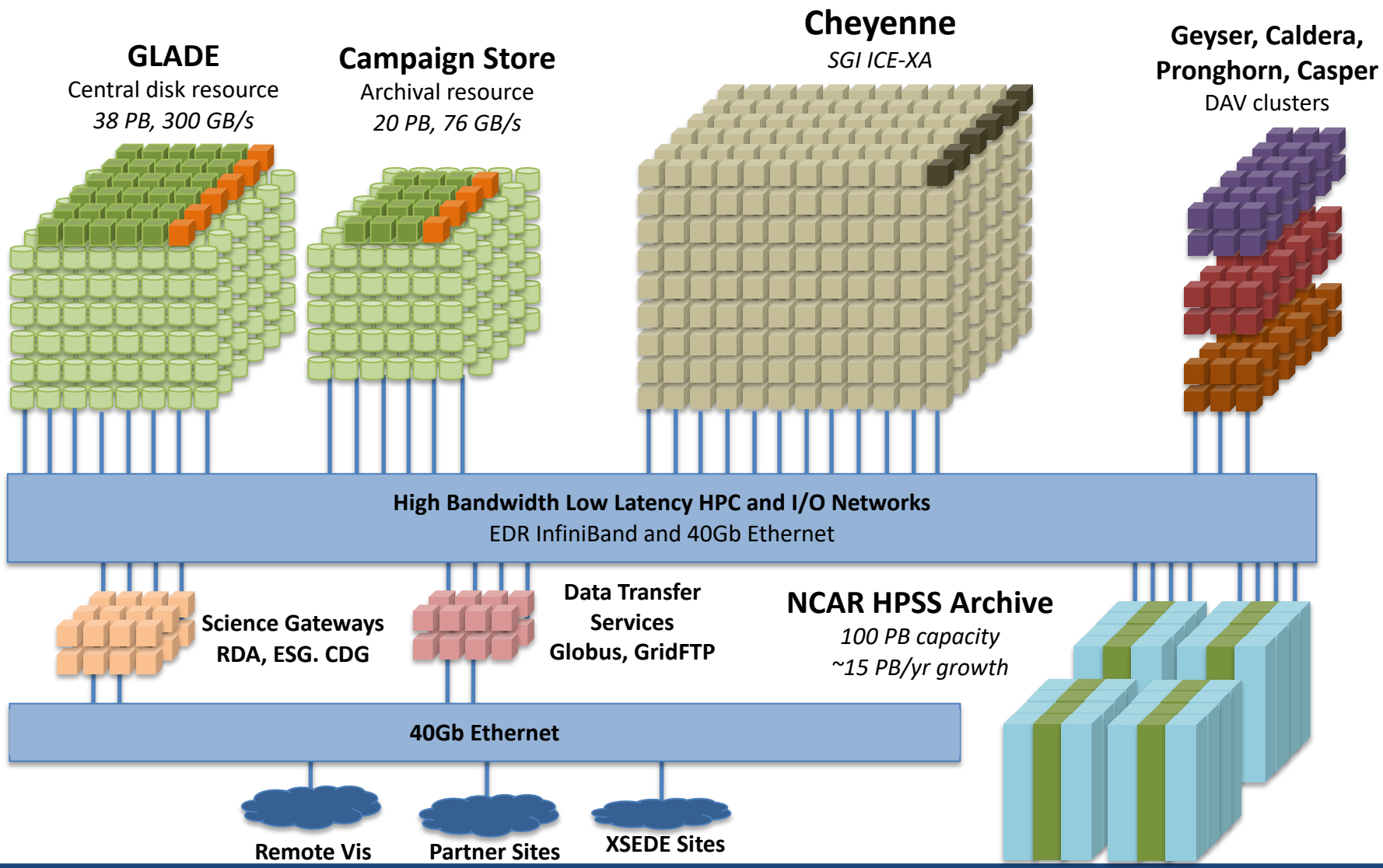
- Challenges in managing large amounts of data:
  - Technology is changing rapidly



Introduction

- National Center for Atmospheric Research (NCAR) has over 100 PB of data storage in its HPC environment
- Have managed large amounts of data for several decades
- Share our experience and best practices we've learned over the years

HPC Architecture



Outline

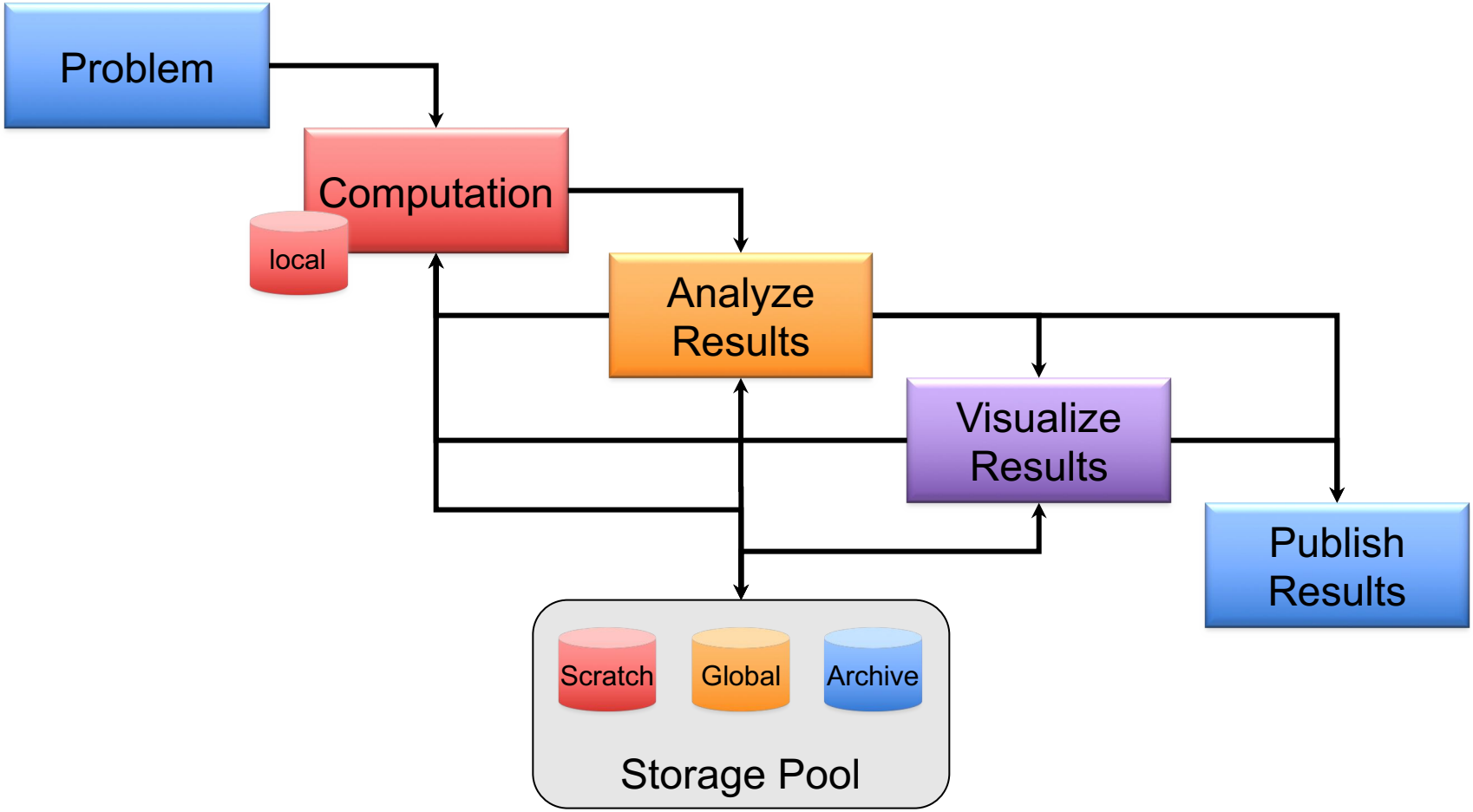
- User Needs and Workflows
- Architecture
- Performance
- Data Integrity
- On-going Maintenance and Support
- Issues when Scaling Systems
- The Future

User Needs and Workflows

User Needs

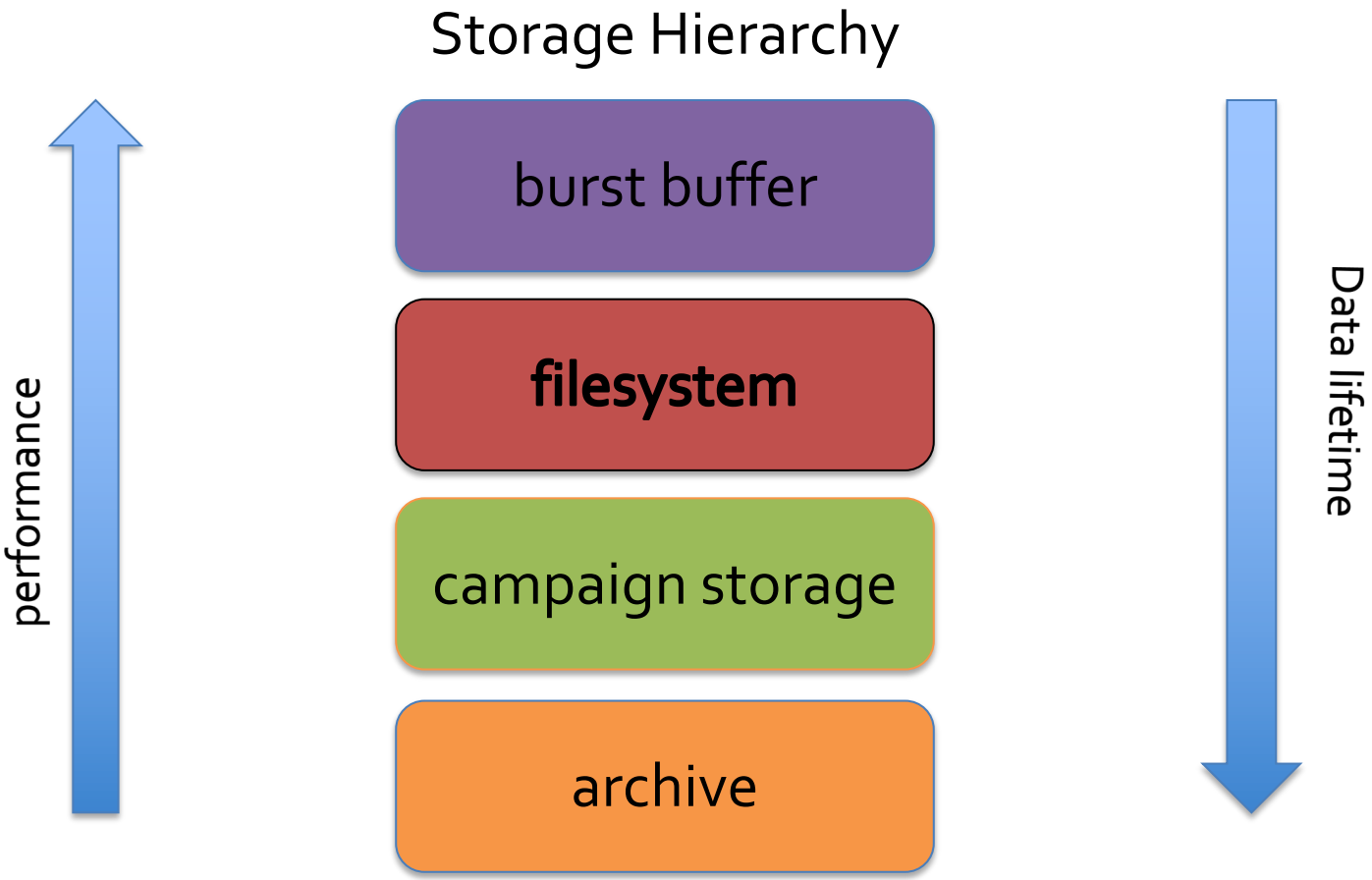
- Understanding user needs is critical, especially for very large systems
- Don't want to spend millions of dollars on something that isn't used
- Some Key Needs:
  - Capacity
  - Data lifetime(s)
  - Performance
  - User interface

User Needs: Workflow



Architecture

Architecture



Architecture

- Interoperability
  - Will everything work together well?
- What are other sites doing?
  - Worth engaging with others

Architecture

- Performance is a key aspect of storage architecture
- How do you architect a system to meet specific performance requirements?
- How do you know if a system is performing as expected?

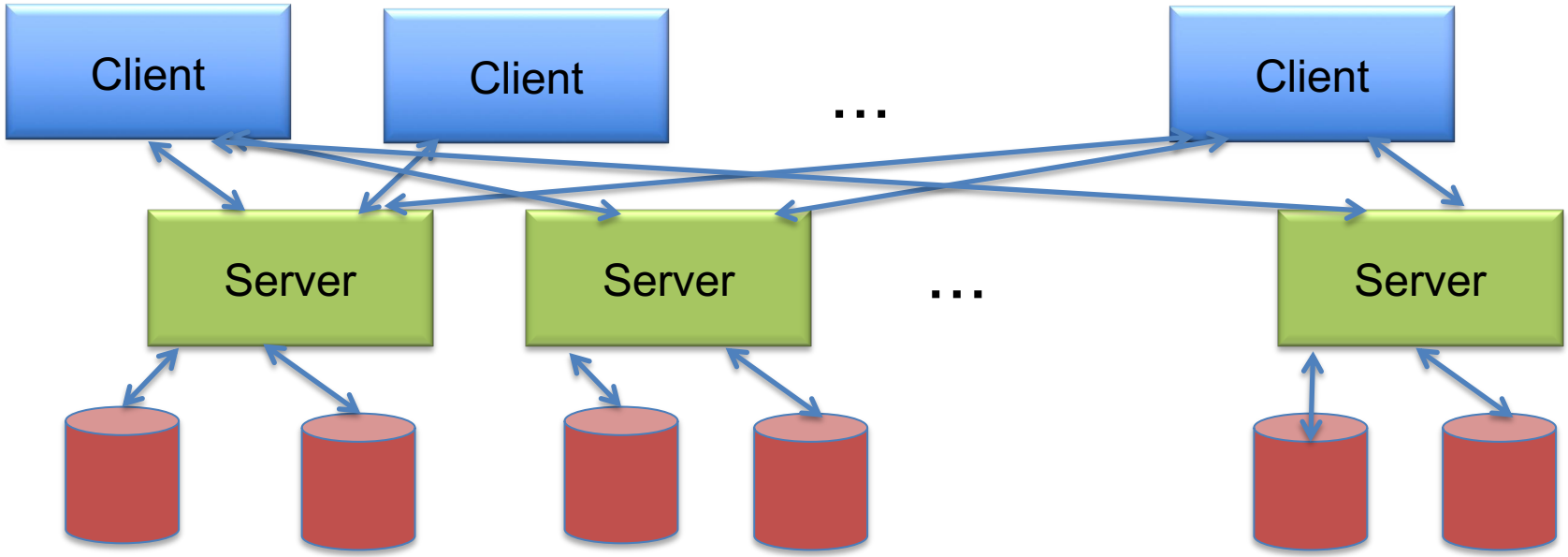
Performance

Performance

- How do you architect a system to meet specific performance requirements?
  - Assumes you have requirements
  - Options:
    - Take advantage of vendor knowledge
    - Use specs for different components to match performance and check overall throughput or IOPS
    - Benchmarks: **specific, measurable, reproducible**
    - All Three

Performance

- Benchmark should be as close as possible to real workload
- Real workload on storage systems can be very complex



Performance

- I/O patterns from multiple processes that change over time can be hard to capture in a benchmark
- Need to approximate
- At NCAR: sustained bandwidth for data transfer and sustained metadata operations per second

Performance

- Benchmark tool examples: *IOR*, *mdtest*

- Can re-run benchmark after upgrades or other changes to ensure that performance still meets requirements

Performance

- Storage benchmarks and tests have helped NCAR over the years:
- We found that one parallel file system had very different performance if tests were run on certain client nodes
- As part of one large supercomputer procurement, found that the parallel filesystem performance was significantly less than expected
- Found data corruption in one vendor's filesystem

Performance

- Common causes of storage performance problems:
  - Too many copies of data being made in the data path
  - Inefficient use of storage system components, including networks (e.g., small I/O request sizes)

Data Integrity

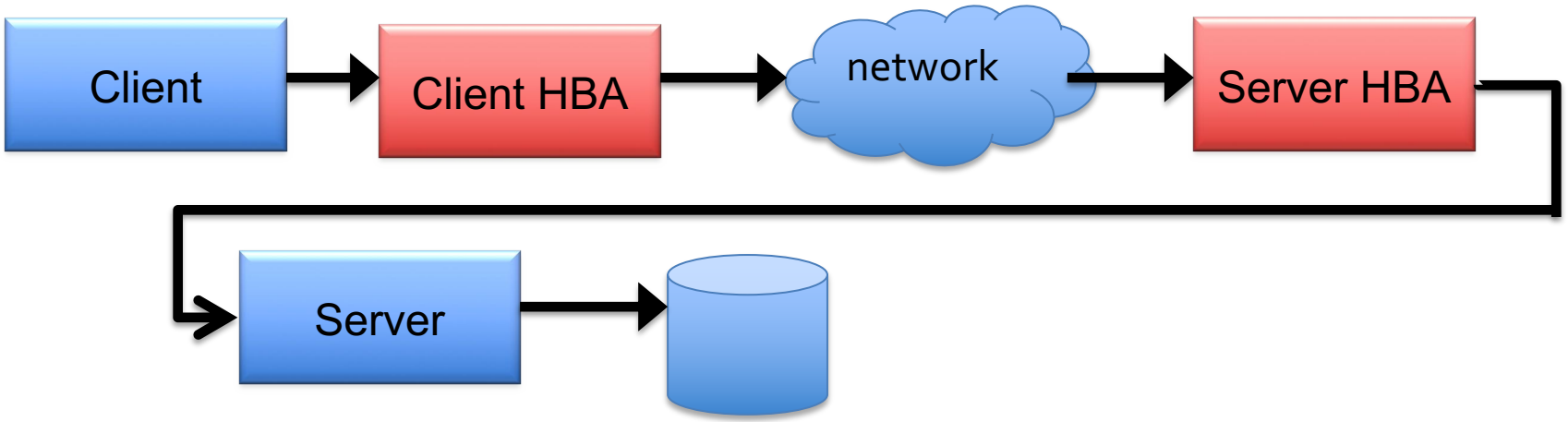
Data Integrity

- Data integrity means that the data read from the system matches what was written to it

- Increased probability of error with larger storage systems

Data Integrity

- Lots of places where data can be corrupted



- Possibility of bit rot

Data Integrity

- Most components have some form of error detection or error correction
- Disks have ways to detect and correct some errors
- Ethernet and TCP have checksums to detect errors
- Some filesystems, like ZFS, use checksums on every data block

Data Integrity

- Yet, errors can still occur
- At one site, bad firmware on network switch caused data to be corrupted
- At NCAR, flaky HBA caused data corruption (data pattern dependent)

Data Integrity

- "Gold standard" is end-to-end data integrity protection
- Can use cryptographic checksum (e.g., MD5, SHA1/2) for each file to achieve end-to-end data protection
- Calculated and verified on client side (T10-PI)
- T10-PI or build your own

On-Going Maintenance and Support

On-Going Maintenance and Support

- Be aware that firmware upgrades can be risky
- Have heard of cases where sites lost filesystems or data

On-Going Maintenance and Support

- Read release notes carefully; will the new firmware really fix a problem you have?

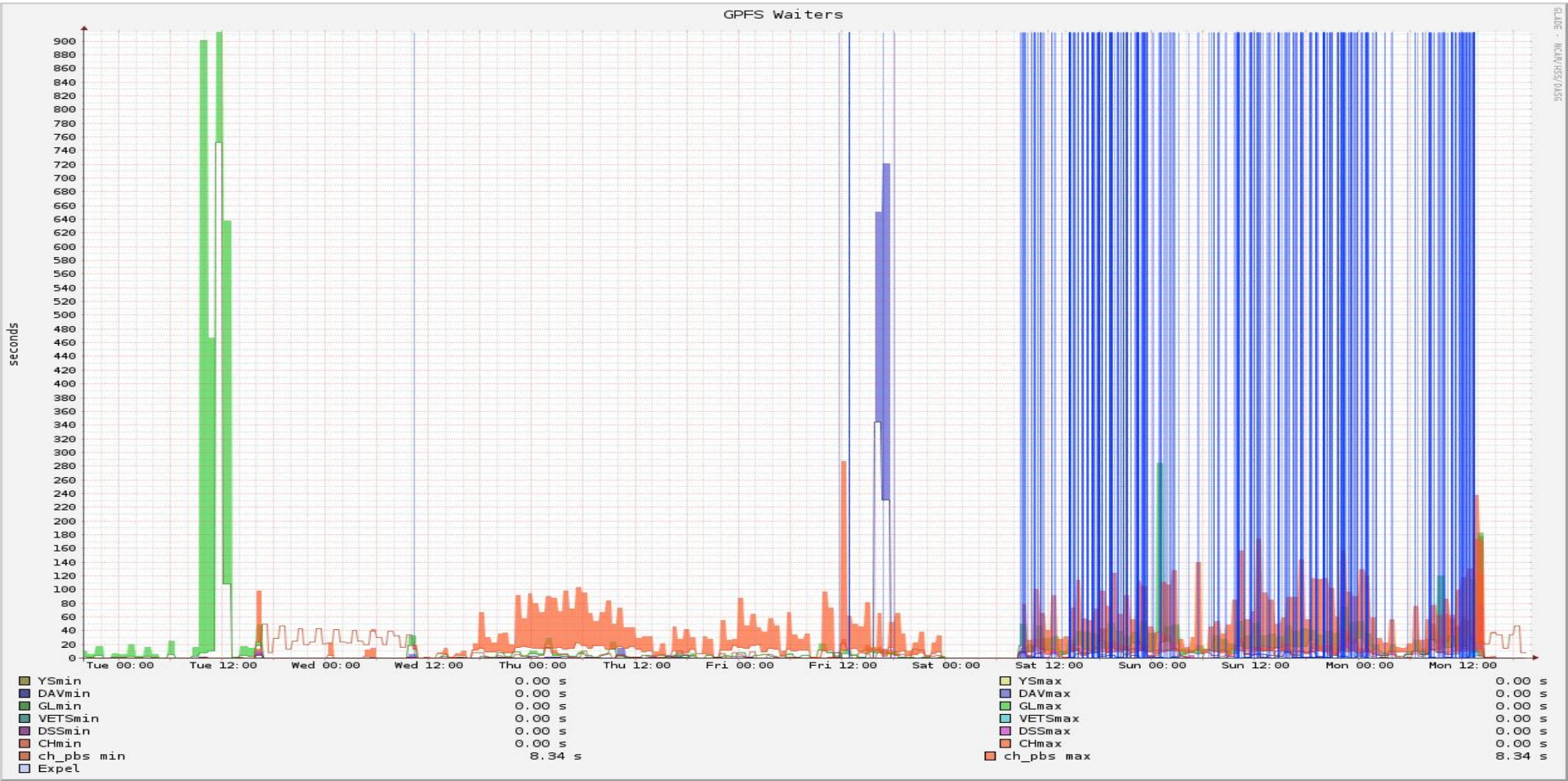
- If need to upgrade:
  - Install on test system
  - Wait until firmware has been out in the field for a few months, if possible

On-Going Maintenance and Support

- Proactive error monitoring – important to identify minor problems before they become major
- We use Nagios – hundreds of checks:
  - Filesystem approaching 90% full
  - Critical backup failed
  - NTP died (needed for kerberos)
- Have detected numerous minor problems and corrected them before they became major

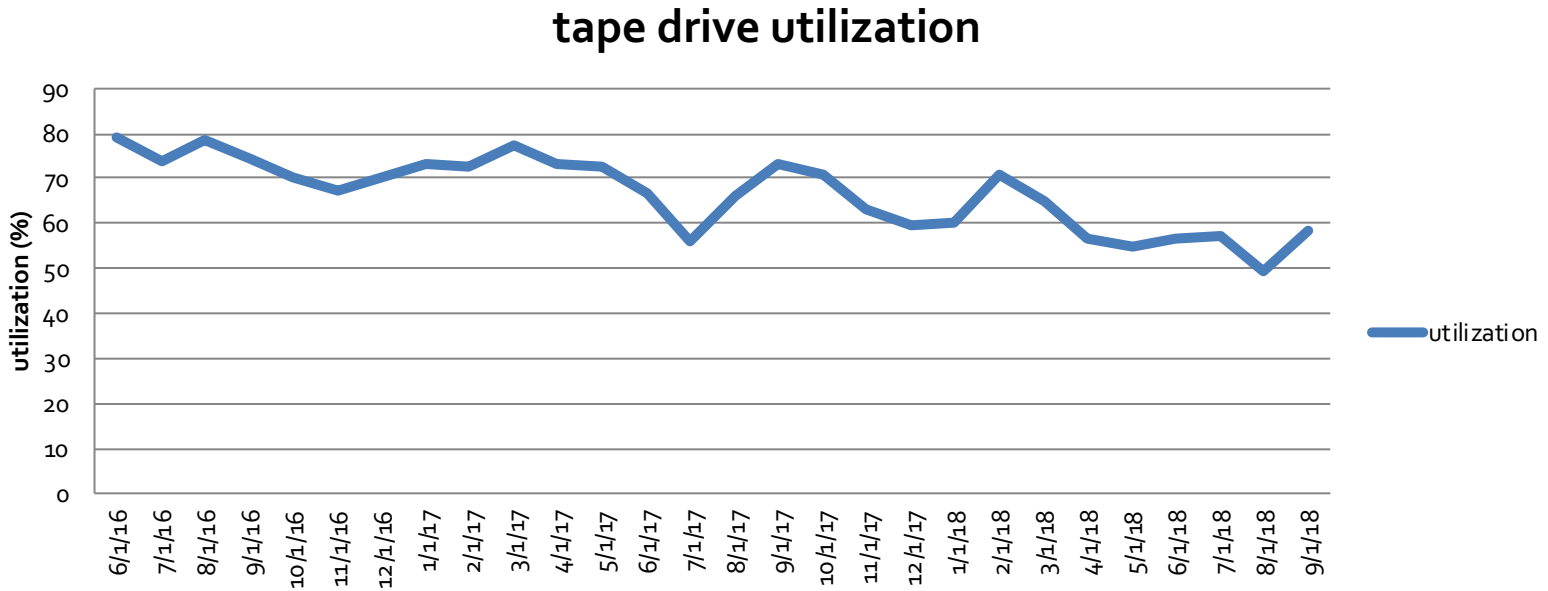
On-Going Maintenance and Support

- Track system metrics



On-Going Maintenance and Support

- Metrics are a longer term investment that can pay big dividends down the road (e.g., for capacity planning)



Issues When Scaling Systems

Issues when Scaling Systems

- *Need to have a full stack test system*
  - Serving many more users; can't have large system down for days to troubleshoot a complex upgrade procedure
  - Need to be able to test patches; a bug in a new version of software caused us to lose a test filesystem rather than the production filesystem

Issues when Scaling Systems

- Plan for a lot more automation
  - Not possible to run commands manually on large numbers of nodes
  - Can't look through 1000's of messages files
  - Can't just run *sar* on hundreds or more nodes and make sense of the output

Issues when Scaling Systems

- Problems scale too
  - More problems
  - Also, new types of problems show up:
    - Software and hardware often aren't tested as much at scale
    - New software (needed to manage larger systems) can lead to new problems

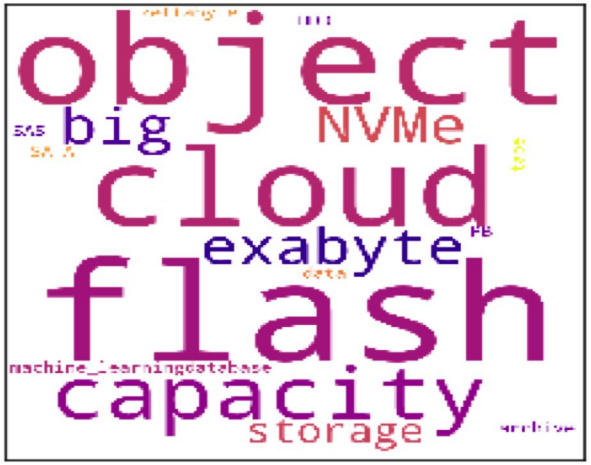
Issues when Scaling Systems

- Coordinating with Others
  - Relying much more on teammates and colleagues
  - Won't be plugging in network cables yourself
  - Security team
  - Operations group
  - Tighter relationship with vendor

The Future

The Future

- "It's tough to make predictions, especially about the future" – Yogi Berra



The Future

- May not be able to predict the future, but can prepare for different possibilities

- NCAR set up an HPC futures lab; great way to explore and test new storage technologies on premises and in the cloud

Summary

- Understand user needs
- Engage with other sites and the community
- Performance benchmarks are worthwhile
- Seek end-to-end data integrity
- Be careful with firmware upgrades
- Track system metrics
- Test and explore latest technologies