# Introduction to Storage challenges in the HPC World

Tom Langborg
Storage expert
NSC
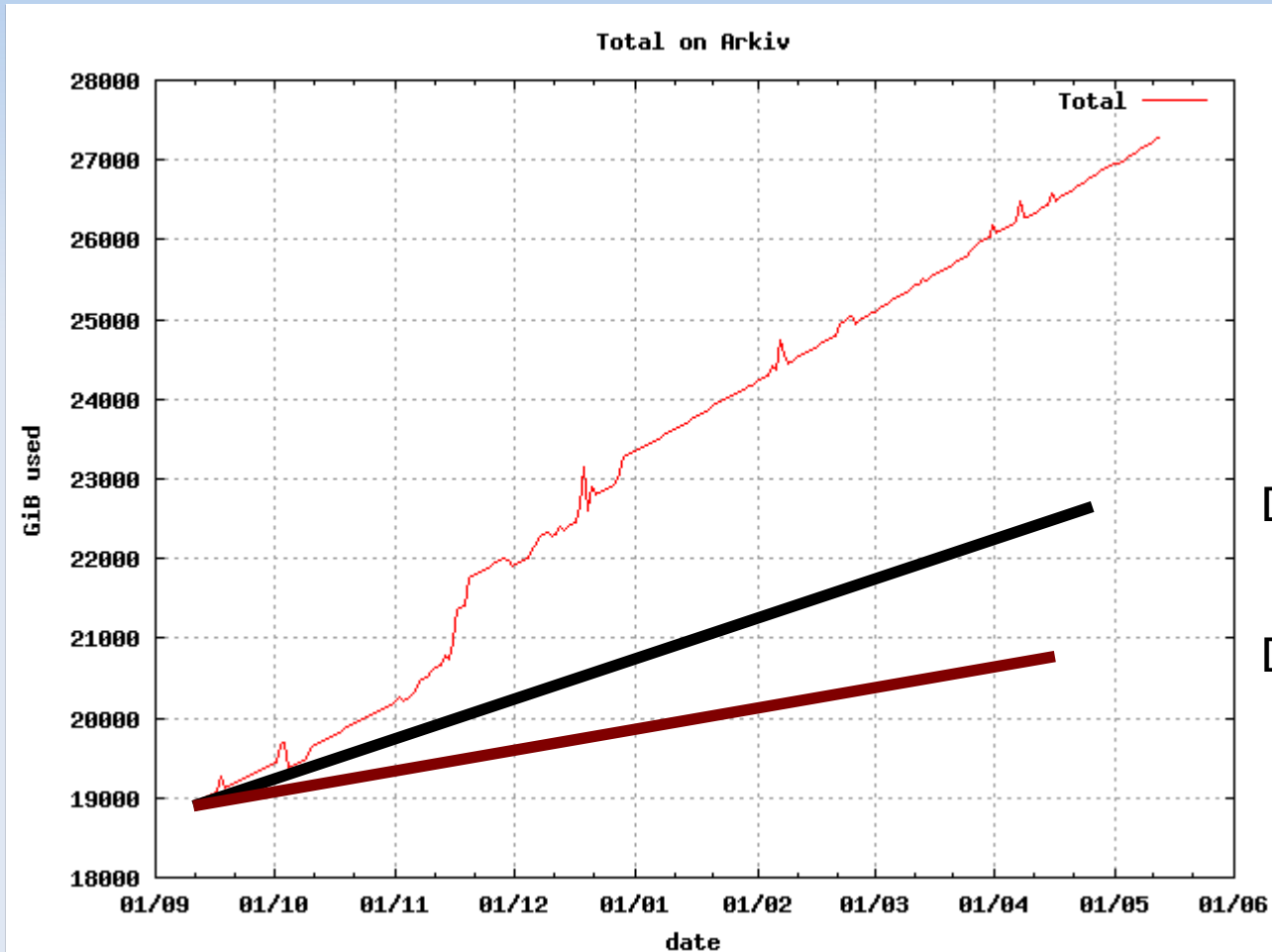Linköping University
Tom@nsc.liu.se

# Challenge

Two types of challenges

volume and speed

# Problem

# More problems

- 100% I/O utilization
  - Rebuild disk
  - Data strip
- Network bandwidth
- Secure data
  - Secure from disaster
  - Secure from user errors
- Secure data transmissions
  - Long time
  - HW errors like silent error on write

# And more problems

- Stable platform
  - Migrate to new hw, even servers
- Management
  - Ownership
  - Search able
- Special filsystem
  - Need modified kernel on clients
- HSM
  - Meta data handling
  - Move data to tape require one more tape robot

# Swedish National Infrastructure for Computing

- Coordinate and develop high end computing capacity for Swedish research.

-  Provide long term funding for high performance computing resources in Sweden

- Coordinate national investments in hardware

- Have a development mission for:

  - Computers

  - Grid

  - Storage

  - Network

# What have we done in SNIC?

- Create a storage group. (2008)
- Work as link between centra and as a storage expert group for SNIC's board.
- Evaluate storage project or storage needs
- Implement and maintain storage infrastructure
- Assimilation and nomenclature definitions

# Type of storage

- **Cluster storage**
  Short-term storage (days)
  Need speed

- **Centre storage**
  Middle-term storage (months)
  Need speed and volume

- **National storage**
  Long-term storage (years)
  Need volume and management

# Cluster storage

- Speed is important

- Storage time is days

- Common to use clusterfile system as GPFS, Lustre...

- Files are personal.

- Volume are about ten times memory.

# Centre storage

- Speed and volume are important
- Virtualize hw
- Storage time is months
- Common to use cluster filesystem and NFS
- Files are mixed personal and project own.
- Volumes are very depending on projects
- Can have access from multiple clusters

# National storage

- Volume and management is important

- Virtualize filsystem

- Storage time is years

- Common to use some type virtualize NDGF's solution with dCache,

- Files are now own by project

# Advice

- Don't mix different type of data
- Work together
- Use same definitions
- Help users with longterm storage
- Handel (some) storage as infrastructure

# Final word

- Take care of your storage problem now!! Next year is your data twice so much