David Ramírez Alvarez

HPC INTEGRATOR MANAGER

WWW.SIE.ES

dramirez@sie.es

ADMINTECH 2016

BeeGFS

Solid, fast and made in Europe

www.beegfs.com

Thanks to Sven for info!!!

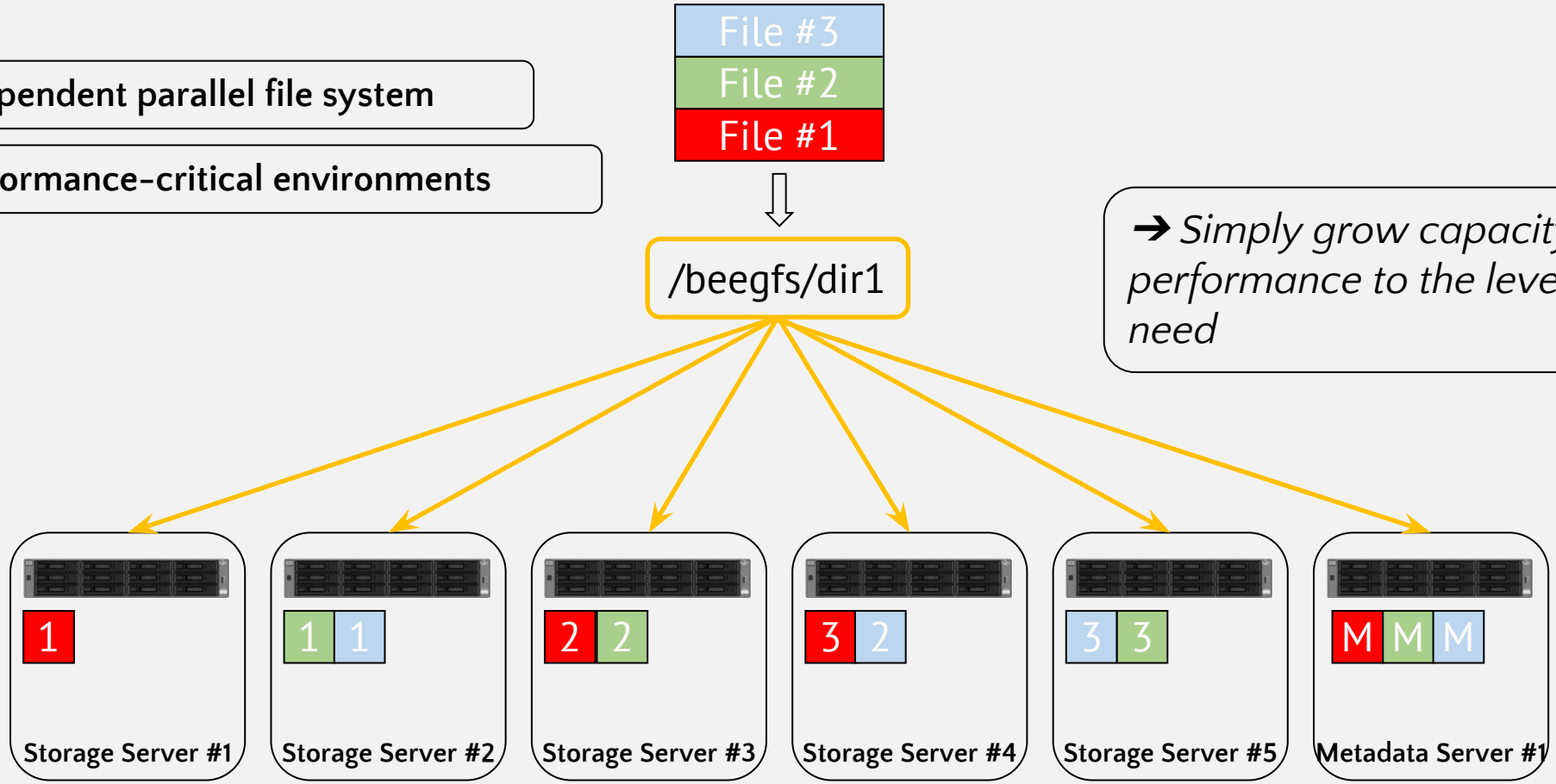
February 2016 | Sven Breuner, CEO, ThinkParQ

What is BeeGFS?

BeeGFS is...

...a hardware-independent parallel file system

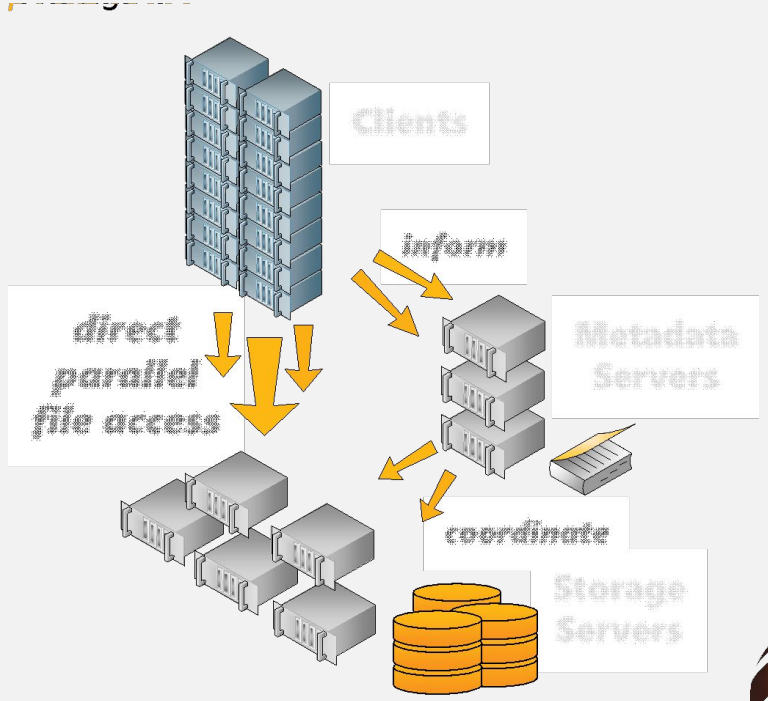
...designed for performance-critical environments



➔ *Simply grow capacity and performance to the level that you need*

BeeGFS Architecture

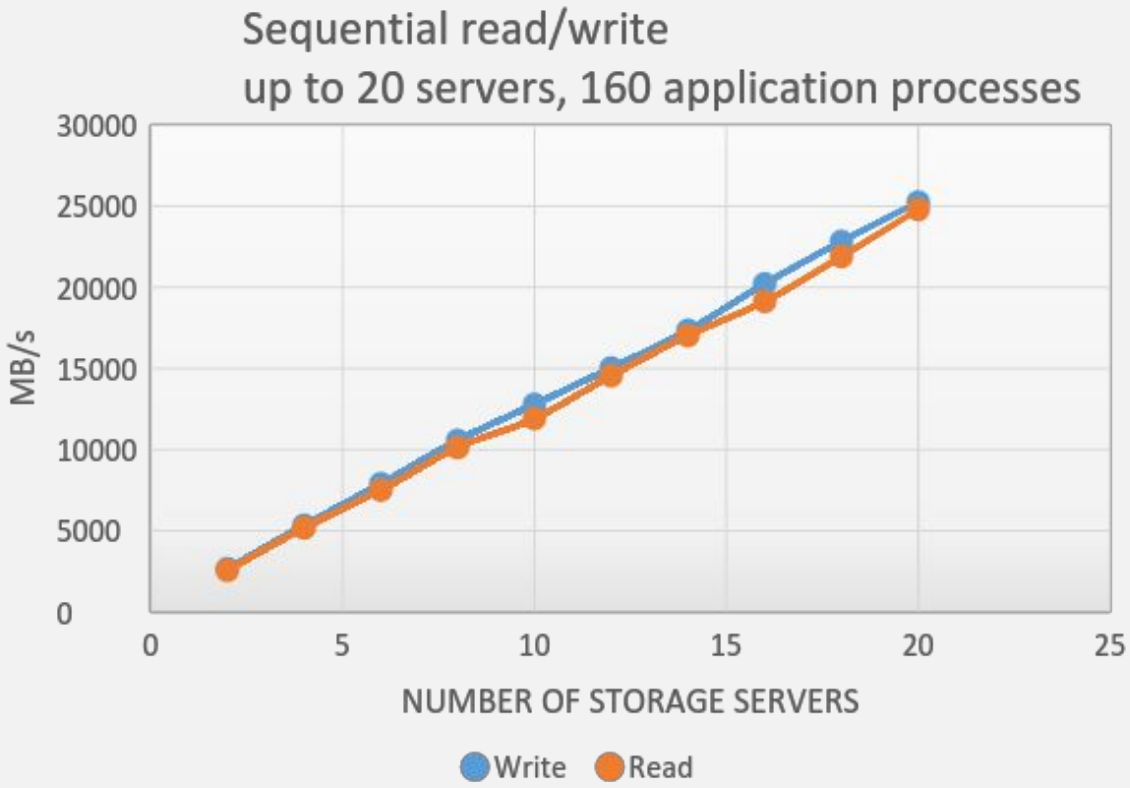
- Client
  - Native Linux module to mount the file system
- Storage Service
  - Store the (distributed) file contents
- Metadata Service
  - Maintain striping information for files
  - Not involved in data access between file open/close
- Management Service
  - Glue everything together and watch services
- Graphical Administration and Monitoring System
  - GUI to perform administrative tasks and monitor system information
    - Can be used for “Windows-style installation“

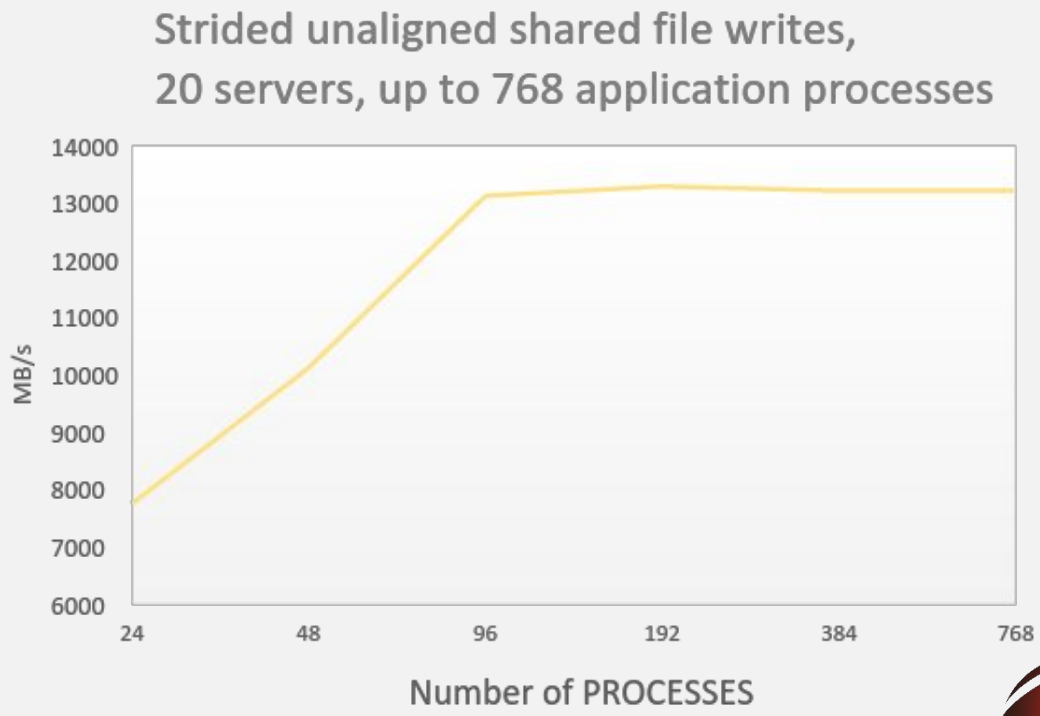


Key Aspects

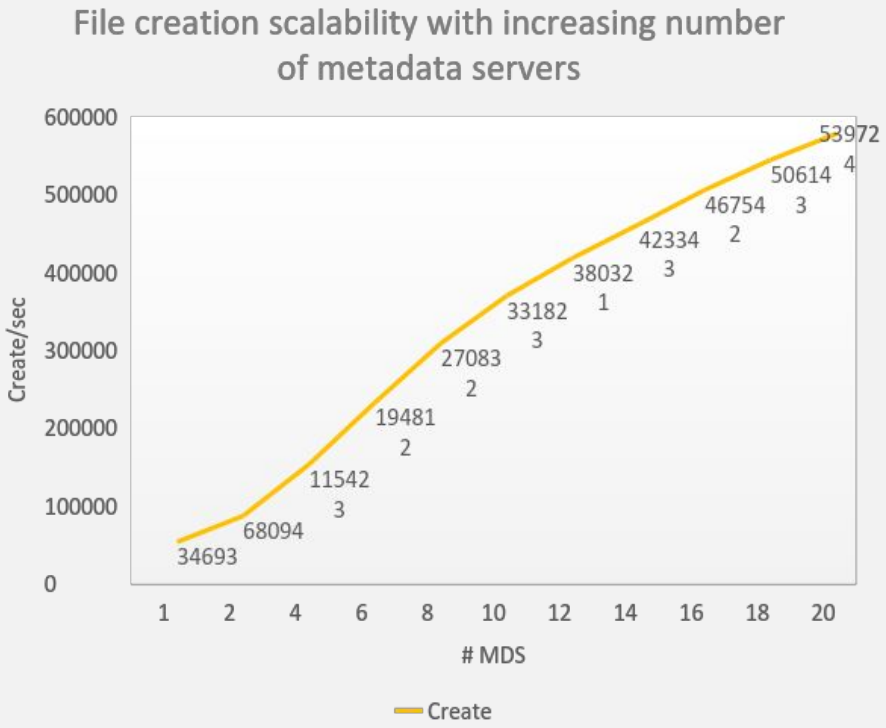
- **Performance & Scalability**
  - Initially optimized for HPC
  - Completely multi-threaded – lightweight design
  - Supports RDMA/RoCE and TCP (Infiniband, 40/10/1GbE, ...)
  - Distributed file contents:
    aggregated throughput of multiple servers
  - Distributed metadata across multiple servers
  - High single stream performance (multiple GB/s)

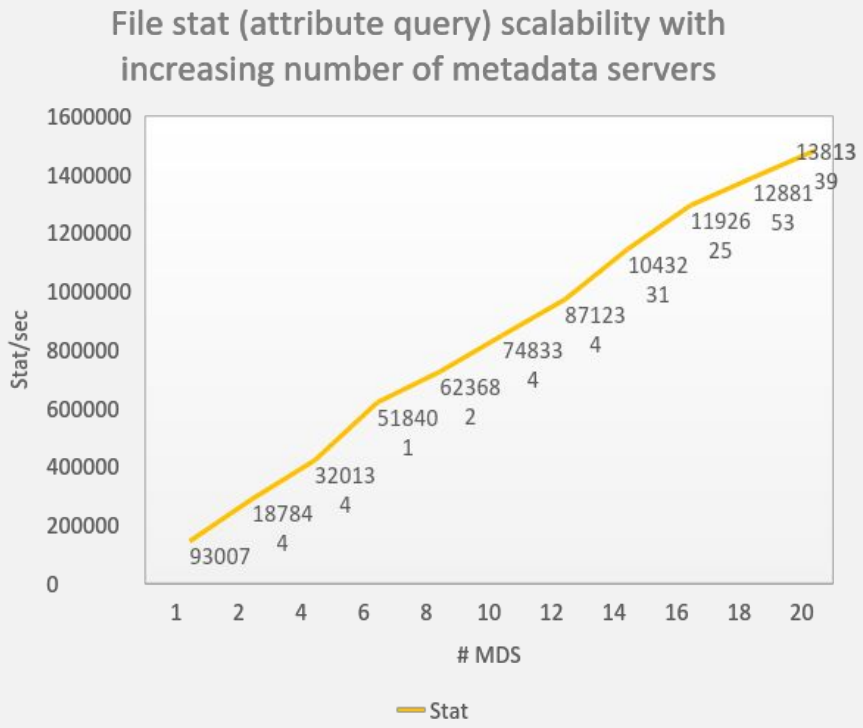
Throughput Scalability





Metadata Scalability





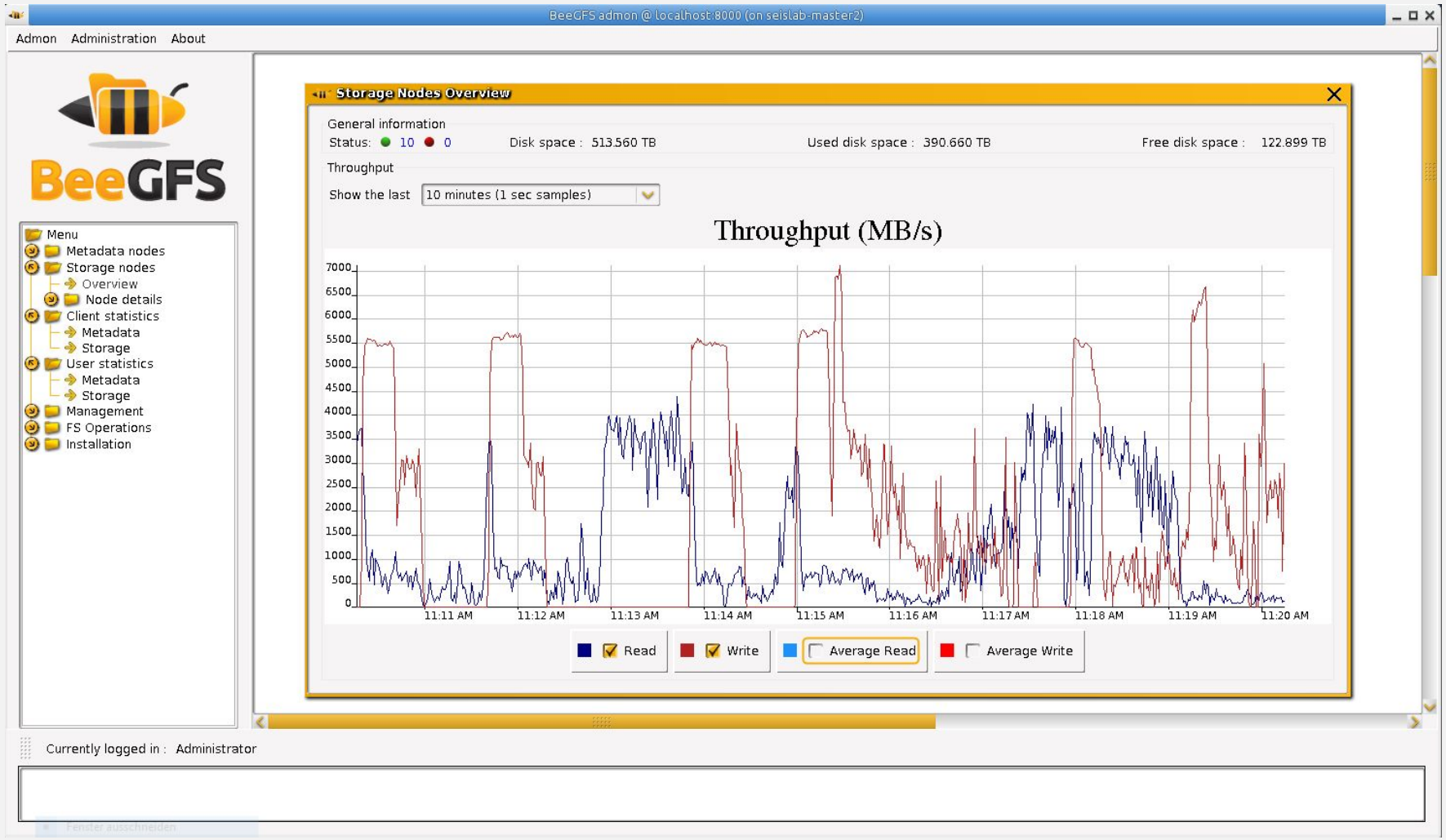
Key Aspects

- Performance & scalability
- **Flexibility**
  - Multiple services (any combination) can run together on the same machine
  - Flexible striping per-file / per-directory
  - Add servers at runtime
  - On demand filesystem „per job“ possible (BeeOND)
  - Runs on ancient and modern Linux distros/kernels
  - Runs on different Architectures, e.g.
    - ARM, Xeon Phi, Power, Tilera, ...
  - NFS & SMB/CIFS re-export possible

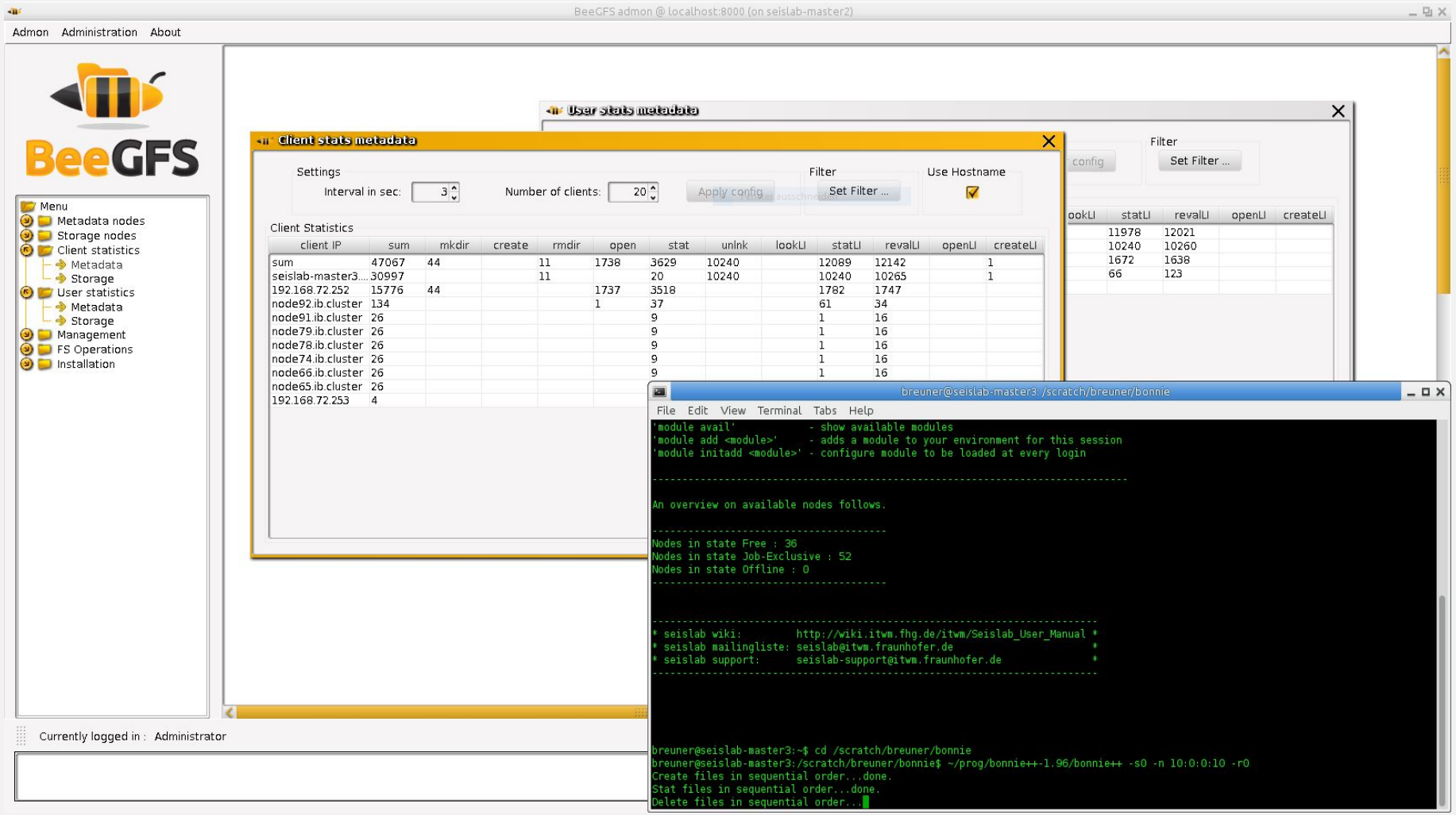
Key Aspects

- Performance & scalability
- Flexibility
- **Robust & easy to use**
  - Applications access BeeGFS as a “normal“ (very fast) file system mountpoint
    - Applications do not need to implement a special API
  - Servers daemons run in user space on top of standard local filesystems (ext4, xfs, zfs, ...)
  - No client kernel patches, kernel updates are trivially simple
  - Packages for Redhat, SuSE, Debian and derivatives
  - Hardware independent (runs on shared-nothing HW)
  - Graphical monitoring tool

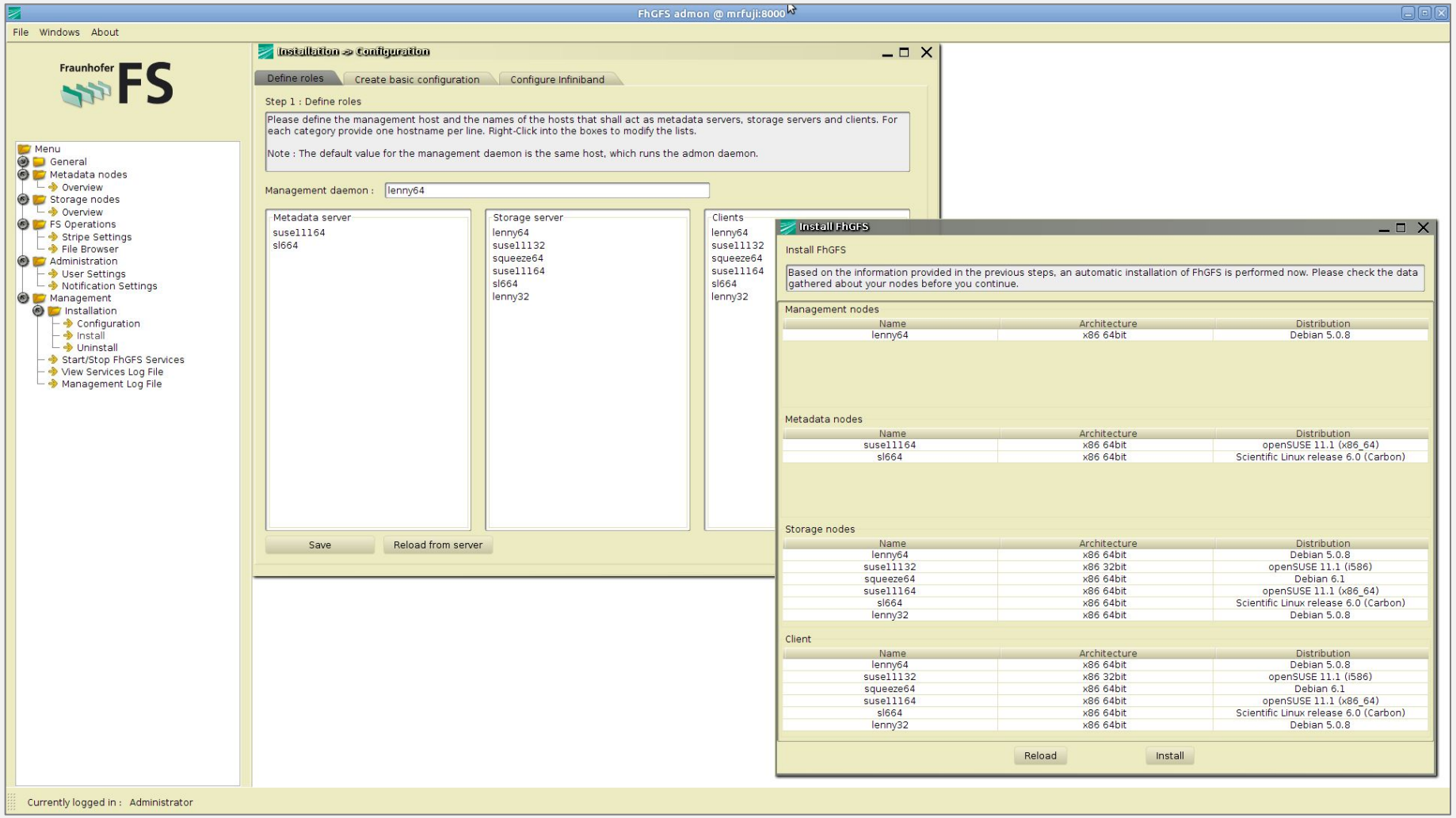
Live Throughput Overview



Live per-Client and per-User Statistics



GUI for Windows-style Installation



The easiest way to setup a parallel FS…

```
# EXAMPLE…
$ beeond start -n $NODEFILE -d /local_disk/beeond -c /my_scratch

Starting BeeOND Services…
Mounting BeeOND at /my_scratch…
Done.

----------------------------------------------

# GENERAL USAGE…
$ beeond start -n <nodefile> -d <storagedir> -c <clientmount>
```

Scemama Anthony
@BeeGFS Wonderful feature I have been dreaming of for years!!! Thank you!!!!

BeeOND – BeeGFS On Demand

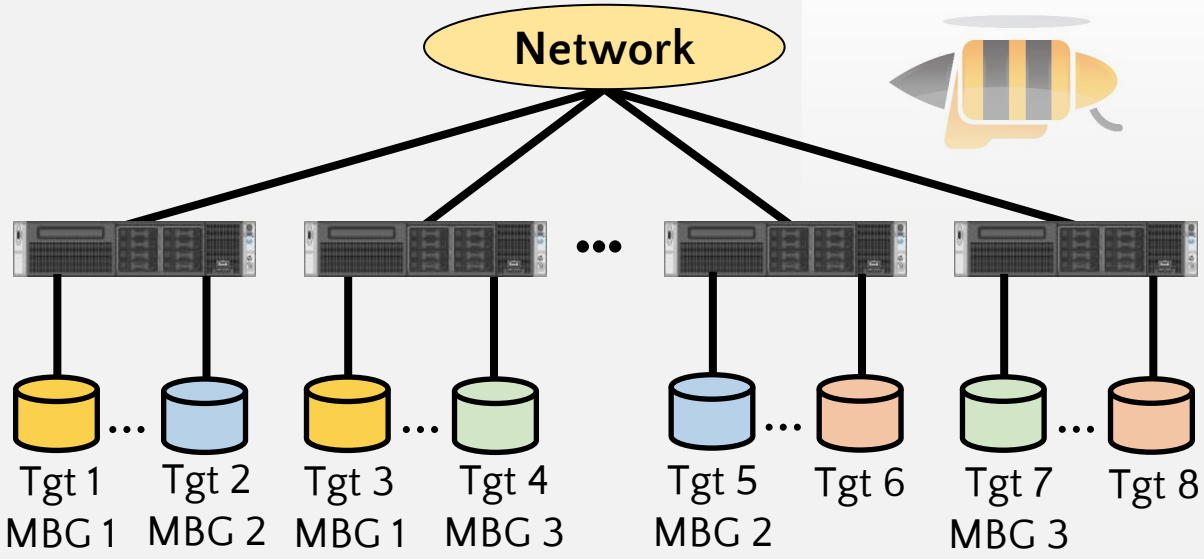
- Create a parallel file system instance on-the-fly
- Start/stop with one simple command
- Use cases: cloud computing, test systems, cluster compute nodes, ...
- Can be integrated in cluster batch system (e.g. PBS)
- Common use case: "per-job parallel file system"
  - Aggregate the performance and capacity of local SSDs/disks in compute nodes of a job
  - Take load from global storage
  - Speed up "dirty" I/O patterns

BeeOND®

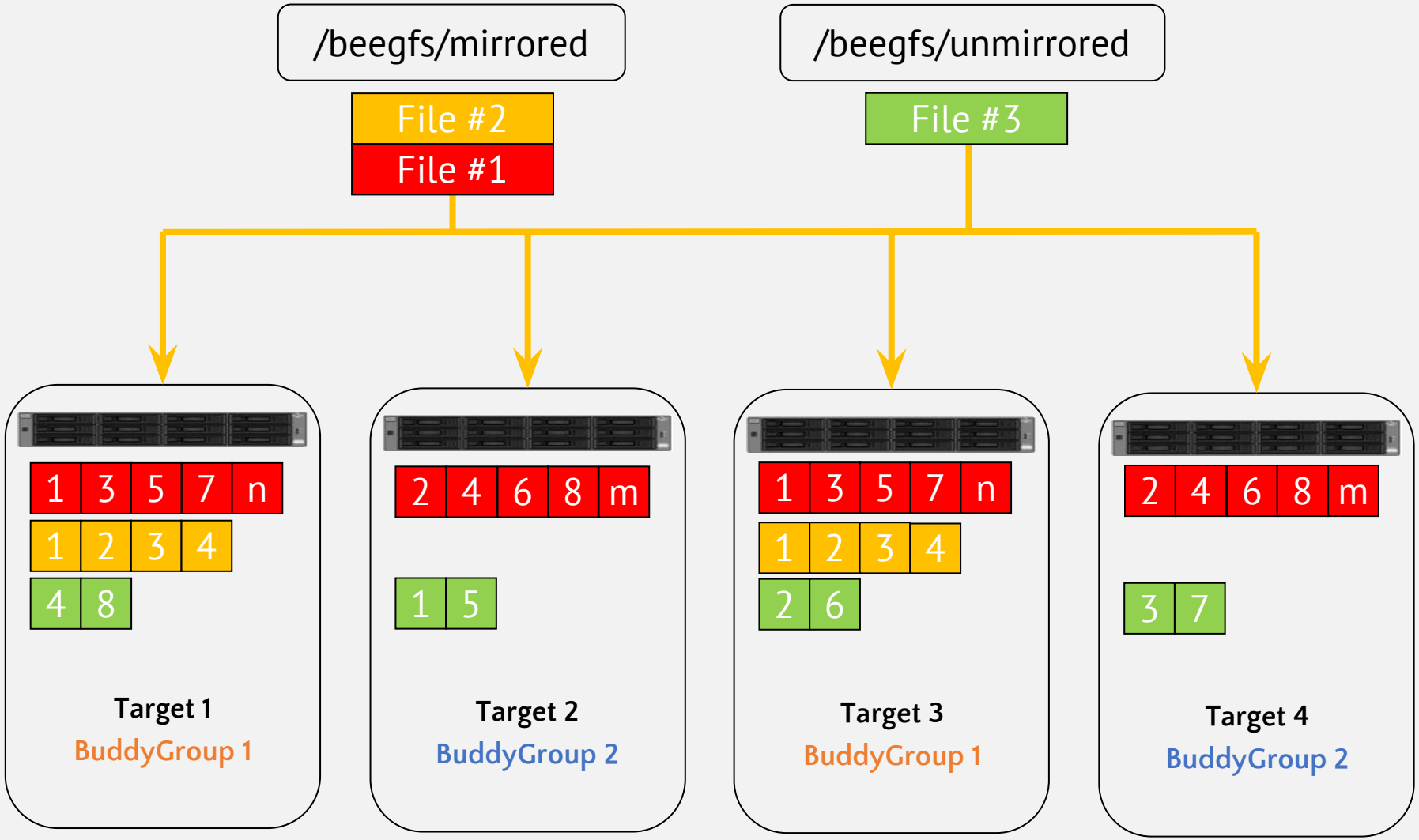
HPC Admintech 2016

Built-in Data Mirroring

- Based on „mirror buddy groups“ of storage targets
  - Primary/secondary target in a buddy group internally replicate chunks
  - But: Targets can still also store non-mirrored chunks
  - Write operations are forwarded for high throughput
  - Read possible from both targets
- Internal failover mechanisms
  - In case primary is unreachable or fails, an automatic switch is performed
  - Self-healing (differential rebuild) when buddy comes back
- Flexible: Can be enabled globally or on a per-directory basis



Buddy Mirroring per Directory



Business Model

- **BeeGFS is free to use for end users: www.beegfs.com/download**
  - Ready-to-install binaries, compelete source code also available

- **System integrators/partners for turn-key solutions**
  - System setup and tuning
  - First point of contact (1st- and 2nd-level support)
  - Partners make back2back contract with ThinkParQ for 3rd-level support

transtec

BeeGFS allows us to easily deliver peta-scale turn-key storage solutions
- transtec

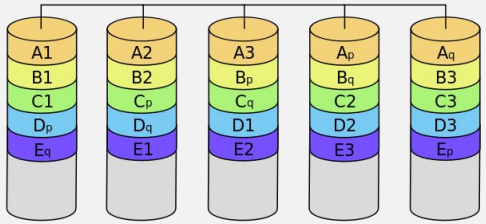
- **Professional 3rd-level support**
  - Pricing based on number of servers and timeframe (e.g. 3 or 5 years)
  - Access to enterprise features (Buddy Mirroring, ACLs, quota enforcement)
  - Special customer website area: www.beegfs.com/customerlogin

What's new in the 2015 Release Series?

- BeeOND (BeeGFS On Demand)
- Trinity:
  - Quota Enforcement
  - Access Control Lists (ACLs)
  - Built-in data mirroring
- Per-User Statistics in Admon GUI
- New manual setup tools (/opt/beegfs/sbin/beegfs-setup...)
- BeeGFS C API

Topics for 2016

- Buddy mirroring for metadata
  - Work in progress, expected Q2/2016
- BeeGFS as a service on Amazon Cloud
  - Received funding from Amazon
  - Also in touch with Microsoft for Azure cloud
- Target pools for different hardware (e.g. fast vs big)
- Striping with parity across servers
  - Tolerate server failures with less capacity overhad compared to mirroring
  - Configurable on a per-directory basis
- Object interface for HTTP put/get style access
  - To support applications that were written for such interfaces



- LIVE TRAINING / DEMO!!!!

Thank You!