Contribution ID: **115** Type: **not specified**

Data To Network: building balanced throughput storage in a world of increasing disk sizes

Wednesday, May 16, 2018 10:50 AM (20 minutes)

The ever-decreasing cost of high capacity spinning media has resulted in a trend towards very large capacity storage 'building blocks'. Large numbers of disks - with up to 60 drives per enclosure being more-or-less standard – indeed allow for dense solutions, maximizing storage capacity in terms of floor space, and can in theory be packed almost exclusively with disks. The result are building blocks with a theoretical gross capacity of about 180 TByte per unit height when employing 12 TByte disks. This density comes at a cost, though: getting the data to and from the disks, via front-end storage management software and through a network, has different scaling characteristics than the gross storage density, and as a result maintaining performance in terms of throughput per storage capacity is an ever more complex challenge. At Nikhef, and for the NL-T1 service, we aim to maintain 12MiB/s/TiB combined throughput, supporting at least 2 read- and 1 write stream per 100TiB netto storage, from any external network source down to the physical disks. Especially this combined read-write operational pattern poses challenges not usually found in commercial deployments. Yet this is the pattern most commonly seen for our scientific applications in the Dutch National e-Infrastructure.

In this study we looked at each of the potential bottlenecks in such a mixed-load storage system: network throughput, limitations in the system bus between CPU, network card, and disk subsystem, at different disk configuration models (JBOD, erasure-encodings, hardware, and software RAID) and the effect on processor load in different CPU architectures. We present the results of different disk configurations and show the limitations of commodity redundancy technologies and how they affect processor load in both x86-64 and PowerPC systems, and how the corresponding system bus design impacts overall throughput.

Combining network and disk performance optimizations we show how high-density commodity components can be combined to build a cost-cutting system without bottlenecks – offering constant-throughput multi-stream performance with over 700TiB netto in just 10U and able to keep a 100Gbps network link full – as a reference architecture for everything from a single Data Transfer Node down to a real distributed storage cluster.

Desired length

20 minutes

Primary author: SUERINK, Tristan (Nikhef National institute for subatomic physics (NL))

Presenter: SUERINK, Tristan (Nikhef National institute for subatomic physics (NL))

Session Classification: Storage and file systems

Track Classification: Storage & Filesystems