NFS: Back to the Future and Holy Cow Agnes, where are we now?

Brian L. Wong
Distinguished Engineer – Sun Microsystems Network Storage

NFS as a Data Sharing Protocol

- ∀ NFS provided file sharing in 1985

- ∀ Data relocation from client to server

- ∀ Itm asked the details of storage management from client system s

    File system –to-disk m appings hidden

    W hat volum e m anager?

    M any sites use H SM transparently

NFS: Successful File Sharing

∀ Simple, reliable, comprehensible

∀ Easy to deploy

∀ Hid the details of storage

∀ Provided transparent redirection

∀ Performance good enough

NFS as a Disk Replacement

- ∀ NFS also replaced disks

- ∀ Diskless clients centralized management

- ∀ Inexpensive (TCO-wise) server disks replaced expensive, hard-to-maintain client disks

NFS: Not a Disk Replacement

- ∀ A commercial failure— why?

  Client disks got *cheap*

  - ∀ But administrators didn't

  Jammed networks = bad performance

  - ∀ Networks were shared, not switched
  - ∀ 10 Mbit/sec Ethernets overpowered by 50 MHz CPUs, 10 Mbyte/sec SCSI

  Boot servers effectively facility SPOFs

  - ∀ No effective, inexpensive HA facilities in 1994

NFS as a Protocol Suite

- ∀ Basic NFS is pretty good

  Simple, consistent semantics

- ∀ Needing improvement

  Global name space

  Interoperable ACL protocol

  Security (trust domain issues)

  Performance

NFS Performance

- ∀ In a word, it sucks, right?

- ∀ Well, no...

  2,000 NFS ops/sec = 200-4000 client systems

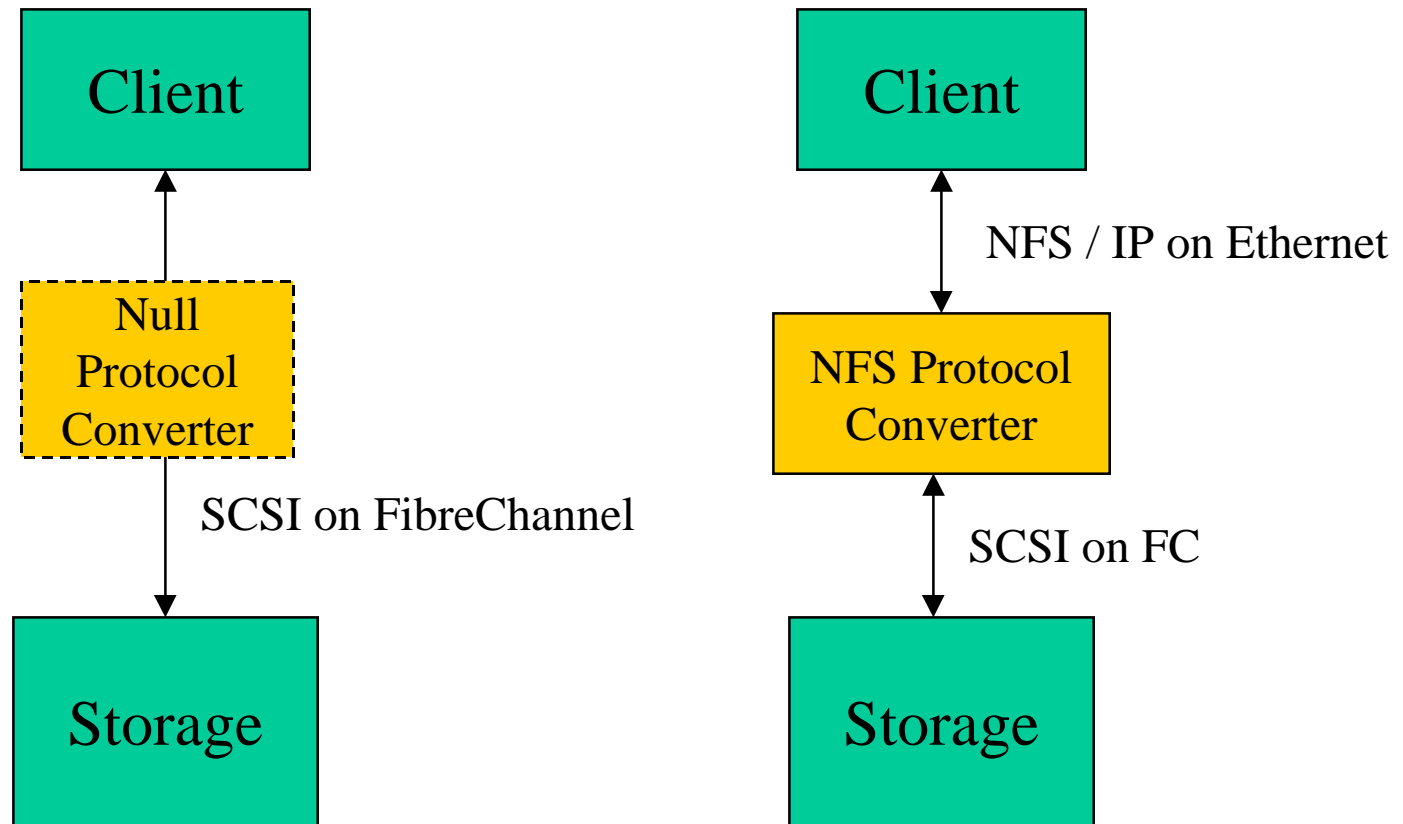 Even fairly large (250 GB+) servers sustain much less than 2,000 ops/sec, peak << 5,000

  High end servers > 100,000 ops/sec (SFS97)

    - ∀ Even SFS97_R1 results exceed 38,000 ops/sec

Major Perception Problems

- $\forall$ Users/admins are not well calibrated

  "NFS is too slow"

  Little or no capacity planning help

- $\forall$ "We can't do that, it's old technology"

  so are TCP/IP, x86 and rotating magnetic disks

- $\forall$ NFS is drastically under-marketed

SAN and NAS Equivalence

Client	Client

NFS / IP on Ethernet

Null Protocol Converter	NFS Protocol Converter

SCSI on FibreChannel

SCSI on FC

Storage	Storage

Storage Big Problem : Semantics

- ∀ Storage blocks have no semantics

  - "a LUN is a bag of bits"

  - Serverless backup problems

  - Limited ability of internal storage services (e.g., point-in-time copy/snapshot)

- ∀ Causes data/storage confusion

  - Data = files

  - Storage = disk blocks

Storage Big Problem : Parallelism

∀ **Parallel physical plant**

   Gigabit Ethernet ~= Fibre Channel

∀ **Parallel administrative skills**

   FC skills required for storage, IP skills for networks

∀ **Parallel software stacks**

   Separate code bases for networking, storage

Storage Transitions

- ∀ Transports: dedicated → commodity

  FibreChannel → IP

- ∀ Semantics: simple → rich

  SCSI blocks → Objects or Files

- ∀ Connectivity: simple → complex

  Mostly DAS → mostly SAN → centralized W SAN

NFS in the Future?

"I don't know what the file sharing protocol of the future will be, but it will be called NFS... "

(with apologies to C A R . Hoare)

Why NFS?

- ∀ Higher level protocol

    Hides storage details

    Exports data semantics

    Easily (relatively) made highly available

- ∀ Protocol has storage semantics

    More parallel to SCSI blocks than most

    Can have affinity with iSCSI efforts

take it to the nth

Why NFS?

∀ "Doesn't" have to be promulgated

   It's already a de facto industry standard

∀ Adapts storage to networking

   Commodity transport

   Amenable to WANs

NFS Performance Issues

- ∀ Two broad categories of NFS usage

  Attribute-intensive (SFS, similar to TPC-C)

  Data-intensive (???, corresponds to TPC-H/R)

  - ∀ "If you haven't run the benchmark, the performance sucks"
  - ∀ No formal data intensive benchmark

- ∀ Cost of TCP/IP

  Efficiency, not throughput or latency

- ∀ Clients, as well as servers

  No formal client benchmark

WhatToDo

- ∀ Complete the protocol suite

    Global name space

    Security

    Integrate with name services

- ∀ Develop data management capabilities

- ∀ Drive up efficiency

    Pay attention to client side

    Seek affinity with offload vendors, esp iSCSI

- ∀ Marketing