



The General-Purpose Storage Revolution

Jeff Bonwick

Sun Fellow and CTO, Storage

Sun Microsystems

"I know you're out there.
I can feel you now.
I know that you're afraid...
you're afraid of us.
You're afraid of change.
I don't know the future.
I didn't come here to tell you
how this is going to end.
I came here to tell you
how it's going to begin."

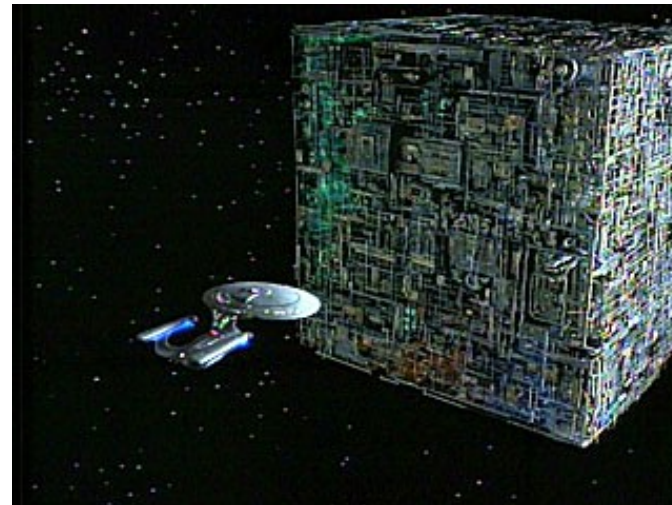
- Neo, "The Matrix" (1999)

General-Purpose Compute Revolution

- It happened in compute over the last two decades
 - Volume CPUs replaced special-purpose hardware



- Driven by economic fundamentals – resistance is futile

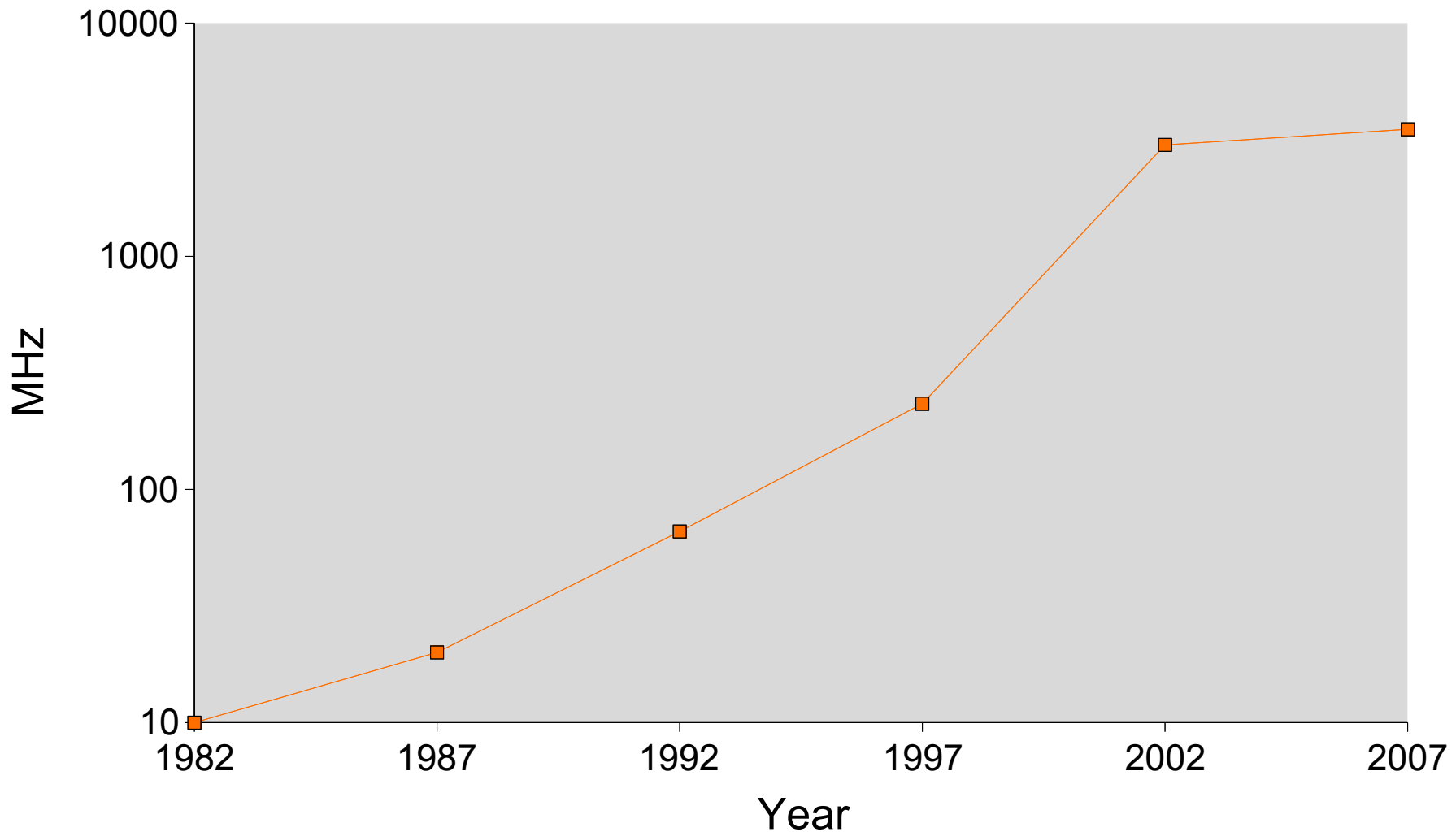


General-Purpose Storage Revolution

- It's happening in storage right now – on three fronts:
 - Hardware
 - Special-purpose hardware vs. volume microprocessors
 - Software
 - Special-purpose kernels vs. volume operating systems
 - Networking
 - Special-purpose interconnects vs. Ethernet
- General-purpose storage is inevitable
 - Market is ripe for disruption
- More than just cost reduction
 - New things become possible

Clock Rate and Core/Thread Trends

Clock rate hit a brick wall in 2002...



Clock Rate and Core/Thread Trends

- ... but cores are proliferating:
 - AMD – 4 cores
 - Intel – 4 cores
 - Sun – 8 cores, 64 threads
- Vertical scale in every socket

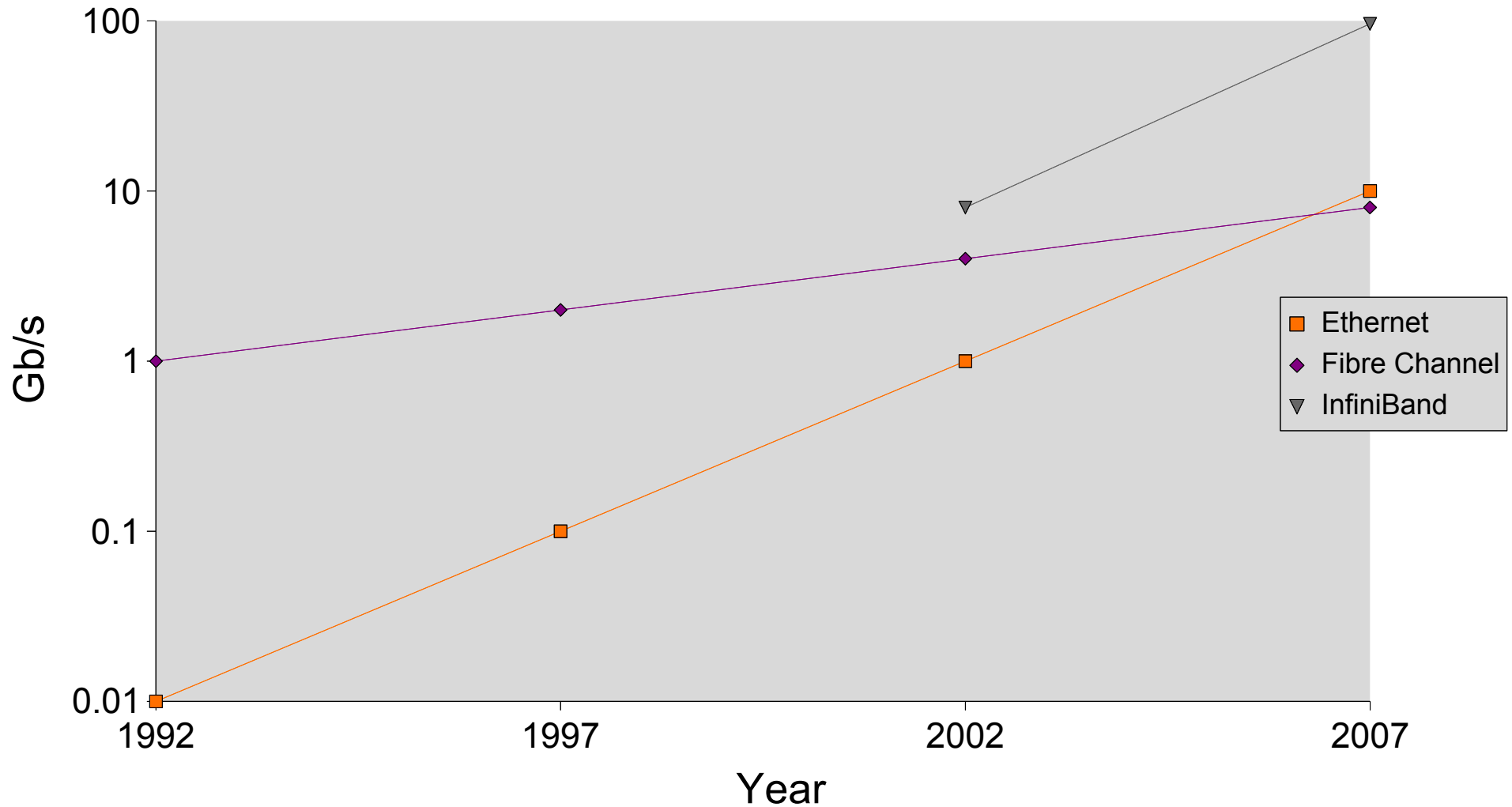


Implications

- Compute becomes cheap – but requires scalable software
 - Multi-core & CMT: vertical scale in every socket – even at the low end
 - Solaris is really good at this
- RAID, crypto, checksums move from controllers to CPUs
 - Eliminates special-purpose hardware
 - Reduces power consumption
 - Enables end-to-end data integrity, stronger security
- Storage will host applications
 - Data mining, indexing – anything you'd rather do near the data
 - Applications can be deployed as virtual appliances
 - But how will we talk to them?

Network Performance Trends

The writing is on the wall...



Implications

- The storage network is the computer
 - Intelligent storage systems join the *real* network
- A new breed of storage systems becomes possible
 - From moving bytes to answering questions (Google, Greenplum)
 - Leverage inside-the-box data speeds

Our Strategy

- Cheap, fast hardware
- Simple, powerful, resilient software
- Horizontal scale on fast, commodity networks
- Virtual appliances providing network data services

Our Special Sauce

- System-level integration, everywhere
- Novel use of general-purpose components
- Building storage developer communities

Thumper: Cheap, Fast, Integrated H/W

- Integrates compute and storage
 - 4 dozen drives + 4 CPUs + 4 GigE ports in 4U; 4 racks per PB
 - 3 GB/s internal bandwidth – platter speed on all 48 drives
 - Lots of cheap, powerful compute near the data
 - No H/W RAID – not needed with ZFS
- No SAN required
- The first in a family
 - Thumper JBOD (SAS)
 - More cores, many more disks
 - Very low structural overhead
- Powered by ZFS



ZFS: Simple, Powerful, Resilient S/W

- ZFS does for disks what VM did for memory
 - You don't manage DIMMs and ECC – why should you manage arrays?
- Integrates filesystem, volume management, RAID
 - Pooled storage – no more volumes, partitions, provisioning
 - Creates reliable storage out of unreliable components
 - Protects the entire data path – disks, HBAs, switches, driver bugs, etc.
 - Self-healing – detects and corrects silent data corruption
 - File, block, and transactional object interfaces
 - Full or thin provisioning
 - Unlimited, constant-time snapshots and clones
 - Built-in compression
 - De-duplication, CDP, remote replication, and encryption in the works

pNFS: Standardized Horizontal Scaling

- Parallel NFS
 - Part of IETF NFSv4.1 specification
 - Namespace and data placement scale horizontally
 - Works with existing, heterogeneous NFS clients
- Commodity network hardware
 - No special interconnect – runs on Ethernet
 - Also runs on FC and IB, if you're so inclined
- Potentially much more than NFS
 - Overall architecture isn't NFS-specific
 - New data services can use the same foundation
 - pCIFS

ADM: Integrated HSM

- Automated Data Migration (next-generation SAM)
 - Simple to manage
 - Scalable to vast datasets
 - Integrated ZFS support
- Policy-based data migration and archiving
 - Plugs into ZFS on top, tape libraries on the bottom
 - Automatically migrates data from disk to tape
 - Replaces traditional backup with continuous archiving
 - With de-duplication in ZFS, replaces VTL as well
- 100% open data formats

FishWorks: Integrated Appliances

- Build turn-key solutions from general-purpose components
 - Special-purpose capabilities
 - Volume economics
- Deep system-level integration
 - Hardware fault domains fully modeled by FMA
 - High-level services like remote replication built on ZFS
 - Business analytics and data forensics built on DTrace
 - Link aggregation, multi-pathing, etc built on Solaris
 - Simple, cheap heartbeat card for active/active HA
 - Beautiful browser-based management built on AJAX and XML-RPC
- Common look and feel across all Sun appliances

Solaris: The Storage OS of Choice

- More scalable than Linux, BSD or Windows
- Lightweight virtualization (zones) ideal for virtual appliances
- High-quality NFS and (very soon) in-kernel CIFS
- Industry-leading tools (DTrace, mdb)
- ZFS – the most active OpenSolaris community
 - ZFS ported to Linux (FUSE), FreeBSD, MacOS X (Leopard)
 - New algorithms for compression, block allocation, etc
 - Amazing effect on pace – communication drives development & adoption
- Sun is reclaiming thought leadership
 - Proprietary, closed-source vendors have no way to respond

The Storage Network is the Computer

