

Patagonia - Ein Mehrbenutzer-Cluster für Forschung und Lehre

Felix Rauch

*Christian Kurmann, Blanca Maria Müller-Lagunez,
Thomas M. Stricker*

Institut für Computersysteme
ETH Zürich



*Eidgenössische
Technische Hochschule
Zürich*

25. März 1999

Forschungscluster



Bild: Distributed ASCI Supercomputer von Henri E. Bal

Charakteristiken eines Forschungsclusters

- Schnelle Prozessoren (1 - 4 pro Knoten)
- Grosser Speicher (Haupt und Massenspeicher)
- Leistungsfähiges Netzwerk (Switches und Gigabit/s)
- Benutzungsmuster:
Tagsüber Entwicklung - Nachts Experimente

Schulungscluster



Bild: Patagonia Cluster ETH Zürich

Charakteristiken eines Schulungsclusters

- Grosse Festplatten für umfangreiche Softwareinstallationen
- Betriebssysteme in Grundkonfiguration und spezieller Konfiguration
- Grosse räumliche Ausdehnung
- Systemsicherheit:
 - Installation und Daten vor Studenten
 - Hardware vor Diebstahl
- Benutzungsmuster: Ausschliesslich tagsüber

Mehrbenutzer-Cluster

Beobachtung:

Beide Cluster haben ähnliche Anforderungen, aber unterschiedliche Nutzungsmuster

These:

Ein einziger Cluster genügt

Inhalt

- Motivation
- Klassen von PC-Betriebssystemen bzgl. Sicherheit
- Hardware des Patagonia-Clusters
- Technologien für Patagonia:
 - Multi-boot / Betriebssysteme
 - Installation durch Klonen
 - Sicherheit / Unterhalt
- Performance Evaluation (Images Klonen)
- Schlussfolgerungen

Klassifikation von PC Betriebssystemen

Oberon
MacOS 8.x
Windows 9x
Windows NT
UNIX

	Oberon	MacOS 8.x	Windows 9x	Windows NT	UNIX
Netzwerkfähigkeit	0	0	0	+	+
Sicherheit	-	-	-	+	+
Getrennter Zustand System/User Config.	-	-	-	-	+

Hardware

Raum mit 24 Maschinen:

- Intel Pentium II, 400 MHz
- 128 MB SDRAM
- Fast Ethernet Netzwerk
- 9 GB Ultra2 SCSI Festplatten

Davon 16 Maschinen mit:

- Zwei Prozessoren
- 256 MB SDRAM
- Gigabit Ethernet

System-Software

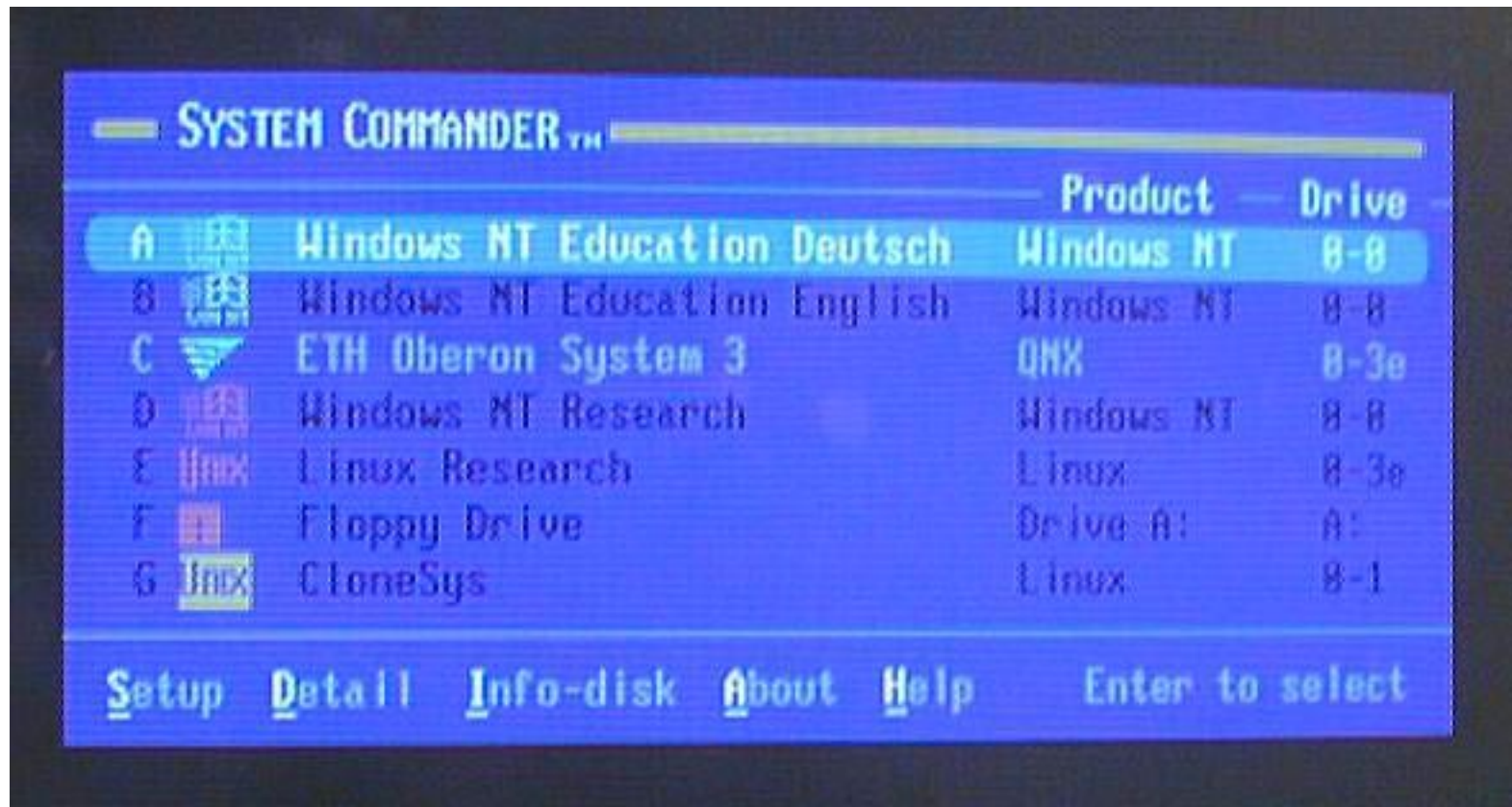
Schulung:

- Windows NT Deutsch
- Windows NT Englisch
- Oberon

Forschung:

- Linux
- Windows NT
- Oberon

Boot-Manager *System Commander*



Installation

Partitionierung

• Boot-Partition	0.020 GB
• Partitionen für Windows NT 4.0 Edu	2 x 2.0 GB
• Partitionen für Windows NT Research	2.5 GB
• Partition für Linux	1.0 GB
• Partitionen für Oberon	2 x 0.1 GB
• Reservepartition (Solaris, Oracle usw...)	1.0 GB
• Kleine LINUX Service-Partition (wünschenswert)	0.25 GB
	<hr/>
	ca. 9 GB

Installation

Replikation durch “Klonen”

- Erstinstallation
 1. Erstellen einer Master-Platte
 2. Blockweises Kopieren der Master-Platte
- Klonen einzelner Images / Partitionen
 1. Service-Betriebssystem Booten
 2. Blockweises Kopieren der Images über Netzwerk

Konfiguration

Konfiguration der maschinenspezifischen Parameter (IP Nummer, Hostname, Hostid)

- Manuell
- Automatisch über DHCP (mit Server)
- Automatisch anhand der Ethernet MAC-Adresse mittels Tabelle

DHCP = Dynamic Host Configuration Protocol

MAC = Media Access Control

Sicherheit

Ziel: Sicherheit ohne Behinderung der Benutzer

Wird erreicht durch drei Stufen:

- 1 Booten mit *System Commander*
- 2 Partitionen voreinander schützen und verstecken
 - Sperren mit *Device Lock* und Neuzuweisung zum Laufwerksbuchstabe C (Windows NT)
 - Mount-Tabellen (UNIX)
 - Schreibgeschützte Partitionen plus Ramdisk (Oberon)
- 3 Verwendung von Zugriffsrechten und Autorisation bei zentralem Server

Unterhalt

- Cluster Administration Tool erlaubt Überblick über Cluster und remote boot
- Schnelles Netzwerk hilft bei Restauration von Partitionen und Updates

Netzwerkkonten für UNIX und Windows NT

Leistungsfähiger Sun-Server für:

- Home-Verzeichnisse über SMB (mit Samba) bzw. NFS vom Sun-Server
- Autorisierung über speziellen Windows NT-Server (NT-Clients) bzw. NIS vom Sun-Server (UNIX-Clients)

Konten-Generierung auf UNIX - Übername auf NT mit Scripts

Passwort-Synchronisation mit kommerziellem Programm *Passync*

SMB = Server Message Block Protocol

NFS = Network File System

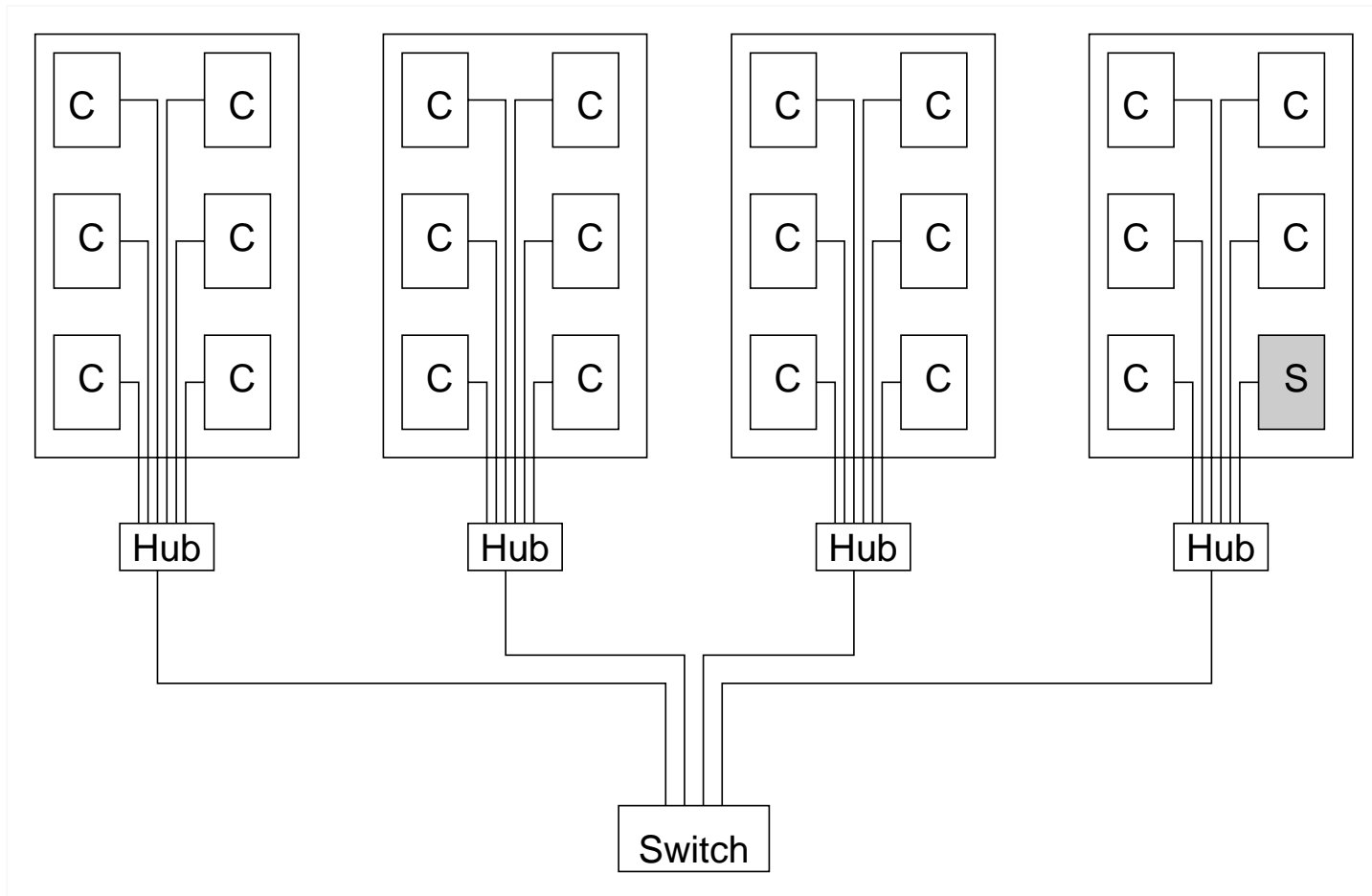
NIS = Network Information Service (früher yellow pages)

Technische Limiten

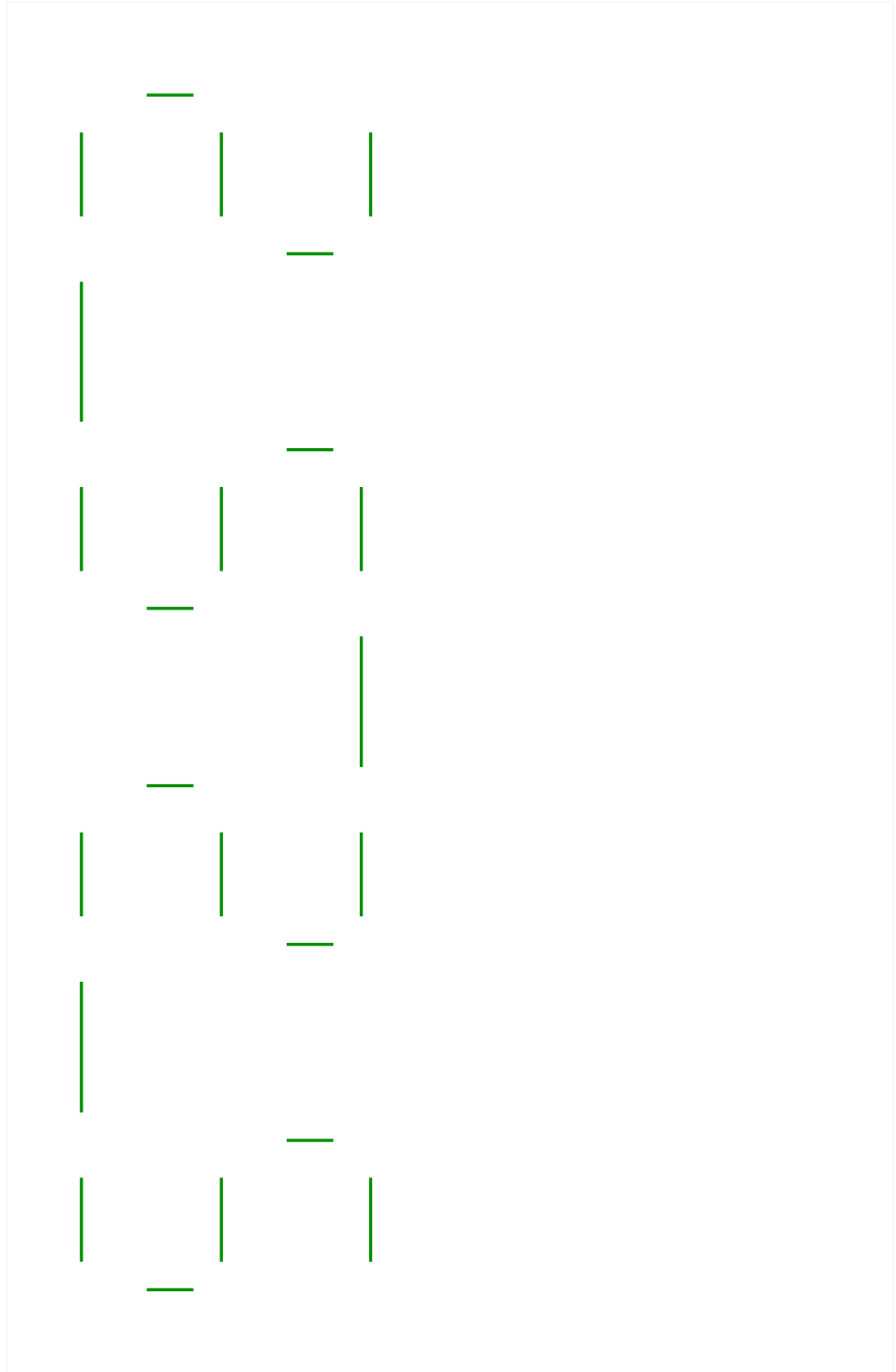
- Schreiben auf Ultra2 SCSI Platte *ca. 20 MByte/s*
(Seagate Cheetah write avg)
- Lesen von lokaler Festplatte *ca. 16 MByte/s*
(Seagate Cheetah read avg)
- Lesen von remote Files (NFS) *21 MByte/s*
(via UDP über Gigabit/s Ethernet)
- Dekomprimieren eines Images *12 MByte/s*
(gunzip >/dev/null mit 400MHz)
- Übertragung Punkt zu Punkt *40 MByte/s*
(Gigabit Ethernet via TCP)
- Totale Kapazität im Ethernet Switch *>3 GByte/s*

Performance Evaluation (Klonen)

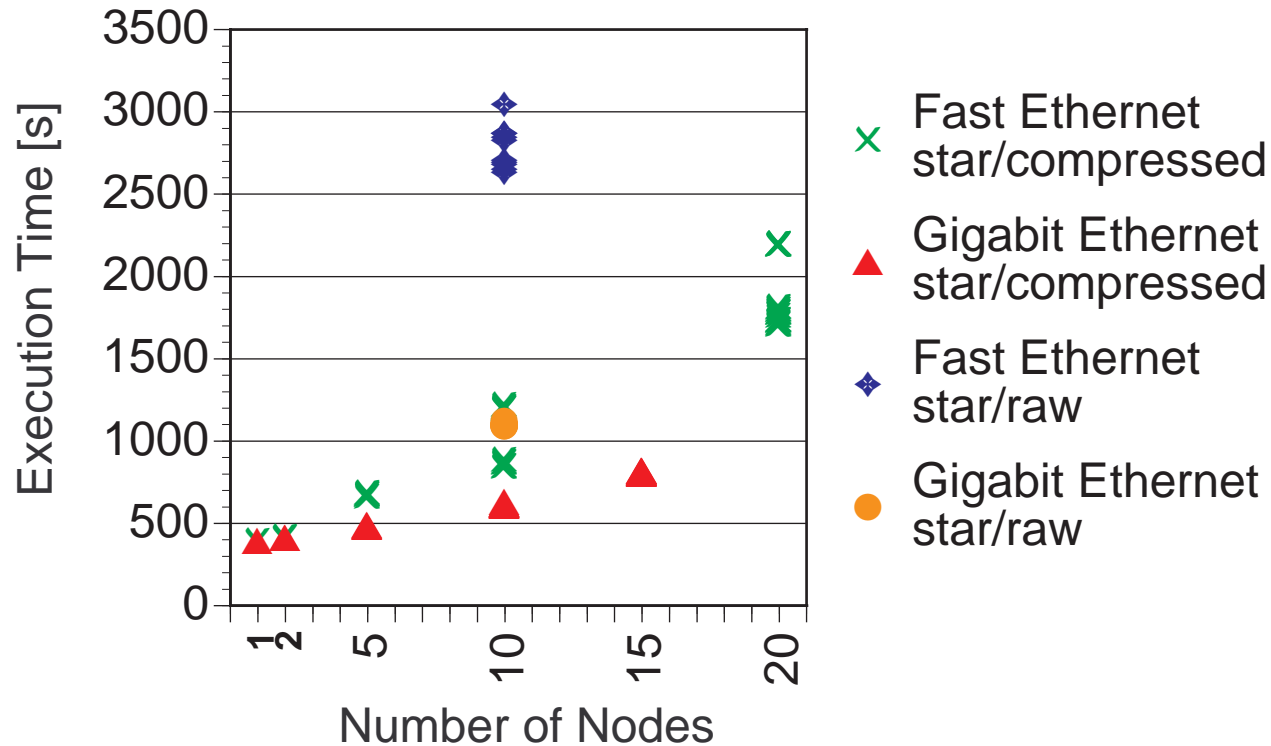
Netzwerk-Topologie Fast Ethernet





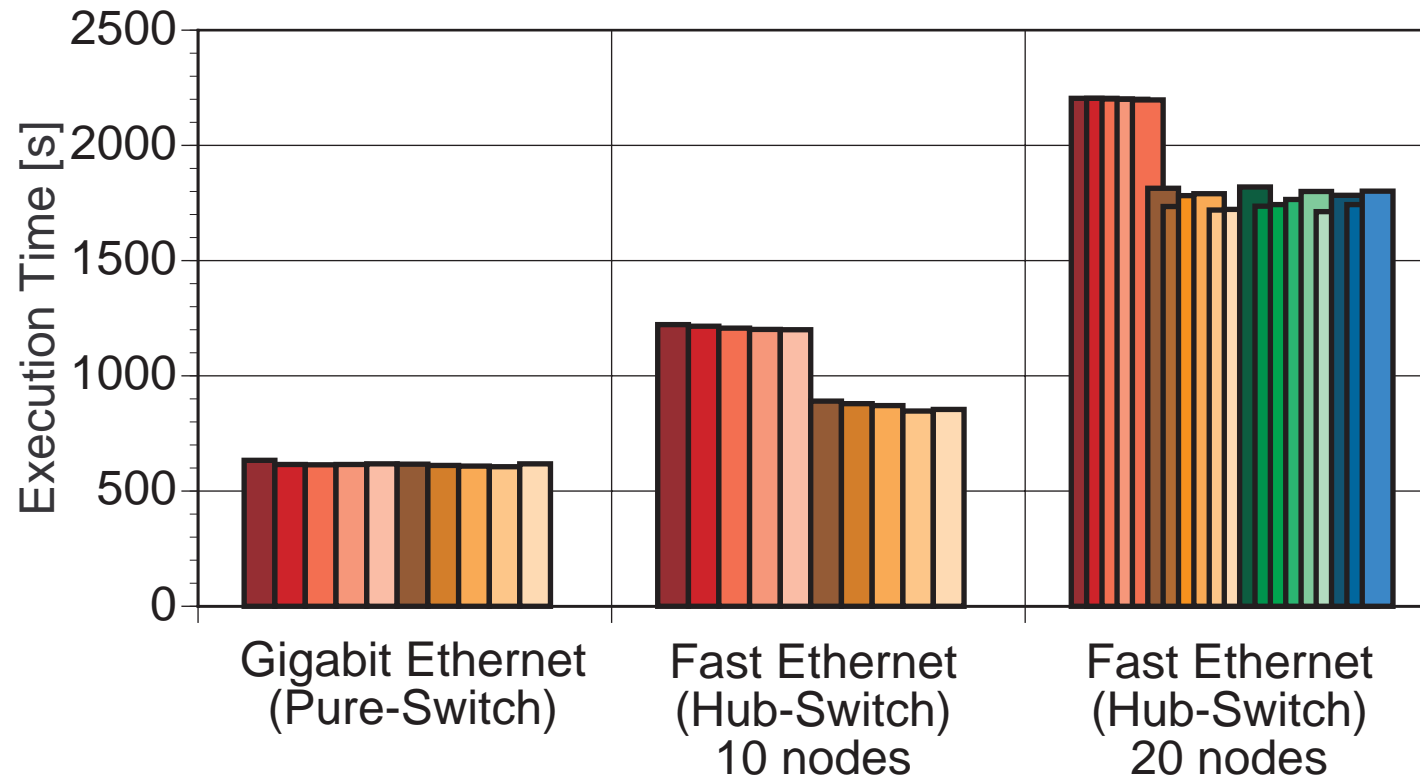


Ausführungszeiten Klonen

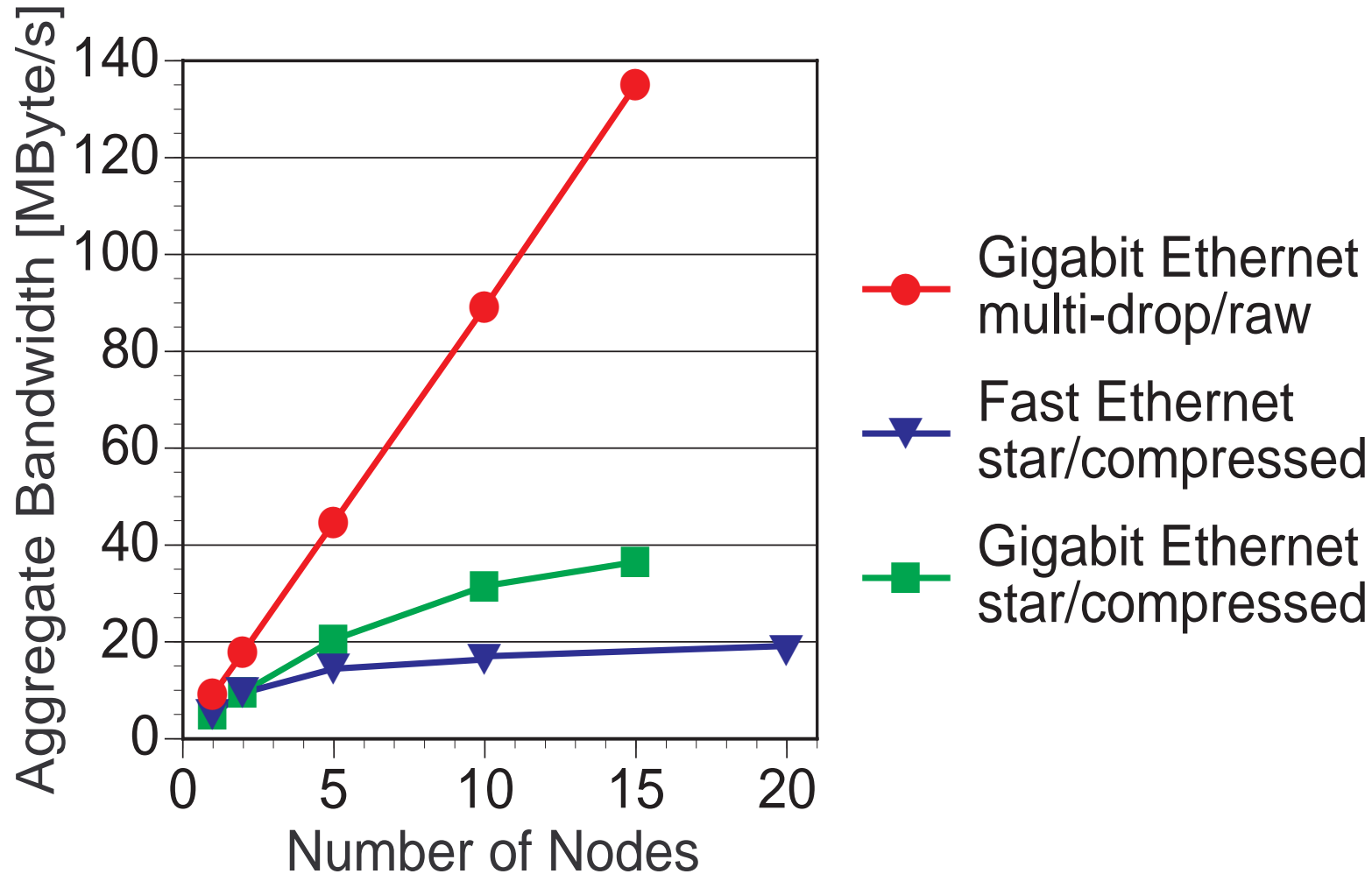


Windows NT Partition über NFS
(2 GB raw, 1 GB compressed)

Auswirkungen der Hub/Switch-Topologie



Totale Schreib-Bandbreite



Schlussfolgerungen

- Erfolgreiche **Installation und Inbetriebnahme** eines universellen Clusters für **Forschung und Lehre**
- Erleichterung von Wartung und Installation durch:
 - Kleines **Service-Betriebssystem**
 - Schnelle, **grosse Festplatten**
 - **Hochleistungs-Netzwerk**
- **Multi-boot** Installationen geben **grosse Flexibilität**
- **Klonen** von ganzen Software Installationen als neue, interessante Anwendung von **Gigabit/s Netzwerken** (ausserhalb des parallelen und verteilten Rechnens)

Performance Evaluation (Klonen)

Netzwerk-Topologie Gigabit Ethernet

