

Genentech's IT Infrastructure

John "Scooter" Morris, Ph.D.



Disclaimer

This will take a very long time

I don't have a very long time

- Aren't you glad?

I might leave out some details here and there



Outline

Setting

Current Infrastructure

- A bunch of details which I will skip

Future Directions



Setting

“Genentech is a pharmaceutical company dedicated to applying recombinant DNA technologies to unmet medical needs.”

Statistics

- ~3,700 Employees
- ~\$1.3B in Revenue
- 8 products
 - Protropin, Nutropin, NutropinAQ, NutropinDepot, Activase, Pulmozyme, Herceptin, Rituxan
- Three major sites
 - South San Francisco, California
 - Vacaville, California
 - Porrino, Spain
 - Several U.S. Sales offices



Setting

This has significant implications for our IT infrastructure

- **Must support research**
 - **Discovery**
 - **Technology**
- **Must get products to market**
 - **Preclinical development**
 - **Clinical development**
 - **Manufacturing**
 - **Partnerships**
- **Must satisfy regulatory agencies**
- **Want to be able sell products also...**
 - **Sales**
 - **Marketing**



Setting - Research

Academic culture

- Open, fast-paced environment
- Need to provide tools as much as solutions
- Build vs. buy is often slanted towards build (or at least extend)

Computational needs are high

- Bioinformatics
- Molecular modeling
- Computational Chemistry

Platform choice

- Based on “BFB” metric
 - BFB: Bang for the buck

Customers often have IT expertise



Setting - Development

Business culture

- Emphasis on repeatable processes
- Need to provide solutions
- Build vs. buy is very slanted towards buy

Regulatory concerns

- Regulatory agencies want to ensure our systems do what we say they do
- This implies a significant increase in documentation
- Security is a major issue

Platform choice

- Based on software vendor support matrix



Current IT Infrastructure

Highly heterogeneous

- Servers: SGI, Compaq, HP, Sun
- Desktops: Mac, PC

Primarily IP-based network

- Lots of AppleTalk, though

Security based on M&M principle

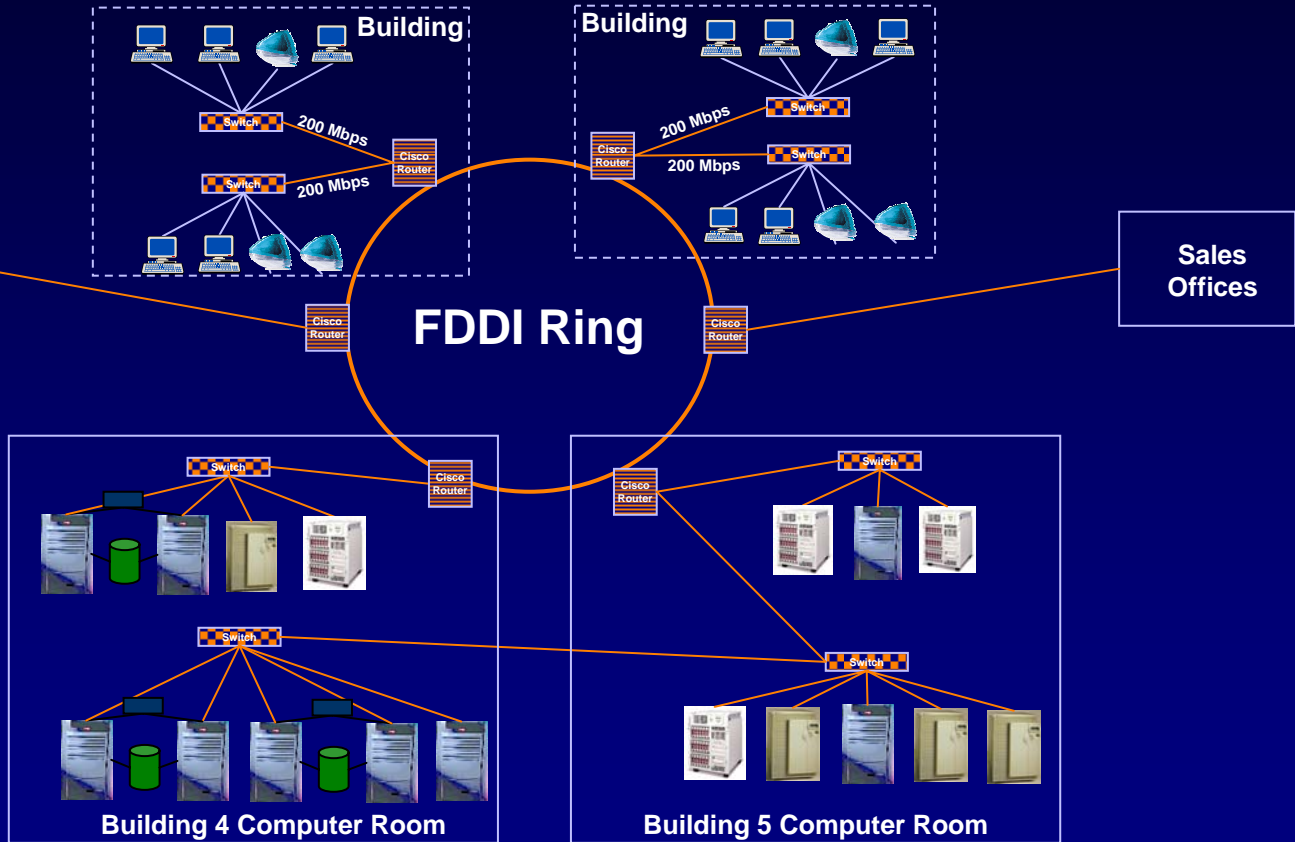
- Hard outer shell, soft inside
- Some “softness” appearing to support collaboration
- Important to maintain open environment



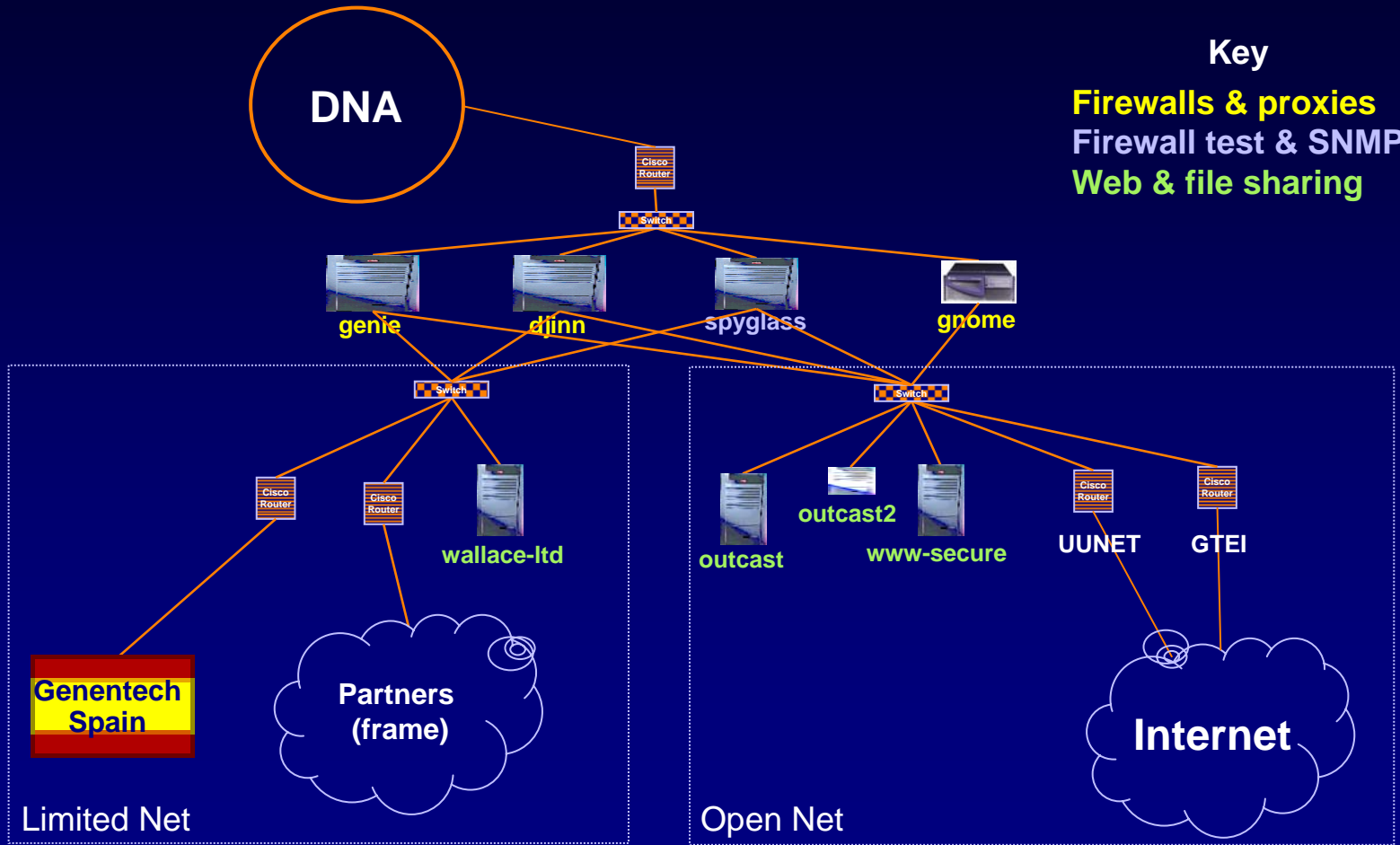
Current IT Infrastructure



Vacaville



Current IT Infrastructure



Key

- Firewalls & proxies
- Firewall test & SNMP
- Web & file sharing



Details

This starts the part I'm going to leave out....



Details - Network

Backbone: 100 Mbps FDDI Ring

- Migrating to switched Gigabit Ethernet

Vacaville link: 200 Mbps SONET Ring

Desktop: Goal is switched 10/100 Ethernet

Routers: CISCO

Addressing: DHCP preferred

Naming: DNS (Bind 8.2.2), WINS

Firewall: SOCKS5 (Aventail)

Monitoring: SNMP (NetView)



Details - Desktop

Compaq

- Windows 95/98/NT 4.0
 - Migrating to Windows 2000

Apple

- Mac OS 8.6/9
 - Migrating to Mac OS X

Primary Applications

- Office 97/98
 - Migrating to Office 2000 after 2001 release
- Netscape Communicator (Browser, Mail, Calendar)
 - Considering a migration to Mozilla
- Norton Antivirus
- FileMaker Pro



Details - Server

Compaq Tru64 Unix

- Web, E-Mail (IMAP), Bioinformatics, Infrastructure (DNS, Firewall, DHCP, backup/restore), General computing, Oracle
 - 4.0D (TruCluster 1.5)
 - Migrating to 5.0A (TruCluster Server 5.0A)

HP/UX

- Manufacturing, Commercial Computing (Lawson, PeopleSoft)
- 10.20, 11.0

Solaris

- Medical Affairs, Infrastructure (Calendar, Remedy, Web Proxy)

SGI

- Molecular Modeling, Computational Chemistry

NT

- Workgroup Computing, Specific Applications



Details - Database

Oracle 7.3.4

- Migrating to 8.1.5

Overall approach is to move to three-tier approach

- Web Browser for presentation
- Application servers for business logic
- Database server for data store



Details - Web

Server: Netscape Enterprise Server 4.0, 4.1

Programming: Perl/CGI, Embedded Perl, Javascript

- Starting to use JSP

Application Servers:

- WebObjects, Oracle Application Server

Distributed Computing:

- Direction is towards Enterprise Java Beans
 - WebLogic, WebObjects
 - Tuxedo in use for Manufacturing applications

Development Tools:

- Dreamweaver, JBuilder, Visual Age



Details - Security

Main account is based on Unix username/password

NT accounts use Unix username, but different password

Oracle accounts often use Unix username, but also lots of application-specific accounts

LDAP is used for Directory services

- Netscape Directory Server 4.0

Serious regulatory restrictions (21CFR Part 11)

Future direction is Kerberos



Details - Internet/Extranet

Firewall is based on SOCKS5 (RFC??)

Totally Proxy-based (*very secure*)

Firewall has three parts:

- Internal
- Internet
- Limited Net

Internet link is redundant

- 9 Mbps link with GTE Internet
- T1 (1.54 Mbps) link with UUNET
- Uses OSPF for dynamic fail-over

Limited Net

- Private connection to partners via T1 and Frame Relay

Investigating SOCKS5-based VPN

Testing reverse-web Proxy

- Allows external partners access to selected internal web sites



Details

Any questions on the details?

- I didn't think so....



Future Directions

Three-tier

- Implementation of highly available Oracle servers
 - TruCluster 5.0A

Kerberos-based security

- *Not* based on Windows 2000
- Single Sign-on through Windows 2000 & MacOS X
Kerberos integration

Distributed Computing

- Enterprise Java Beans
- Heavy use of Web
- Embedded perl, JSP

New Research Cluster



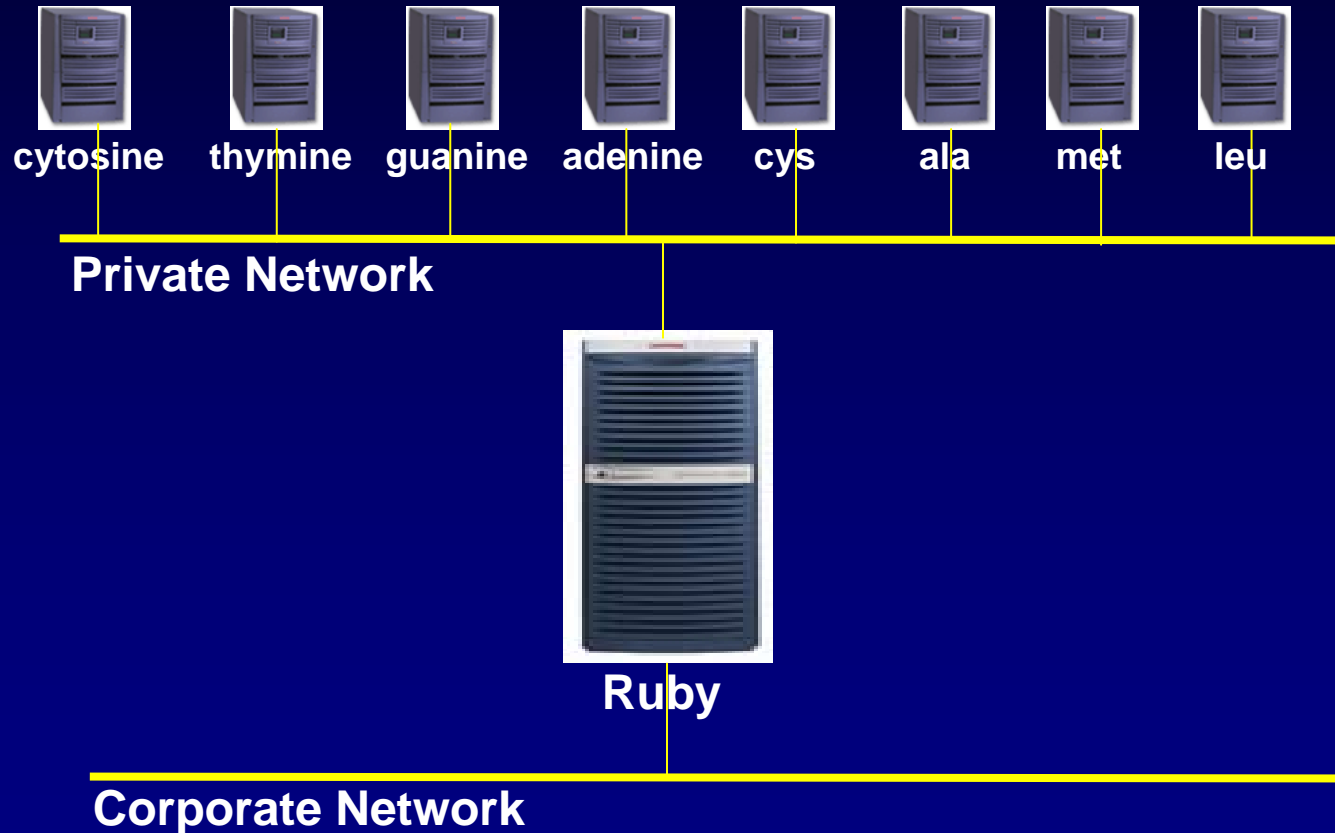
Bioinformatics Computing Environment

Nine Nodes

- 1 8400 (ruby)
 - 10 666MHz EV56 processors
 - 700GB UltraSCSI disk subsystem
 - Tru64 Unix 4.0D
- 2 ES40's (adenine, thymine)
 - 4 500MHz EV6 processors
 - 75GB Ultra2 disk subsystem
 - Tru64 Unix 4.0F
- 6 ES40's (cytosine, guanine, cys, met, ala, leu)
 - 4 500MHz EV6 processors
 - 75GB Ultra2 disk subsystem
 - Tru64 Unix 5.0



Bioinformatics Computing Environment



Functionality

Ruby is used for all user access

- Database
- Web
- Terminal access

ES40's used for back-end processing

- Selected programs modified to distribute load
- Utilizes rsh
- Input files transferred to back-end machine, output files transferred back
- User permissions maintained



Issues

Difficult to administer

- Manual database synchronization
- Shadow “.rhosts”
- Private net

Inflexible

- Would like to have more flexibility in offloading ruby, etc.

Not highly available

- If ruby is down, all 9 nodes are unavailable



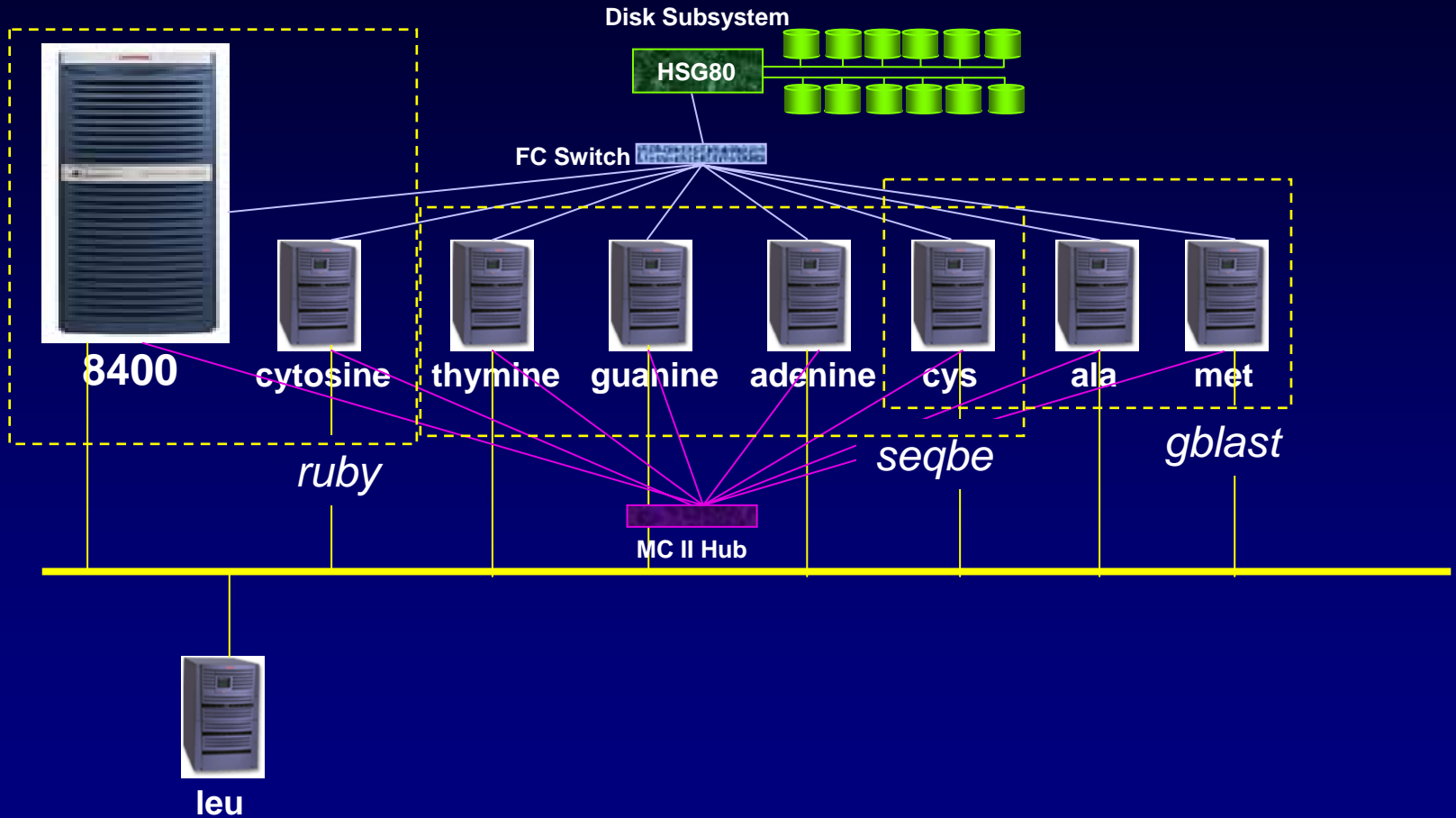
Proposed Architecture

8 node cluster

- 1 8400, 7 ES40s
- FiberChannel disk subsystem (100MB/s)
- Multiple Cluster aliases
 - Ruby: preferred node --> 8400
 - Seq(?): distributed amongst ES40s
- 8th ES40 could be used for Oracle, development system



Proposed Cluster



Advantages

Simplified maintenance

- Single, shared disk subsystem
- Single system image (one root)

More flexibility

- All 8 nodes would be on the corporate network
- Ruby alias could move to an ES40 during 8400 downtime
- Additional aliases could provide for flexible groups and distribution of CPUs

Better performance

- Disk subsystem should be substantially faster



Implementation

Phased approach

- Start with 2 ES40's
- Add additional 5 ES40s when stable
- Add 8400 (ruby) at last step

Issues

- Need to carefully assess performance and scalability of the Cluster File System



Alternatives

Why not LSF or PBS?

- Both were considered
- Want to maintain the user model
- Wanted to avoid using NFS or database synchronization
- Lots of cluster experience for other projects
- 8 node limit not a problem
- Get high availability “for free”



Questions?

