Herding virtual workstations at Google

Michael Hanselmann

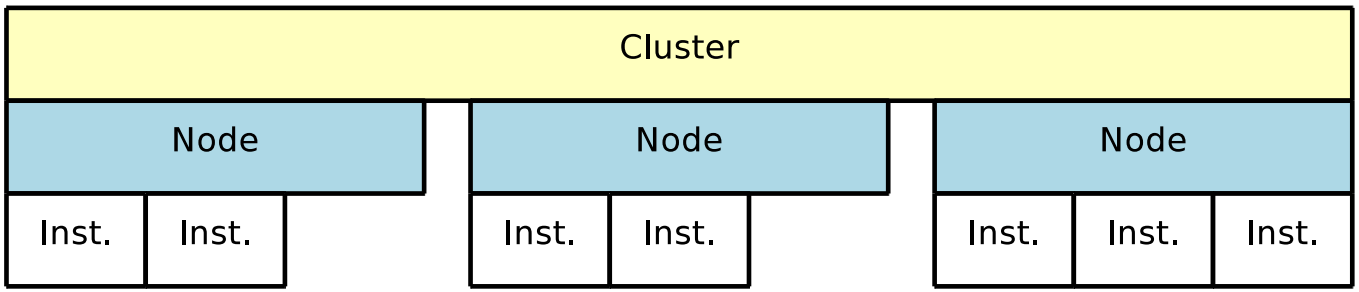
Google Inc.

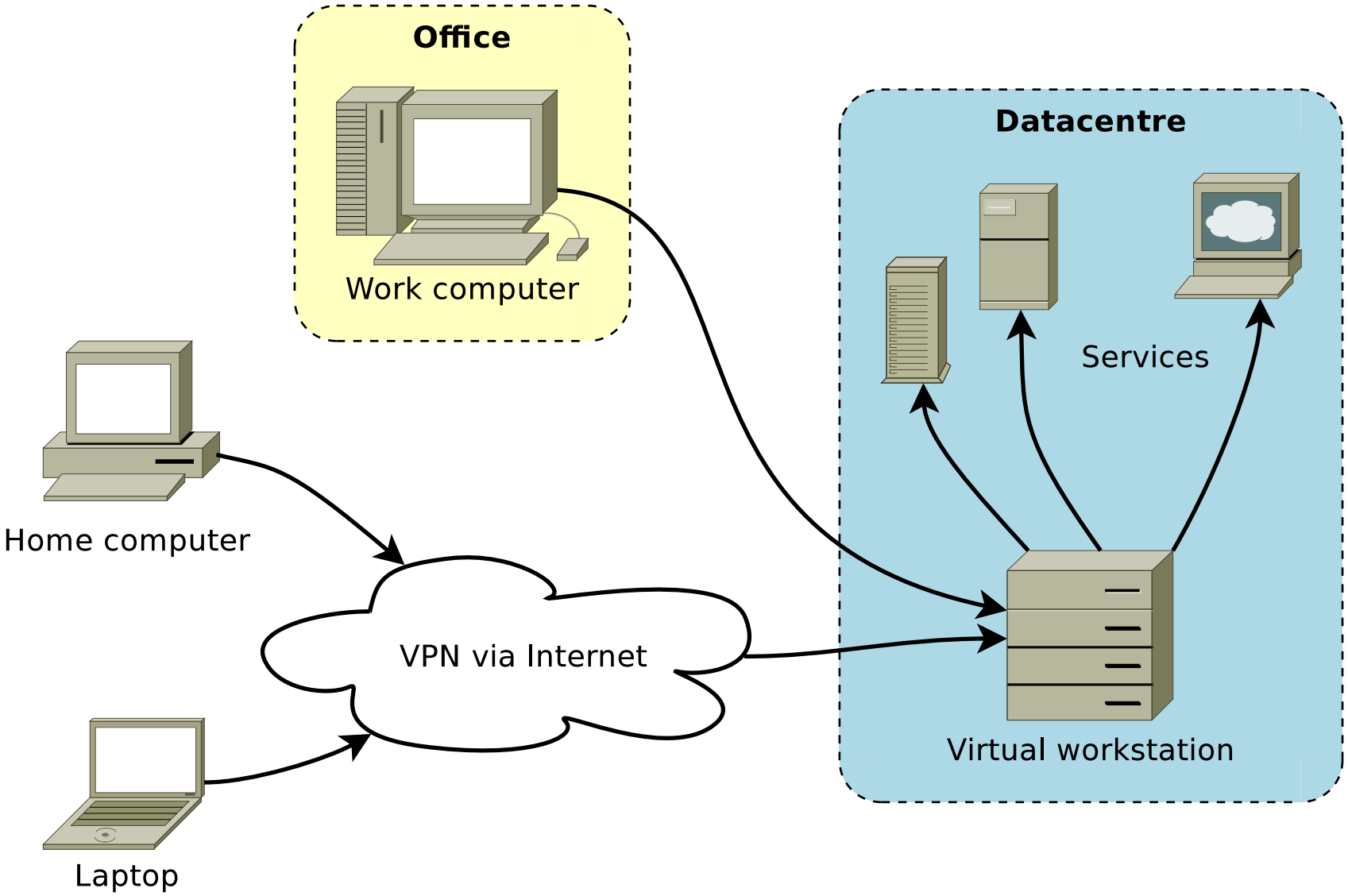
Fórum Internacional de Software Livre 10
Porto Alegre, Brazil
June 24-27, 2009

- `/usr/bin/whoami`
- Talking about implementation at Google
- Concept can be used in other environments
- Terminology
  - Virtualization
  - Cluster
  - Node
  - Instance = machine = virtual machine

- What are virtual workstations?
  - One machine, one owner
  - root like on normal workstation
- Can be implemented on various platforms
  - Bare metal
  - Virtualized systems
    - Partial resilience against hardware failure

- Ganeti 2.0
  - Software to manage clusters of virtual machines
    - Automation allows you to scale easily
    - Makes it simple to manage 10s of nodes and 100s of instances
  - Combines virtualization and data replication
    - Integrated in a unified interface
    - Virtual machines are portable between nodes
  - Hypervisor backends
    - Abstraction layer
    - Backends for Xen and KVM included
  - http://code.google.com/p/ganeti/




Office
Work computer
Datacentre
Services
Home computer
VPN via Internet
Virtual workstation
Laptop

- User perspective
  - Work from office
  - Work from home
    - VPN endpoint near services
  - Work from remote office
    - Slow connection
    - NFS latency
    - Other services
  - Resuming sessions
    - One environment regardless of location

- Reduced operational costs
  - Less support costs
  - Datacentres are more cost efficient than office space
  - Reduce infrastructure in office
- Improved energy efficiency
- Transparent for most use cases
- More with less, more with more

- Upgrade cluster machines and all instance owners on the cluster profit
  - More people per upgraded machine profit than with individual machines
- Monitoring
  - Hardware and software
- Failover/migration
  - Migration transparent to user

- Disaster recovery (DR)
- Easy to turn up/reinstall
  - Infrastructure needs to be in place
- Latency is important
  - 100s of kilometers still possible
  - Clusters in several locations
- Bandwidth not so important

- Requires network connection
  - But so do many applications today
  - Working offline requires copying data
- Reliability (Single point of failure)
  - DR helps
- I/O performance

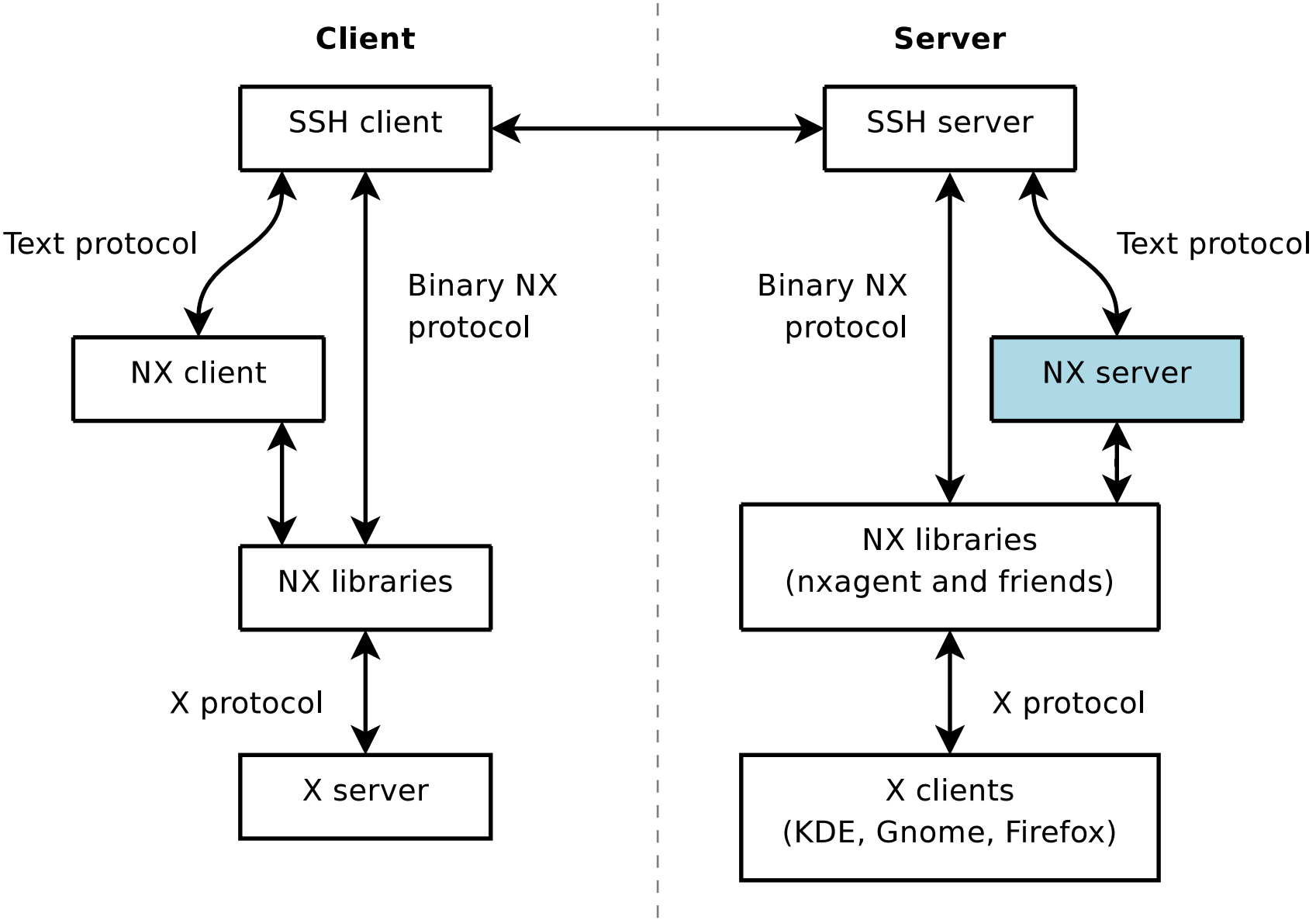
- Usability
  - Partially application's fault
- Local storage and devices not easily accessible from virtual machine
- Multimedia applications
  - 3D
  - Audio I/O
  - Video I/O

- Resource allocation on clusters
  - Part of Ganeti
  - Plugin framework
  - ganeti-htools for balancing
- Network outages

- Low-powered machines (e.g. laptops)
- SSH
- NX

- 1980's: X Window System
  - Many round-trips, latency-sensitive
  - Large amount of redundant data transferred
- 1990's: RFB protocol (used in VNC)
  - Bitmap-based

- 2003: NoMachine NX
  - Diffs instead of full requests
  - Compressed on X11 protocol level
  - Round-trip suppression
  - Proxy on each side
  - Restore session from home or office, keep session state
  - Client available free of charge


Client
Server
SSH client
SSH server
Text protocol
Binary NX protocol
Binary NX protocol
Text protocol
NX client
NX server
NX libraries
NX libraries (nxagent and friends)
X protocol
X protocol
X server
X clients (KDE, Gnome, Firefox)

- Commercial NX server
  - Licence fees
- FreeNX
  - Thousands lines of BASH and Expect scripts

- Neatx
  - Uses NoMachine's Open Source libraries
  - Built on NX technology
  - Written from scratch in Python (and some C)
  - Simpler to set up than commercial NX or FreeNX
  - Not all features implemented
  - Source code available under GPL licence
    - http://code.google.com/p/neatx/
  - Volunteers are welcome!

Questions & Answers