Johan Seland

Cloud Computing and Future Trends

SINTEF Petroleum Development Workshop – Session 3

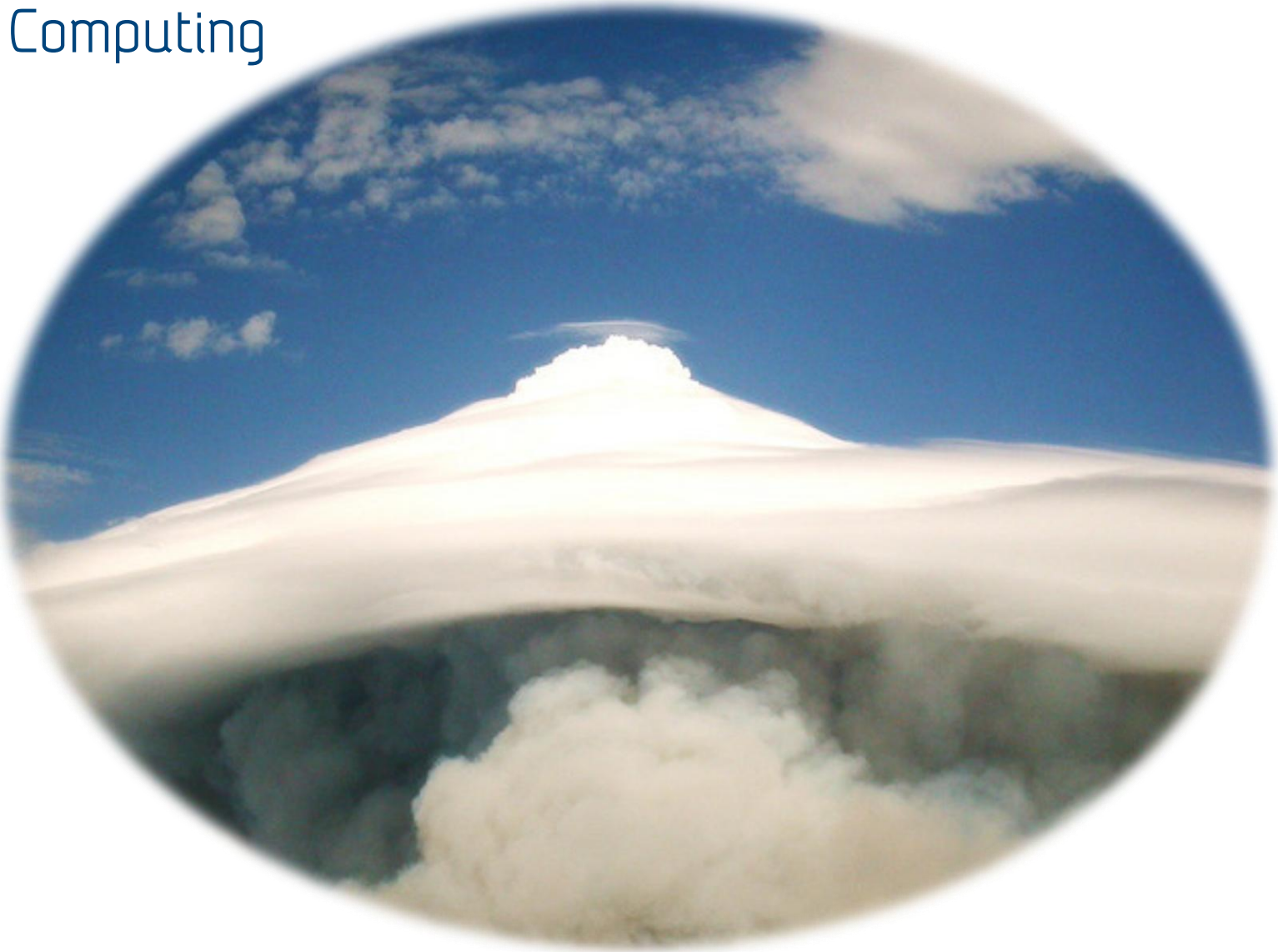
Trondheim - 9. December 2010

Overview

- Cloud Computing
  - Amazon Web Services
  - Case study: CloudSCORE
- Future Trends
  - Computer Architectures
  - Languages and tools
  - You

The best way to predict the future is to invent it.
Alan Kay

Cloud Computing

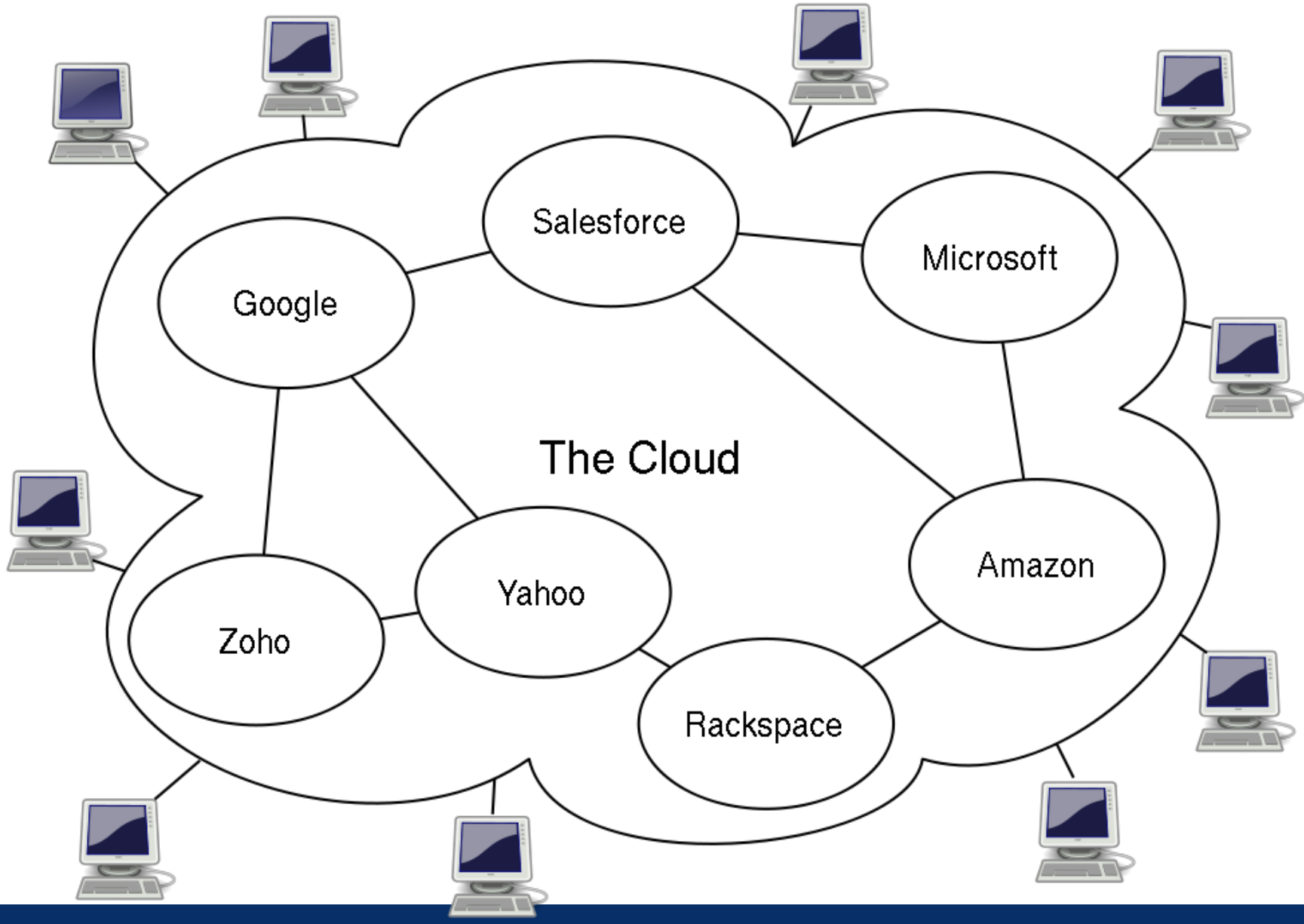


What is it really?

Cloud computing is web-based processing, whereby shared resources, software, and information are provided to computers and other devices on demand, like the smart phones

Keywords:

- Web-based
- Shared resources
- On-demand


Salesforce
Microsoft
Google
The Cloud
Amazon
Yahoo
Zoho
Rackspace

Benefits of cloud computing

- "Infinitely" scaleable
- Pay-as-you go
  - Same price:
    - 1000 hours on one node
    - 1000 nodes for one hour
- Deploy on browser
  - No requirements on users hardware
  - Make GUI work on tablets and cell phones

- Potentially licensing

Problems with Cloud Computing

- Security
  - Who has access to our data?
  - Virtual Private Clouds
- Latency
  - Gaming as a service is coming (OnLive)
- Deployment - Development
  - Yet a new set of tools and APIs
- Licensing
- Lock-in
- Lock-out (Wikileaks)

- What if the cloud goes "down"?
  - Compare to power supply lines

What's different this time?

Is this any different than

- Mainframes, X-terminals and Thin-clients?
  - Scalability (elasticity)
  - Pay-as-you go
  - Web delivery (not tied to vendors client)
- Grid computing?
  - Definitely some overlap (Foster/Kesselman)
  - Non-interactive
  - Batch based
  - Grid has no business model
  - Lot's of cloud technology was developed as grid-technology
  - Friday 17/12-2010 – Andre Brodtkorbs trial lecture (Oslo):
    - *Cloud Computing – How is it different from Grid Computing?*

Everything as a service

The cloud can be seen as the combination of:

- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Data as a Service (Daas)
- Utility as a service (UaaS)

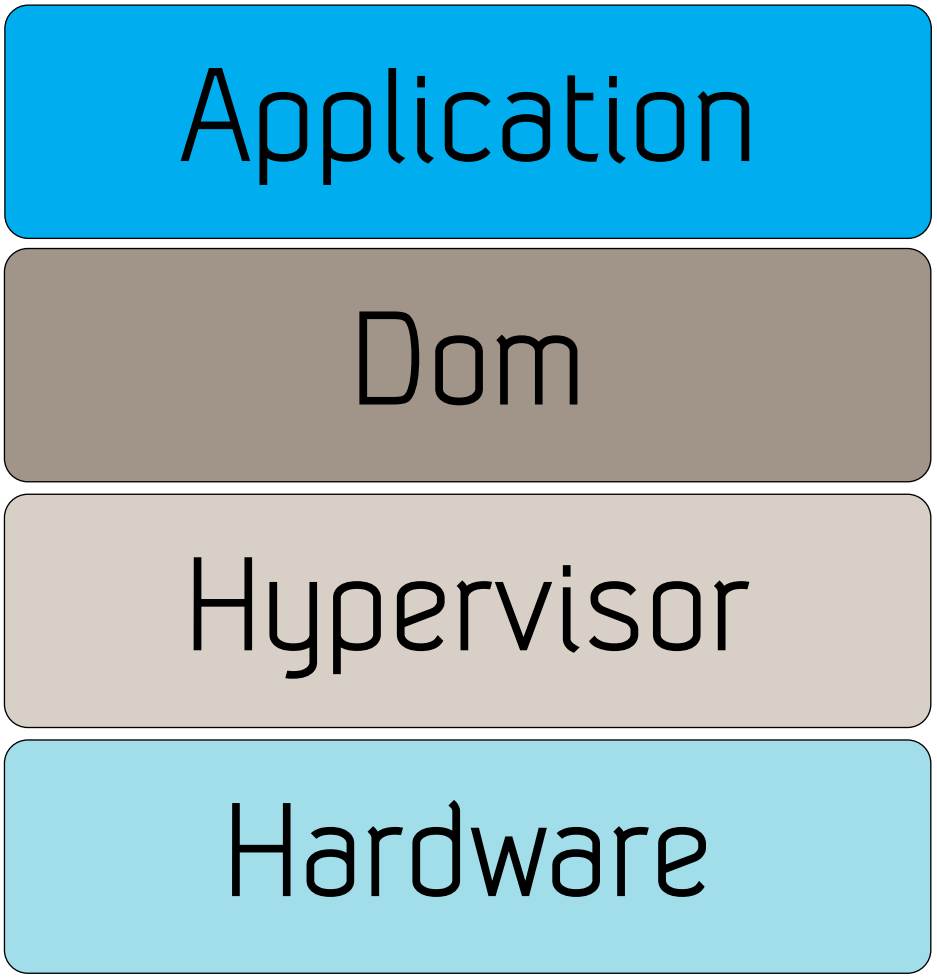
- Is the enterprise ready for this?
  - SINTEF/ERGO
  - A larger Norwegian Oil Company

- Can it afford not to?
  - Economy of scale
  - Is our enterprise large enough to host a cloud?

Virtualization – "Abstracted Hardware"

- One physical server runs multiple OS
- Allows higher utilization of servers
- Useful in development environments
  - Virtualbox, VMWare

- Typically 97%++ barebones performance
- Dedicated drivers can give HW access



AMAZON WEB SERVICES

AWS Service Umbrella

- Compute
- Messaging
- Storage
- Content Delivery
- Monitoring
- Database
- Networking
- Web Traffic
- E-Commerce
- Payments
- Workforce

Most interesting to us

- Elastic Compute Cloud (EC2)
  - On-demand servers
- Amazon Elastic Block Store (EBS)
  - Persistent off-instance storage
- Simple Storage Service (S3)
  - HTTPS-based interface for loads/stores
- Databases?
  - SimpleDB (NoSQL) and RDS (Relational)

Example Pricing

Service	Price
Micro Instance	$0.02 / hour
Large CPU Instance (7.5 GiB RAM, 2 cores)	$0.34 / hour
High CPU Instance (7.0 GiB Ram, 8 cores)	$0.68/ hour
Cluster Compute Instance (23 GiB, 8 cores, 10GiB Ethernet)	$1.60/ hour
Cluster GPU Instance (22 GiB, 8 cores, 2 Tesla GPUs, 10GiB Ethernet)	$2.10/hour
High Redundancy S3 Storage	$0.14 / GiB / Mnth
Reduced Redundancy S3 Storage	$0.093 / GiB / Mnth
EBS Storage	$0.10 / GiB / Mnth
Small MySQL Instance	$0.22 / hour

EC2 Overview

- Interfaces
  - Web based console
  - Command line tools
  - APIs
- API Bindings
  - Java
  - PHP
  - Python
  - Ruby
  - .NET
- No C/C++/Fortran

EC2 Terminology

- Instance
  - A running virtual machine
- Instance Type
  - Which "hardware" to run on
- AMI
  - Amazon Machine Image
- Region
  - Physical location of instance
- Key pair
  - A public/private key pair used to login to instances
- Security Groups
  - Manages the firewall settings of instances

CLOUDSPH

CloudSPH

Main Features:

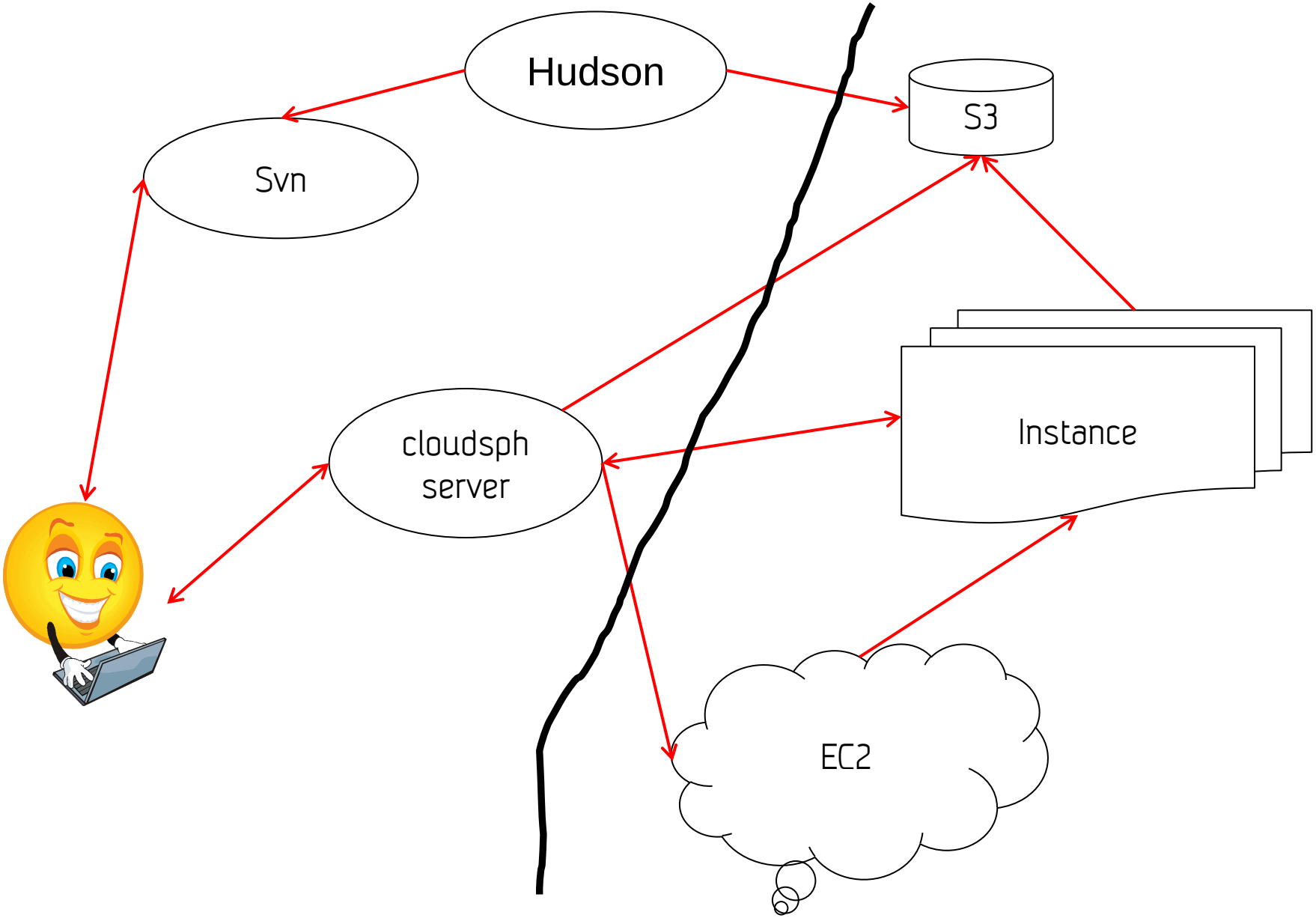
- SPH Simulations take a long time
- Start instances on EC2 from browser
- Download result in background
- Automatic building of simulator from SVN

- ≈ 5000 lines of code

- **AVAILABLE TODAY!**

Technical info

- Mashup of many technologies
  - SCORE simulator in C++/OpenMP/CUDA
  - Java for server side logic
  - Generated Javascript for web GUI
  - Ubuntu Linux on EC2-instances
  - SSH for communication with instances
  - Shell scripts on instances
  - A sparkle of XML for static data
  - Hudson to CI server to build binaries – push into S3
- Encrypted communication (https and SSH)


Hudson
S3
Svn
cloudsph
server
Instance
EC2

Google Web Toolkit

- Tools that compile Java into Javascript
  - Not full Java library on client
- Develop server and client code simultaneously

- Async calls between client and server

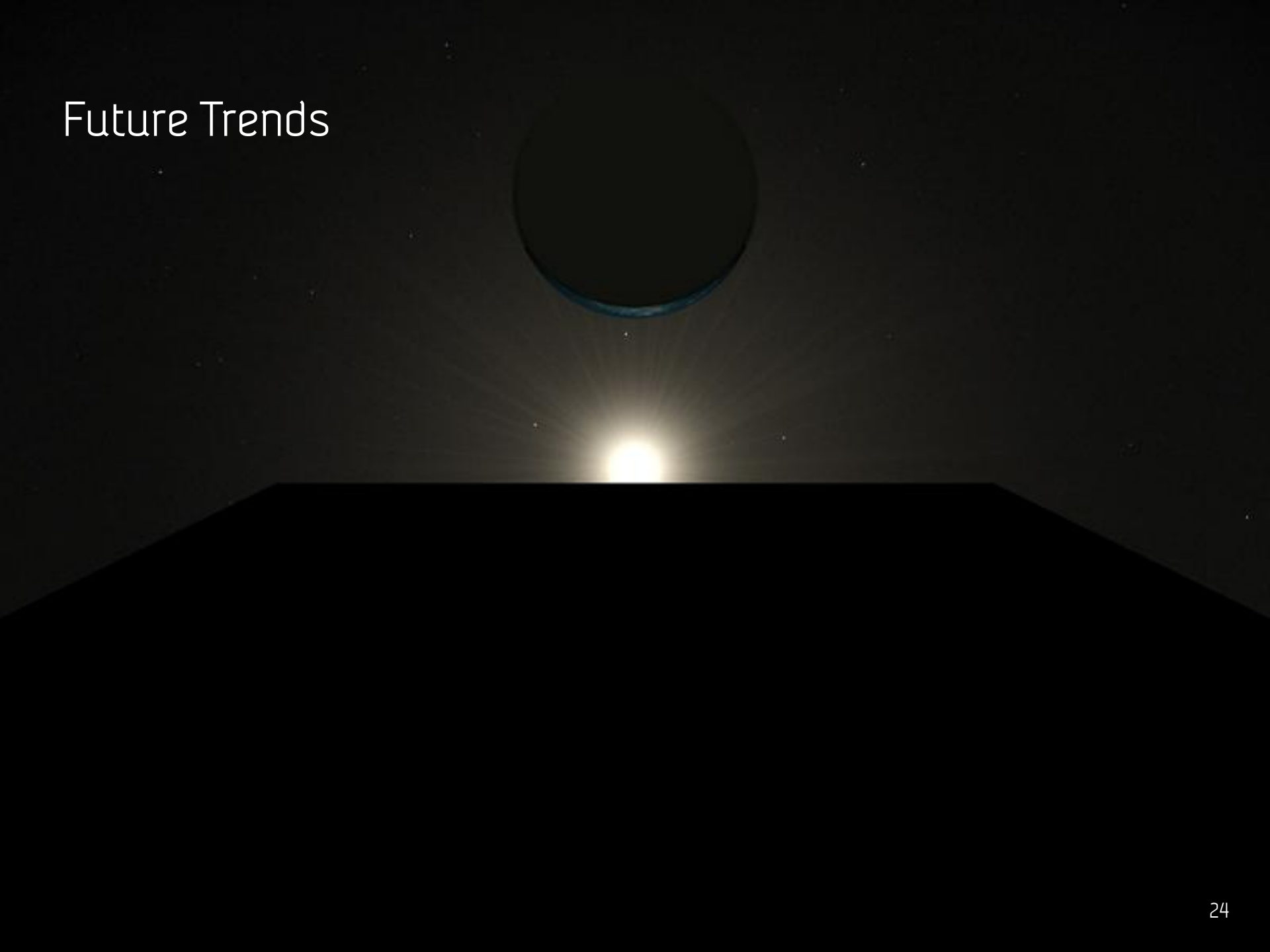
Experiences

- Easy to forget instances
  - $$$$
- Relatively slow to transfer data out
  - Uses US region
- APIs from Amazon are good and well documented
- GWT is relatively pleasant
- Synchronization issues

Cloud Questions and Discussion

- How to get started:
  - aws.amazon.com
- Time to start playing with web GUIs

Future Trends



Future Trends

- Computer Architectures
- Languages and tools
- You

CPU Roadmaps (Server/Workstation)

- 2011: Intel Sandy Bridge, 16-core AMD Bulldozer, AMD Fusion
  - On-chip low/midrange GPU
  - Dynamically scale power between CPU and GPU
- 2012: 16-core Ivy Bridge, AMD NG Bulldozer/Fusion
  - NUMA
- 2013: Haswell
  - 22 nm process
  - New cache design
  - Fused Multiply-Add
- 2014: Rockwell
  - 16 nm die shrink of Haswell

Speculation next 5-10 years

- Moores law will hold (Intel has plans to 2029)
- Focus on TDP (thermal design power) rather than performance
- Number of cores will continue to increase
- Dynamic clocking
- No jump in clock frequency on the horizon
  - Exotic cooling solutions in server rooms
  - Cooling of server rooms already problematic
- New memory hierarchies
- More vector units
  - CPU/GPU fusion

GPU Roadmaps and speculation

What we know:

- 2011: Nvidia Kepler (28nm, 1.4 TFlops?)
- 2013: Nvidia Maxwell (22nm, 3.9 TFlops?)
  - 330 TFlops in a 42U rack
  - 3 racks will make a petaflop machine

Speculation:

- "Big-iron" features in high end GPUs
  - NUMA
  - Virtual Memory
  - Virtualization

Languages, compilers and curriculums

- Hardware is 5++ years ahead of mainstream languages
  - C++/Java/C# is getting good support for task-parallelism
  - Data-parallelism in libraries (90/10-rule)
  - Network support is good in most languages except Fortran and C/C++
- Lots of research languages

- Full auto-parallelizition is a dream
  - Even if compiler research has seen a boost lately

- Parallelization is still an "advanced" topic in CS curriculums
  - Ongoing debate in ACM/IEEE

Language speculation

- Programmer should expose parallelism
  - Compiler backends for various hardware
  - Language VMs (JVM, CLI) will optimize just-in-time
- Functional languages might make a comeback
- HW optimized libraries
  - BLAS, FFTs for vendors
  - Experts (Applied Maths ☺) should write the kernels

- Scalability more important than optimal algorithms?
- Cloud constructs in languages?

Long Range Speculations

Scary to make predictions more than a few years

- Moores law will probably hold (Krauss/Starkman predict 600 years)
- Memristors
  - Combine HDD and RAM first (HP 2013?)
  - Memory and logic on the same chip
  - Many-to-many communication
- Everything will be networked, cheap and small
  - Networked vessels in bloodcells?
- Optical or quantum computers
  - A new jump in "frequency"
  - 500 GHz supercooled, single transistor has been demonstrated (IBM)

- 2100++ Monoliths, Matrioshka brains?