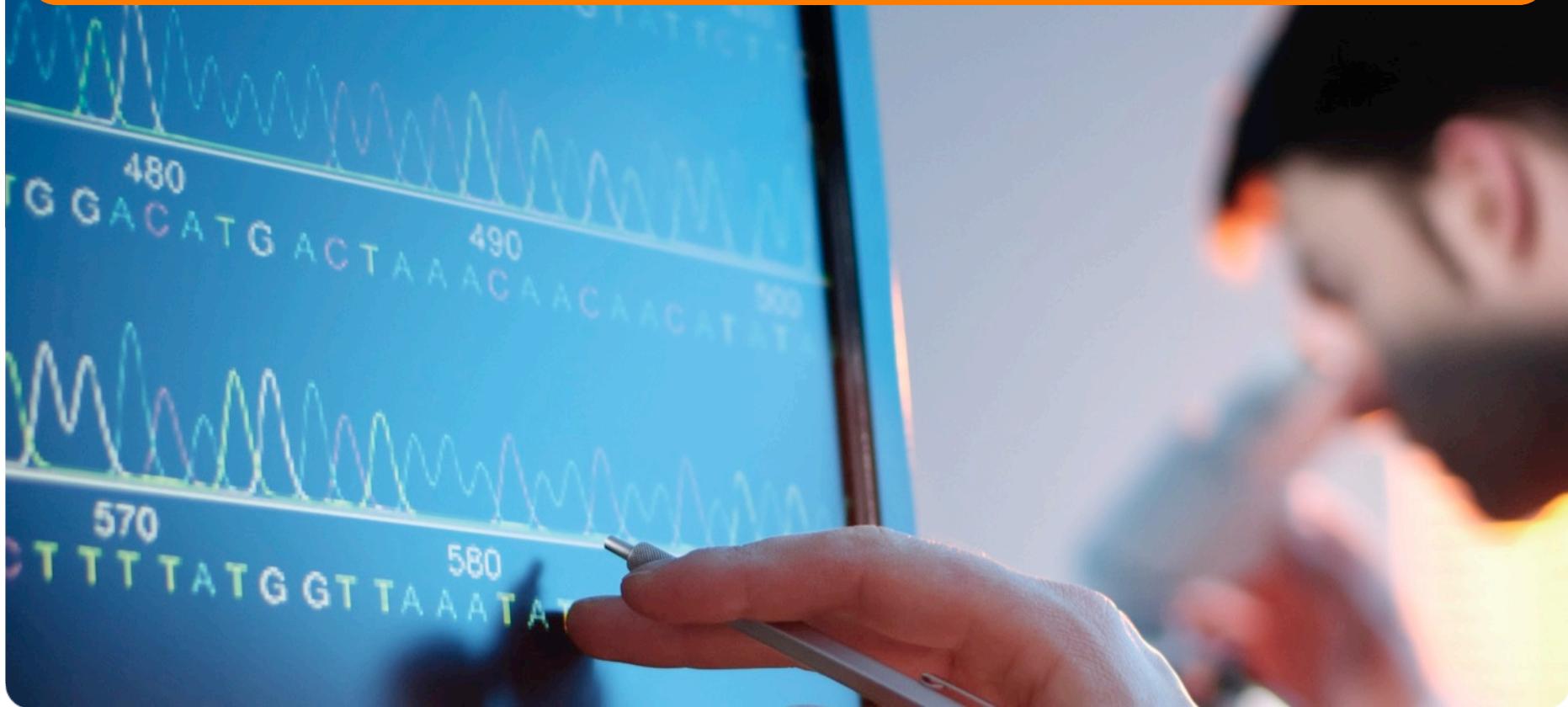


# HPC CLOUD INTRODUCTIE



Jan Bot, SURFsara  
15 oktober 2015

**SURF**

# lets over jullie



# Doel van vandaag

- Introductie van het nieuwe HPC Cloud
- Kweken van een gezamelijk vocabulaire
- [S4R] Gelegenheid tot netwerken

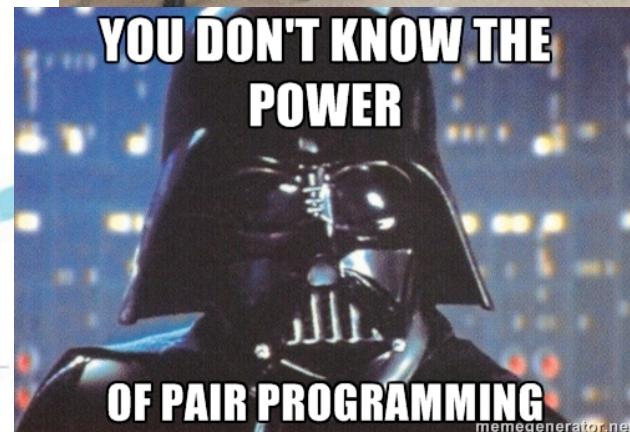
# De hands-on

## PROGRAMMING



ONE  
MACHINE

EFFICIENT  
TEAM



SURF

# Het programma

10:15 Inloop

10:30 Introductie Cloud Computing (← you are here)

11:00 Introductie HPC Cloud

11:30 HPC Cloud demo

12:15 Lunch

13:00 Hands-on HPC Cloud

15:15 Wrap-up & vragen

**15:30 Borrel**

16:00 Eind

# CLOUD: EEN INTRODUCTIE



Jan Bot, SURFsara  
5 augustus 2015

**SURF**

# Dit praatje ...

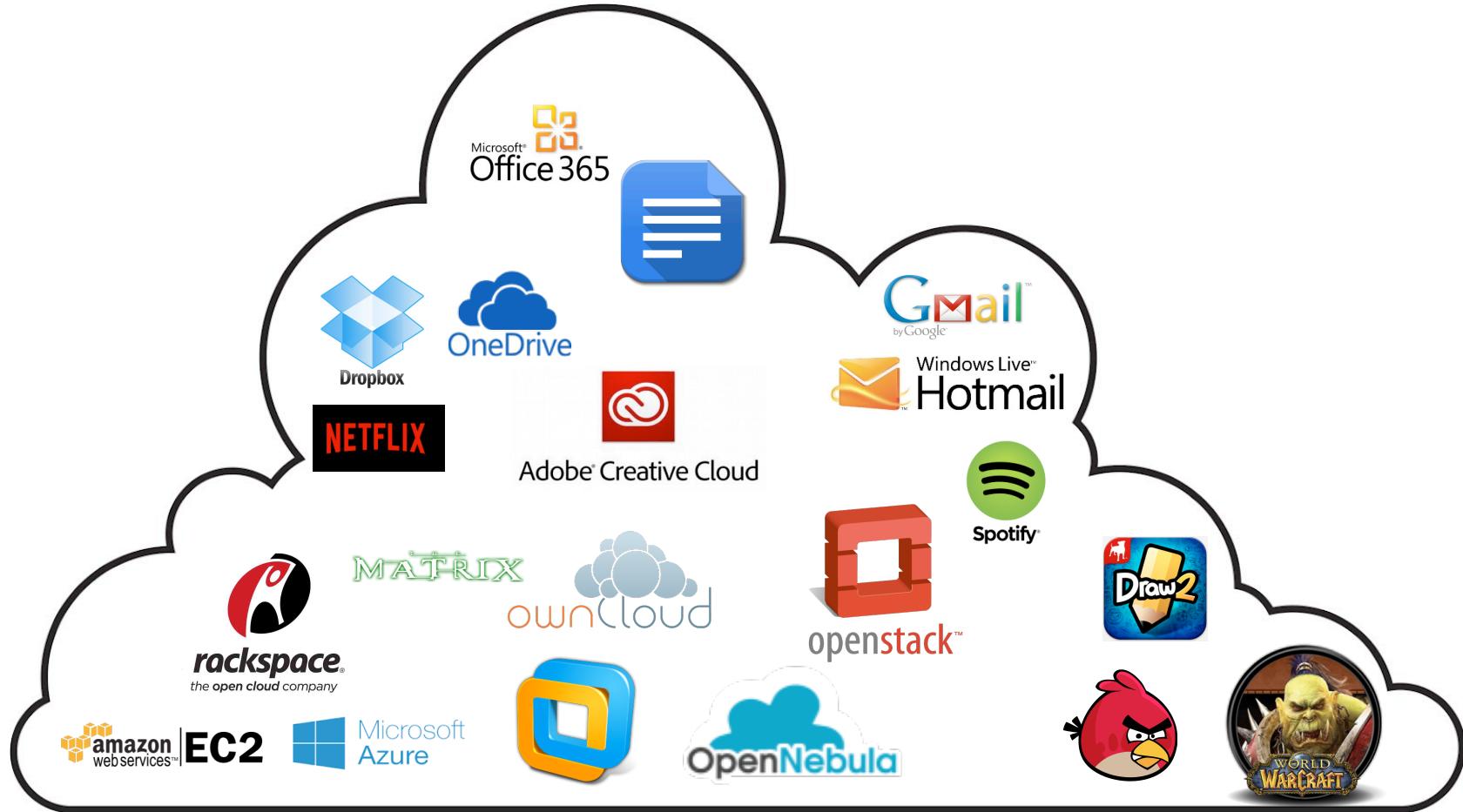
- bevat de basis informatie over cloud computing
- moet zorgdragen voor een gedeeld vocabulaire
- geeft een overzicht van de verschillende vormen van cloud
- plaats de SURF(sara) services in perspectief
- laat zien wat (hpc) cloud *niet* is

# Compensatie endorfine



SURF

# Say 'Cloud' one more time...



SURF

# Waarom Cloud?

Ontlasten van de eigen ICT organisatie door een dienst bij een derde partij af te nemen.

## Voordelen

- Geen instap kosten
- Geen eigen beheer
- Dynamisch schaalbaar: betaal wat je gebruikt

## Nadelen

- Weinig tot geen controle over je data
- Slechte controle over hardware
- Verborgen kosten

## Cloud types

- Private / community / public
- Dedicated / overcommitted
- Single / multi tenant

# \* as a Service

- **Infrastructure:** het huren van hardware voor het opzetten van eigen virtuele machines
- **Platform:** een software platform om je eigen applicaties op te ontwikkelen
- **Software:** diensten die direct aan eindgebruikers aangeboden kunnen worden

# Cloud lagen: \* as a Service



**SaaS**  
(software)



**PaaS**  
(platform)



**IaaS**  
(infrastructure)



# Orde in de chaos



SURF

**SaaS**  
(software)



SURF DRIVE

?

**PaaS**  
(platform)



BoD

**IaaS**  
(infrastructure)



HPC Cloud

Kubus

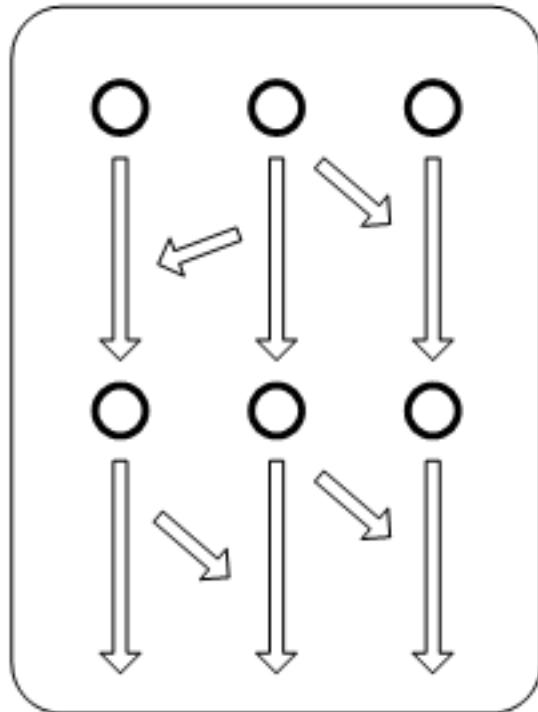
PopCo

SURF

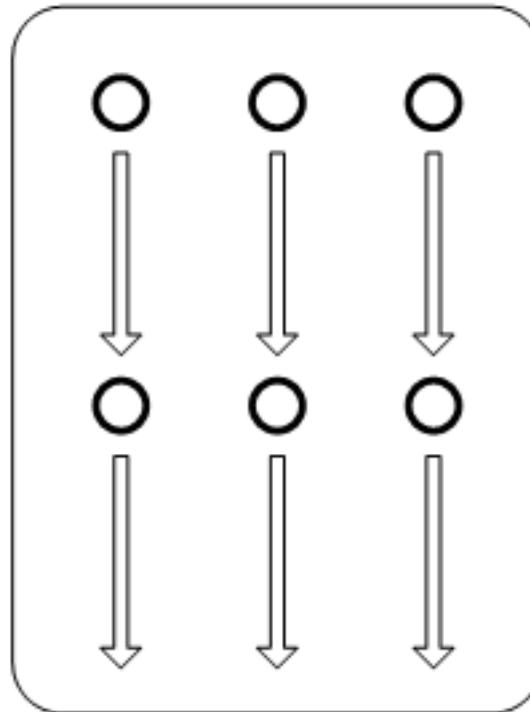
# De SURF rekeninfrastructuur

	Cartesius	NRC/LISA	Grid	HPC Cloud	Hadoop
The machine					
Cores	40,960	7856	5000	2408	1370
Memory	117 TB	26 TB	40 TB	21 TB	4 TB
*aaS	PaaS / SaaS	PaaS / SaaS	PaaS	IaaS	Paas
Unique	<ul style="list-style-type: none"><li>• GPGPU</li><li>• Infiniband</li></ul>	<ul style="list-style-type: none"><li>• Infiniband</li><li>• Customizable workspaces</li></ul>	<ul style="list-style-type: none"><li>• Scale up to European or world wide resources</li><li>• Large local scratch space</li></ul>	<ul style="list-style-type: none"><li>• Own OS</li><li>• GPGPU (soon)</li></ul>	<ul style="list-style-type: none"><li>• Map/Reduce or Spark driven</li><li>• Local storage only</li></ul>

# Parallel Computing

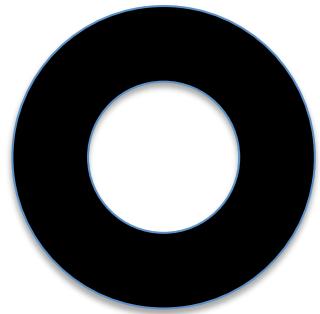


Supercomputing



Cluster  
Cloud

Grid  
Hadoop



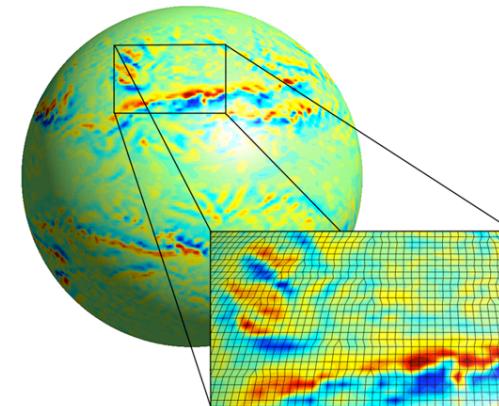
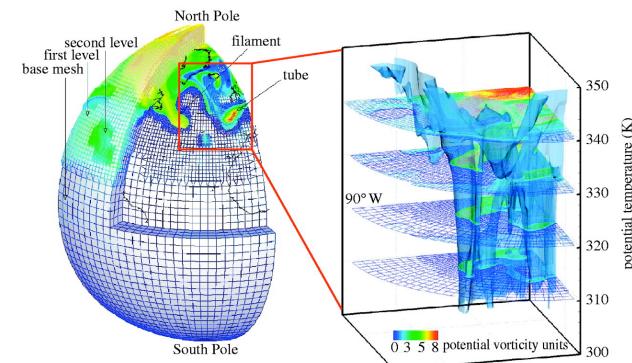
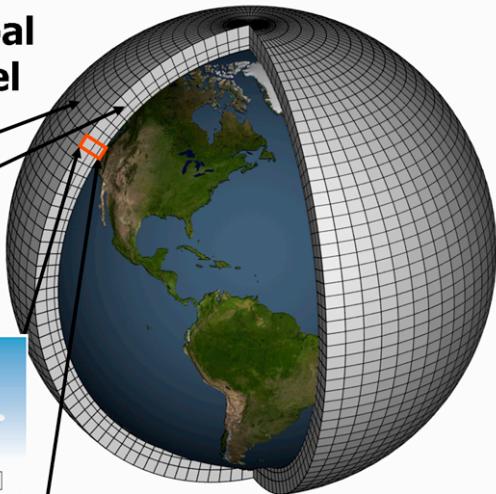
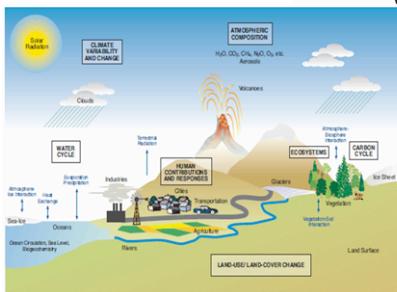
Cloud

# Communication dependent tasks

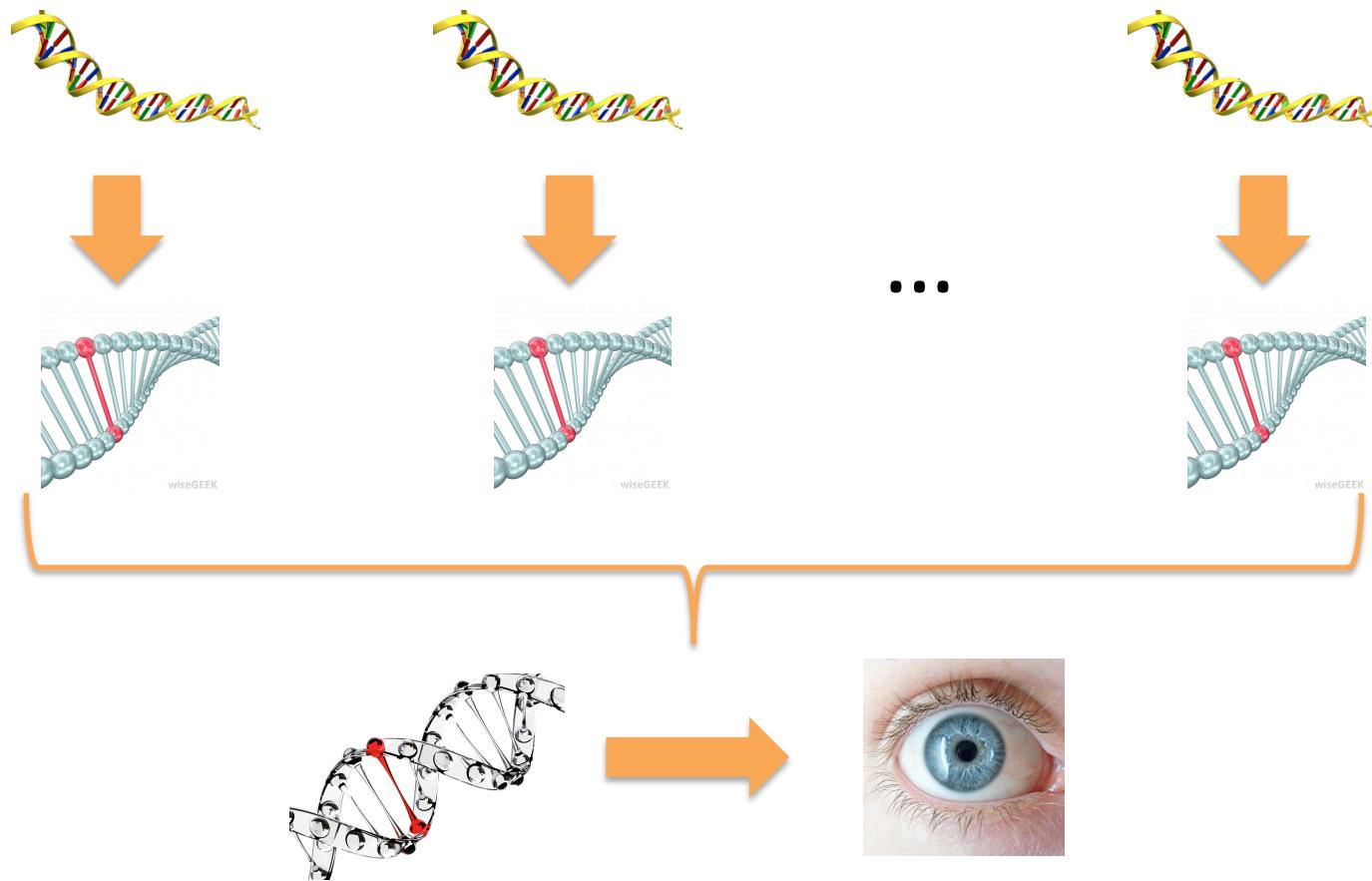
## Schematic for Global Atmospheric Model

Horizontal Grid (Latitude-Longitude)

Vertical Grid (Height or Pressure)



# Data parallel tasks



# Typisch e-infrastructuur gebruik

Infrastructure	Typical usage	Typical users	Use cases
Cartesius	<ul style="list-style-type: none"><li>• Capability computing</li><li>• MPI applications</li></ul>	<ul style="list-style-type: none"><li>• Research groups</li><li>• Consortia</li></ul>	<ul style="list-style-type: none"><li>• Climate modeling</li><li>• Fluid dynamics</li><li>• Galaxy simulation</li></ul>
NRC/LISA	<ul style="list-style-type: none"><li>• Job farming</li><li>• MPI applications</li></ul>	<ul style="list-style-type: none"><li>• Research groups</li><li>• Individual researchers</li></ul>	<ul style="list-style-type: none"><li>• GWAS studies</li><li>• Magnetic field research</li></ul>
GRID	<ul style="list-style-type: none"><li>• Job farming</li><li>• I/O heavy applications</li></ul>	<ul style="list-style-type: none"><li>• Consortia</li></ul>	<ul style="list-style-type: none"><li>• LHC experiments</li><li>• LOFAR</li><li>• NGS data analysis</li></ul>
HPC Cloud	<ul style="list-style-type: none"><li>• 3<sup>rd</sup> party PaaS</li><li>• Courses</li><li>• High Memory apps</li><li>• Microsoft Windows (&amp; other licensed software)</li></ul>	<ul style="list-style-type: none"><li>• Individual researchers</li><li>• Consortia</li></ul>	<ul style="list-style-type: none"><li>• Galaxy platform</li><li>• AMC compute course</li><li>• <i>De novo</i> sequence alignment</li><li>• Stock exchange ticker data</li></ul>
Hadoop	<ul style="list-style-type: none"><li>• Pattern recognition / data mining</li><li>• Web crawls</li><li>• Information retrieval</li></ul>	<ul style="list-style-type: none"><li>• Individual researchers</li></ul>	<ul style="list-style-type: none"><li>• Twitter mining</li><li>• TwiNL</li><li>• SETI analysis</li></ul>

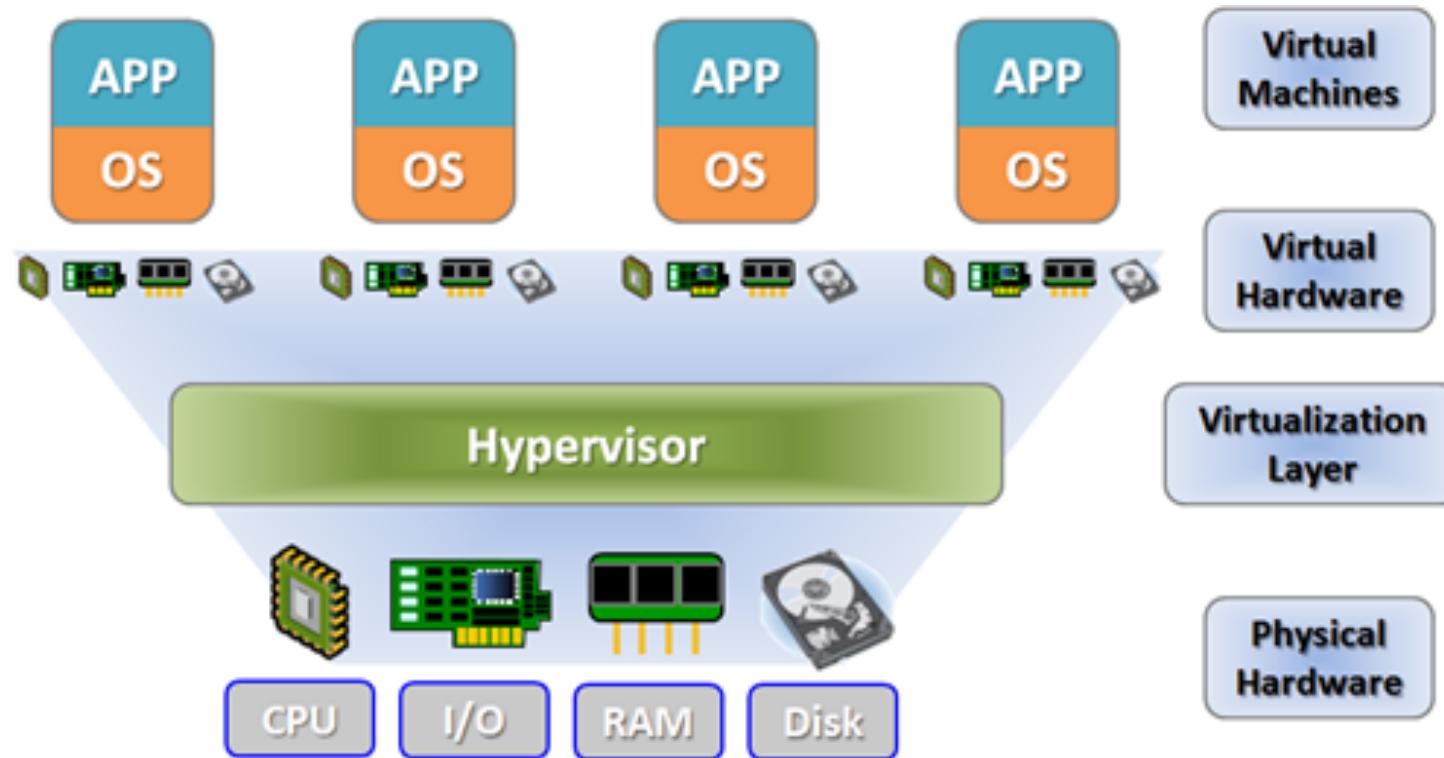
# Infrastructuur overwegingen

- Technisch
  - Theoretische geschiktheid infrastructuur
  - Huidige geschiktheid infrastructuur
  - Beschikbaarheid infrastructuur
  - Toekomstige ontwikkelingen
  - Learning curve
- Organisatorisch
  - Beschikbaarheid infrastructuur
  - Beschikbaarheid support
  - Samenwerkingsverband & omgevingseisen
  - Veiligheidseisen
- Gebuikers
  - Aantal
  - Ervaring
- Etc.

# HPC Cloud

- Een flexibele, schaalbare infrastructuur voor het draaien van simulaties of het verwerken van data
- Beschikbaar vanaf een single core machine tot multi-node cluster
- Web-interface om machines mee te managen

# Virtualisatie: Hoe IaaS werkt



[www.definethecloud.net](http://www.definethecloud.net)

# Waarom HPC Cloud

- Generiek
  - Data in NL
- Tov andere aanbieders
  - Ongelimiteerd dataverkeer
  - Geen overcommit
- Tov andere e-infrastructuren
  - Grafische interface
  - Volledige controle over omgeving
  - Gescheiden netwerk
  - Geen maximale procestijd
  - Relatief eenvoudig in gebruik
  - Eenvoudige aanvraag procedure

# Waarom niet HPC Cloud

- Onderhouden van VMs
  - Updates
  - Account maintenance
- Beperkte schaalbaarheid
  - Beperkt binnen HPC Cloud
  - Geen automatische verdeling van werk
  - Geen Infiniband
- Reken-specifiek
  - Geen hosting service

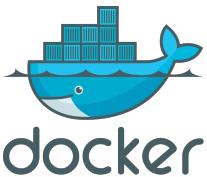
# (HPC) Cloud misvattingen

- Magische versnelmachine: de huidige e-infrastructuur is, per machine, niet sneller dan jouw laptop. Opschalen werkt door meerdere machines te gebruiken.
- Makkelijk verdelen van taken: zit niet in HPC Cloud ingebouwd, dat zul je zelf moeten regelen.
- Oneindig schaalbaar: zoals iedere infrastructuur heeft het HPC Cloud een beperkte omvang. Opschalen over verschillende IaaS platformen heen is lastig.

# HPC Cloud conclusie

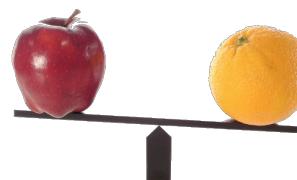
- HPC Cloud is een flexibele, gebruiksvriendelijke omgeving die breed inzetbaar is voor wetenschappelijke rekentaken
- Van alle SURF infrastructuren is HPC Cloud het meest flexibel
- Gebruikers zijn zelf verantwoordelijk voor het inrichten van hun machines, hier geldt ‘with great power comes great responsibility’
- Het is geen panacea, voor grote probleem-sets kun je beter één van de andere platformen kiezen
- Voor een aantal scenarios, waaronder het draaien van Microsoft Windows, is geen alternatief beschikbaar

# Waar we het niet over gehad hebben

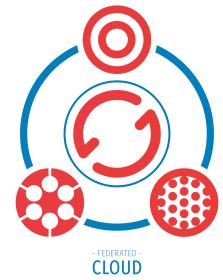


LXD

The Linux container hypervisor  
Fast, dense and secure container management for Ubuntu



EUROPEAN GRID INFRASTRUCTURE  
EGI SOLUTIONS



esgi  
WWW.EGI.EU



SURF

WHAT SURF CAN DO