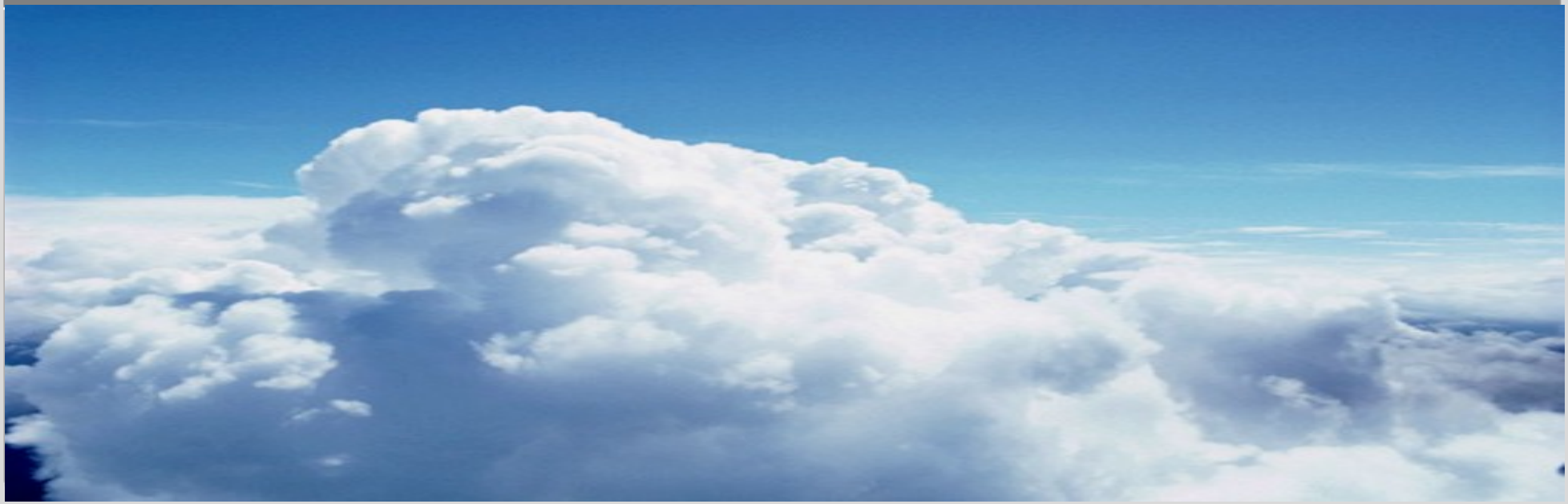
Karlsruhe Institute of Technology

Grid as a Service and Cloud Systems Research

Marcel Kunze, Steinbuch Centre for Computing



The Cloud is just a new Computing Platform

- 70's: Mainframe
- 80's: PC
- 90's: Workstation
- 00's: Grid
- 10's: Cloud

From http://blogs.zdnet.com/Hinchcliffe

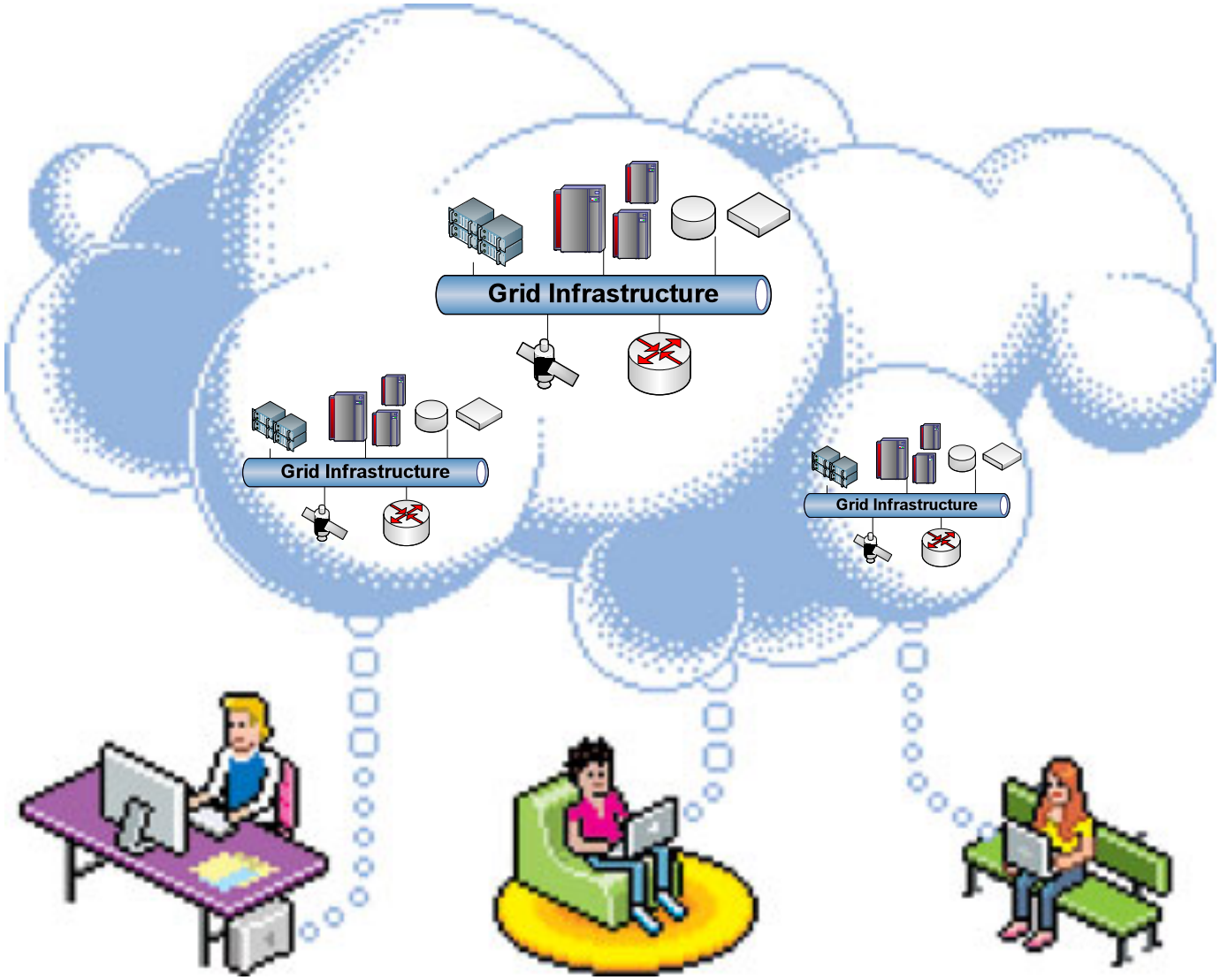
Cloud Computing: A possible Definition

Karlsruhe Institute of Technology

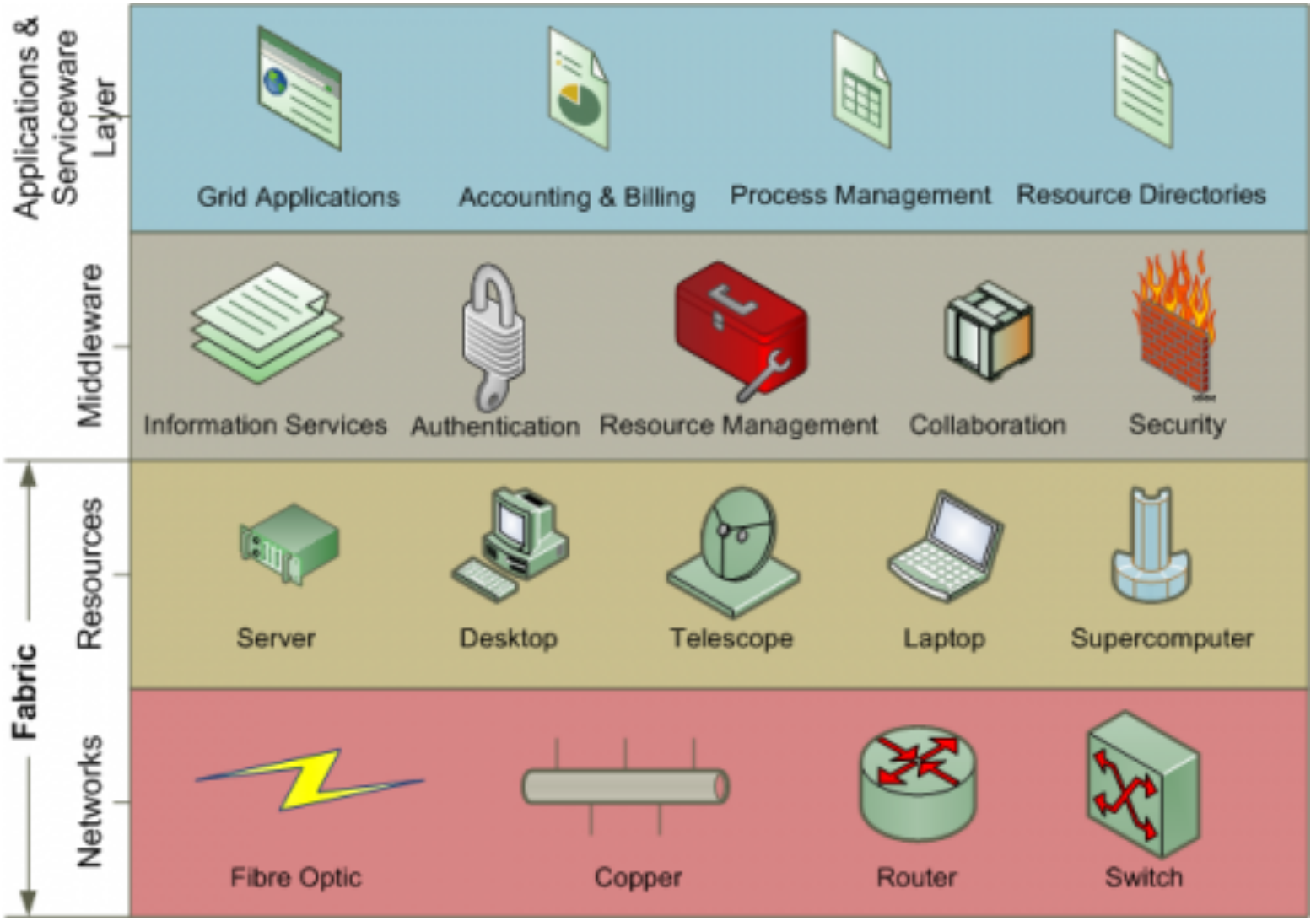
"Building on compute and storage virtualization, and leveraging the modern Web, **Cloud Computing** provides scalable, network-centric, abstracted IT infrastructure, platforms, and applications as on-demand services that are billed by consumption."

C.Baun, M.Kunze, J.Nimis, S.Tai: Cloud Computing, Springer 2009

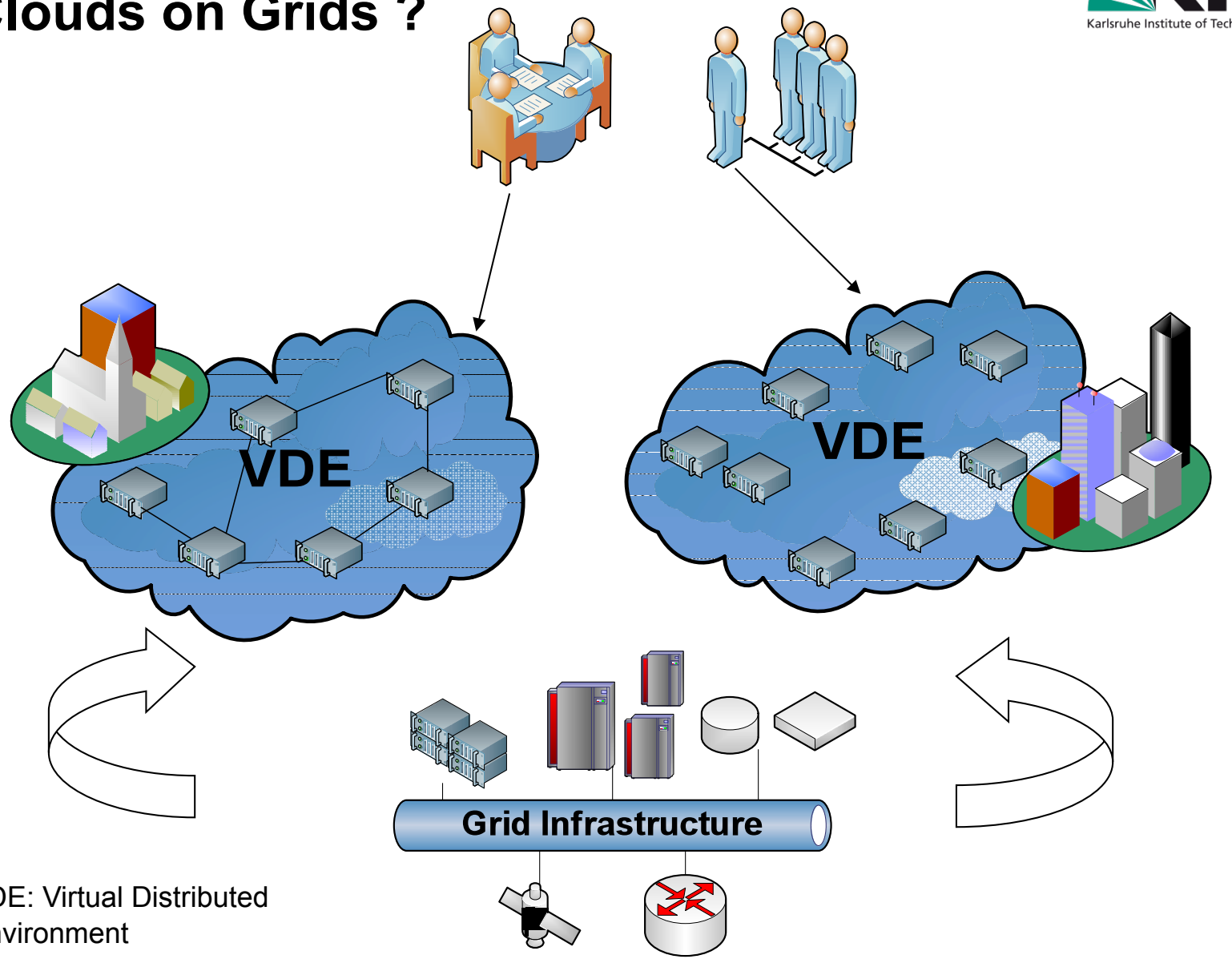
Part I: Grid as a Service



Conventional Grid Architecture

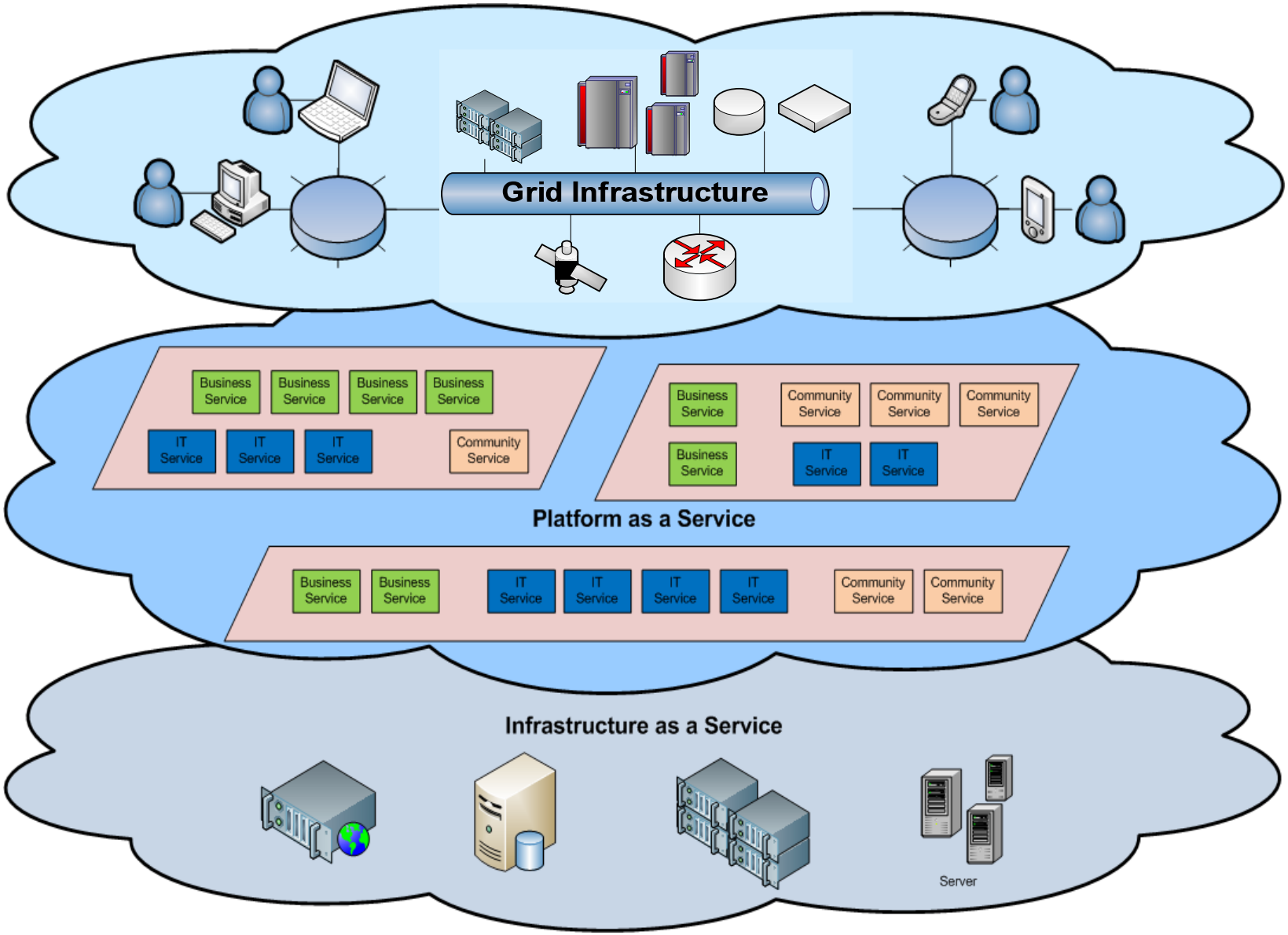


Build Clouds on Grids ?

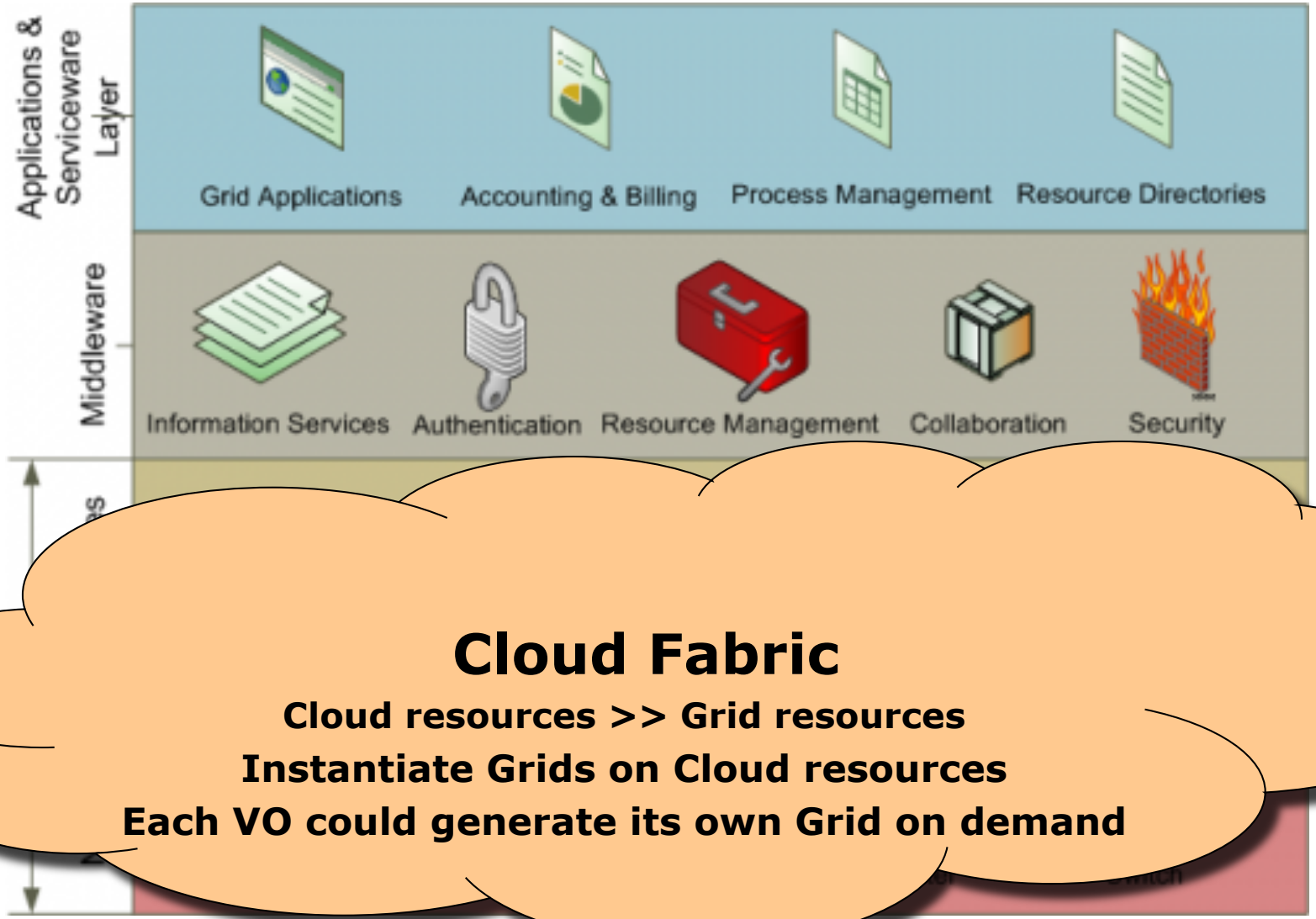


VDE: Virtual Distributed Environment

Build Grids on Clouds (Grid as a Service) ?



Grid as a Service looks promising



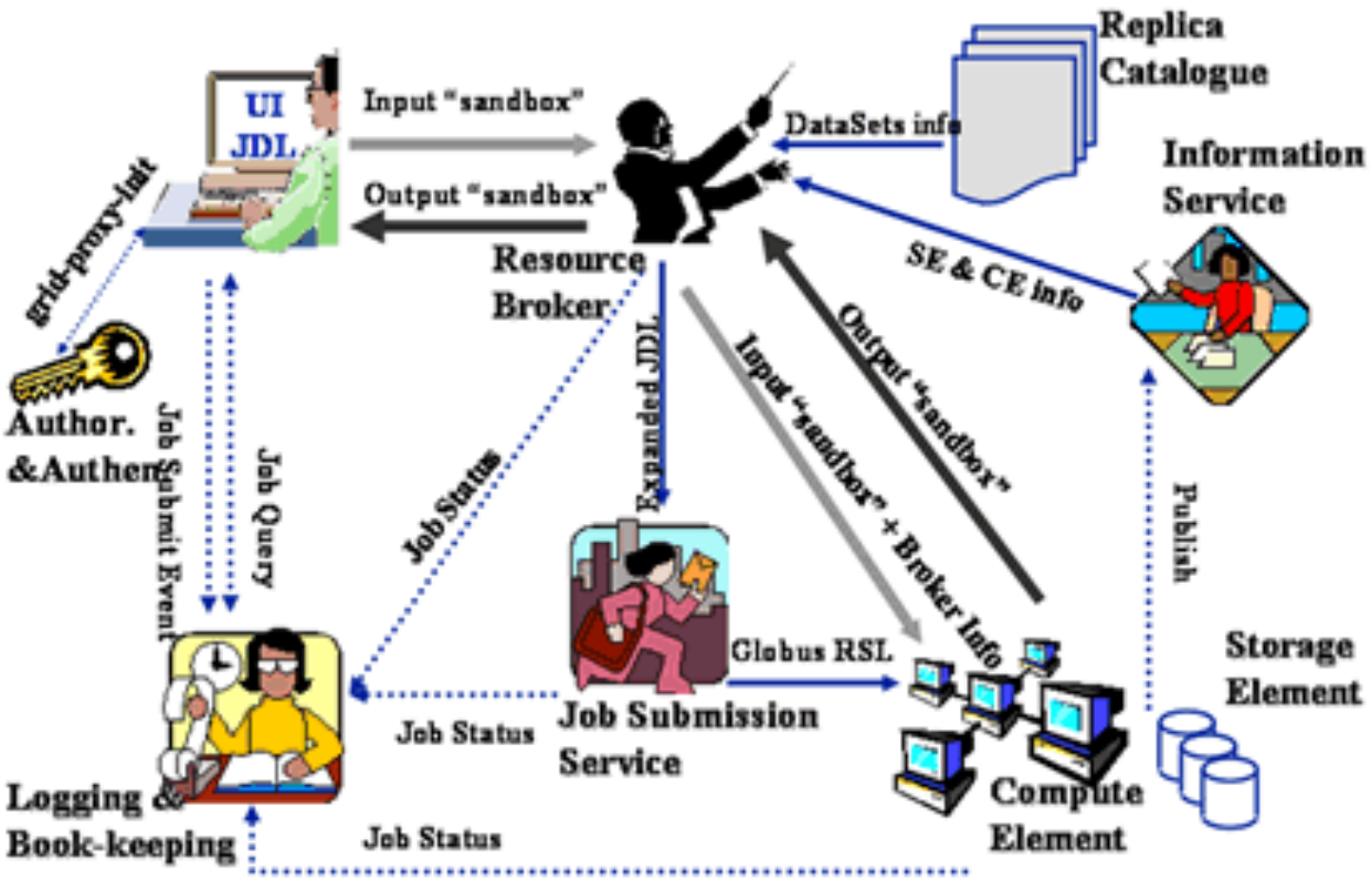
Cloud Fabric

Cloud resources >> Grid resources

Instantiate Grids on Cloud resources

Each VO could generate its own Grid on demand

New Job Flow Scenarios: Grid vs. Cloud

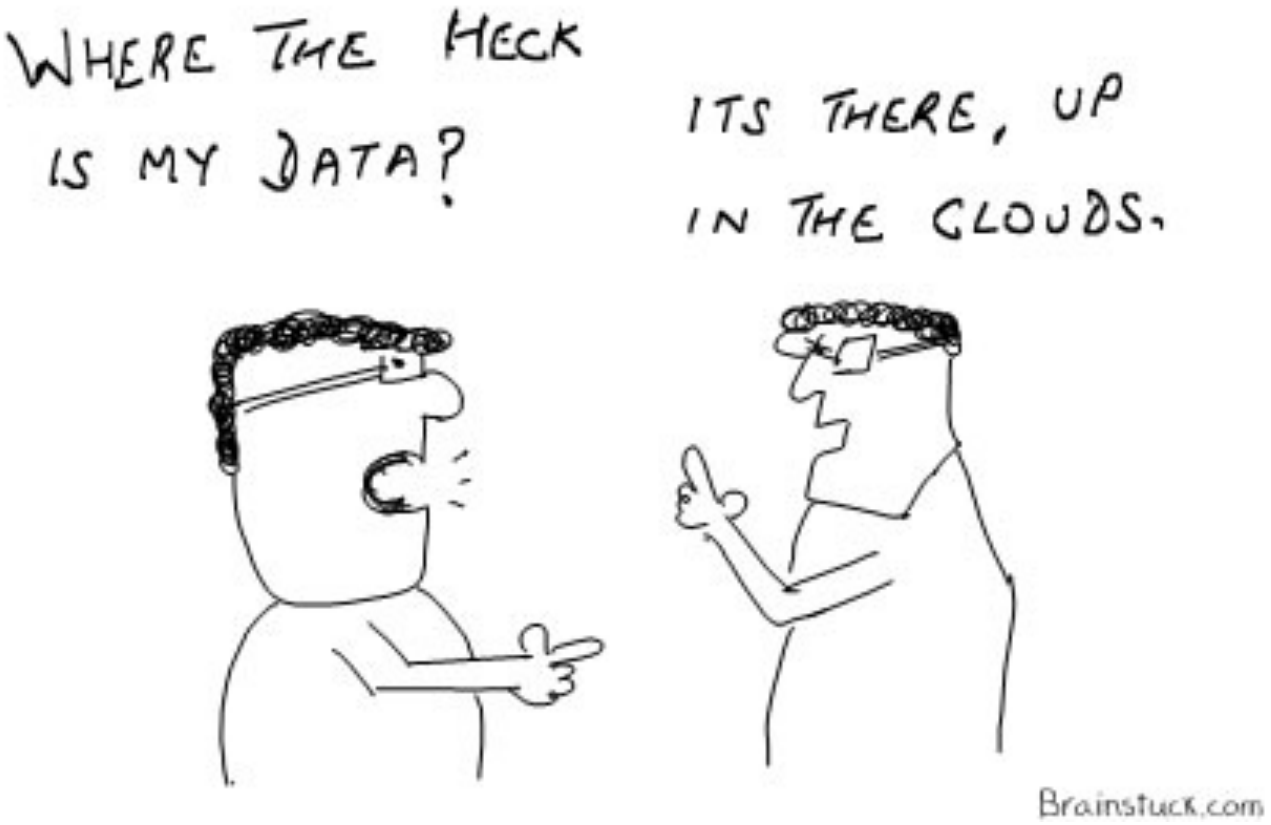


- Grid example: Run 1000 Jobs to generate 1 million MC events
- The Grid is a complex system

A Cloud Job Flow looks smarter

- **Works the same like on a local computer**
  - Develop and debug the program locally
- **Bundle an image with the following characteristics**
  - Automated program start after machine launch (e.g. EC2)
  - Do the data processing (e.g. generate 1000 MC events)
  - Write output to persistent cloud storage (e.g. S3)
  - Automated shut down once finished to stop accounting
- **Just instantiate the suiting number of machines in the cloud**
  - Simply launch 1000 machines to produce 1 million events
  - Cost is the same:
    - Run 1 machine 1000 hours
    - Run 1000 machines 1 hour
- No need for batch queues and scheduling (“unlimited resources”)

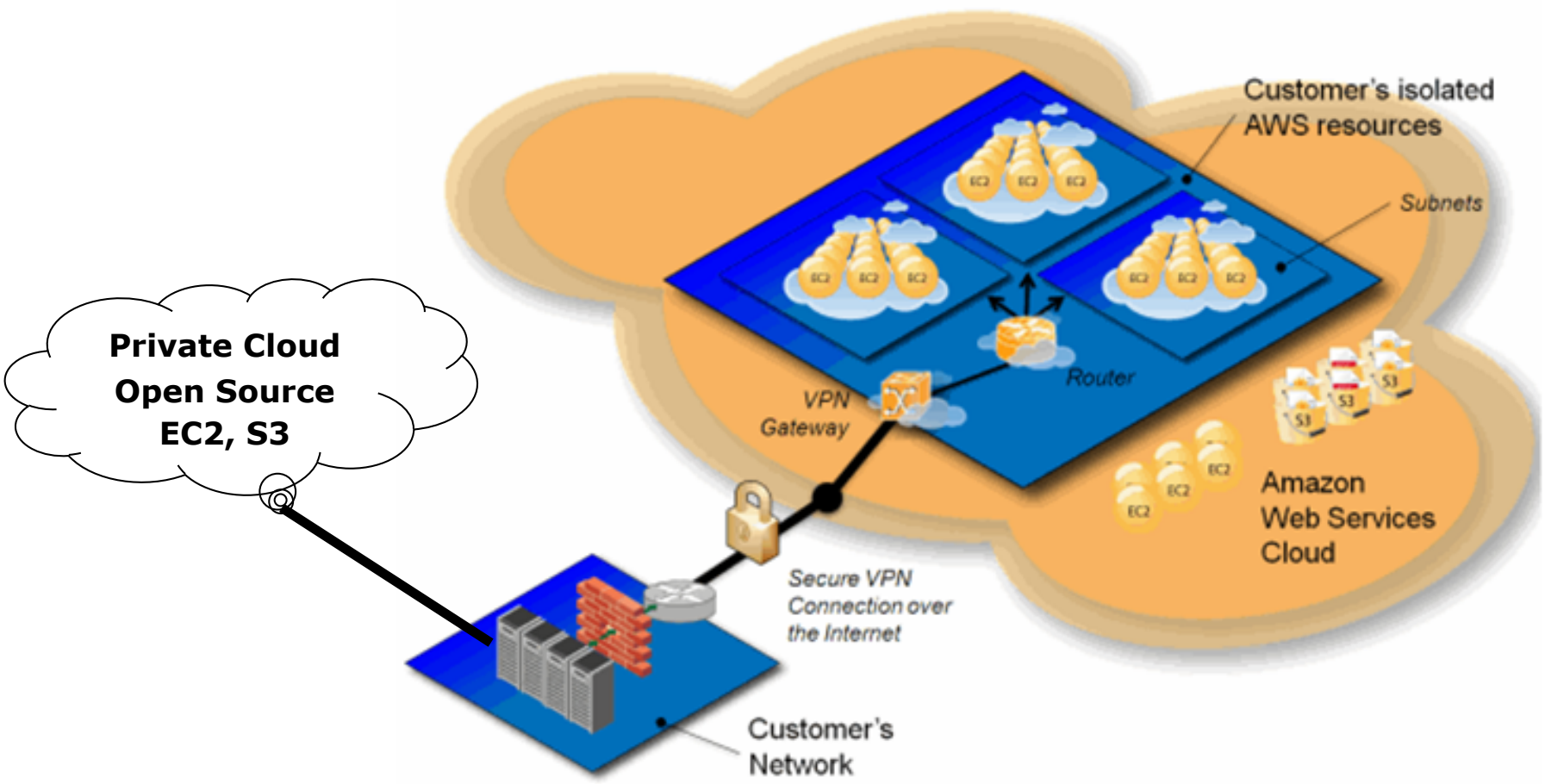
Part II: Cloud Systems Research



The Role of Open Source

- Cloud computing is primarily a commercial endeavor
- What are the options for open source to stay competitive with cloud computing?
- Cooperation
  - Create open source infrastructure for clouds
  - Return innovation and best practices to the community
  - Apply open business models to all components
  - Everybody can participate as resource provider and consumer
  - Possibly evolving into the largest single cloud ecosystem in the world
- Federation
  - Take advantage of standards
  - Interoperability between similar data centers
  - Enterprise virtual private clouds
  - Open source drives standardization and interoperability
  - Overcome vendor lock-in and create a market

Hybrid Clouds: Cloud Bursting



- Transfer workloads and data transparently between clouds
- Example: Amazon Virtual Private Cloud

Cloud Systems Research

- **Simple, transparent, controllable cloud computing infrastructure**
  - What types of interfaces are appropriate for clouds?
  - How should cloud networks be constructed/managed?
  - How are security concerns addressed in “the cloud”?
  - How are various workloads most efficiently transferred?
  - What types of applications can run in clouds?
  - What types of service level agreements are appropriate/possible?

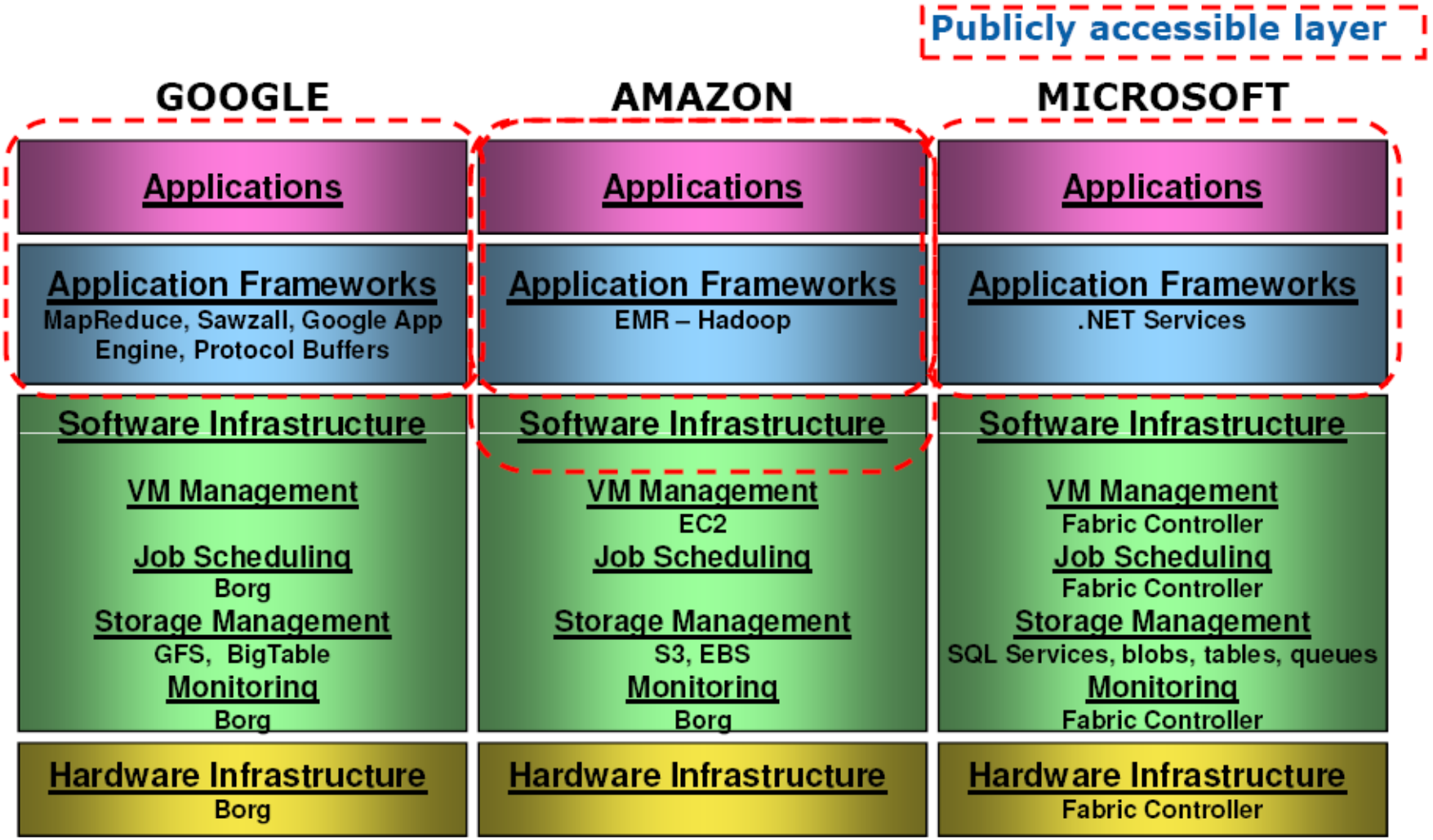
- **Research requirements**
  - Perform experiments also on a low system level
  - Flexible cloud computing framework
  - Compare different methodologies and implementations

OpenCirrus Cloud Computing Research Testbed

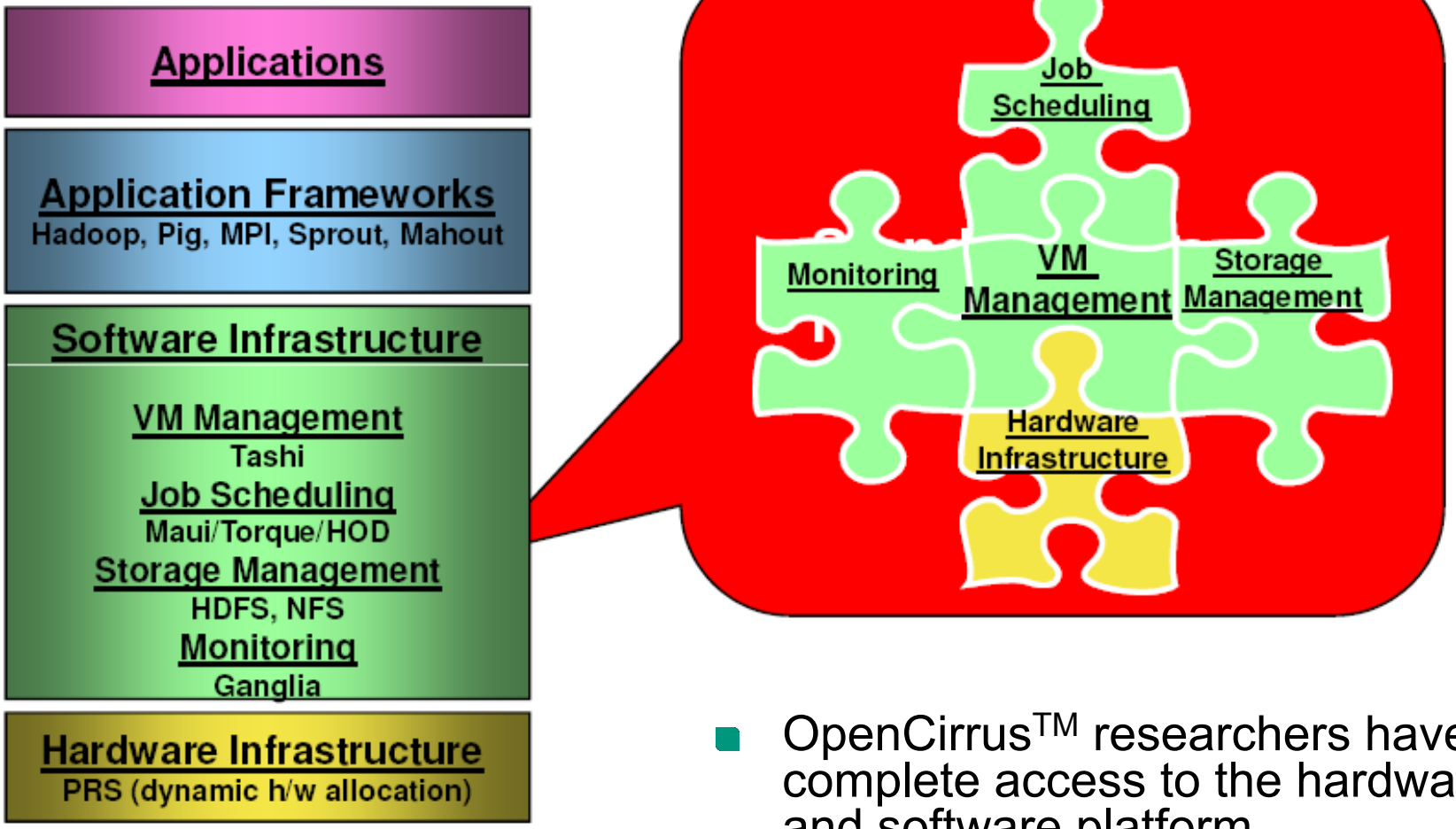
http://opencirrus.org

- **An open, internet-scale global testbed for cloud computing research**
  - a tool for collaborative research
  - focus: data center management & cloud services
- **Resources**
  - Multi-continent, multi-datacenter, cloud computing system
  - Federated “Centers of Excellence” around the globe
    - each with 100–400+ nodes and up to ~2PB storage
    - and running a suite of cloud services
- **Structure**
  - Sponsors: HP Labs, Intel Research, Yahoo!
  - Partners: UIUC, Singapore IDA, KIT, NSF
  - New partners: ETRI, MIMOS, RAS
  - Members: System and application development
- **Great opportunity for Cloud R&D**
  - Accepts proposals for cloud systems research
  - Apply through website

Proprietary Cloud Computing Stacks

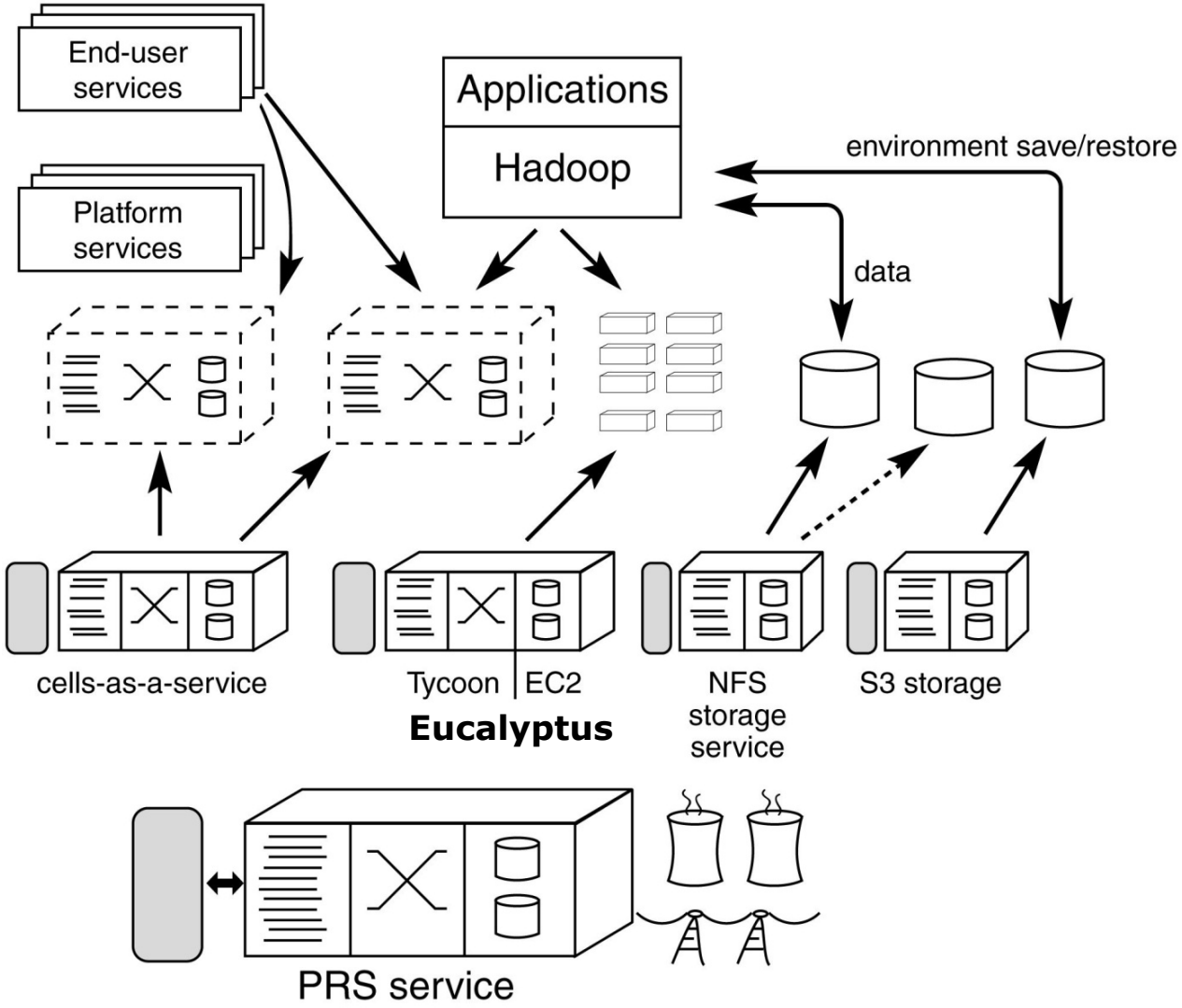


Open Source Cloud Stack



- OpenCirrus™ researchers have complete access to the hardware and software platform

Open Cirrus™ Blueprint



Cloud Application Services

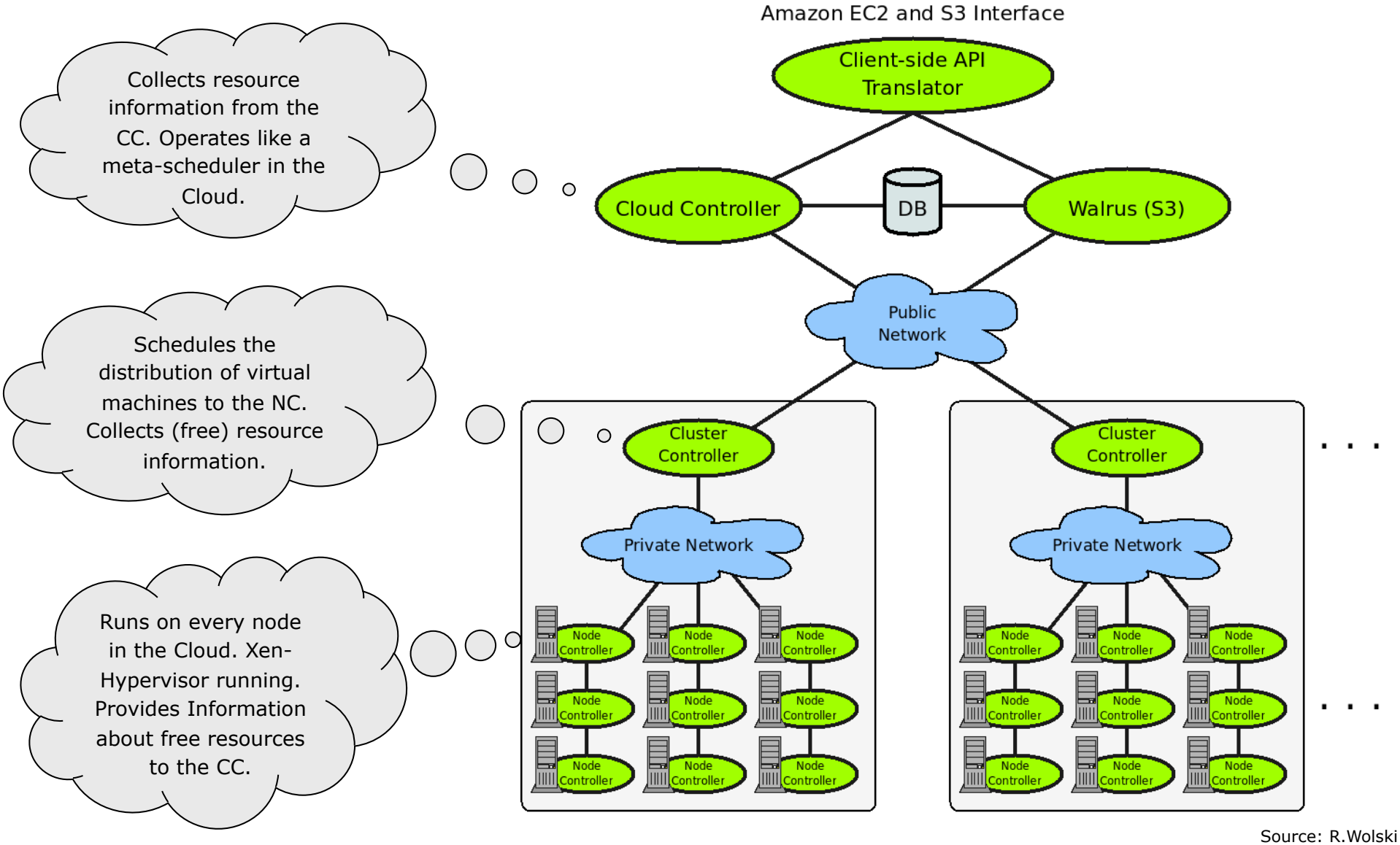
Virtual Resource Sets

Cloud Infrastructure Services

IT-Infrastructure Layer (Physical Resource Sets)

Eucalyptus: A potential VRS

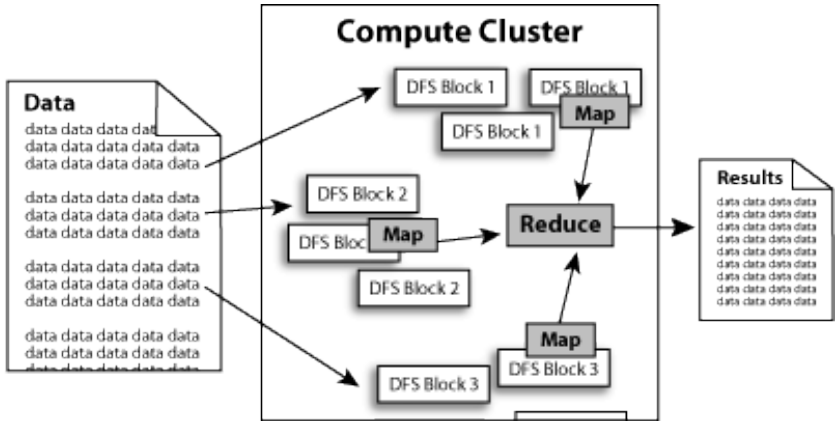
http://eucalyptus.cs.ucsb.edu



Source: R.Wolski

Programming the Cloud: Hadoop

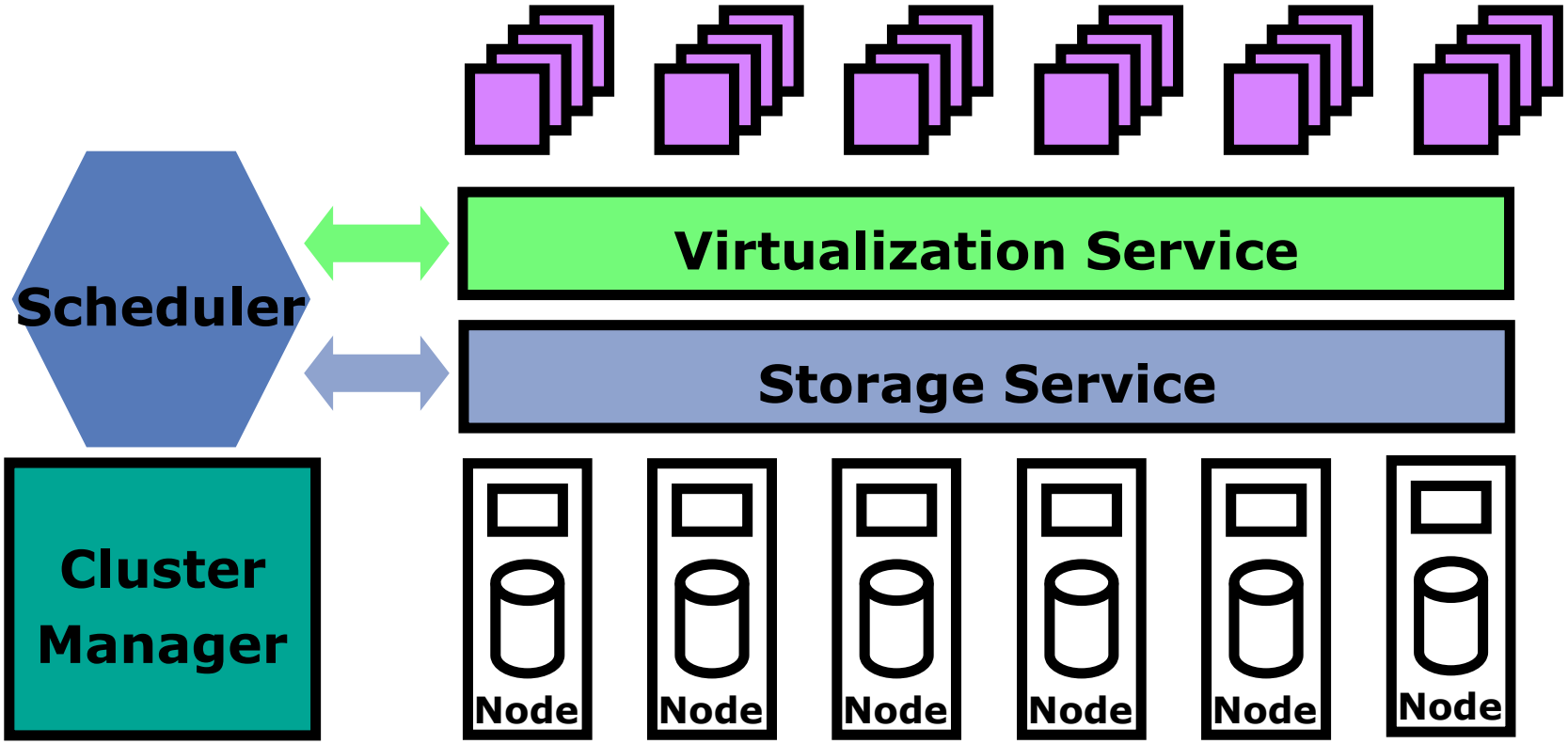
- **An open-source Apache software foundation project sponsored by Yahoo!**
  - Reproduce the proprietary software infrastructure developed by Google
  - http://wiki.apache.org/hadoop/ProjectDescription
- **Provides a parallel programming model (MapReduce), a distributed file system, and a parallel database**
  - http://en.wikipedia.org/wiki/Hadoop
  - http://code.google.com/edu/parallel/mapreduce-tutorial.html



Tashi

http://wiki.apache.org/incubator/TashiProposal

OpenCirrus™ at
Intel Research Pittsburgh
Karlsruhe Institute of Technology



- Co-Scheduling of CPU and data

How do users get access to OpenCirrus™ sites?

- **Project PIs apply to each site separately**
- **Contact names, email addresses, and web links for applications to each site will be available on the OpenCirrus™ Web site**
  - http://opencirrus.org
- **Each OpenCirrus™ site decides which users and projects get access to its site**
- **Planning to have a *global sign on* for all sites**
  - Users will be able to login to each OpenCirrus™ site for which they are authorized using the same login and password.

What kinds of research projects are OpenCirrus™ sites looking for?

- **Open Cirrus™ is seeking research in the following areas:**
  - Datacenter federation
  - Datacenter management
  - Web services
  - Data-intensive applications and systems
  - Hadoop map-reduce applications

- **The following kinds of projects are not of primary interest:**
  - Traditional HPC application development.
  - Production applications that just need lots of cycles.
  - Closed source system development.

Summary

- **Promising to implement Grid services on cloud infrastructures**
  - Fully automated operations (Centralized management)
  - Economy of scale
  - Business models lead to efficient use of resources
- **OpenCirrus™ offers interesting R&D opportunities**
  - Cloud systems development
  - Cloud application development
  - Accepting research proposals
- **Open Source cloud stack**
  - Federation of clouds (Cloud bursting)
  - De-facto standards are around (e.g. S3)
  - Standards evolving (OCCI)