



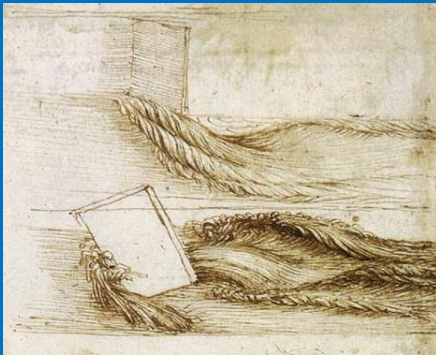
Windows Azure for Research

Dr Kenji Takeda (kenjitak@microsoft.com)
Microsoft Research Connections

[@azure4research](#) [#azureresearch](#)

The Nature of Scientific Discovery

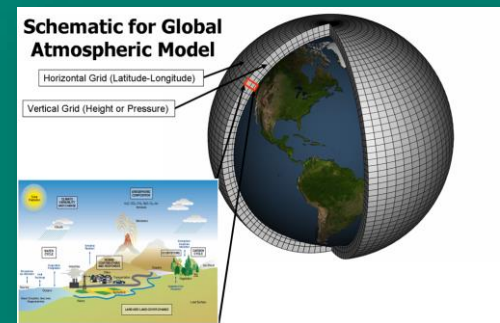
Experiment



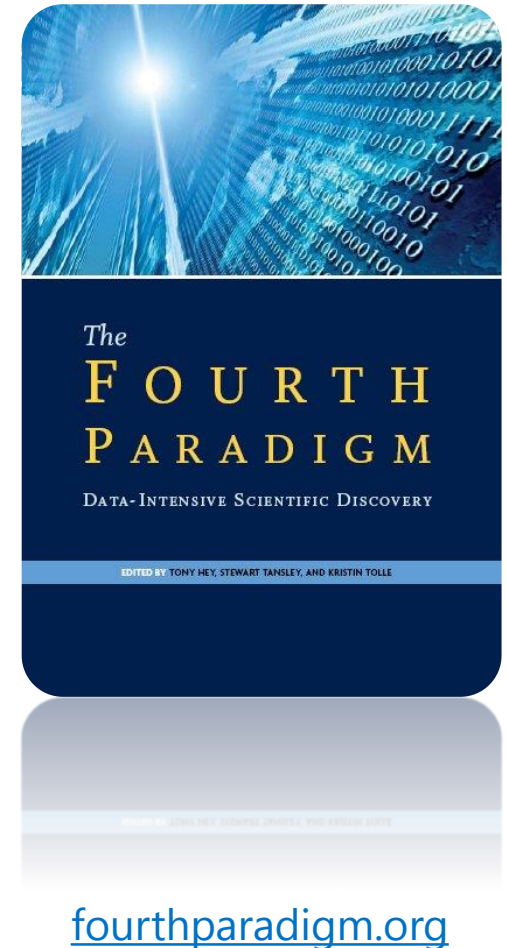
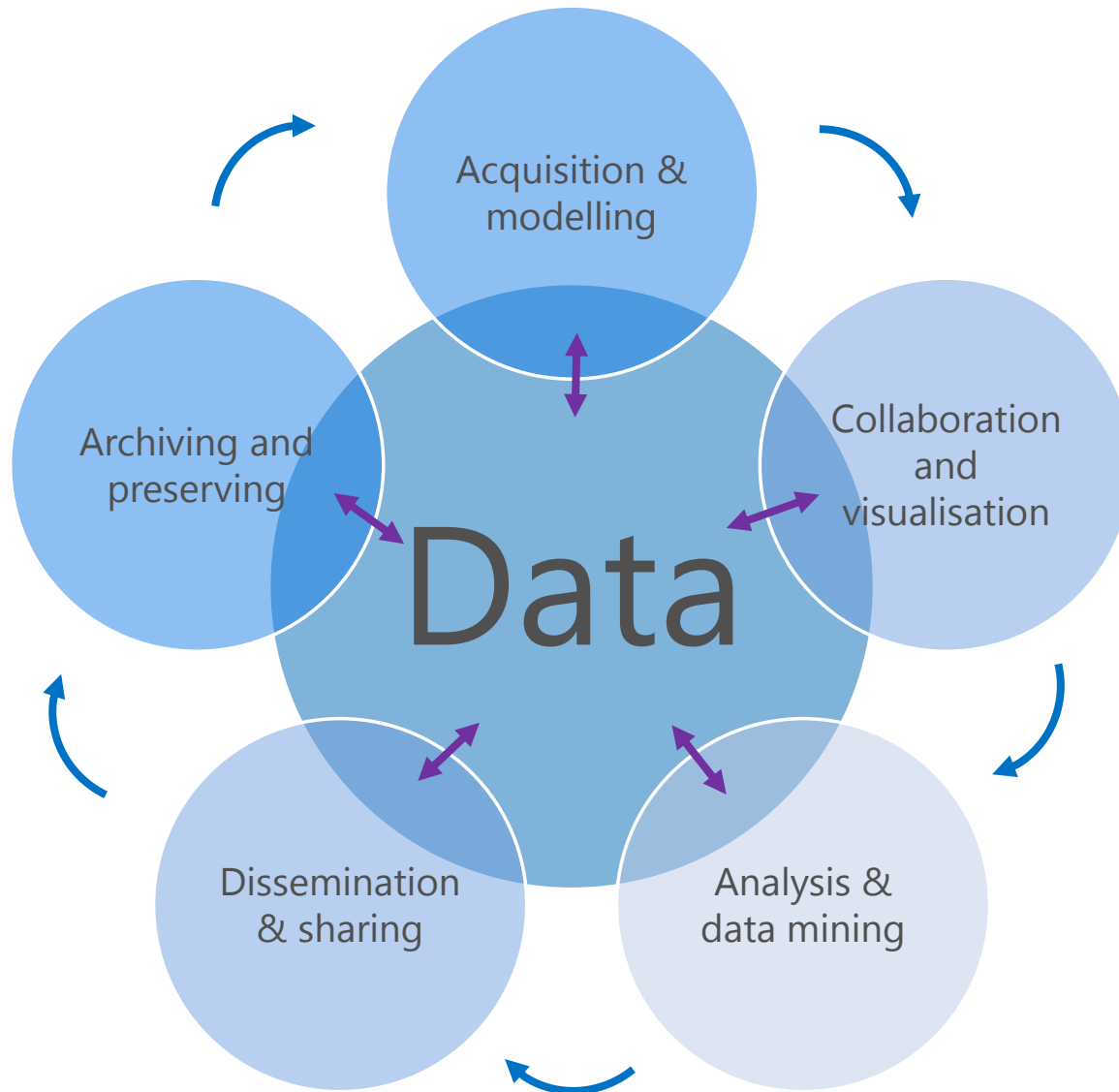
Theory

$$\rho \frac{Dv}{Dt} = -\nabla p + \nabla \cdot T + f$$

Computation



Data-intensive Research



Let scientists be scientists....

Today

Majority of Researchers

Use laptops &
desktop computers

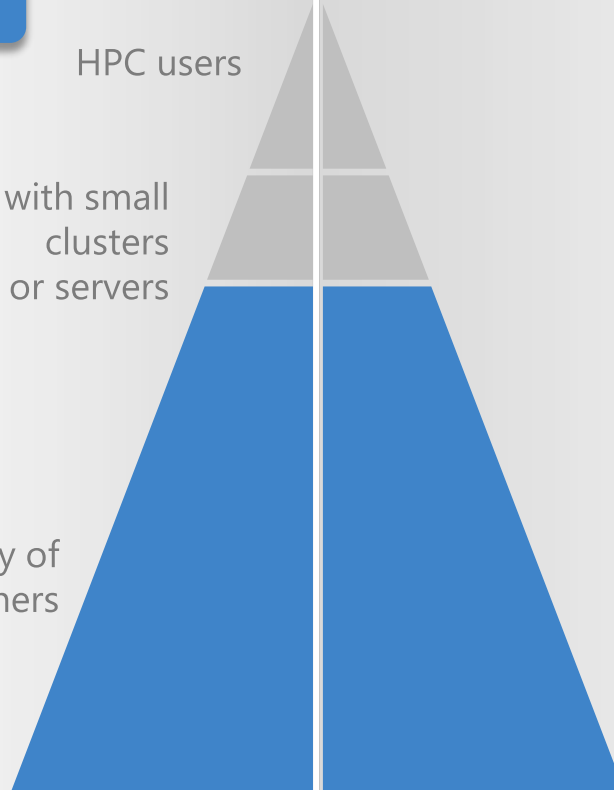
Overwhelmed by
data

Finding analysis
ever more difficult;
sharing even
harder

HPC users

Those with small
clusters
or servers

Majority of
Researchers



Let scientists be scientists....

Today

Majority of Researchers

Use laptops & desktop computers

Overwhelmed by data

Finding analysis ever more difficult; sharing even harder

HPC users

Those with small clusters or servers

Majority of Researchers

Tomorrow?

Paradigm Shift

Powerful tools

Data and analysis tools in the cloud
Cycles, storage, support

Building communities around research results

The ability to marshal needed resources on demand
Without caring or knowing how it gets done...

Accelerating discovery



Let scientists be scientists....

Azure for Research

Azure Research Awards

Windows Azure for Research
Training two-day workshops

Technical resources &
curriculum

Research community
engagements

Tomorrow?

Paradigm Shift

Powerful tools

Data and analysis
tools in the cloud
Cycles, storage, support

Building communities
around research results

The ability to marshal needed
resources on demand
Without caring or knowing
how it gets done...

Accelerating discovery

A Unified
Research
Community

kenjitak@microsoft.com

www.azure4research.com

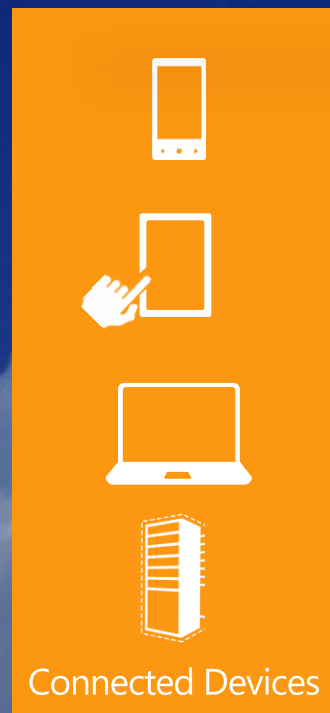
Cloud Computing

Cloud computing provides...

On-demand services,
delivered over the network

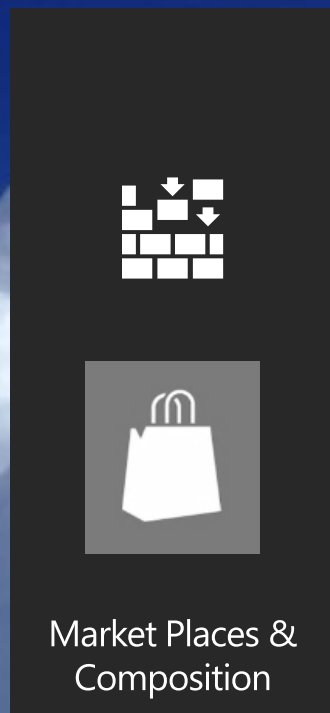
The Cloud

Development Model, Deployment, Management & I/O



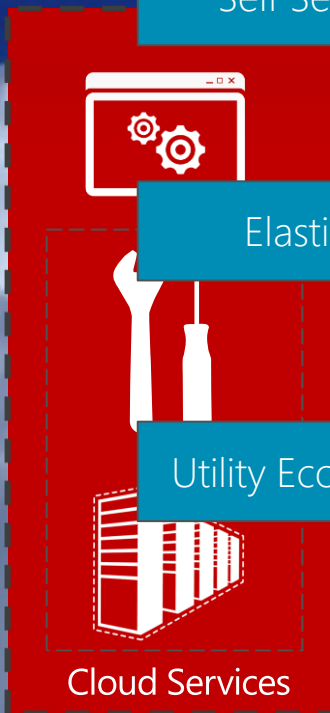
Connected Devices

This orange vertical panel contains four white icons: a smartphone, a hand touching a tablet, a laptop, and a server rack.



Market Places & Composition

This dark grey vertical panel contains two white icons: a grid of squares with arrows and a shopping bag.



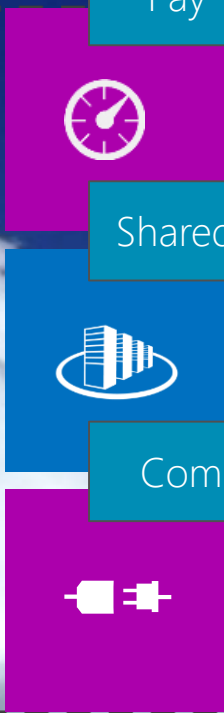
Cloud Services

This red vertical panel contains three white icons: a browser window with gears, a wrench and screwdriver, and a server rack.



Elasticity

This blue vertical panel contains two white icons: a person at a presentation board and a double-headed arrow.



Shared Resources

This magenta vertical panel contains three white icons: a clock, a server rack, and a plug.

Self Service

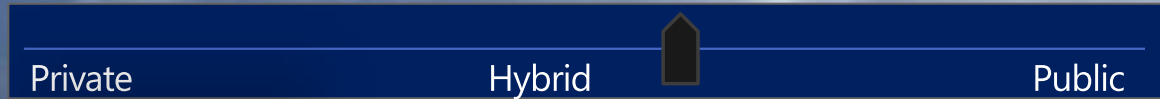
Pay-by-Usage

Elasticity

Shared Resources

Utility Economics

Commoditized



Youngrel Ryu – Environmental Scientist



Youngrel with his PC
Needed to run
calculations against
NASA MODIS data
Problems with data
and compute

Cloud computing makes a difference...

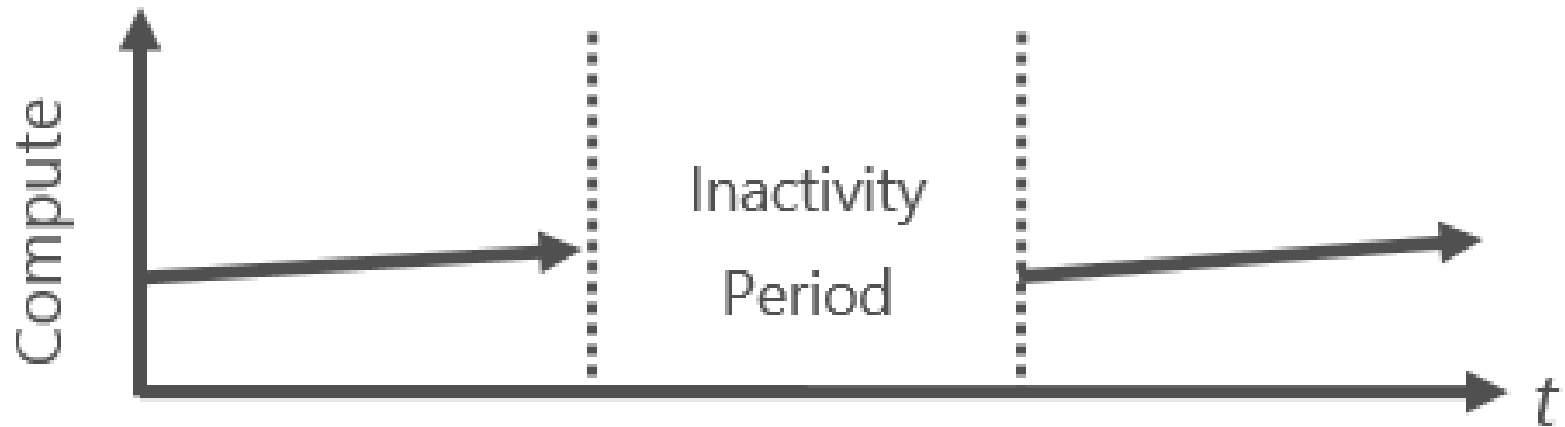
Cloud computing makes a difference

- Scale beyond your PC/laptop
- On-demand, no waiting
- Big data computations without fuss
- Share data with your team, and globally
- Build research services easily that can be accessed widely
- Pay for what you need

Cloud computing is good for..

Getting what you need,
when you need it

Patterns: On and Off



On & off workloads (e.g. batch job)

Compute only when you need it

No queuing

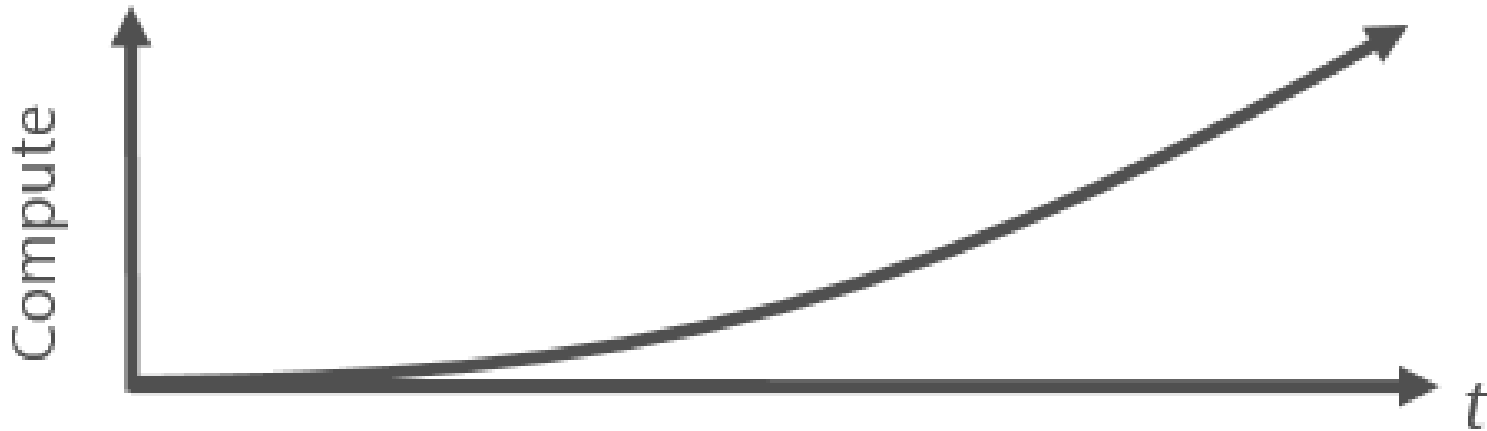
Over provisioned capacity is wasted

Difficult to handle conventionally

Stratospheric Research



Patterns: Growing Fast



Need to grow/scale quickly

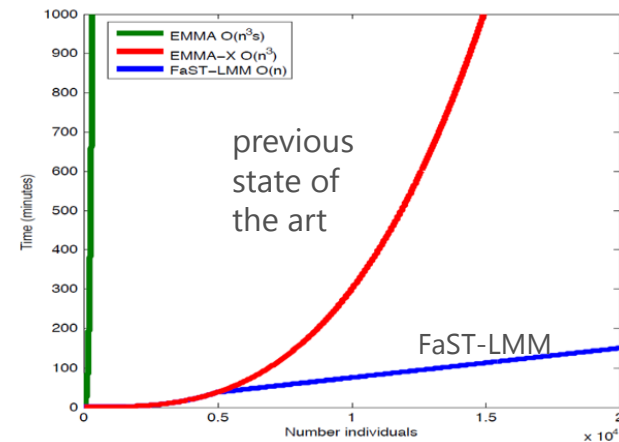
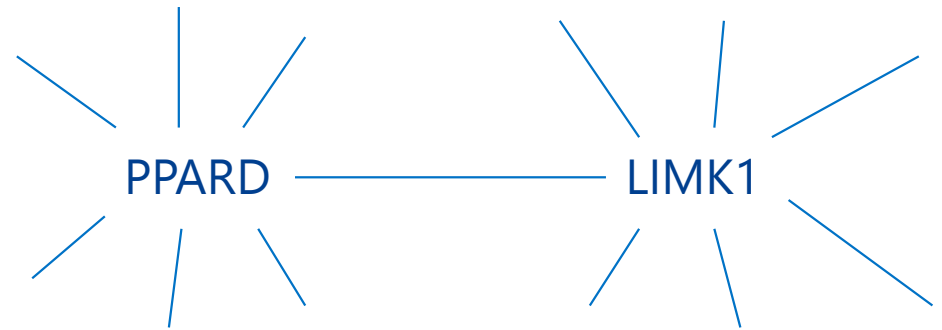
Equally important for data

Keeping up is difficult

Cannot provision hardware fast enough

GWAS: Cardio-Vascular Disease

- Wellcome Trust data for seven common diseases
 - On Azure Marketplace
- With FaST-LMM and Azure, can look at all SNP pairs (about 60 billion of them)
- 1,000 compute years; 20 TB output – we did it in 13 days
- 27k cores on-demand

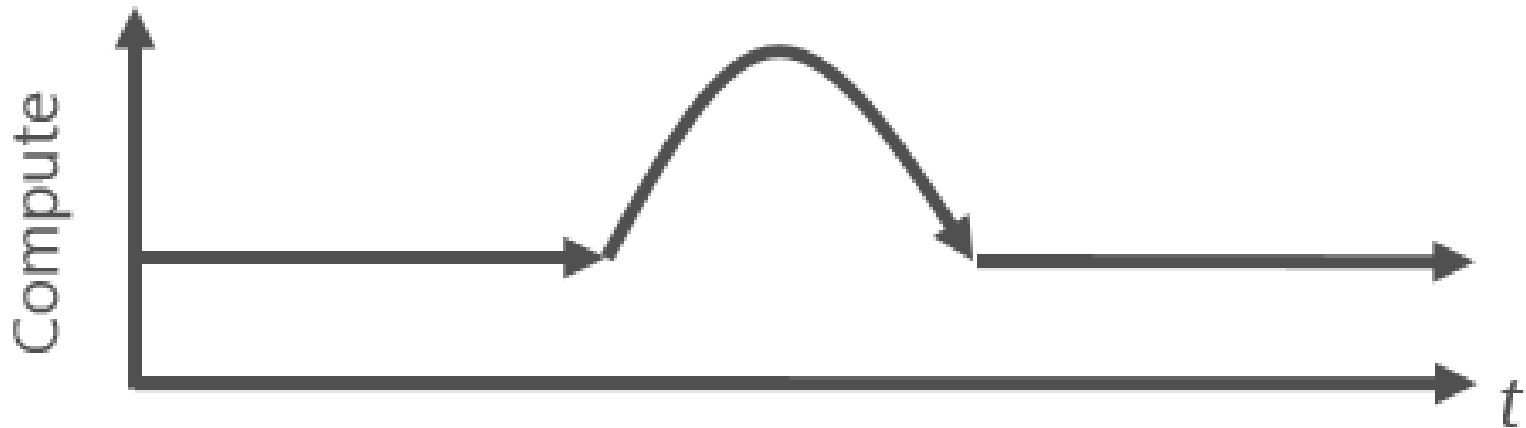


SCIENTIFIC REPORTS

An Exhaustive Epistatic SNP Association Analysis on Expanded Wellcome Trust Data

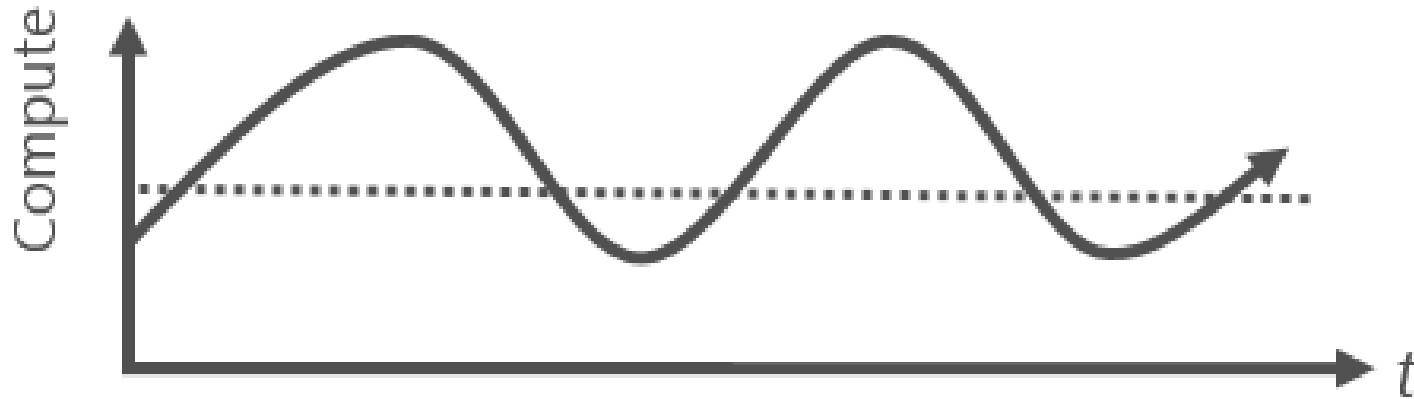
Christoph Lippert, Jennifer Listgarten, Robert I. Davidson, Jeff Baxter, Hoifung Poon, Carl M. Kadie & David Heckerman

Patterns: Unpredictable Burst



- Unexpected/unplanned peak in demand
- Unable to satisfy demand (paper deadlines)
- Sudden spike impacts performance
- Cannot provision for extreme cases

Patterns: Predictable Burst



Predictable trends (e.g. daily, weekly)
Data processing from experiments
Wasted capacity to accommodate

Why Cloud Computing?

- Burst scalability
- Sustainable scalability
- Access anywhere, on any device
- Global collaboration
- Data distribution and de-duplication
- Economy of scale
- Different business models



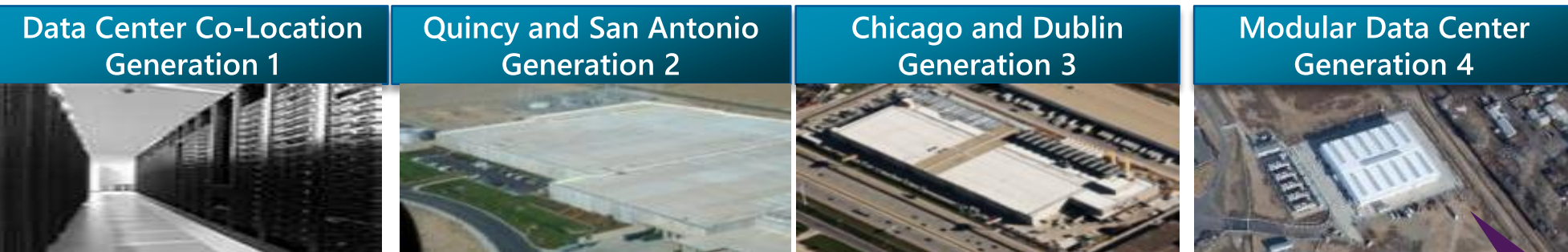
Windows

Azure

The Microsoft Cloud



Microsoft's Data Center Evolution



Deployment Scale Unit

Server



Capacity

Rack



Density &
Deployment

Containers



Scalability &
Sustainability

IT
PAC



Time to Market
Lower TCO

<http://www.globalfoundationservices.com/>

Cloud Computing



IaaS

Infrastructure-as-a-Service
with Persistent remote disks

host



PaaS

Platform-as-a-Service
Stateless, easy to scale, manage

build

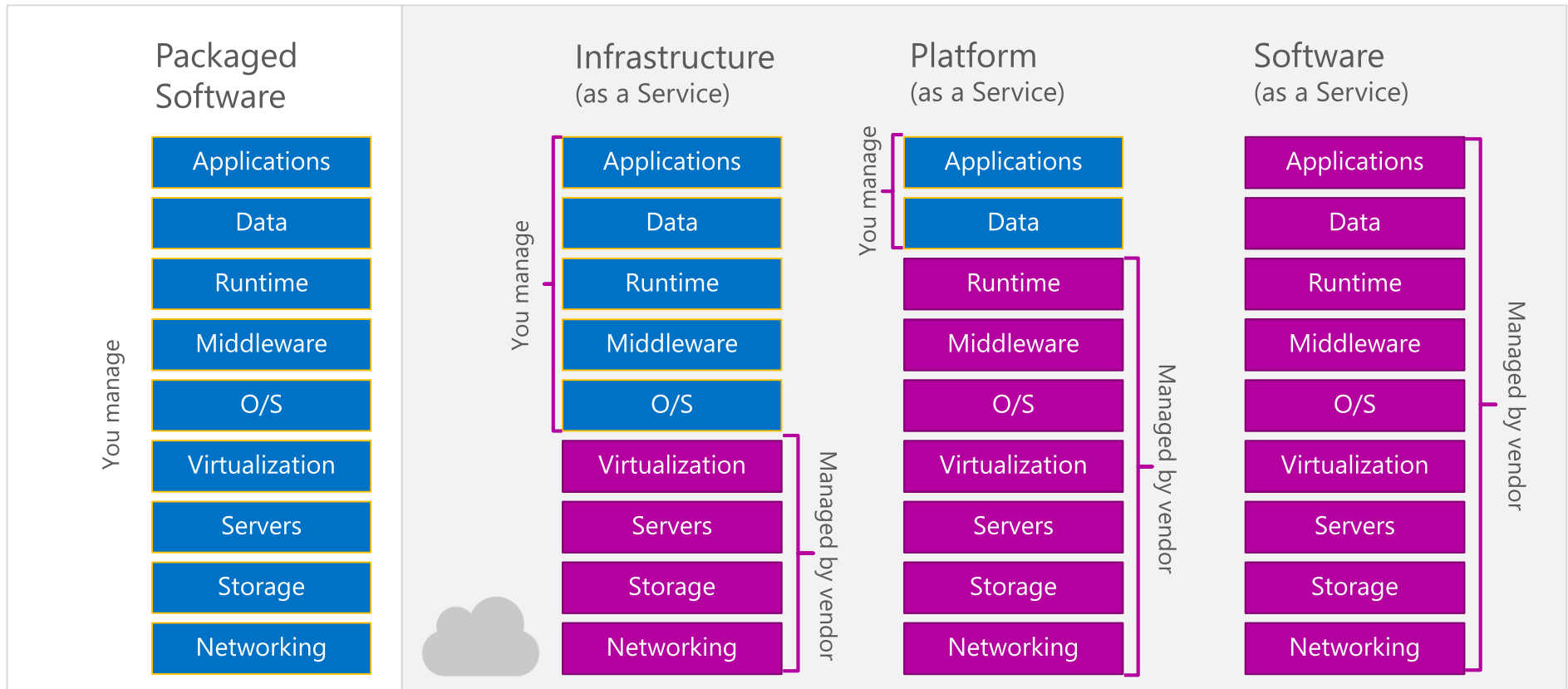


SaaS

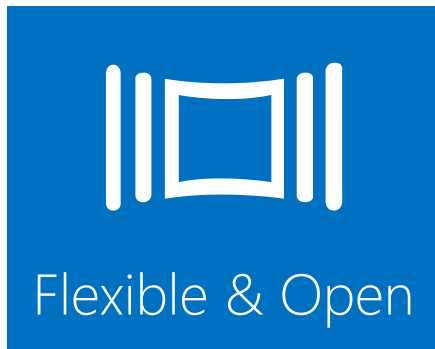
Software-as-a-Service

consume

Cloud Computing Variants



Windows Azure

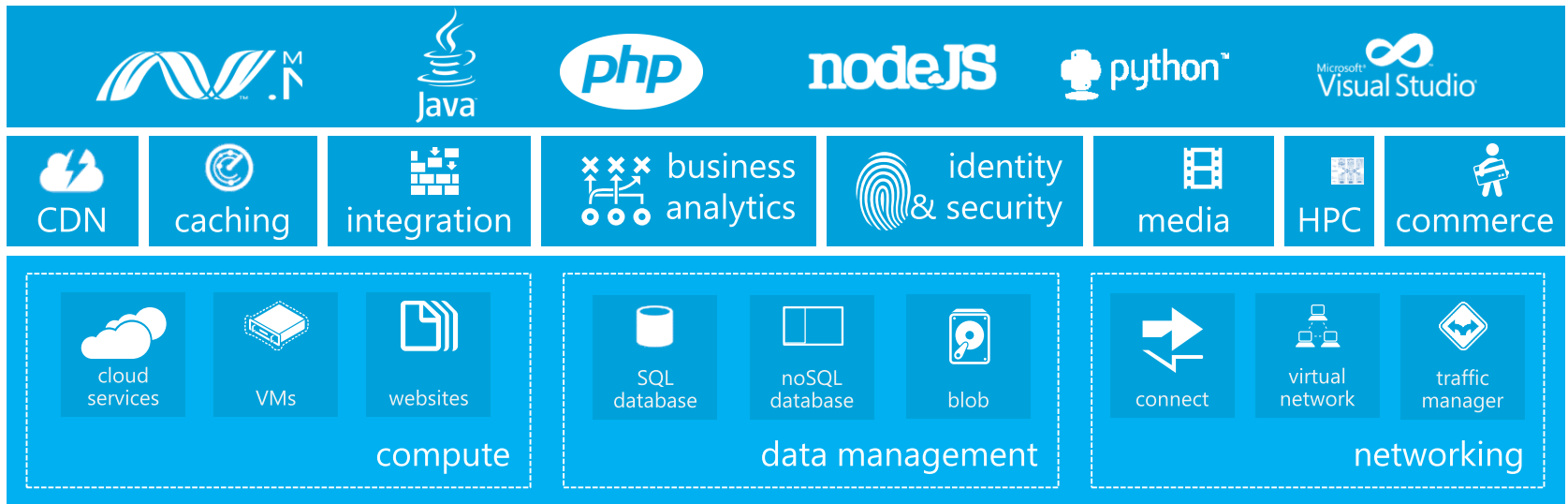


Choose from multiple runtimes and languages for your applications: Python, Java, PHP, .NET, Node.js

Run Linux on Windows Azure Virtual Machines (VHD)

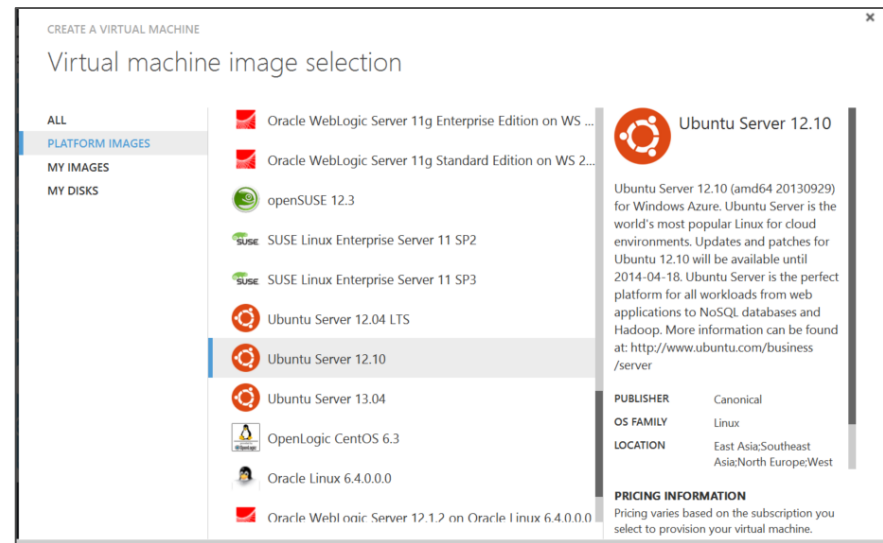
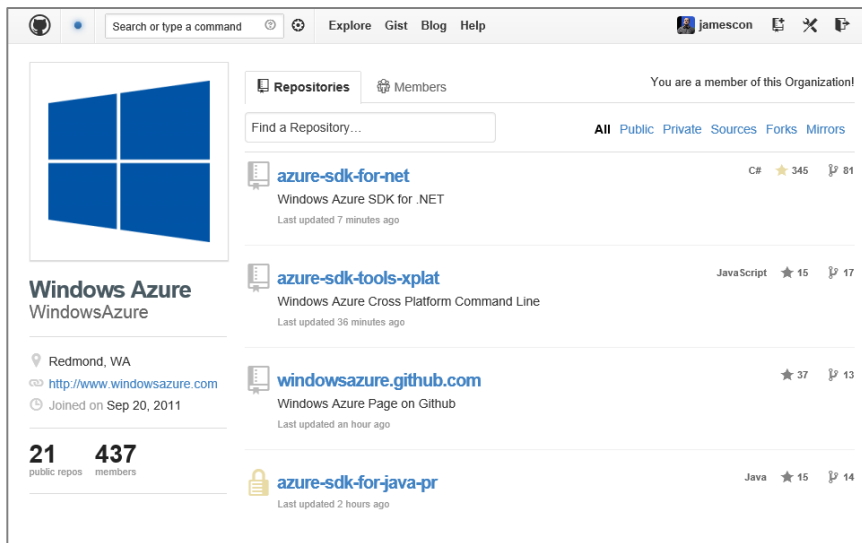
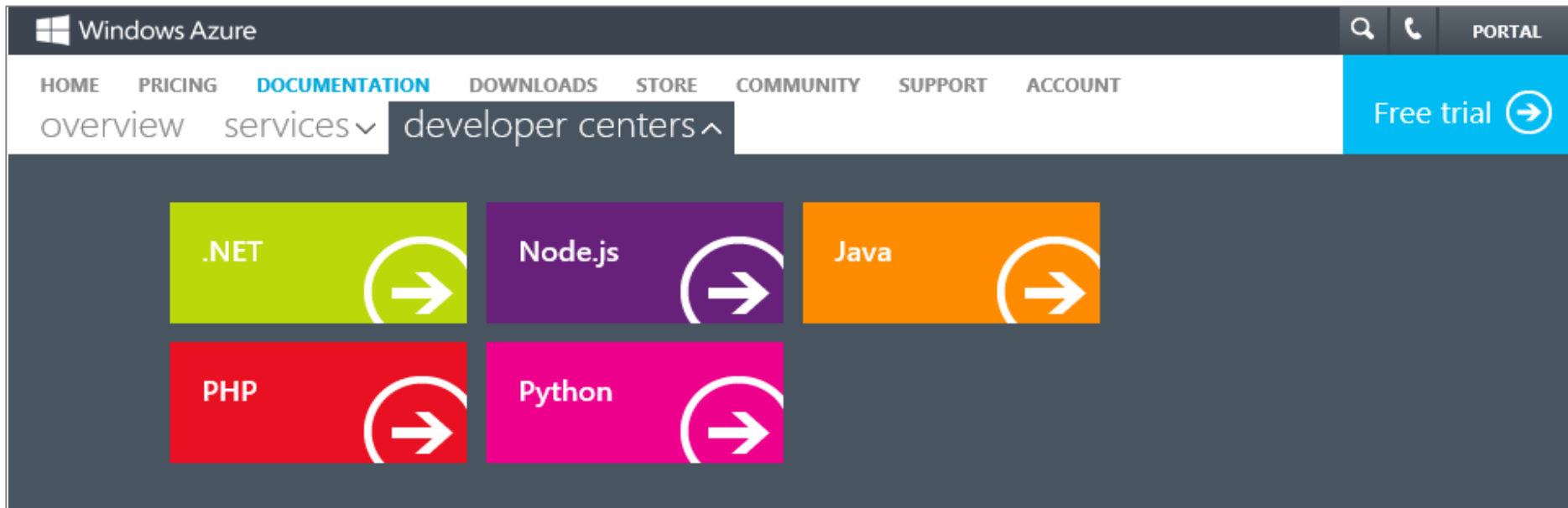
Support multiple frameworks and popular open source applications with Windows Azure Web Sites

HDInsight Hadoop for Big Data analysis



<http://github.com/windowsazure>

Flexible & Open



Virtual Machines

The screenshot shows a web browser window displaying the VM Depot website. The page features the VM Depot logo, navigation links for 'Back', 'DEPLOYMENT SCRIPT', 'PUBLISH', 'SHARE', and 'HELP', and a 'Subscribe' button. The main content area displays the 'biolinux' image with a 5-star rating, 0 comments, and a thumbnail image of a workstation. A description below the image states: 'Bio-Linux 7 is a fully featured, powerful, configurable and easy to maintain bioinformatics workstation. Bio-Linux provides more than 500 bioinformatics programs on an Ubuntu Linux 12.04 LTS base. There is a graphical menu for bioinformatics programs, as well as easy access to the Bio-Linux bioinformatics documentation system and sample data useful for testing programs.'

Image name	biolinux
Publisher	Hyungro Lee

Virtual Machines

The image shows a browser window displaying the VM Depot website. The page features the VM Depot logo and navigation options like 'Back', 'Browse Images', and 'My Account'. The main content area highlights the 'CKAN 2.1 DB' virtual machine image, which has a 5-star rating and 0 comments. Below the image, there is a detailed description of CKAN as open source software and instructions for its deployment on Azure. At the bottom, a table lists the image name as 'CKAN 2.1 DB'.

VM Depot PREVIEW
by Microsoft

VM Depot PREVIEW
by Microsoft Open Technologies


Subscribe
Sign in and join the community

Back DEPLOYMENT SCRIPT PUBLISH SHARE HELP

Browse Images My Account

CKAN 2.1 DB

★★★★★ (1)
0 comments



CKAN is open source, open data portal software used by many governments and organizations around the world to publish their data (incl. data.gov and data.gov.uk). For details on CKAN support visit: <http://ckan.org/services/> IMPORTANT: This VM should be used in conjunction with CKAN 2.0 Web VM. These two VMs need to be connected and will not work without each other (make sure the DB vm is made first). Connecting VMs is done when creating them in the Azure Management Portal, see <http://i.imgur.com/48j2dZ5.png>. It can also be done with the -c flag on the 'azure vm create' command of the Azure CLI tool making sure the "DNS-name" is the same for both. For a step-by-step guide to deployment: <https://github.com/okfn/ckan/wiki/How-to-install-CKAN-2.0-on-Azure-from-VM-Depot>

Image name	CKAN 2.1 DB
------------	-------------

Linux Virtual Machines

The image shows a stack of three browser windows displaying the VM Depot website. The top window shows the homepage with the VM Depot logo and navigation links. The middle window shows the 'Azure Data Science Core' image page. The bottom window shows the details for the 'Azure Data Science Core' image, including a description and a table of image names.

VM Depot PREVIEW
by Microsoft

Back
Browse Images
My Account

Back
Browse Images
My Account

Back
Browse Images
My Account

DEPLOYMENT SCRIPT
PUBLISH
SHARE
HELP

Azure Data Science Core

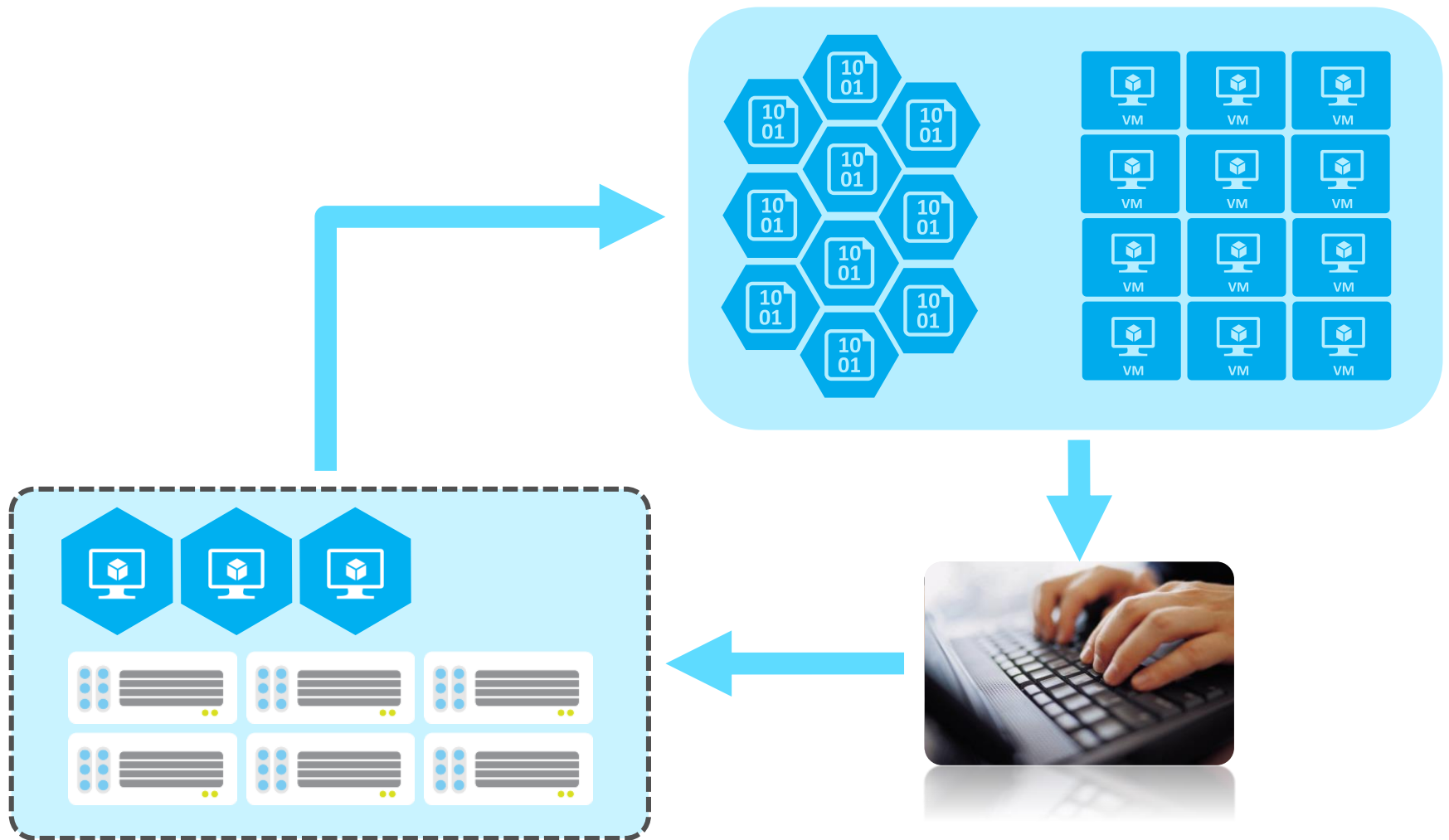
★★★★★
0 comments

A special version of HPC Linux is now available as a Windows Azure virtual machine image. Based on OpenSUSE 12.3, this image includes all the packages you need for big data and data science computations including: - iPython - NumPy - SciPy - pandas - SymPy - scikit-learn - scikit-image - StatsModels - matplotlib - PyTables - NetworkX - Cython - NLTK This image has been specially customized for Windows Azure, so it includes the Windows Azure Agent for Linux and the Azure SDK for Python. Together, these packages give you easy access to Windows Azure's storage and queue services directly from iPython.

Updated for Kafka, Storm, Node.js, and Redis.

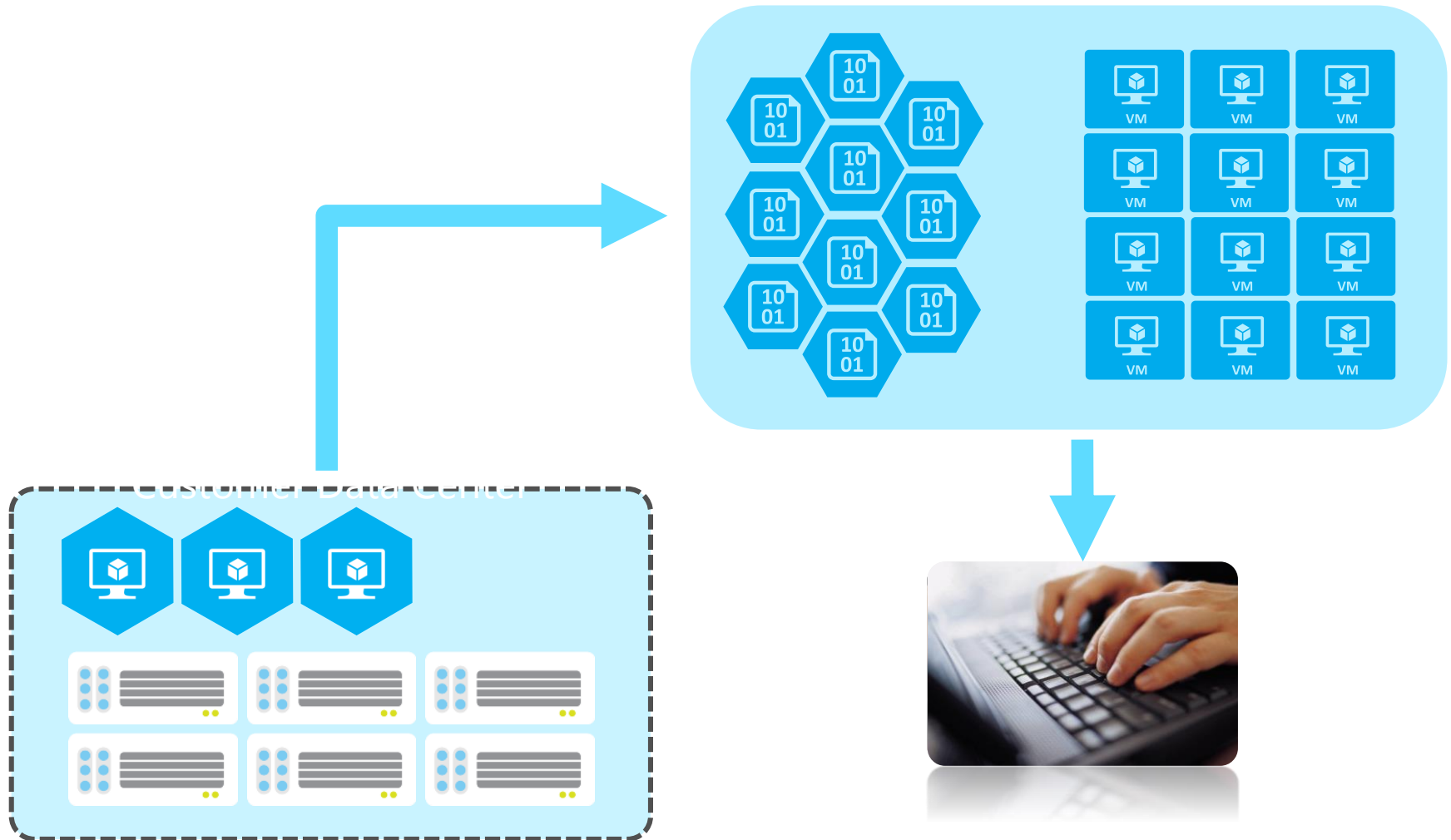
Image name	Image name
	Azure Data Science Core

Hybrid Cloud for Science



Local Data Centre

Hybrid Cloud for Science



Local Data Centre

Over 18 million students and staff to benefit from faster, more secure cloud-computing

Over 18 million students and staff to benefit from faster, more secure cloud-computing

Posted on May 21st 2013



More than 18 million students, staff and researchers at institutions across the UK could start to benefit from a faster and more secure connection when using their institution's cloud-based IT services, thanks to a new peering arrangement between Microsoft and Janet, the UK's research and education network.



Accelerating Research

Youngrel Ryu – Environmental Scientist

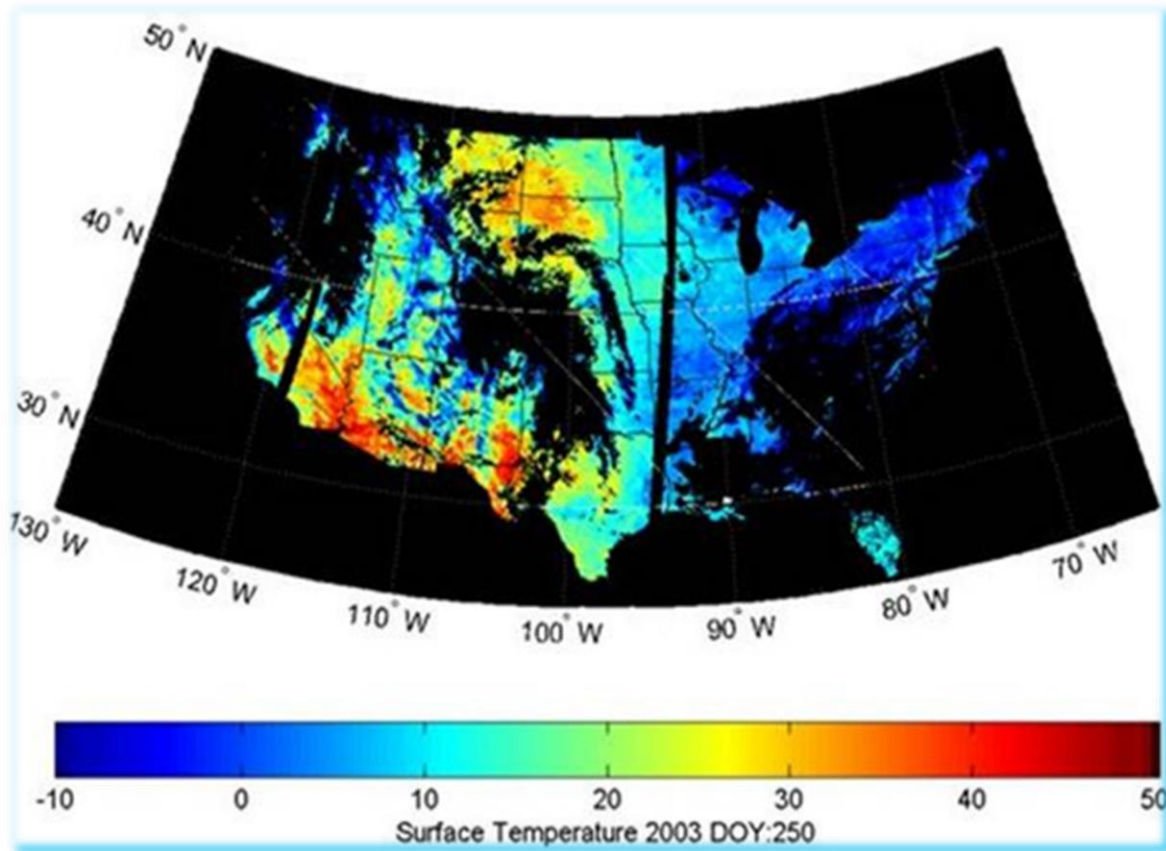


Youngrel with his PC
Needed to run
calculations against
NASA MODIS data
Problems with data
and compute

Cloud computing makes a difference...

MODIS Azure:

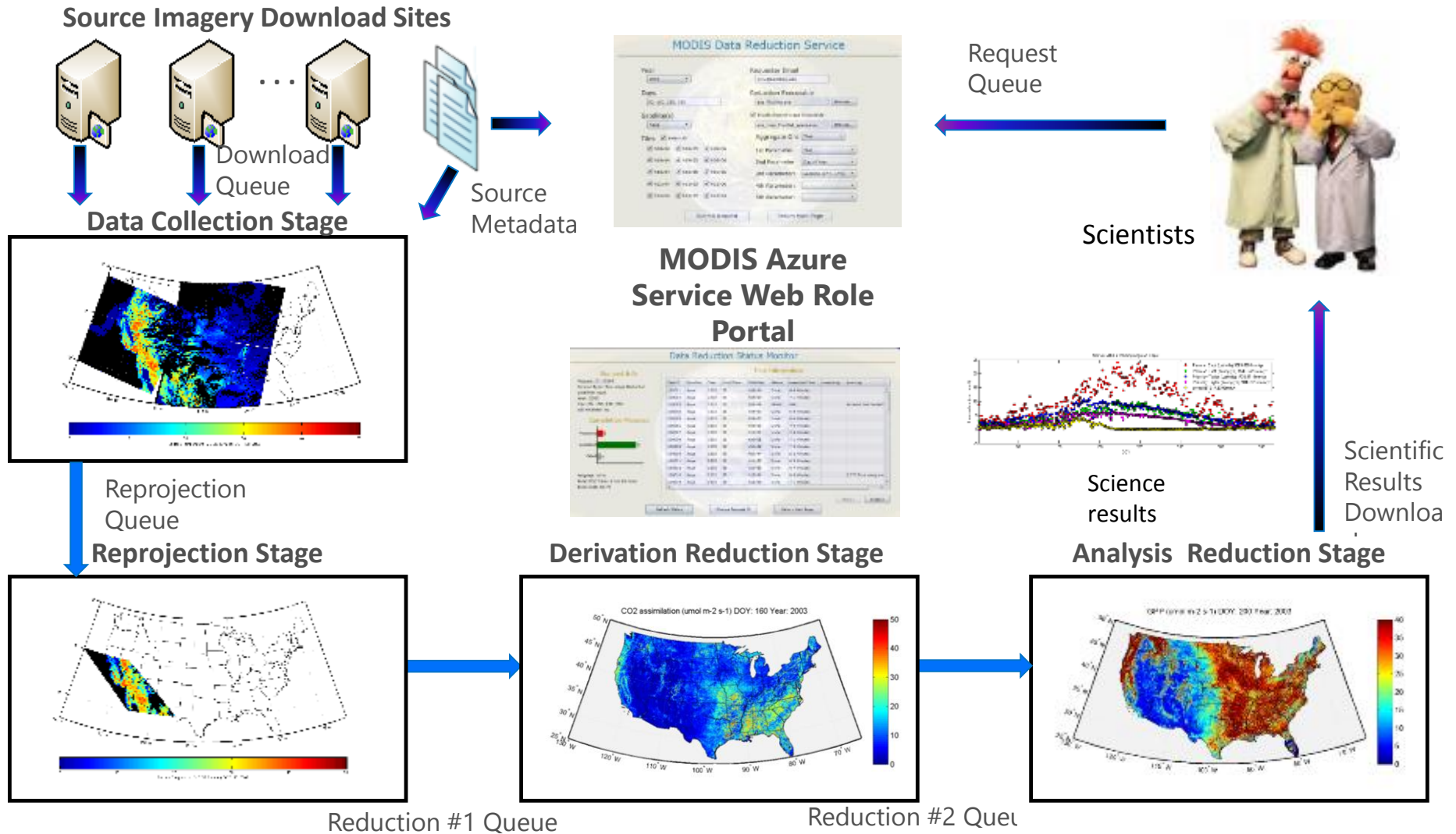
Computing Evapotranspiration (ET) in the Cloud



A pipeline for
download,
processing, and
reduction of
diverse NASA
MODIS satellite
imagery

Catharine van Ingen (Microsoft Research), Jie Li, Marty Humphrey (UVA), Youngryel Ryu (UCB), Deb Agarwal (BWC/LBL), Keith Jackson (BL), Jay Borenstein (Stanford), Team SICT: Vlad Andrei, Klaus Ganser, Samir Selman, Nandita Prabhu (Stanford), Team Nimbus: David Li, Sudarshan Rangarajan, Shantanu Kurhekar, Riddhi Mittal (Stanford)

MODIS Azure Service

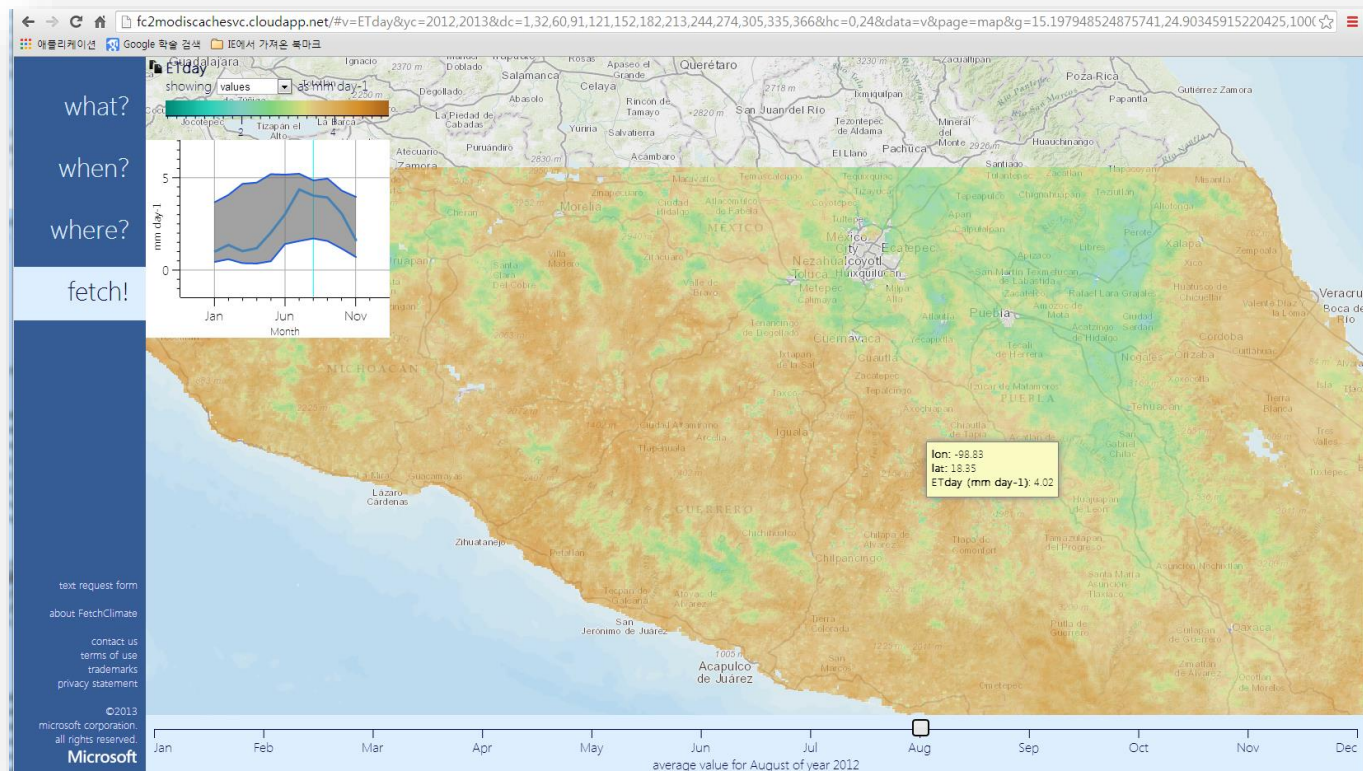


Catharine van Ingen (Microsoft Research), Jie Li, Marty Humphrey (UVA), Youngryel Ryu (UCB), Deb Agarwal (BWC/LBL), Keith Jackson (BL), Jay Borenstein (Stanford), Team SICT: Vlad Andrei, Klaus Ganser, Samir Selman, Nandita Prabhu (Stanford), Team Nimbus: David Li, Sudarshan Rangarajan, Shantanu Kurhekar, Riddhi Mittal (Stanford)

FetchClimate^{Beta}

- Intelligent environmental information service
 - Selects best data source to answer the query
 - Regrids results
 - Calculates uncertainty

Powered by
Windows Azure



Open UK Weather Forecast Data

- Daily forecast + 5 days
- 3 hourly forecast + 5 days
- Observation data

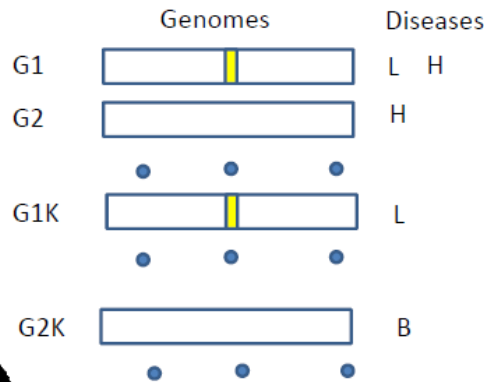
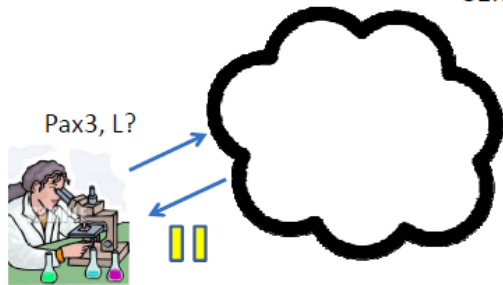
Web search:
"open weather data azure"

The screenshot shows a web browser window with two tabs. The active tab is 'http://data.gov.uk/metoffice-data-arch' and the address bar shows 'http://datamarket.azure.com/dataset/dalagovuk/metofficeweatheropen...'. The page is titled 'Windows Azure Marketplace' and displays the 'UK Met Office Weather Open Data' dataset. The page includes a search form with fields for 'Query Type' (set to 'Daily Forecast'), 'Site Code' (set to 'All Prediction sites'), 'Date of Issue' (set to '05/10/2013'), and 'Time of Issue' (set to '0000'). A 'Search' button is visible. The page also features a 'Like' button with a count of 4. At the bottom, there is a map of the United Kingdom with weather icons overlaid on various regions.

Interactive Genomics

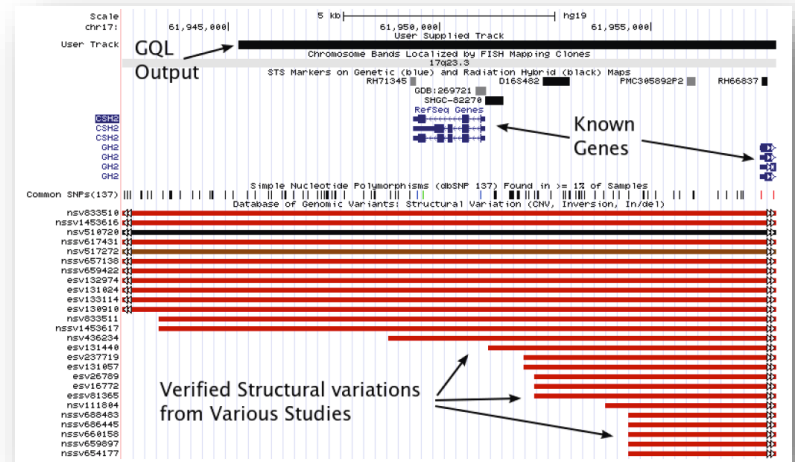
Rapidly Querying Genomes in the Cloud

Christos Kozanitis(UCSD), Vineet Bafna(UCSD), Ravi Pandya(MSR), George Varghese(MSR)



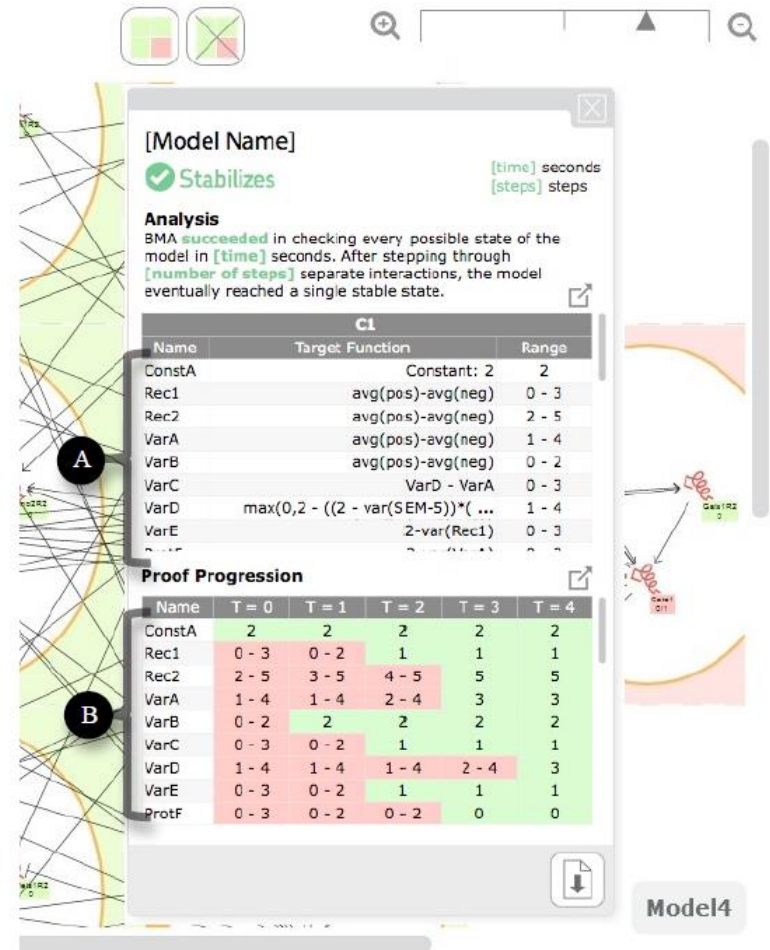
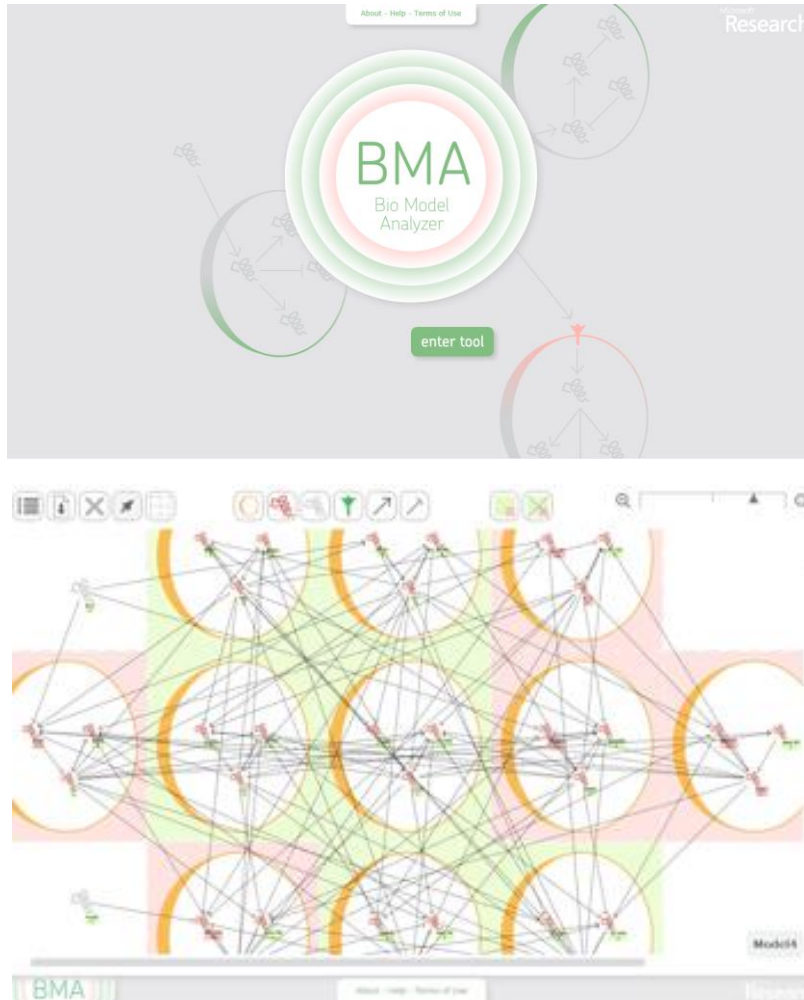
Location, Disease • Gene Text

Browsing



Genome Query Language (GQL)

Computing Cancer




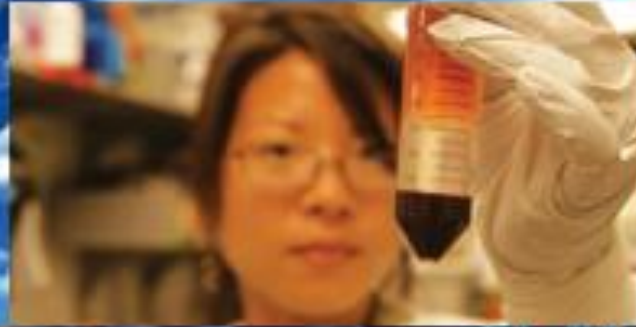
<http://biomodelanalyzer.research.microsoft.com/>

Jasmin Fisher, Microsoft Research Cambridge

What next?

Bringing cloud computing to researchers

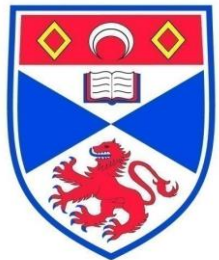
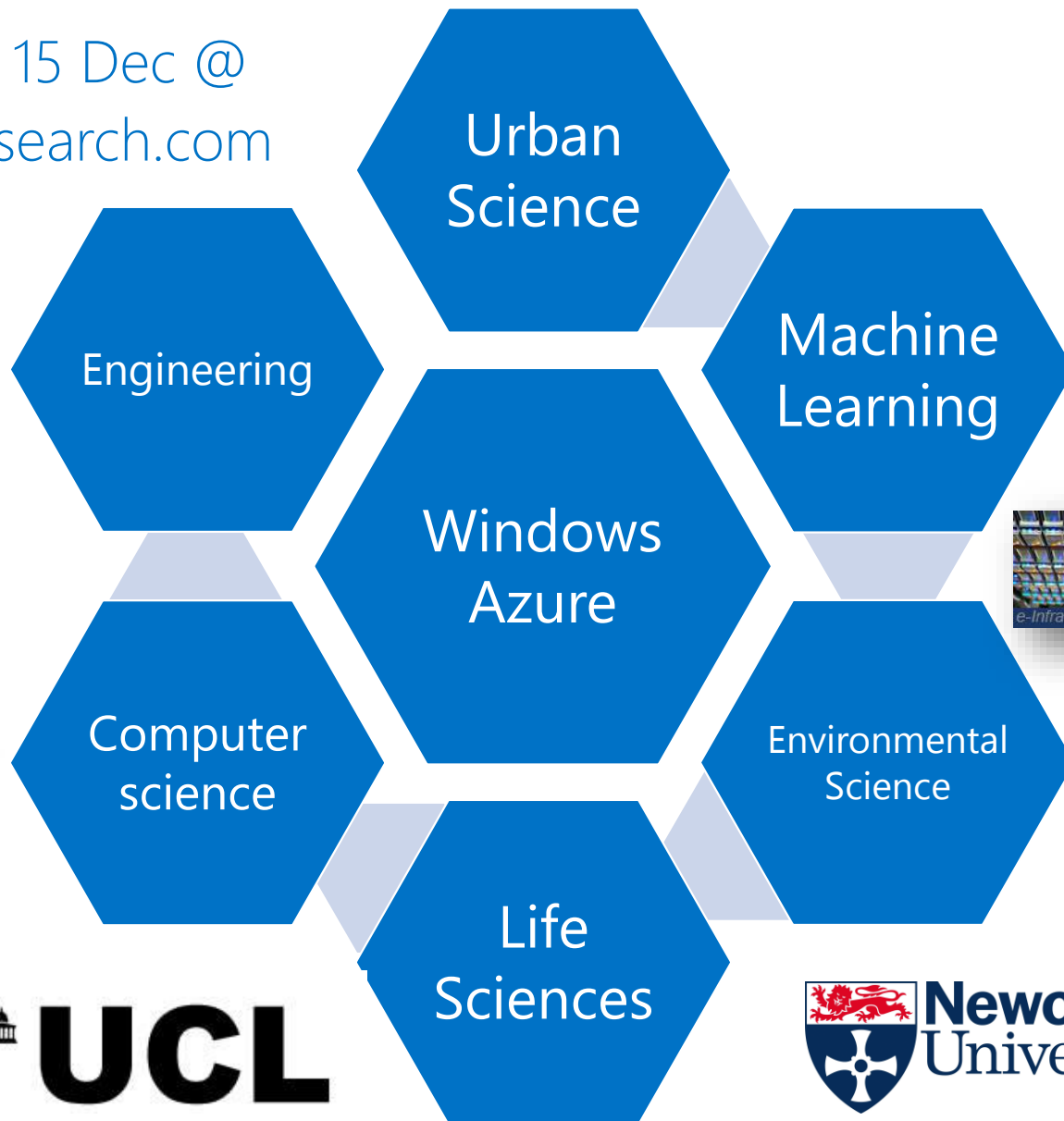
Learn about the new Windows Azure for Research program 



www.azure4research.com

Windows Azure Awards


Apply by 15 Dec @
azure4research.com



University
of
St Andrews



Bringing cloud computing to researchers

Learn about the new Windows Azure for Research program 



Azure Research Awards
Windows Azure for Research
Training (*kenjitak@microsoft.com*)
20-21 Jan'14, Oxford
3-4 Apr'14, Manchester
Webinars (Linux VMs, 4th Dec'13)
Technical papers & curriculum
Research community engagements



[Windows Azure for Research Group](#)



[@azure4research](#) &
[#azureresearch](#)

www.azure4research.com

Thank you!

Microsoft
Research Connections

