Microsoft

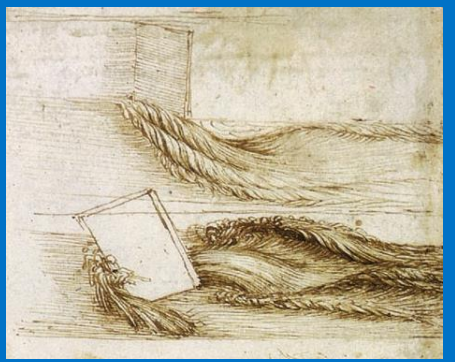
Windows Azure for Research

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

@azure4research #azureresearch

The Nature of Scientific Discovery

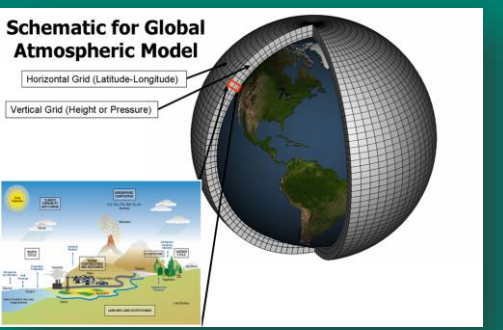
Experiment



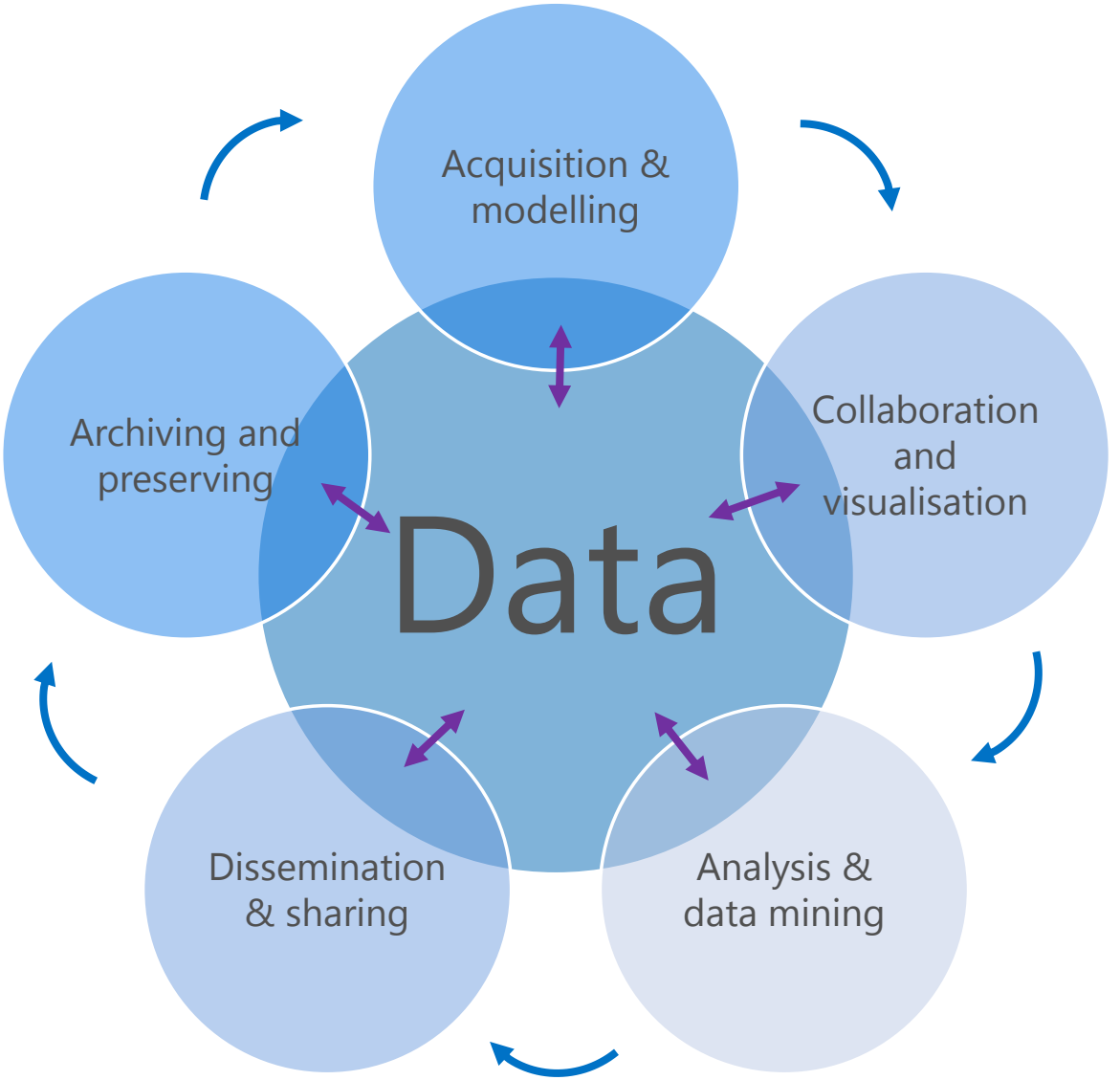
Theory

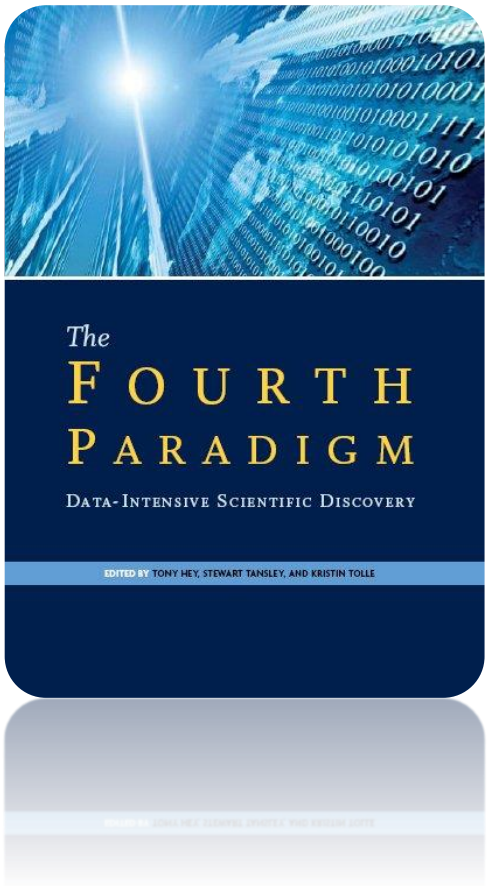
$$\rho \frac{Dv}{Dt} = -\nabla p + \nabla \cdot \boldsymbol{T} + \boldsymbol{f}$$

Computation



Data-intensive Research





fourthparadigm.org

Let scientists be scientists….

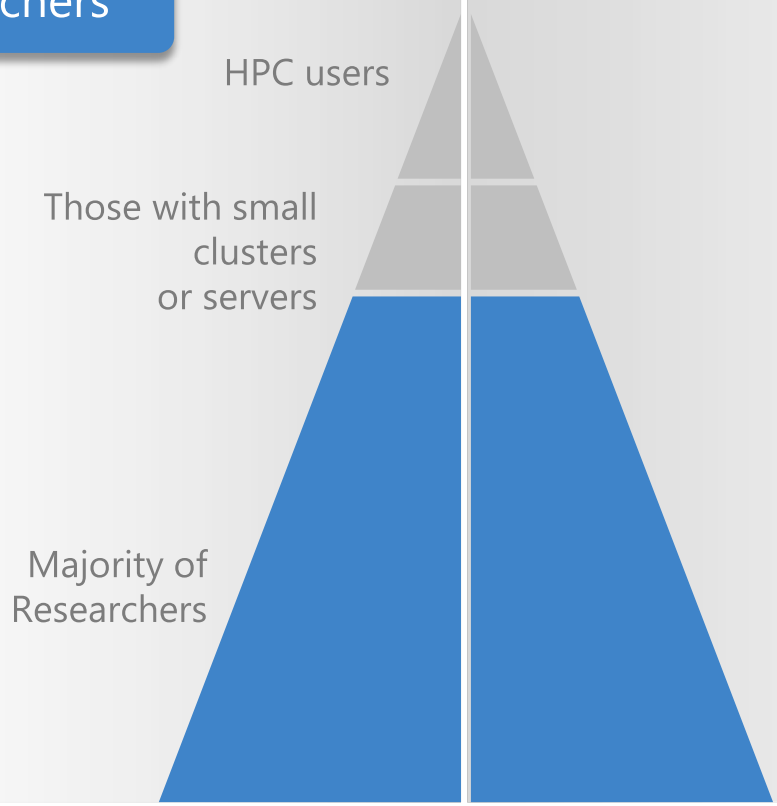
Today

Majority of Researchers

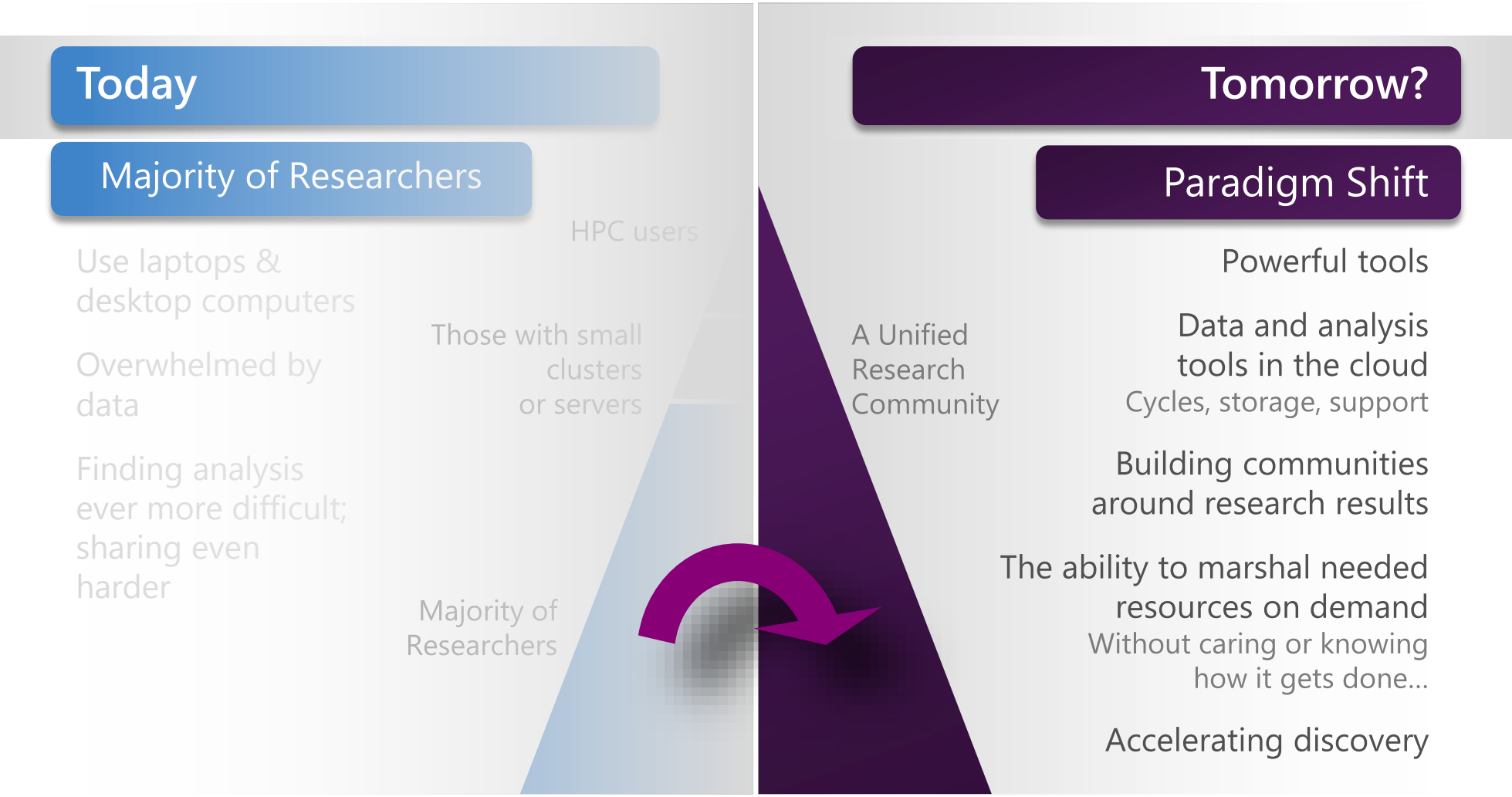
Use laptops & desktop computers

Overwhelmed by data

Finding analysis ever more difficult; sharing even harder



Let scientists be scientists….



Let scientists be scientists….

Azure for Research

Azure Research Awards

Windows Azure for Research Training two-day workshops

Technical resources & curriculum

Research community engagements

A Unified Research Community

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

kenjitak@microsoft.com

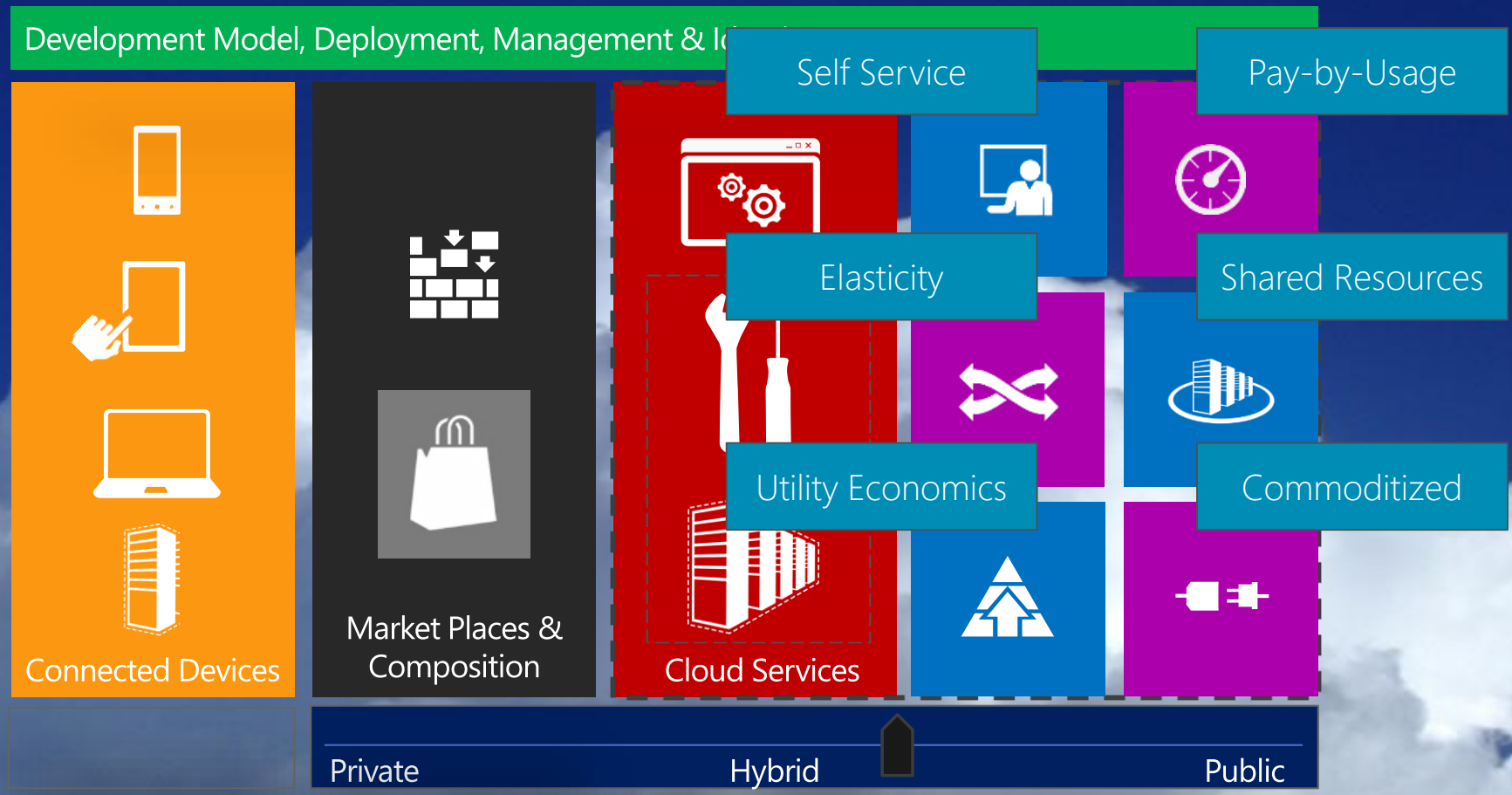
www.azure4research.com

Cloud Computing

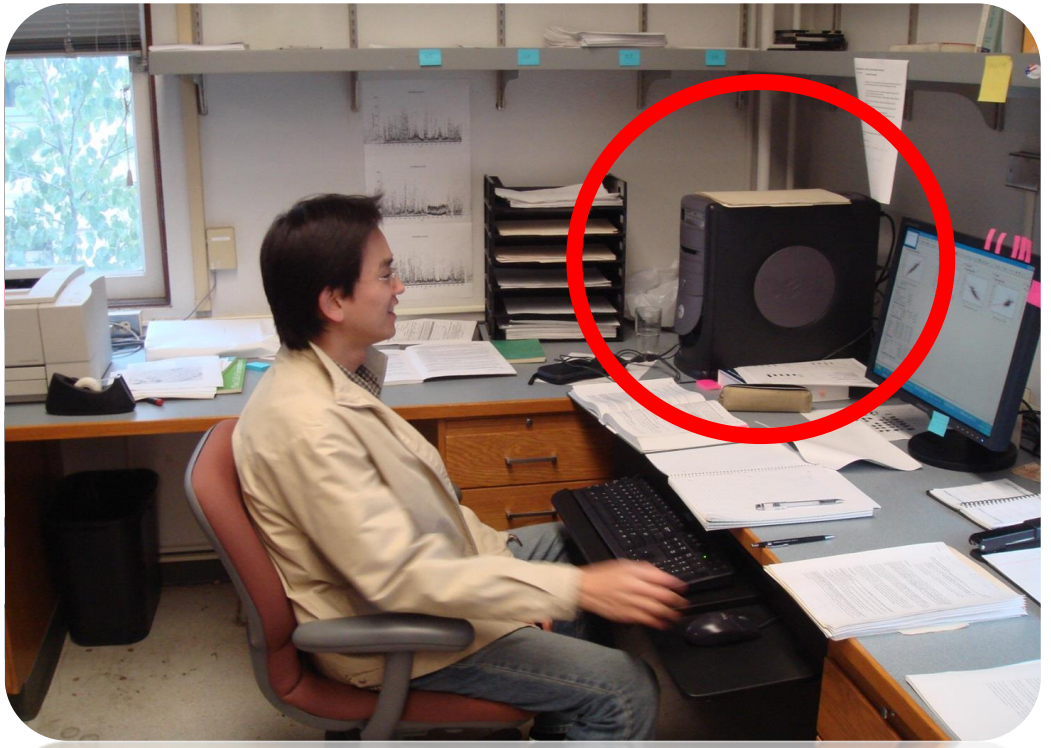
Cloud computing provides…

On-demand services,
delivered over the network

The Cloud


Development Model, Deployment, Management & I
Self Service
Pay-by-Usage
Elasticity
Shared Resources
Utility Economics
Commoditized
Connected Devices
Market Places & Composition
Cloud Services
Private
Hybrid
Public

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…

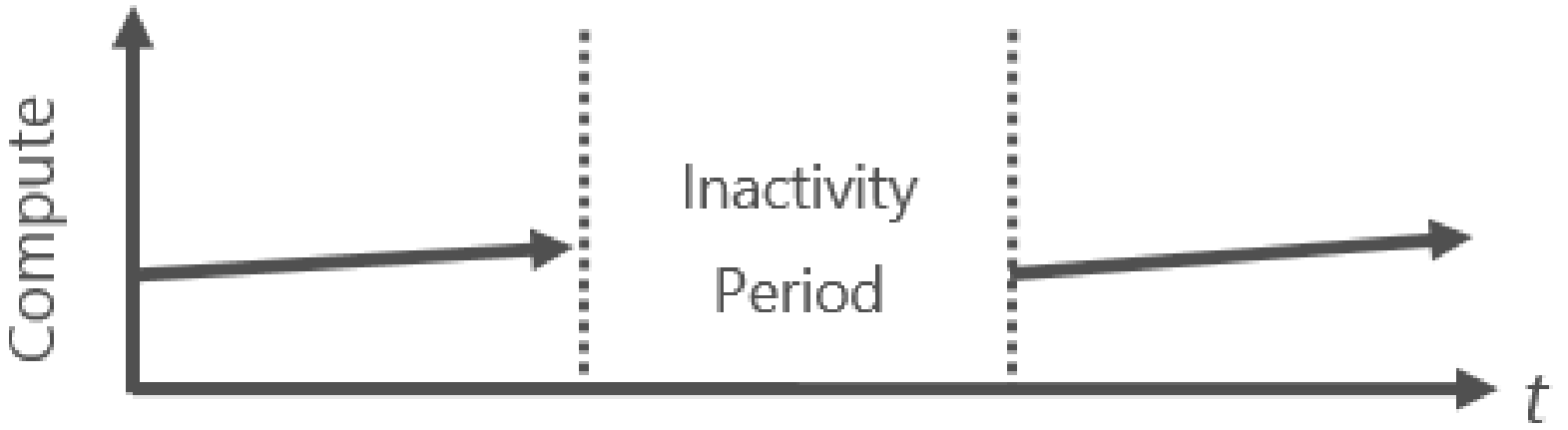
Cloud computing makes a difference

- 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...

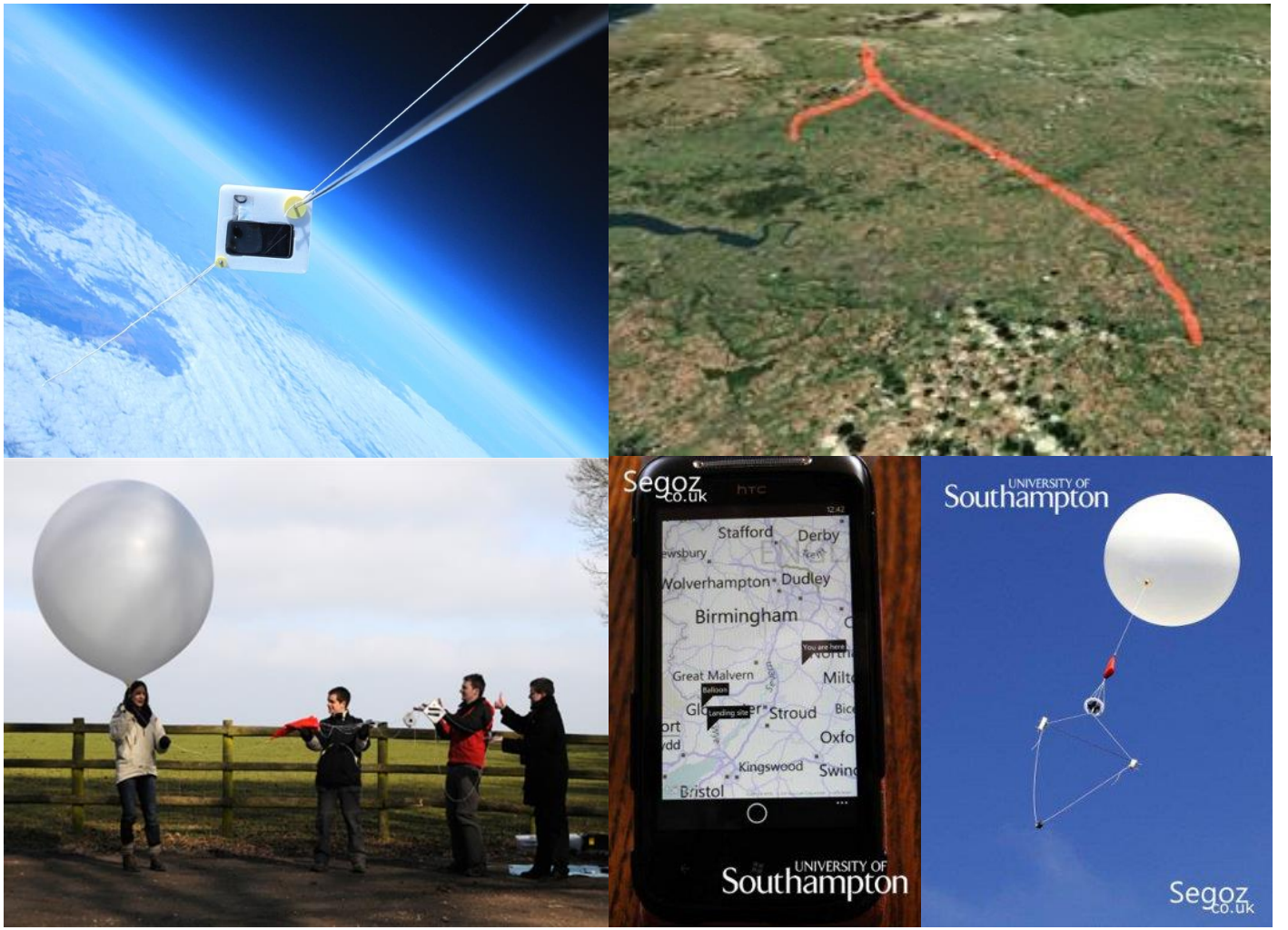
Getting what you need,
when you need it

Patterns: On and Off

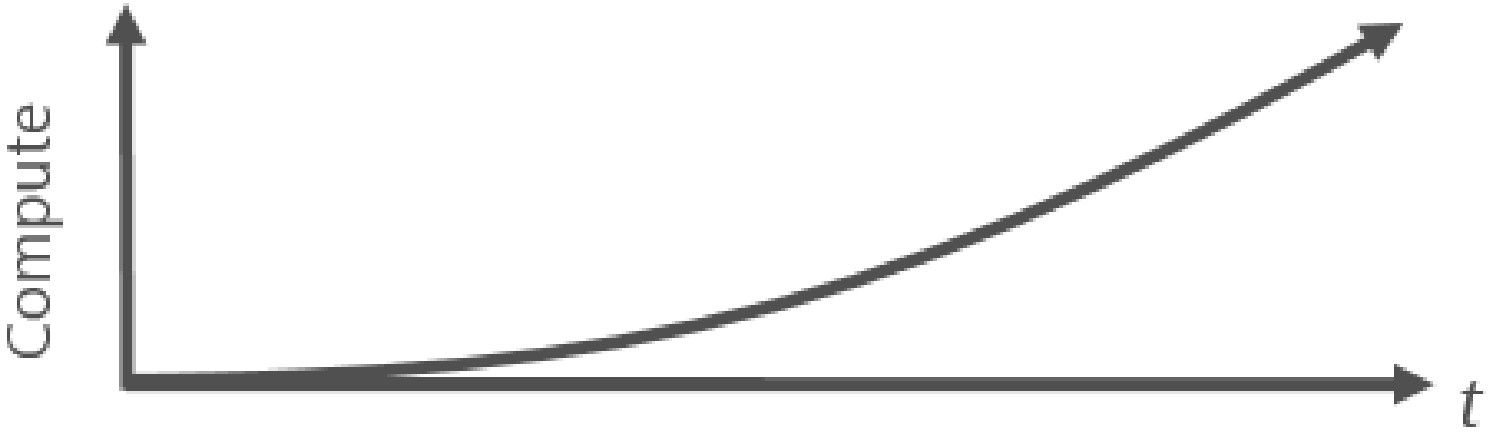


On & off workloads (e.g. batch job)
Compute only when you need it
No queuing
Over provisioned capacity is wasted
Difficult to handle conventionally

Stratospheric Research



Patterns: Growing Fast



Need to grow/scale quickly

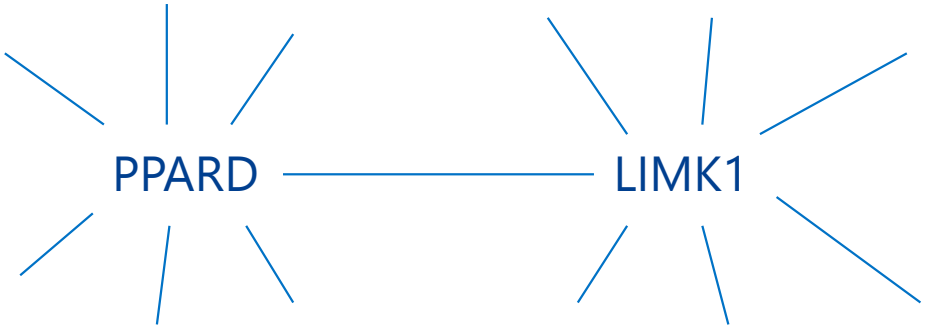
Equally important for data

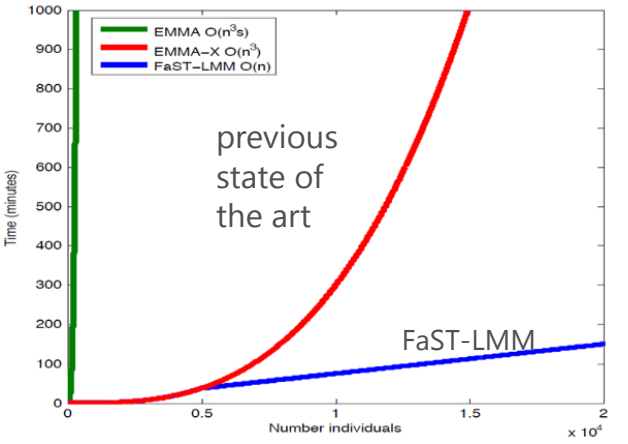
Keeping up is difficult

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



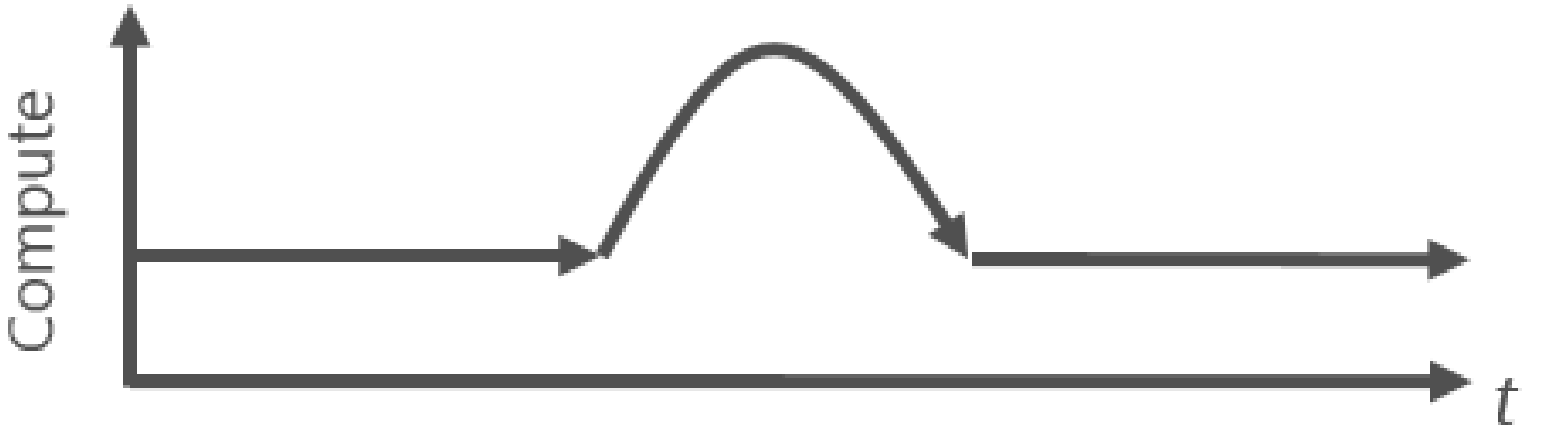


SCIENTIFIC REPORTS

An Exhaustive Epistatic SNP Association Analysis on Expanded Wellcome Trust Data

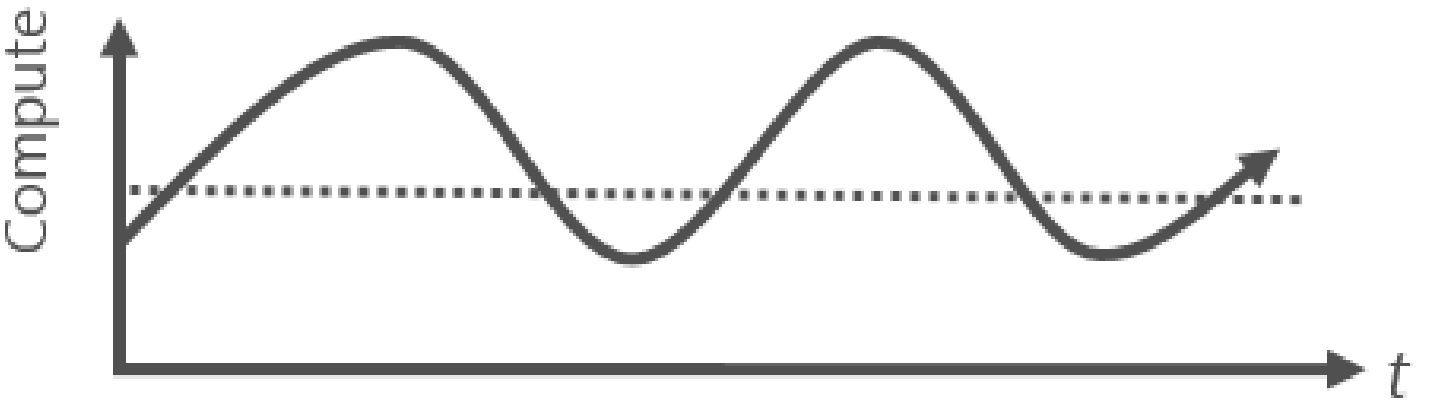
Christoph Lippert, Jennifer Listgarten, Robert I. Davidson, Jeff Baxter, Hoifung Poon, Carl M. Kadie & David Heckerman

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

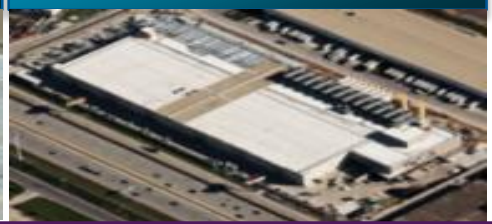
Why Cloud Computing?

- 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

Data Center Co-Location Generation 1	Quincy and San Antonio Generation 2	Chicago and Dublin Generation 3	Modular Data Center Generation 4
			

Deployment Scale Unit

Server	Rack	Containers	IT PAC
			
Capacity	Density & Deployment	Scalability & Sustainability	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

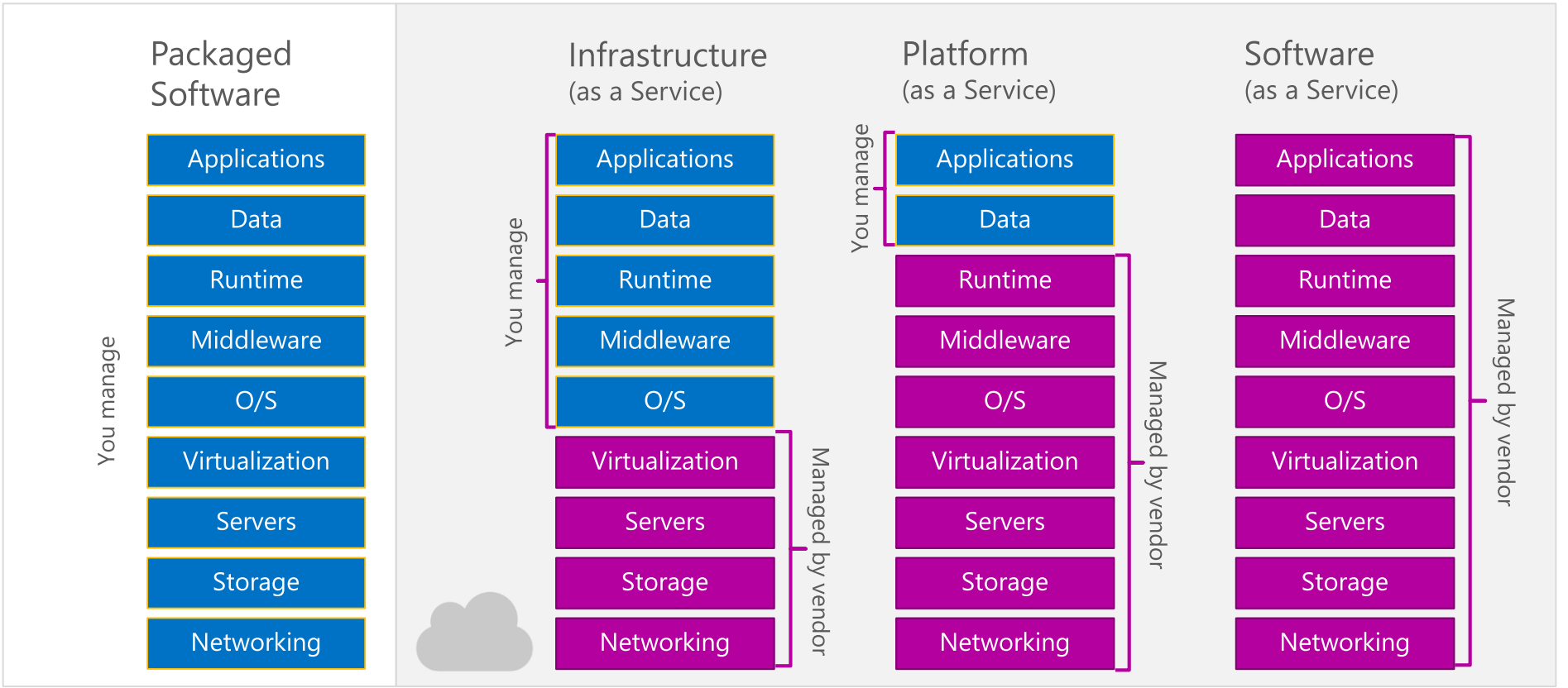
build

SaaS

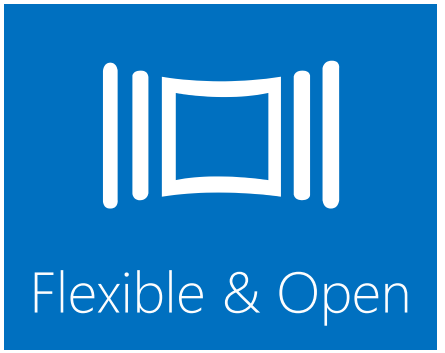
Software-as-a-Service

consume

Cloud Computing Variants



Windows Azure

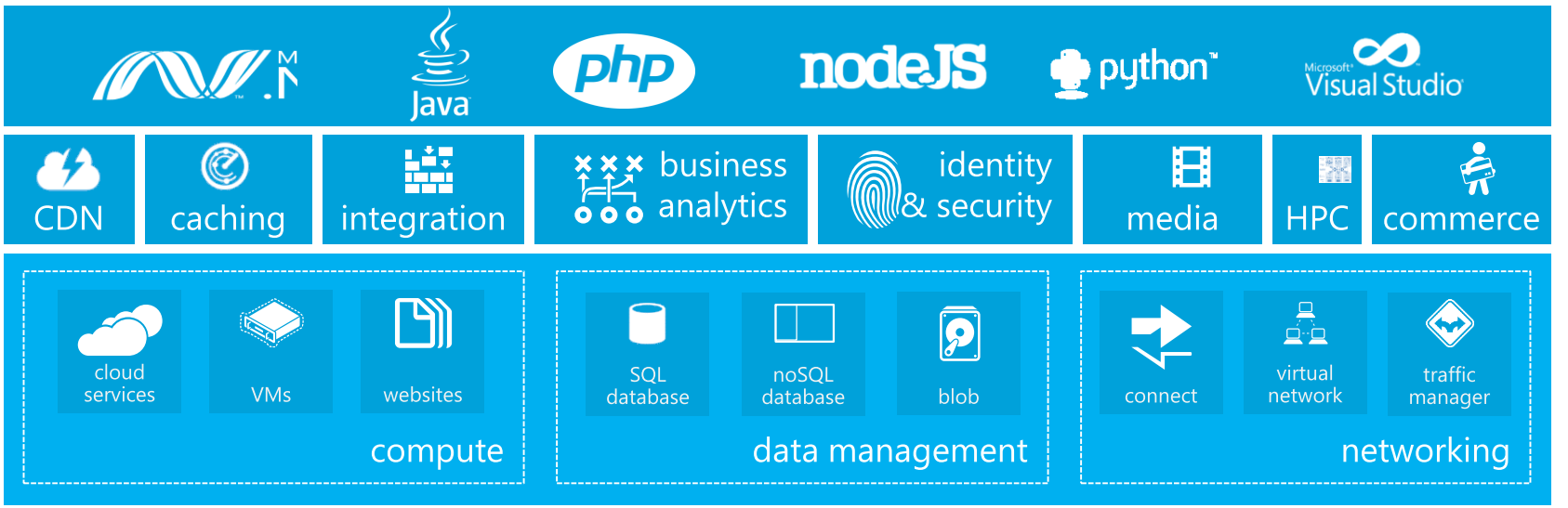


Choose from multiple runtimes and languages for your applications: Python, Java, PHP, .NET, Node.js

Run Linux on Windows Azure Virtual Machines (VHD)

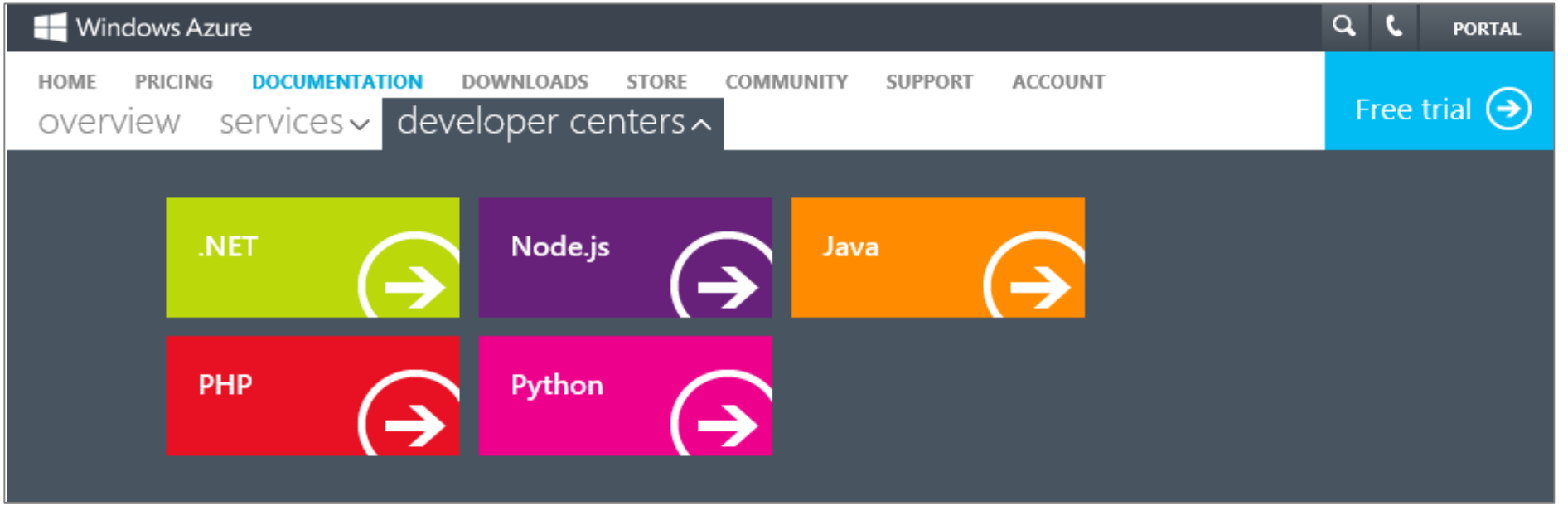
Support multiple frameworks and popular open source applications with Windows Azure Web Sites

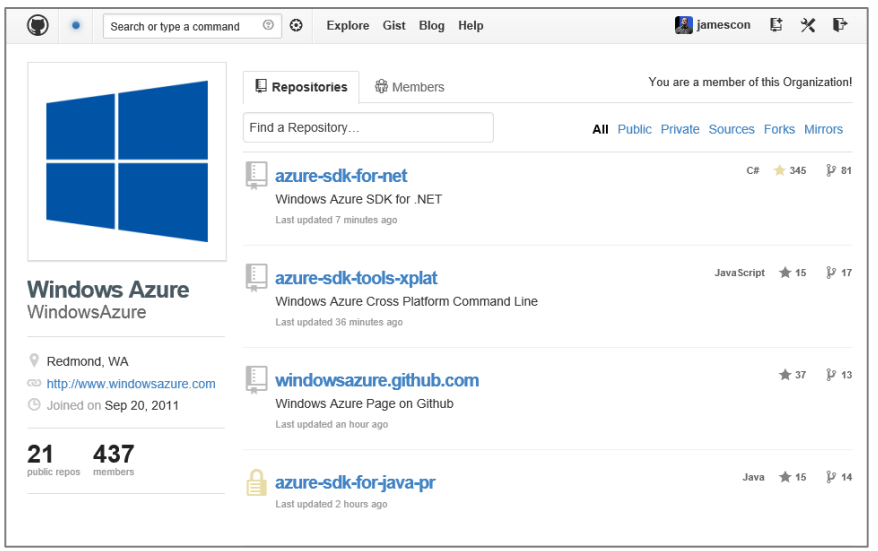
HDInsight Hadoop for Big Data analysis

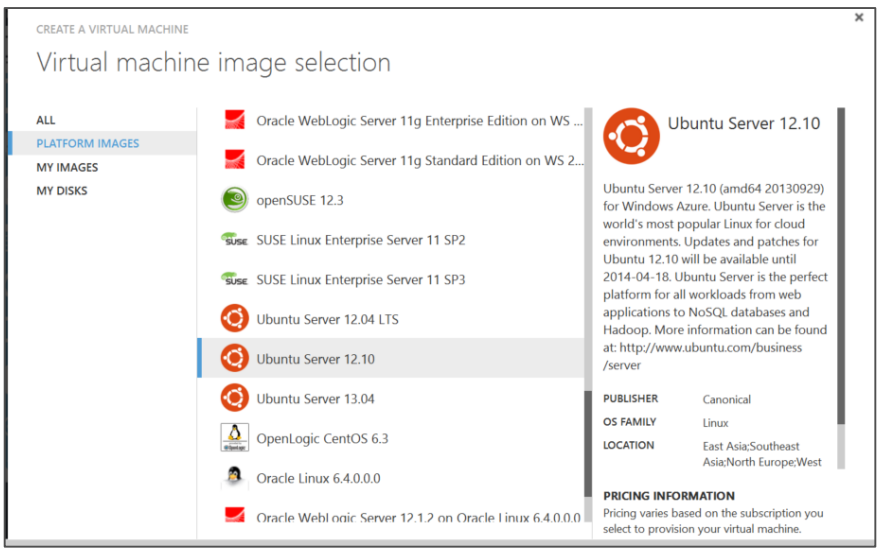


http://github.com/windowsazure

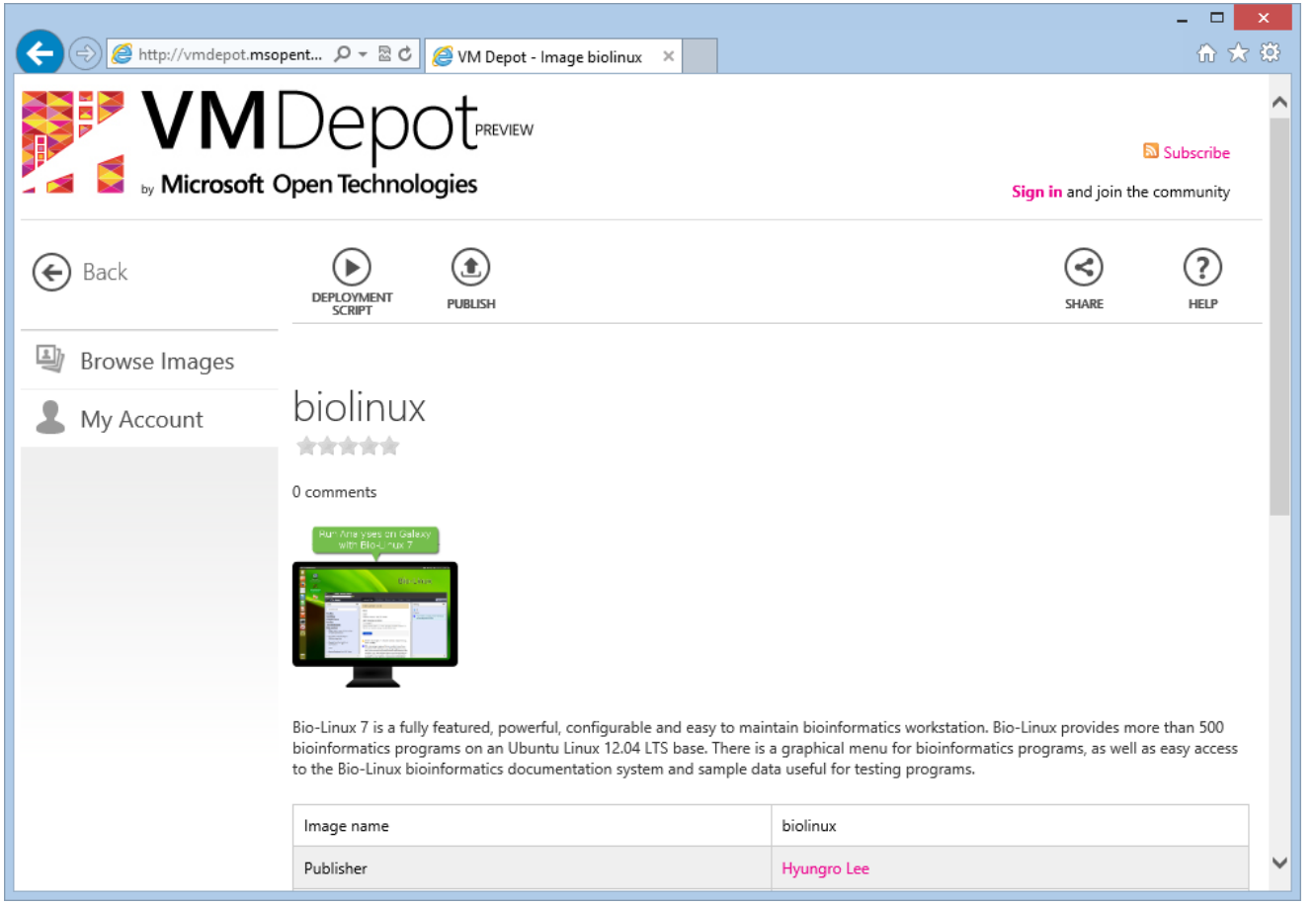
Flexible & Open



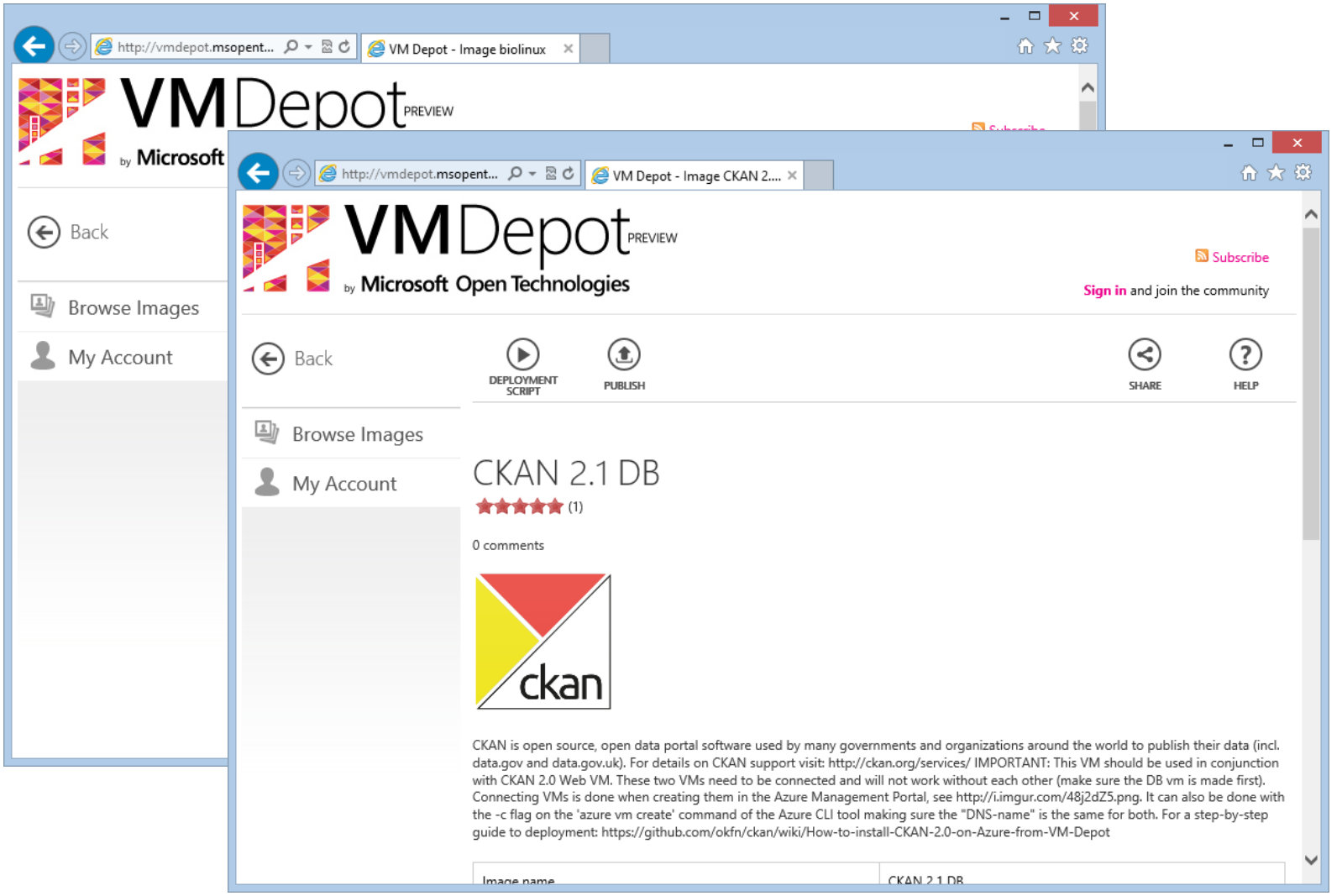




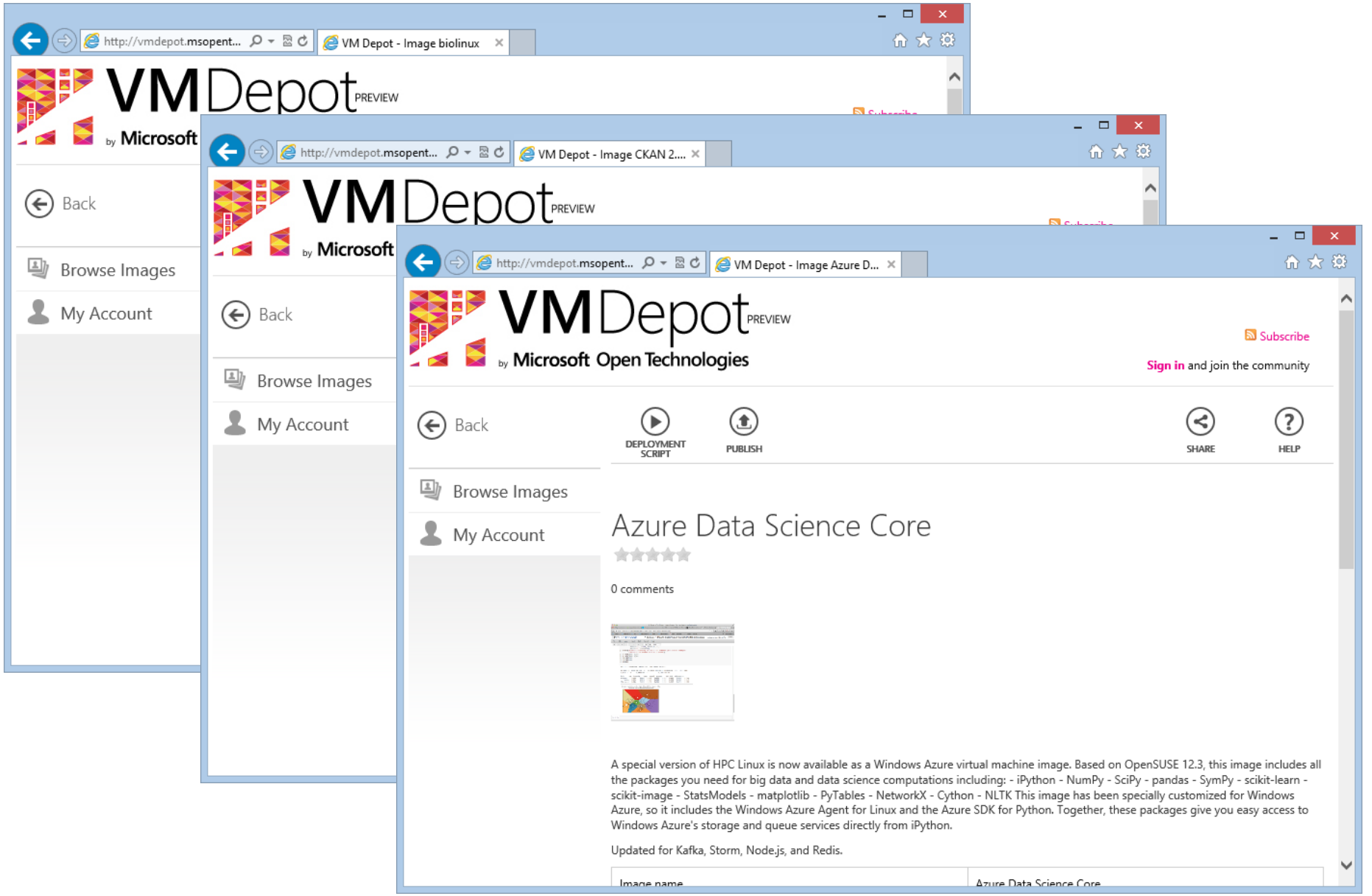
Virtual Machines



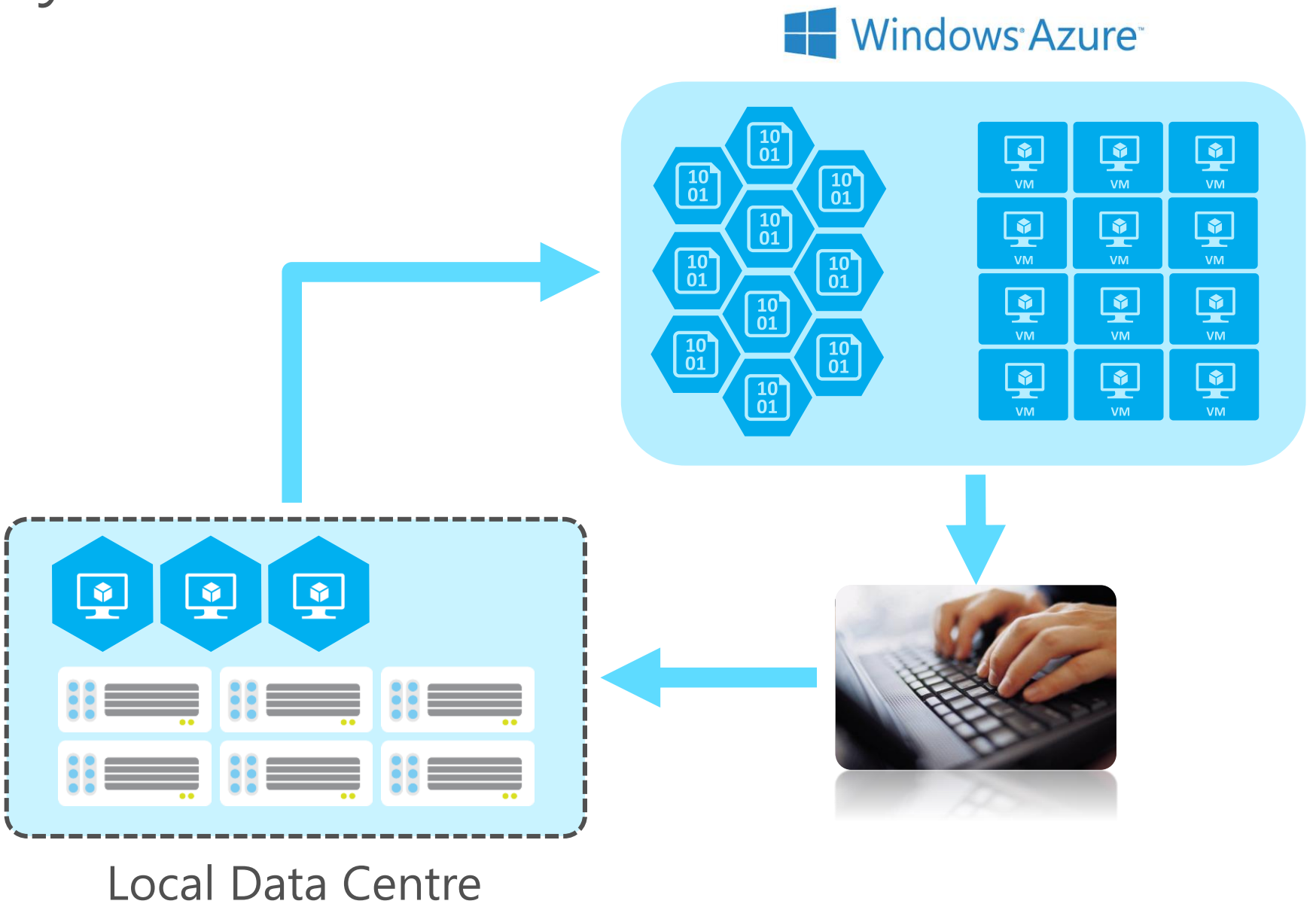
Virtual Machines



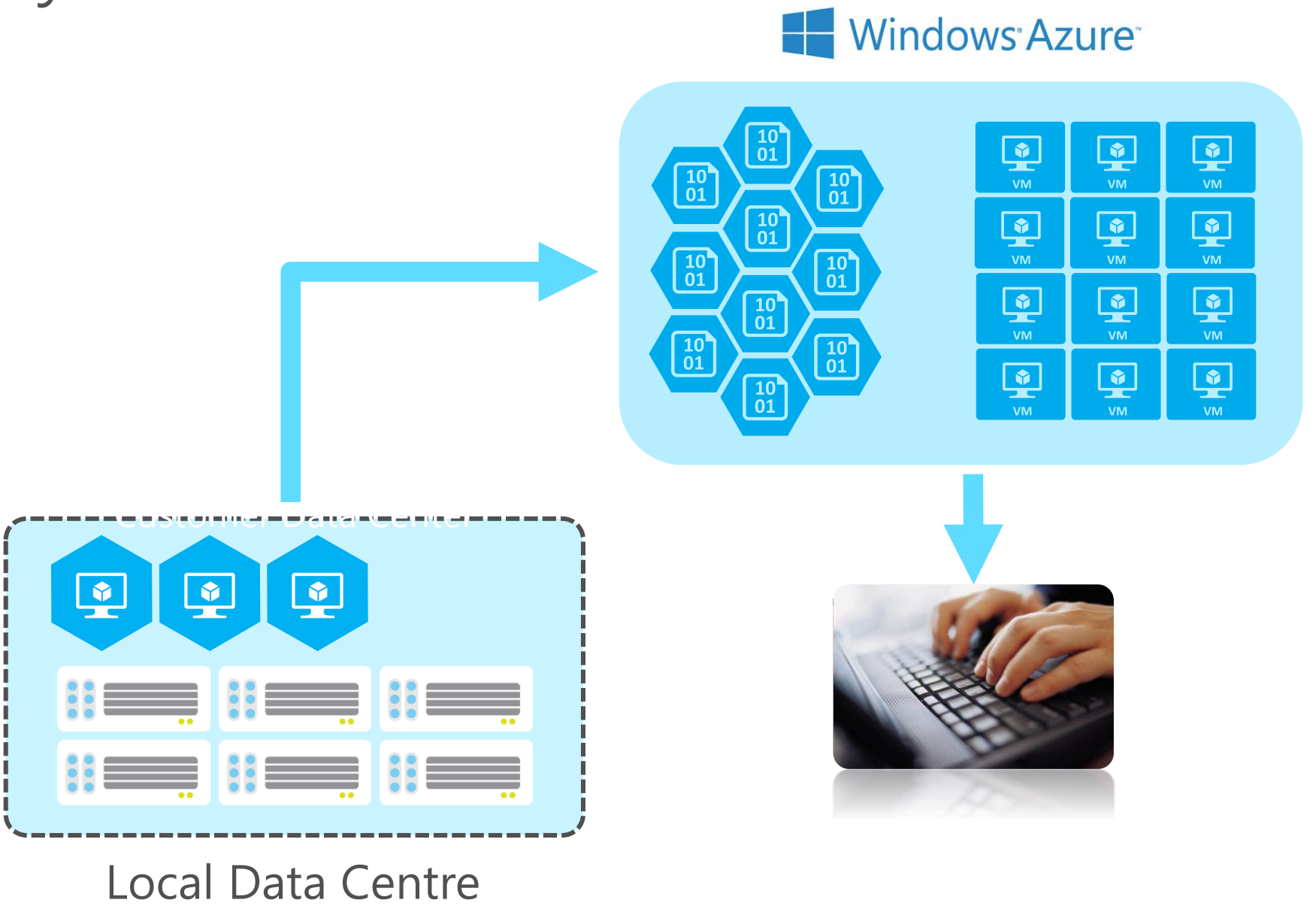
Linux Virtual Machines



Hybrid Cloud for Science



Hybrid Cloud for Science



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

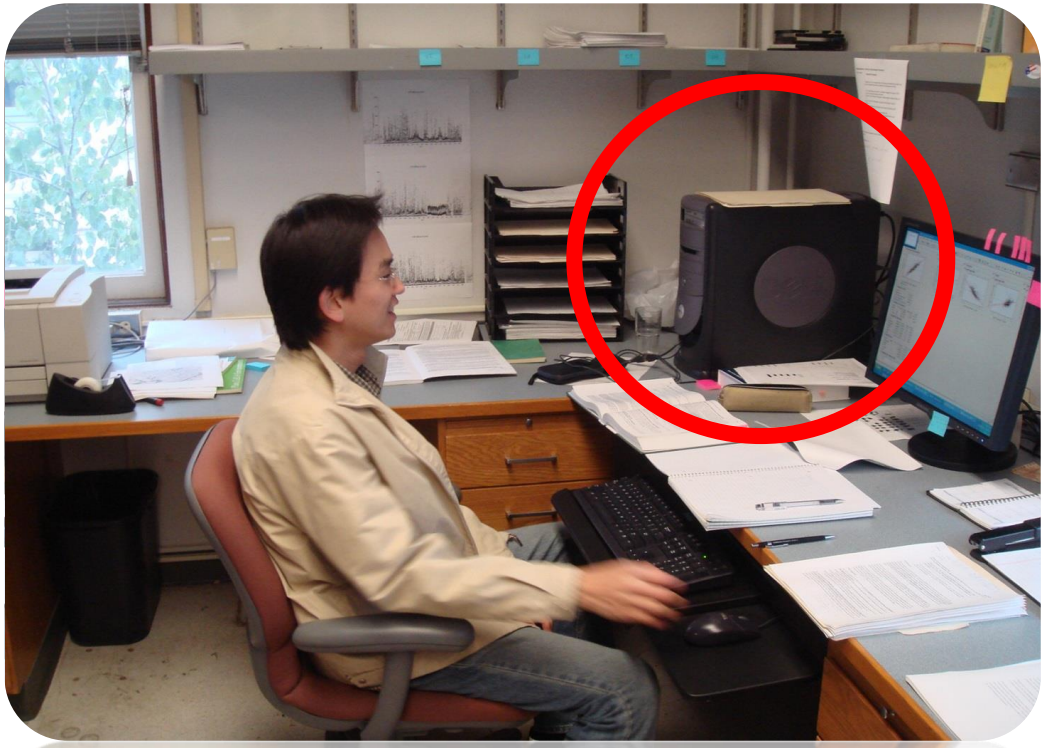
Over 18 million students and staff to benefit from faster, more secure cloud-computing

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

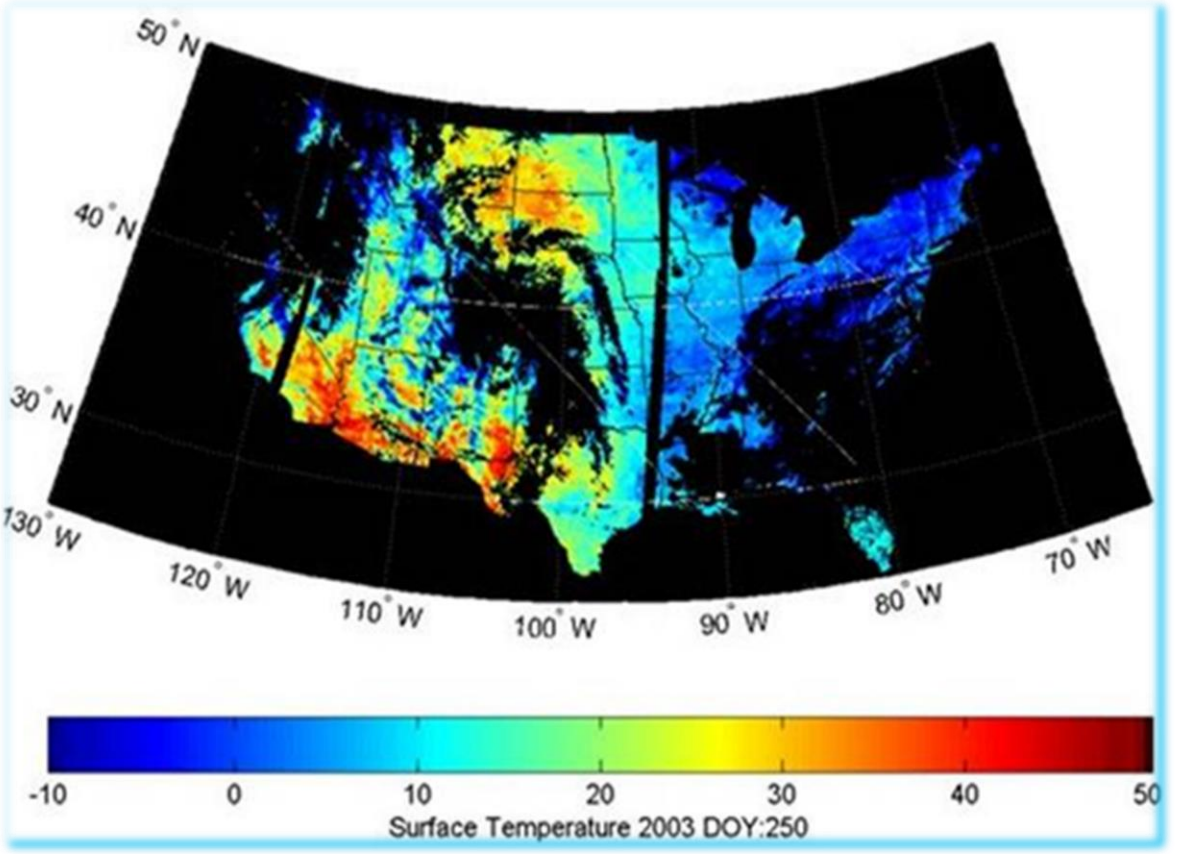
Needed to run calculations against NASA MODIS data

Problems with data and compute

Cloud computing makes a difference…

MODIS Azure:

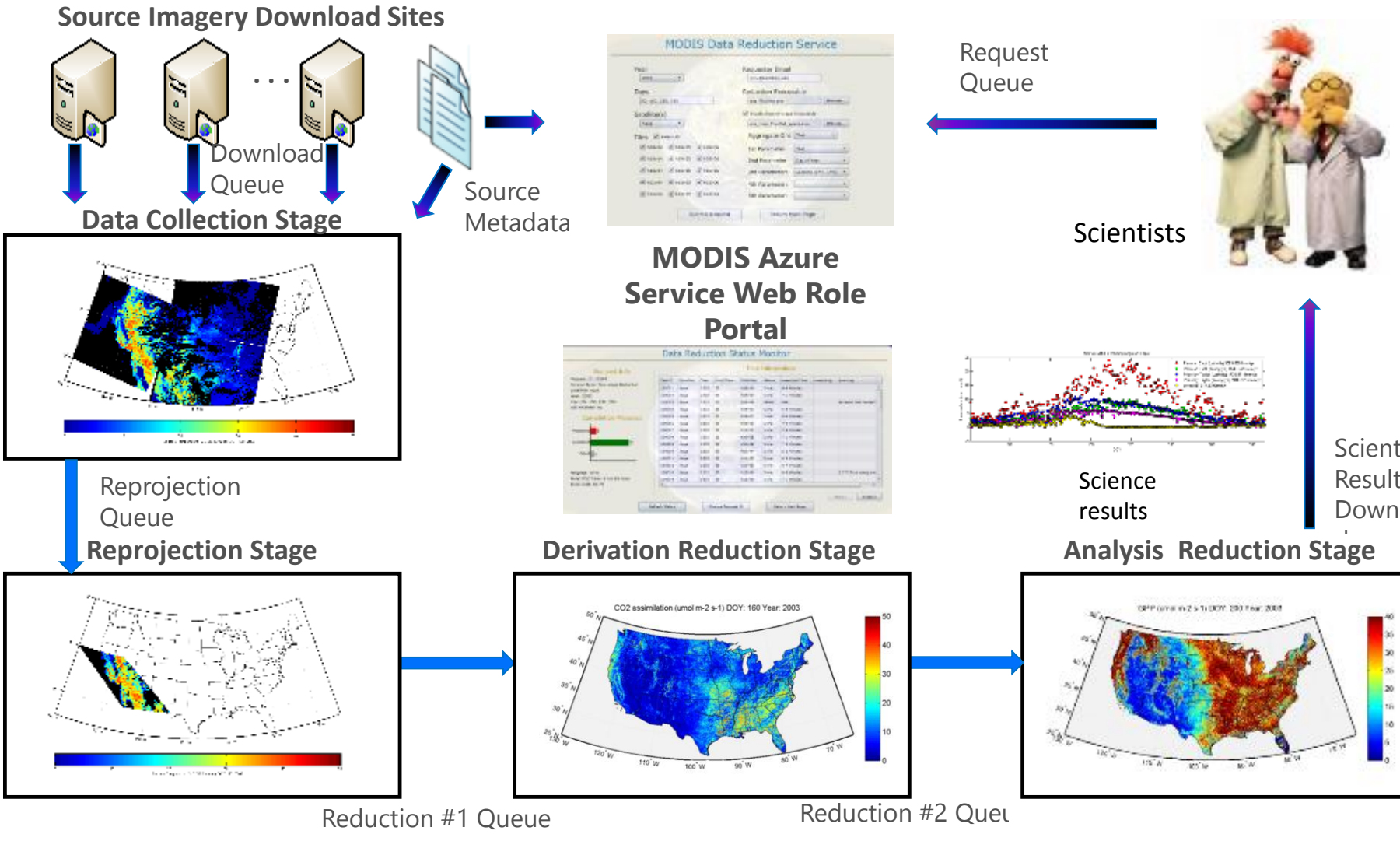
Computing Evapotranspiration (ET) in the Cloud



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

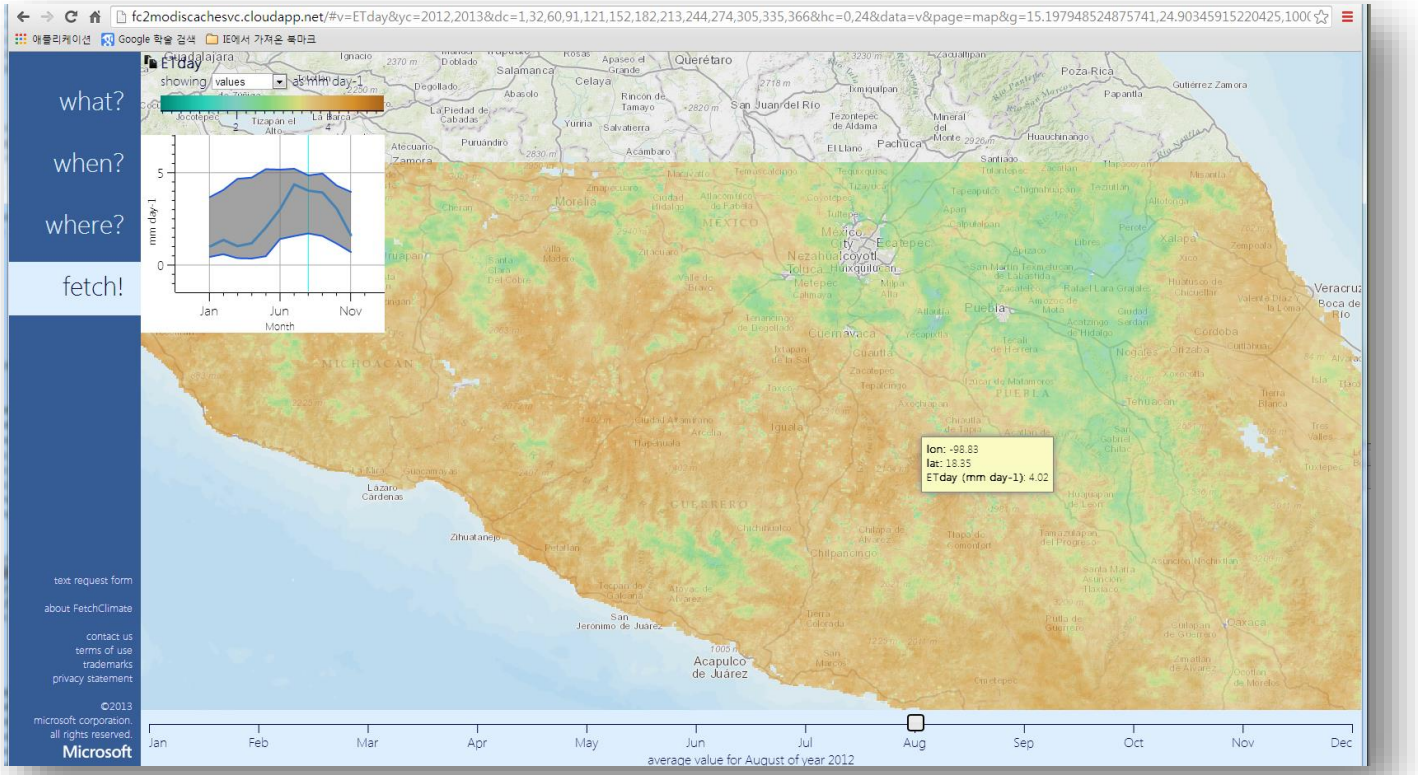


Catharine van Ingen (Microsoft Research), Jie Li, Marty Humphrey (UVA), Youngryel Ryu (UCB), Deb Agarwal (BWC/LBL), Keith Jackson (BL), Jay Borenstein (Stanford) , Team SICT: Vlad Andrei, Klaus Ganser, Samir Selman, Nandita Prabhu (Stanford), Team Nimbus: David Li, Sudarshan Rangarajan, Shantanu Kurhekar,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

FetchClimate[Beta]

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

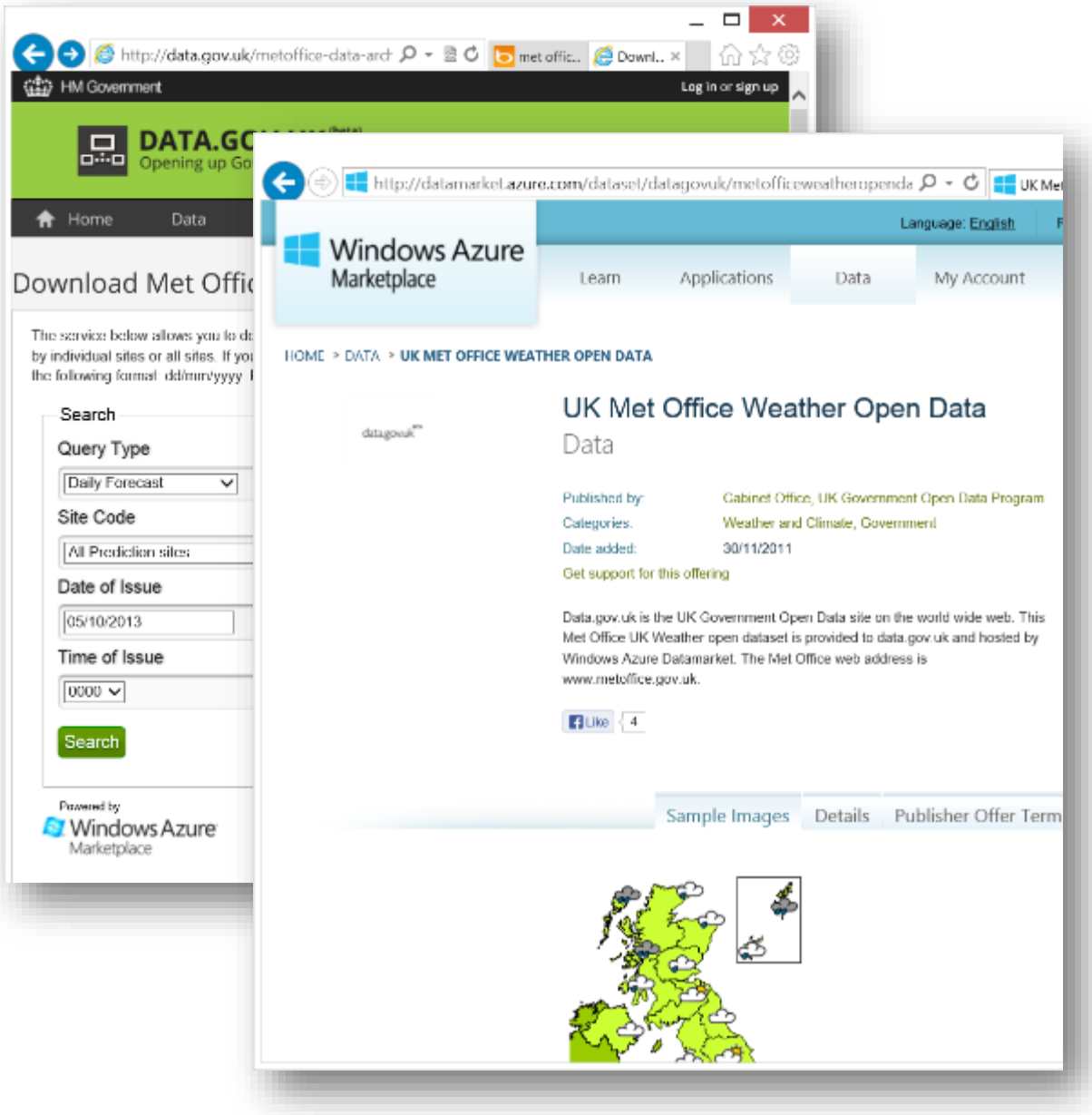
Powered by Windows Azure



Open UK Weather Forecast Data

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

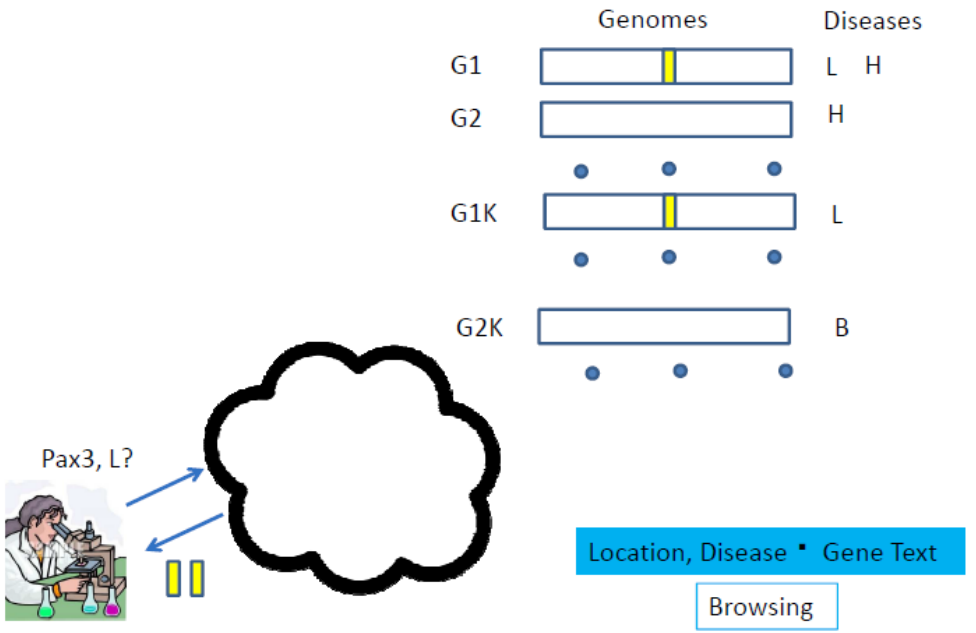
Web search:
"open weather data azure"

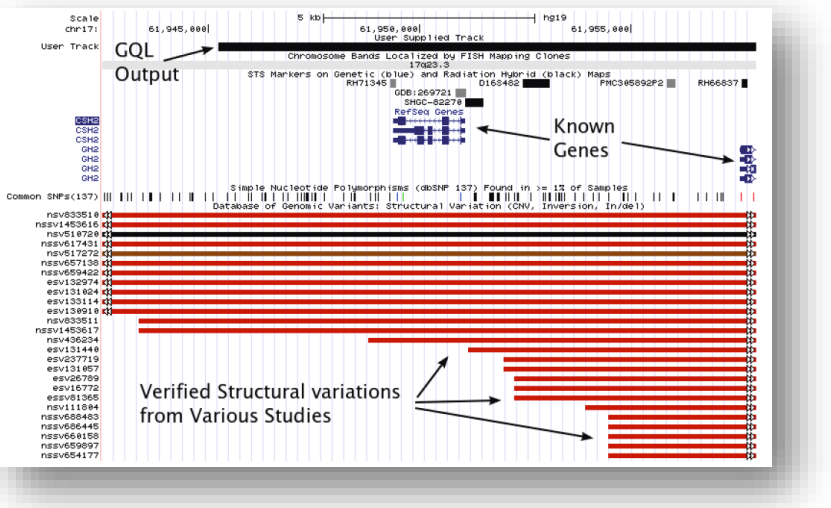


Interactive Genomics

Rapidly Querying Genomes in the Cloud

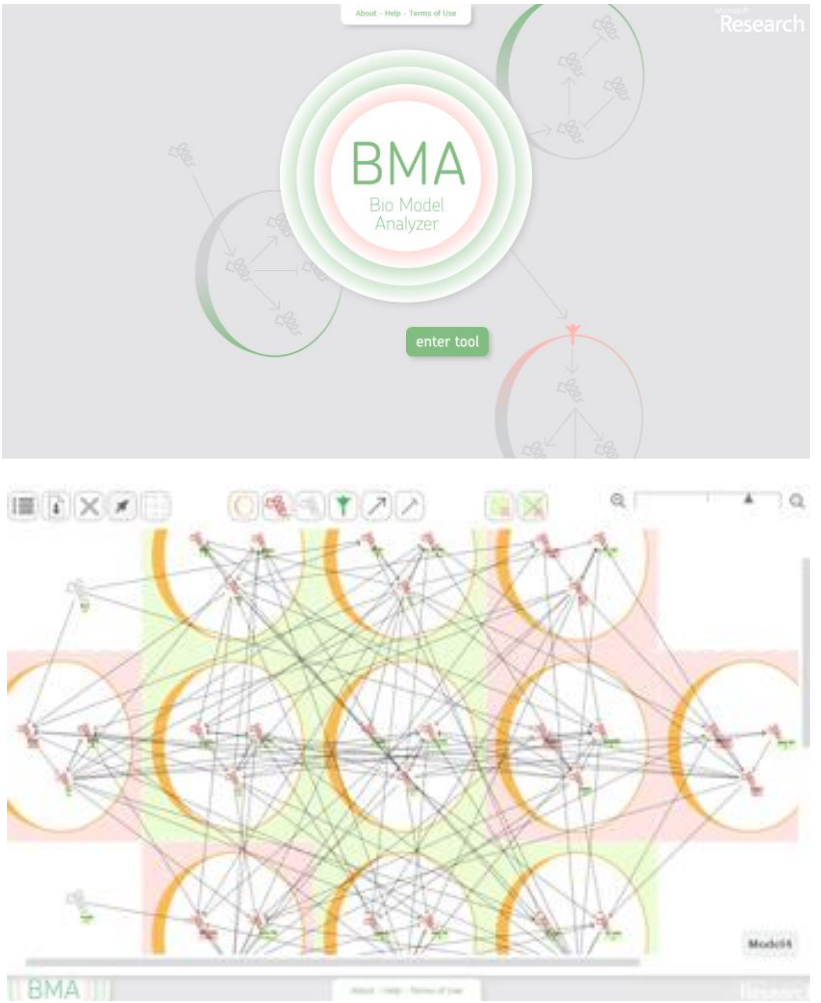
Christos Kozanitis(UCSD), Vineet Bafna(UCSD), Ravi Pandya(MSR), George Varghese(MSR)





Genome Query Language (GQL)

Computing Cancer



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

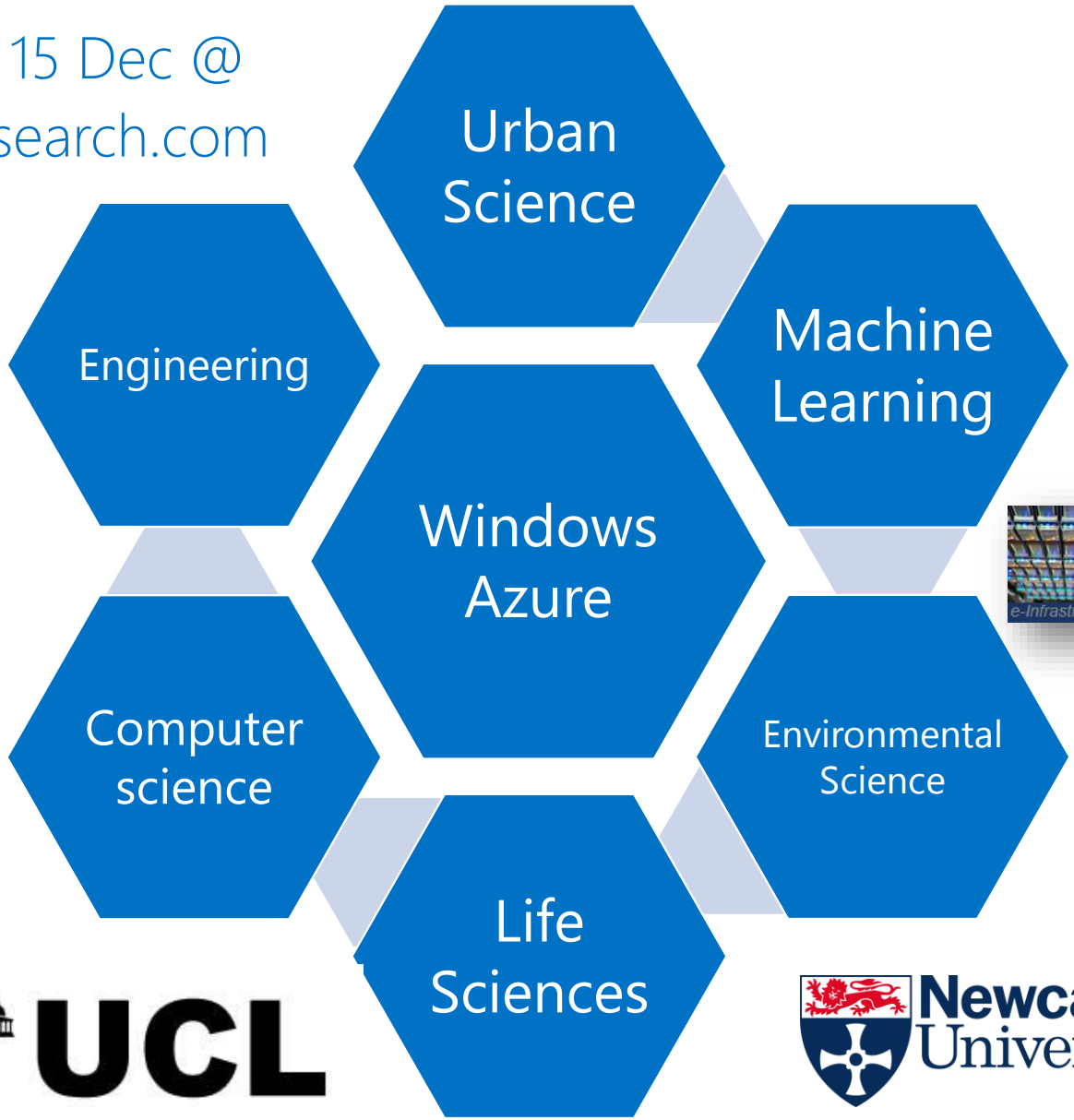
Jasmin Fisher, Microsoft Research Cambridge

What next?

www.azure4research.com

Windows Azure Awards

Apply by 15 Dec @
azure4research.com



BRITISH LIBRARY

University of St Andrews

jasmin
e-Infrastructure for climate and earth system science

UCL

Newcastle University

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