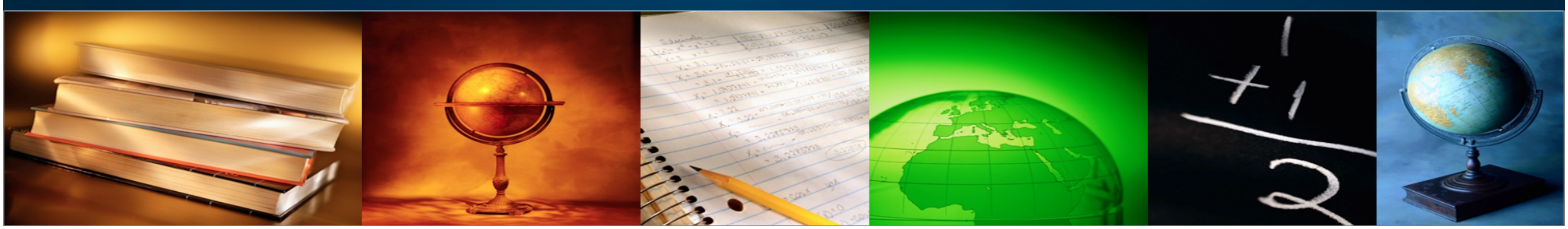
Data Services for Scientific Computing

Tony Hey
Corporate Vice President
Microsoft Research



Scientific Data

In 2000 the Sloan Digital Sky Survey collected more data in its 1st week than was collected in the entire history of Astronomy

By 2016 the New Large Synoptic Survey Telescope in Chile will acquire 140 terabytes in 5 days - more than Sloan acquired in 10 years

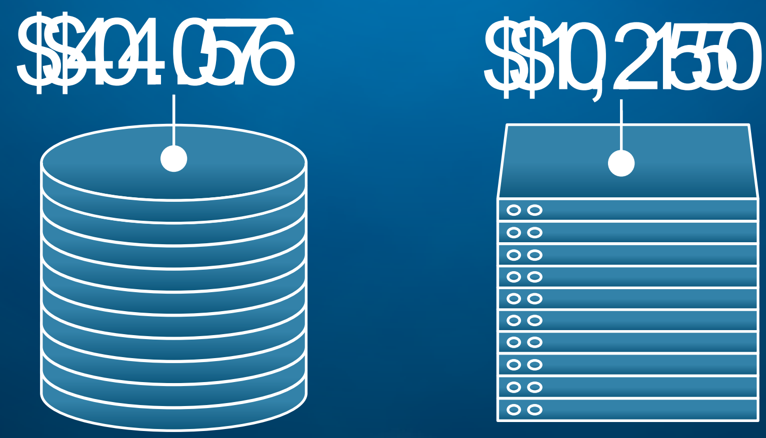
The Large Hadron Collider at CERN generates 40 terabytes of data every second

Sources: The Economist, Feb '10; IDC


The Economist
FEBRUARY 27TH–MARCH 5TH 2010
Economist.com
Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs
The data deluge
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

Economics of Storage

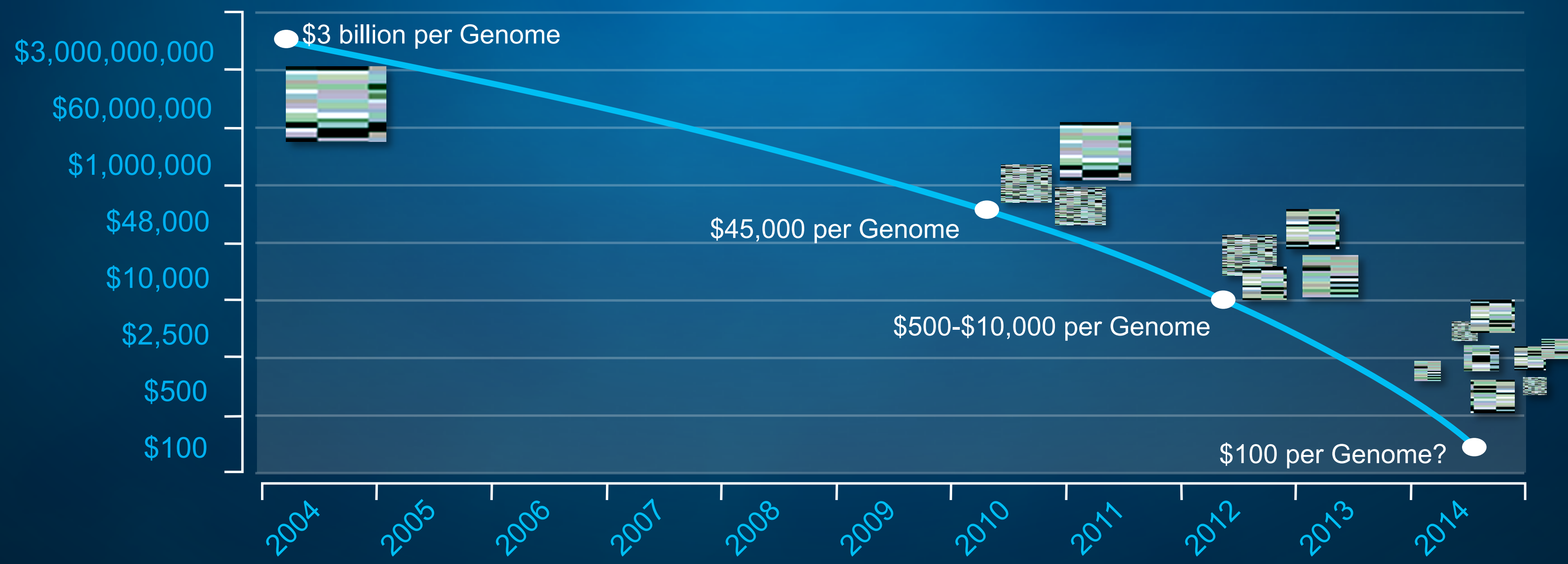
Hard Drive Storage
(per gigabyte)



Web Storage
(per gigabyte)

Source: Wired Magazine April 2010; Figures represented in USD

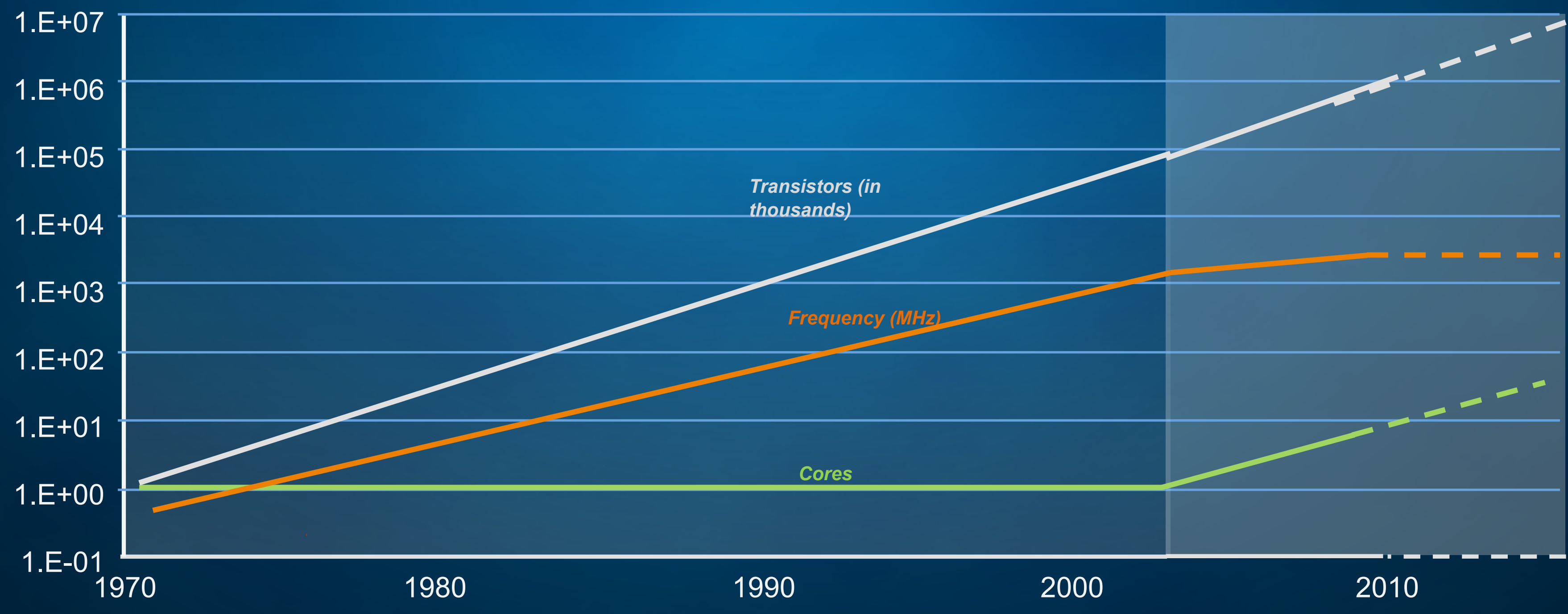
Cost per Genome



Source: George Church, Harvard Medical School, as reported in IEEE Spectrum, Feb '10. Figures represented in USD

Moore's Law is alive and well...

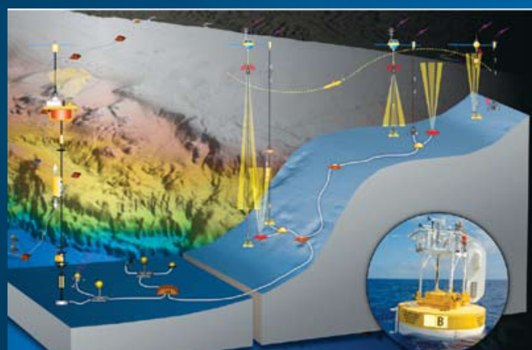
…but a hardware issue just became a software problem



Source: Jack Dongarra, Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, Krste Asanovic, and Kathy Yelick

Computing Tools for Big Data

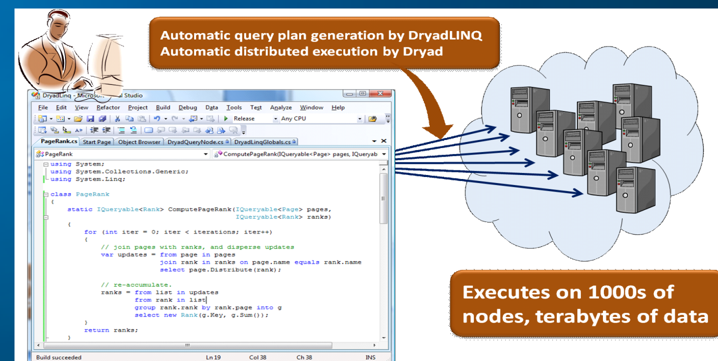
Scientific Workflow Workbench (Trident)



- Built on top of Windows Workflow Foundation
- Visually program workflows with the use of libraries of activities and workflows
- Scale from desktops to HPC clusters
- Distribution: Moving work closer to the data source
- Workflow sharing in myExperiment social Web site for researchers

Version 1.2 available for download on CodePlex (Apache 2.0 open source)

Dryad and DryadLINQ



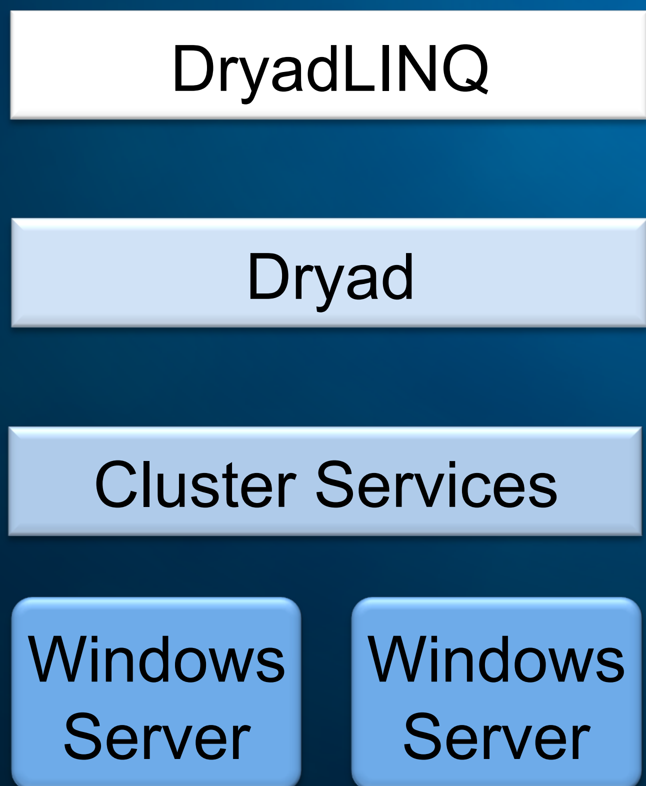
- Programming models for writing distributed data-parallel applications that scale from a small cluster to a large data-center.
- A DryadLINQ programmer can use thousands of machines, each of them with multiple processors or cores, without prior knowledge in parallel programming.

Academic release available for download

Dryad

- Continuously deployed since 2006
- The execution engine for Bing analytics
- Running on >> 10^4 machines
- Runs on clusters > 3000 machines
- Sifting through > 10Pb data daily

Dryad & DryadLINQ



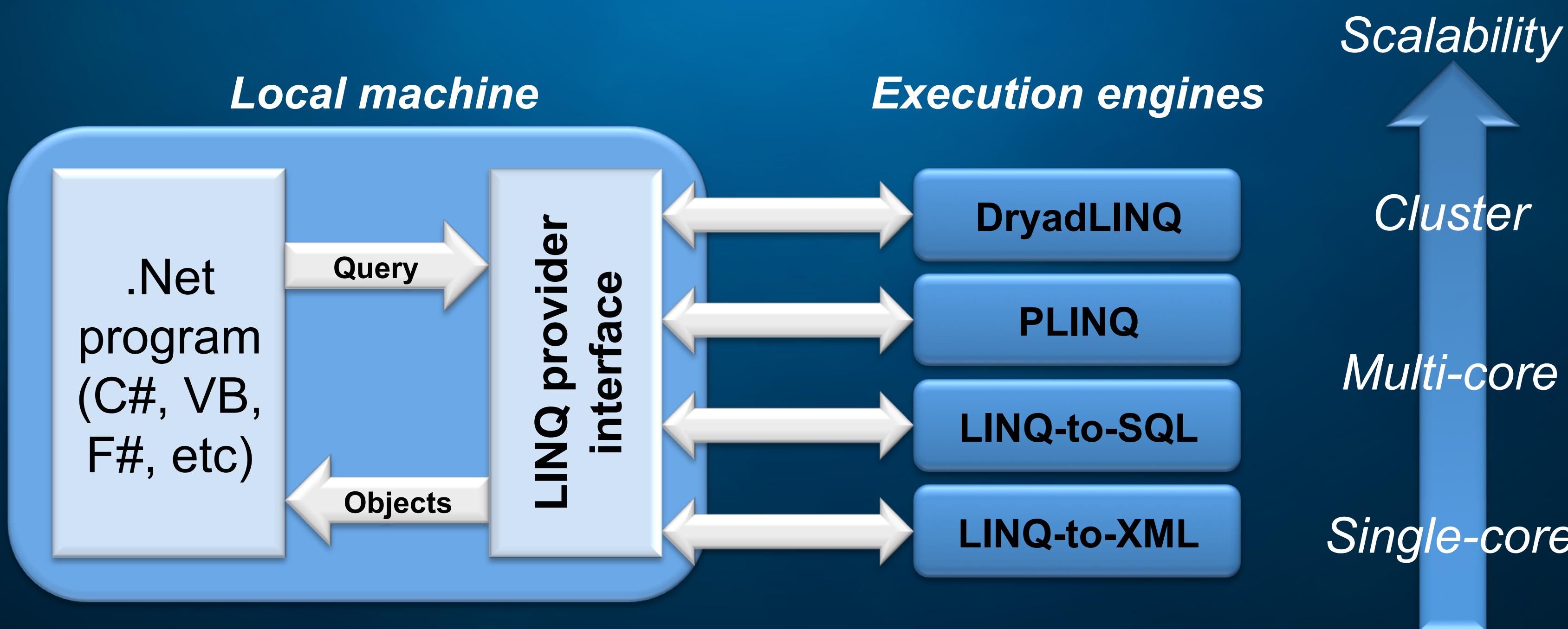
High-level language API (C#)

Dataflow graph as the computation model, distributed execution, fault-tolerance, scheduling

Remote process execution, naming, storage

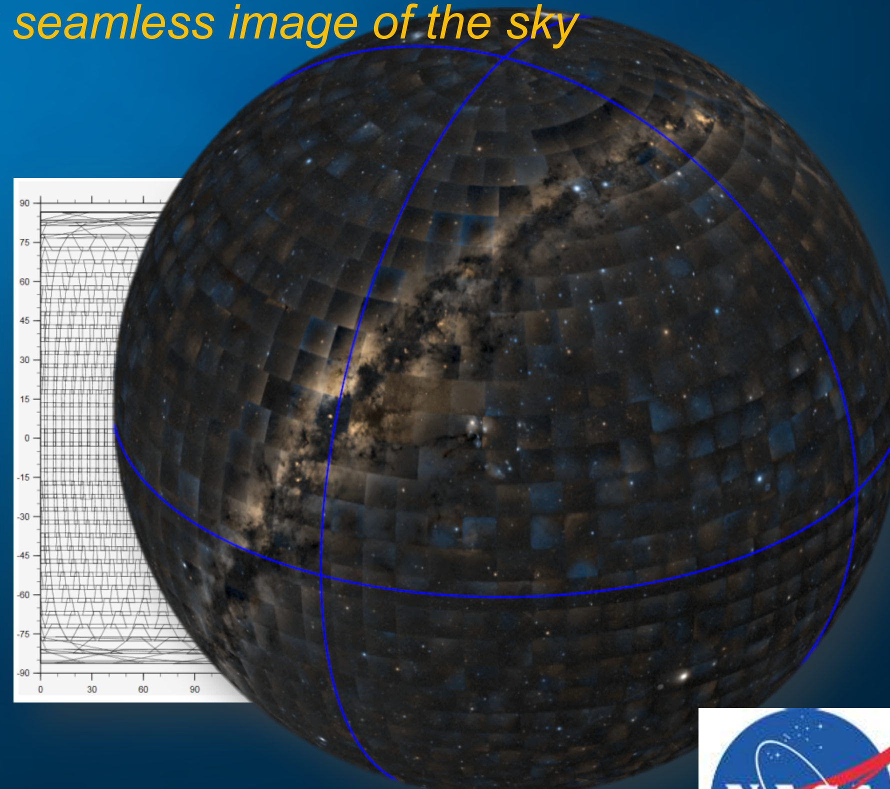
DryadLINQ leverages LINQ's extensibility

LINQ - Microsoft's Language INtegrated Query
Released with .NET Framework 3.5, extremely extensible



WorldWide Telescope - TeraPixel

Challenge: *Create the largest, clearest seamless image of the sky*



Digitized Sky Survey (DSS)

- Produced photographic plates of overlapping regions of the sky
- 1,791 pairs of red-light and blue-light images acquired from two telescopes
- Scanned over 15 year period into3,120,100 files, 417 GB

Create Spherical Image

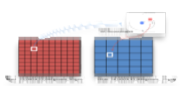
1. Create color plates from DSS data
2. Stitch and smooth images
3. Create sky image pyramid for WWT

WorldWide Telescope - TeraPixel

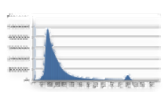
Computational and Data Intensive

1. Create RGB color plates from DSS data

Vignetting Correction (Red, Blue)



Astrometric Alignment



Statistical Analysis (Saturation & noise floor)

Colored Plate Creation

2. Stitch and smooth images

Project Sphere Image onto Plane

Distributed gradient-domain processing

3. Create sky image pyramid for WWT

Tiled Multi-resolution

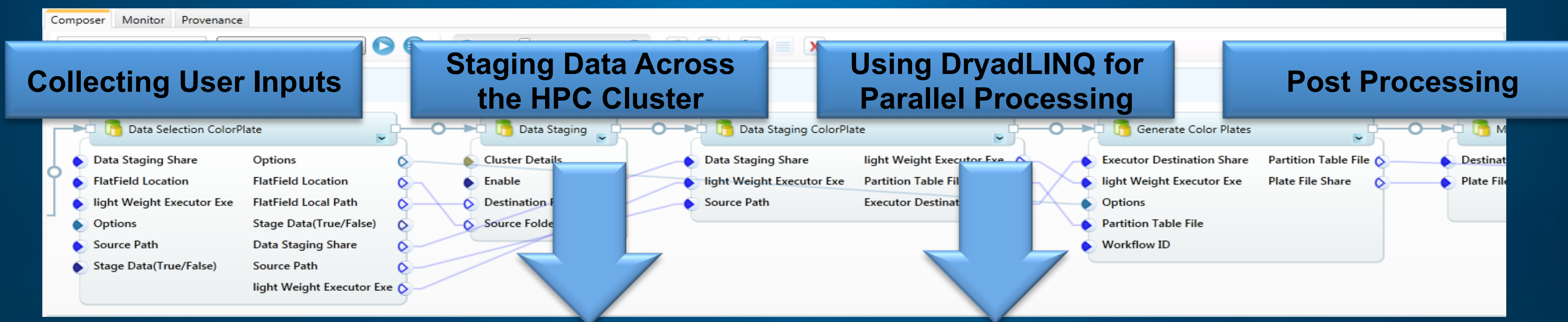
Large-scale data aggregation easily performed with integrated set of technologies

- **DryadLINQ** => concise code
- **.NET Parallel Extension** => faster decompression of DSS data
- **DryadLINQ + Windows HPC** => Efficient and robust execution

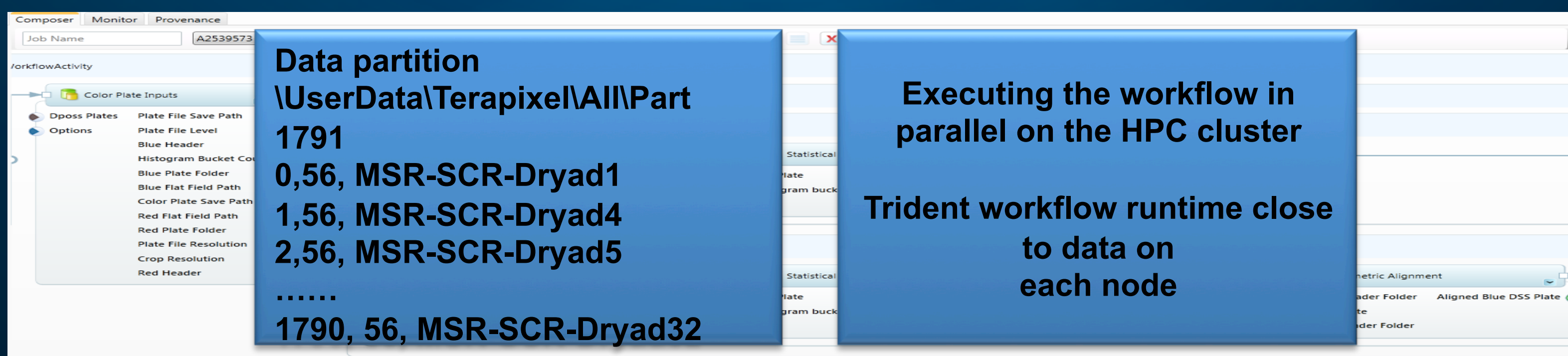
Managed and Coordinated by **Project Trident**: A Scientific Workflow Workbench

Workflows for Processing Data in Parallel

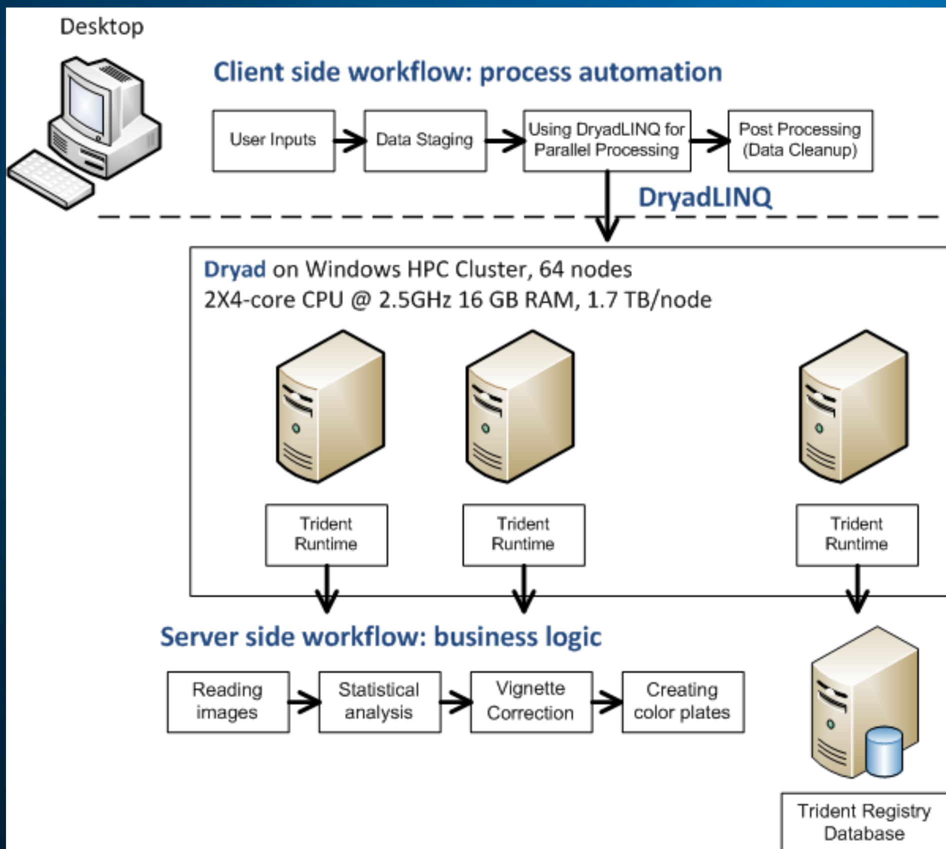
Local Desktop Machine (process automation and reruns)



HPC Cluster (processing data in parallel – e.g. generating color images)



Deployment Architecture

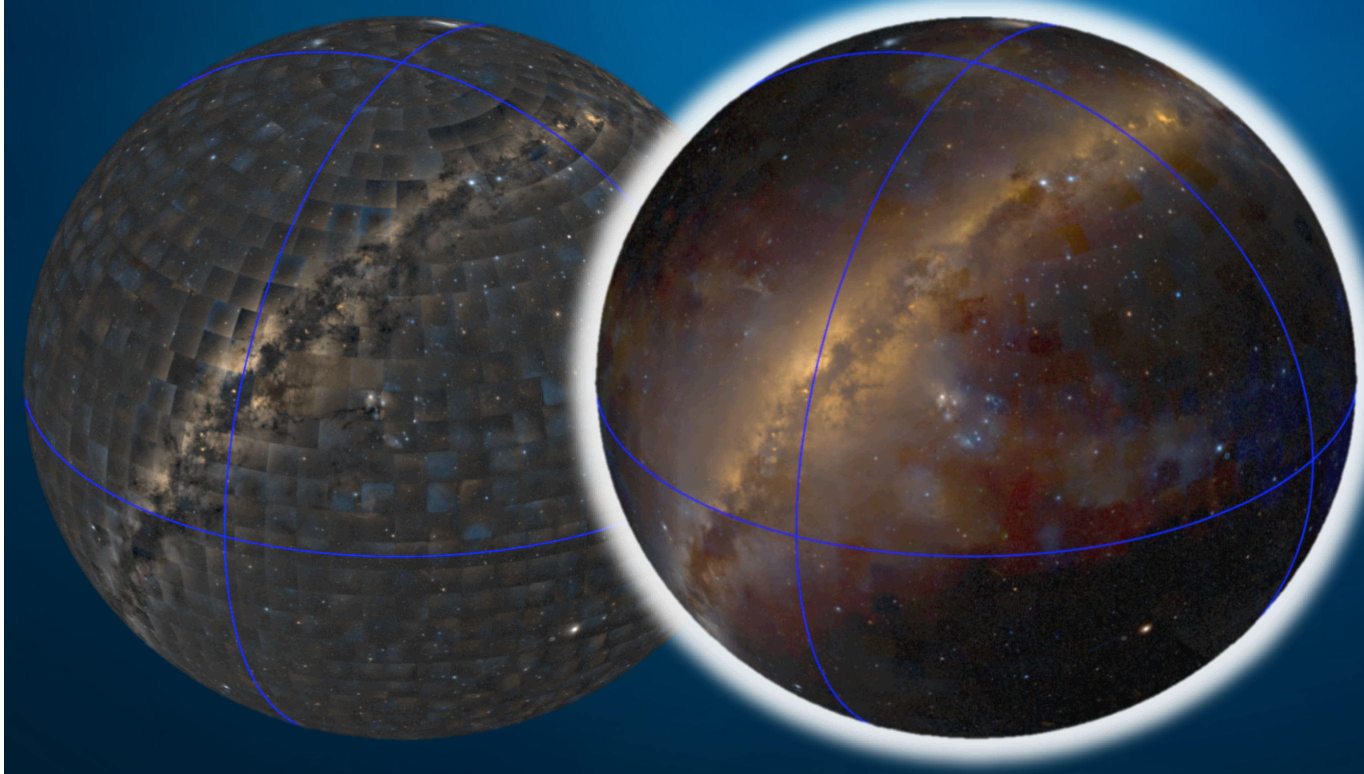


Generating RGB color plates

- Generation of 1,791 plates with 64 compute nodes
- Processing time: 5 hrs.
- Input: 417 GB (compressed, 4 TB uncompressed)
- Output: 790 GB (approx. 450 MB/plate)

WorldWide Telescope - TeraPixel

Result: *Largest, clearest, and smoothest sky image in the world*



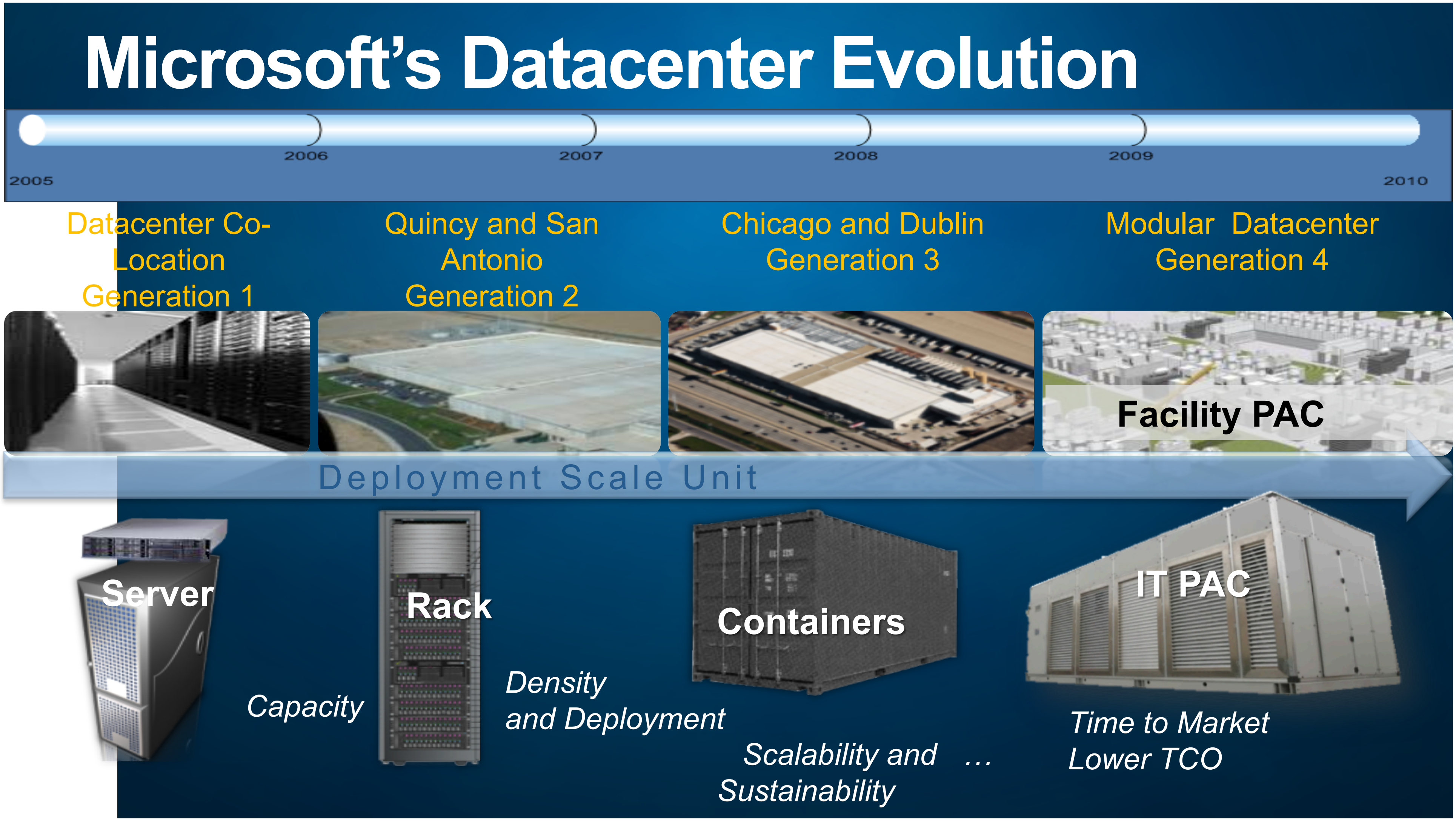
Special Thanks to

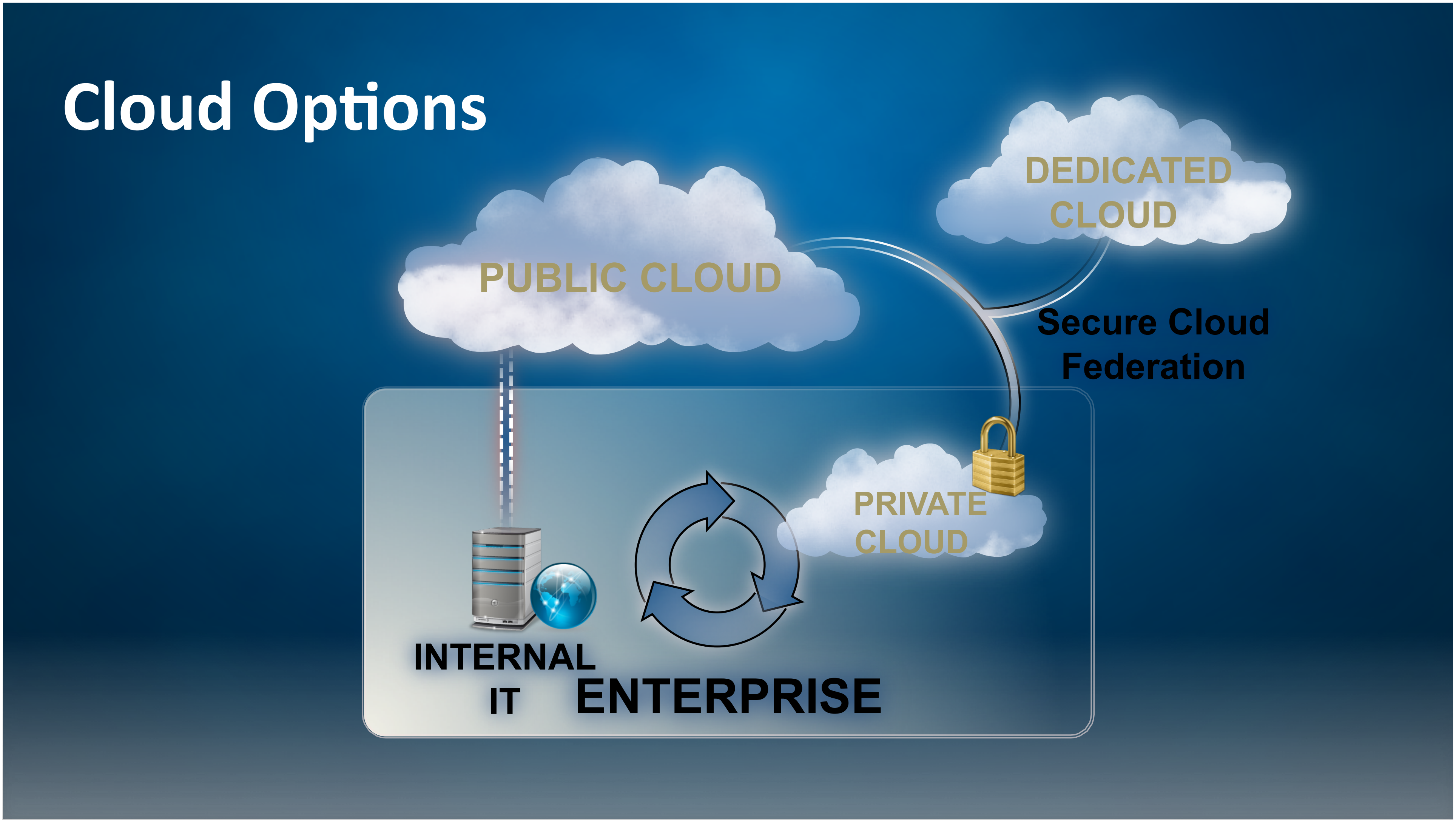
- Brian McLean (Space Telescope Science Institute),
- Misha Kazhdan (Johns Hopkins University), Hugues Hoppe (MSR), and Dinoj Surendran (MSR)
- Dean Guo (MSR), Christophe Poulain (MSR)
- Aditi Team

Cloud Computing: One Definition

For the US National Institute of Standards and Technology (NIST), Cloud Computing means:

- On-demand service
- Broad network access
- Resource pooling
- Flexible resource allocation
- Measured service


Microsoft's Datacenter Evolution
2005
2006
2007
2008
2009
2010
Datacenter Co-Location
Generation 1
Quincy and San Antonio
Generation 2
Chicago and Dublin
Generation 3
Modular Datacenter
Generation 4
Facility PAC
Deployment Scale Unit
Server
Capacity
Rack
Density
and Deployment
Containers
Scalability and
Sustainability
…
IT PAC
Time to Market
Lower TCO


Cloud Options
DEDICATED CLOUD
PUBLIC CLOUD
Secure Cloud Federation
PRIVATE CLOUD
INTERNAL IT
ENTERPRISE

Cloud Services

Infrastructure as a Service (IaaS)

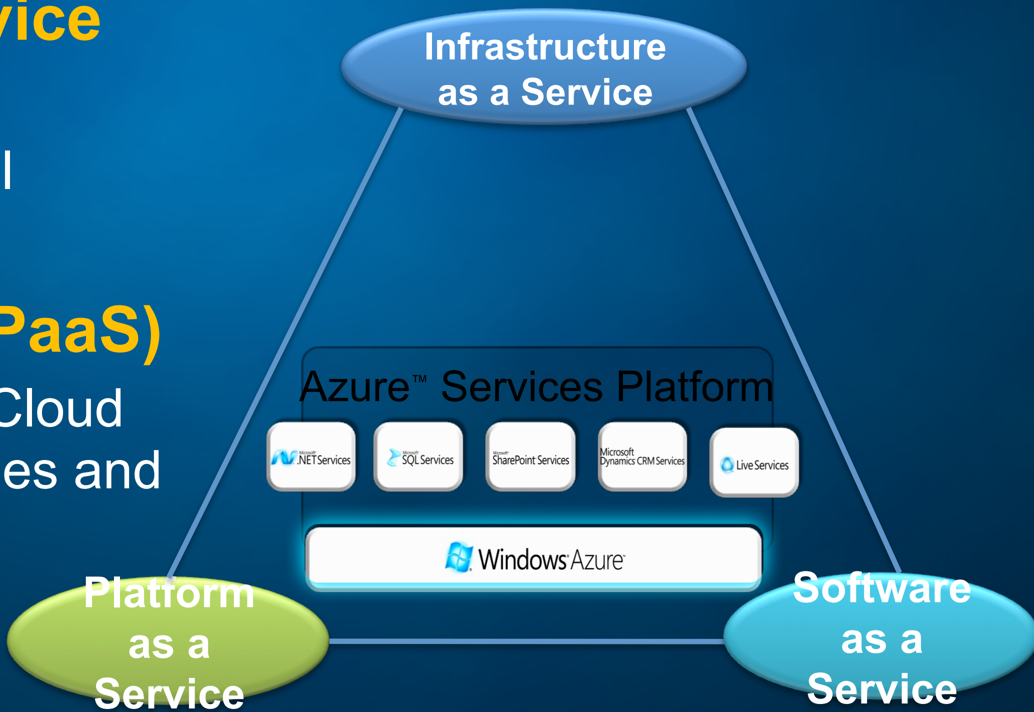
– Provide a way to host virtual machines on demand

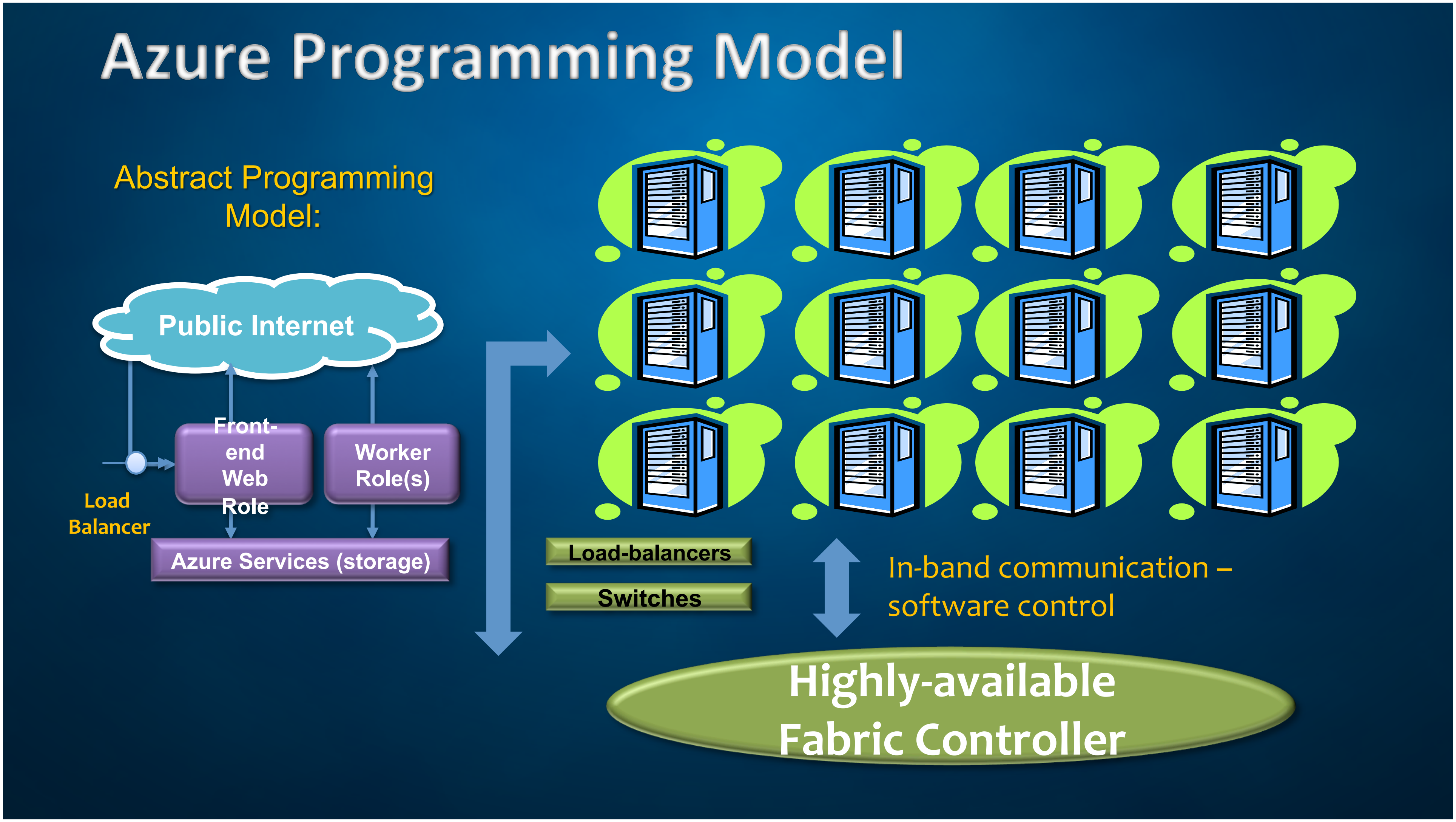
Platform as a Service (PaaS)

– You write an Application to Cloud APIs and the platform manages and scales it for you.

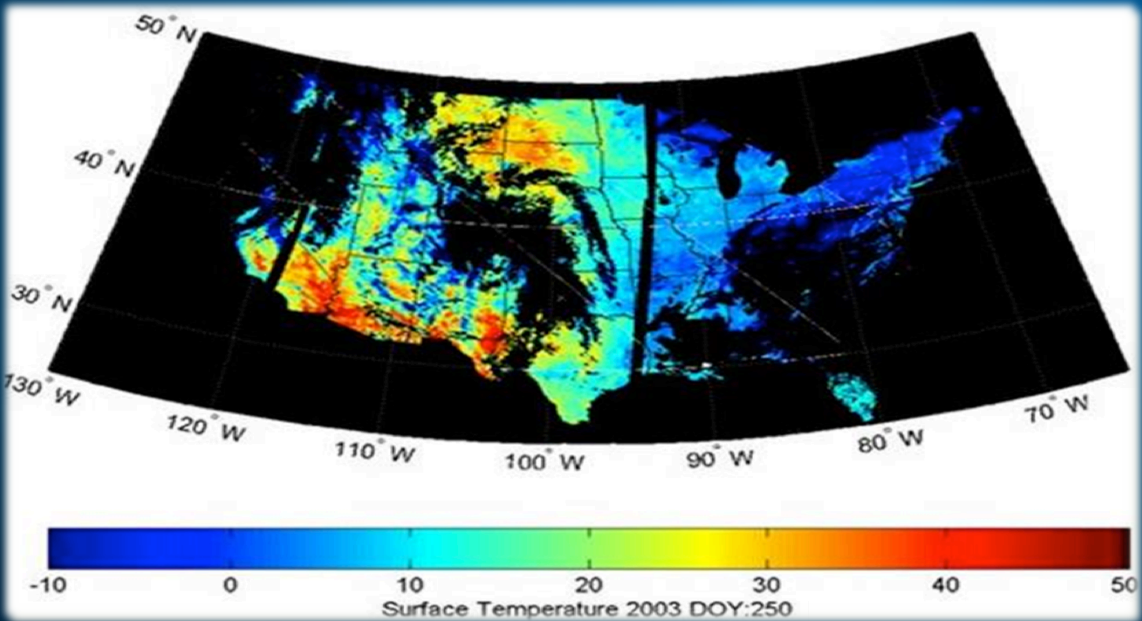
Software as a Service (SaaS)

– Delivery of software to the desktop from the Cloud




Azure Programming Model
Abstract Programming Model:
Public Internet
Front-end Web Role
Worker Role(s)
Load Balancer
Azure Services (storage)
Load-balancers
Switches
In-band communication – software control
Highly-available Fabric Controller

MODIS Azure: *Computing Evapotranspiration (ET) in the Cloud*

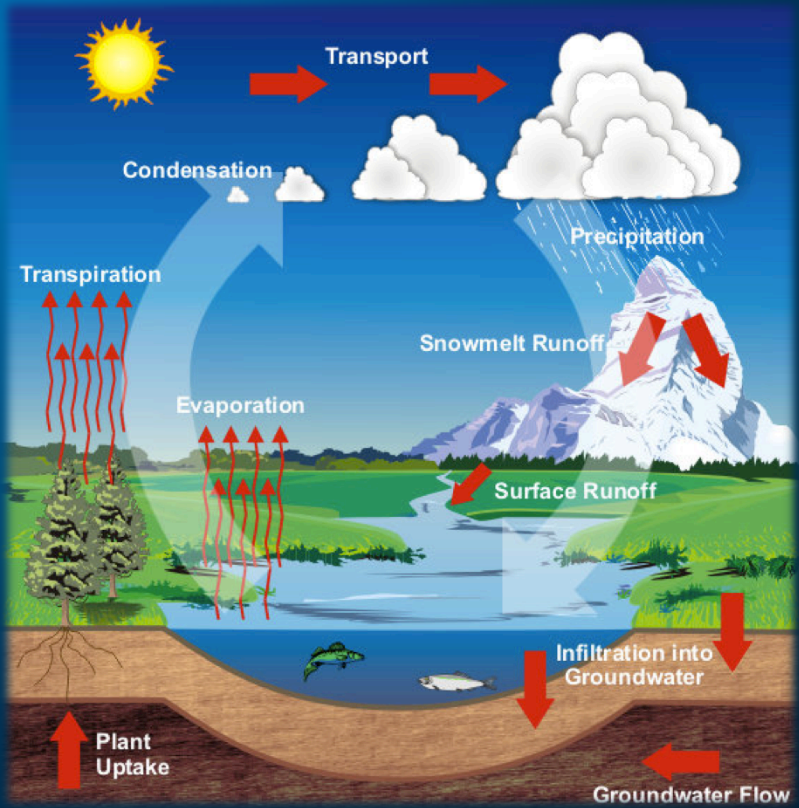


A pipeline for download, processing, and reduction of diverse NASA MODIS satellite imagery.

Contributors: Catharine van Ingen (MSR), Youngryel Ryu (UC Berkeley), Jie Li (Univ. of Virginia)

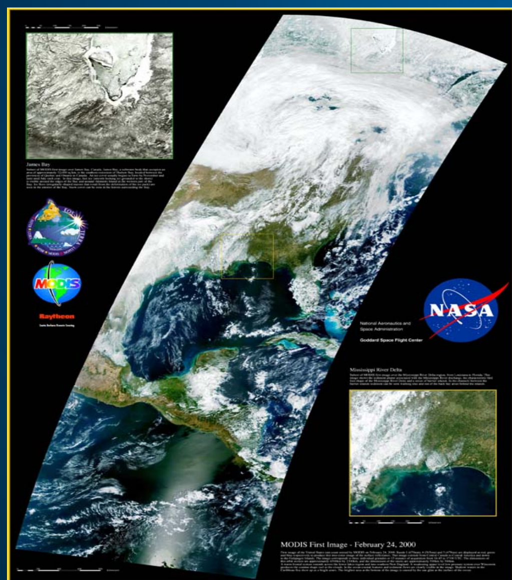
MODIS Azure

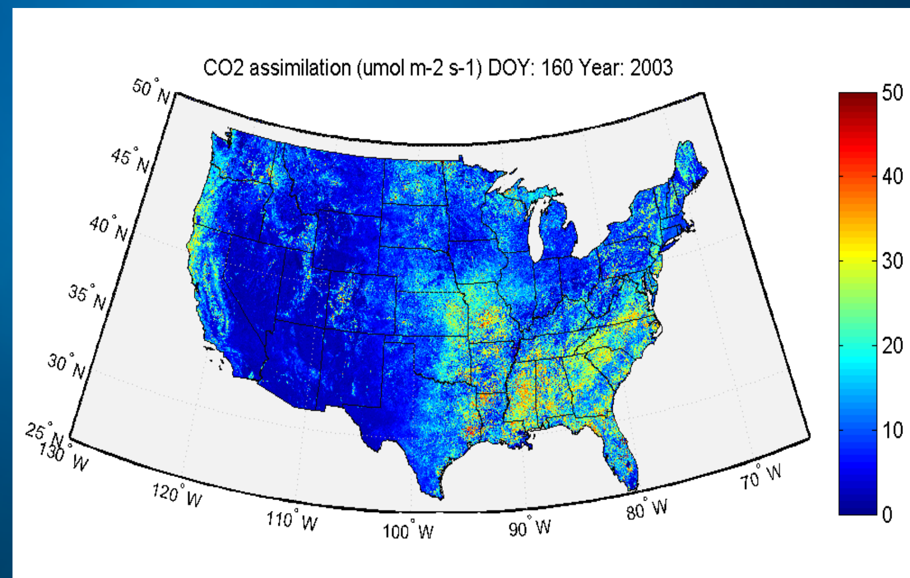
- Evapotranspiration (ET) is the release of water to the atmosphere by evaporation from open water bodies and transpiration, or evaporation through plant membranes, by plants.
- Climate change isn't just about a change in temperature, it's also about a change in the water balance and hence water supply which is critical to human activity.



Source: Youngryel Ryu's PhD project

MODIS Azure





Aqua, Terra: Time series raster data, 36 spectral bands, 1-2d

- Over some period of time at some time frequency at some spatial granularity over some spatial area
- Conversion from L0 data to L2 and beyond as well as reprojection

MODIS Azure: Four Stage Image Processing Pipeline

Data collection stage

- Downloads requested input tiles from NASA ftp sites
- Includes geospatial lookup for non-sinusoidal tiles that will contribute to a reprojected sinusoidal tile

Reprojection stage

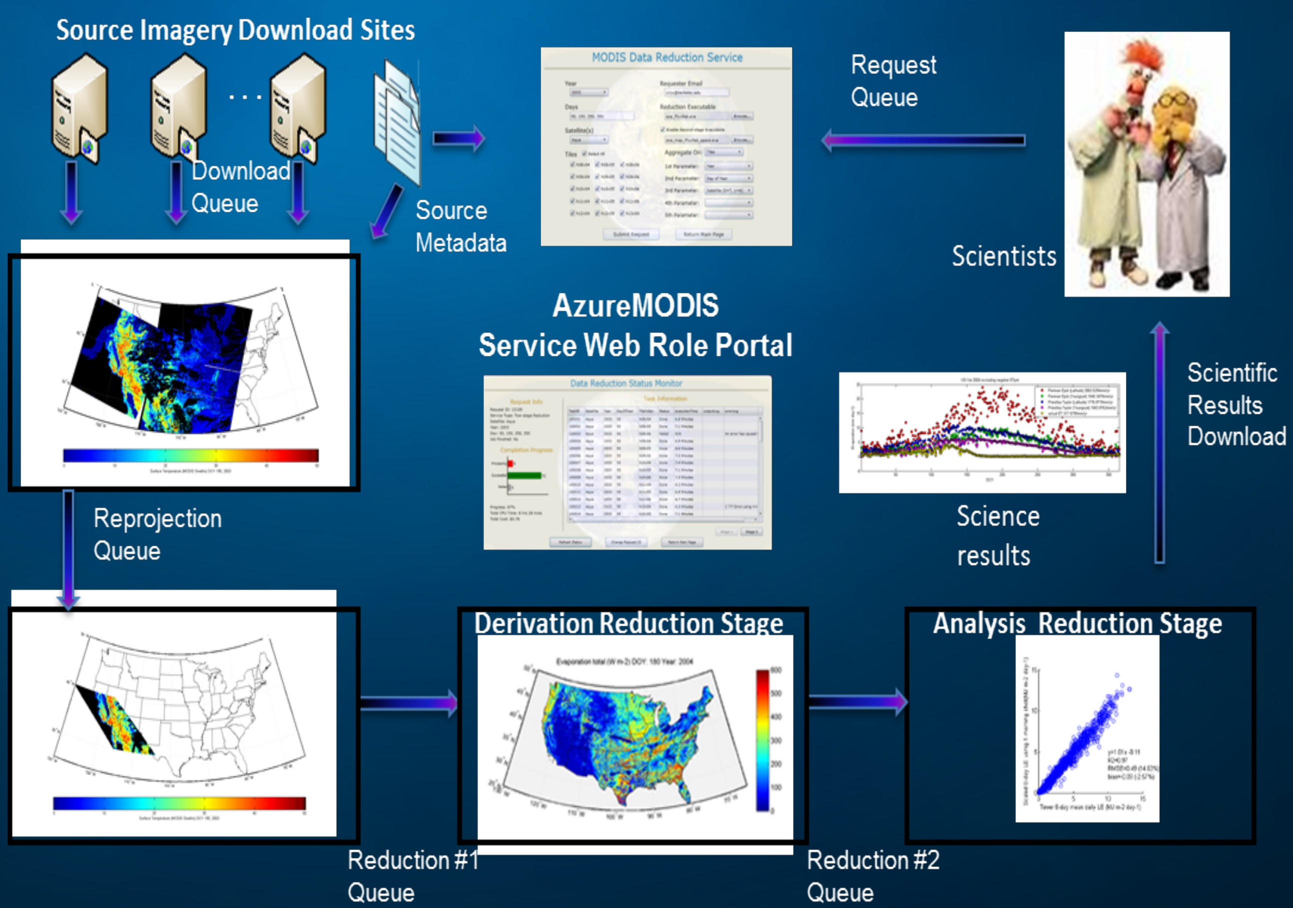
- Converts source tile(s) to intermediate result sinusoidal tiles
- Simple nearest neighbor or spline algorithms

Derivation reduction stage

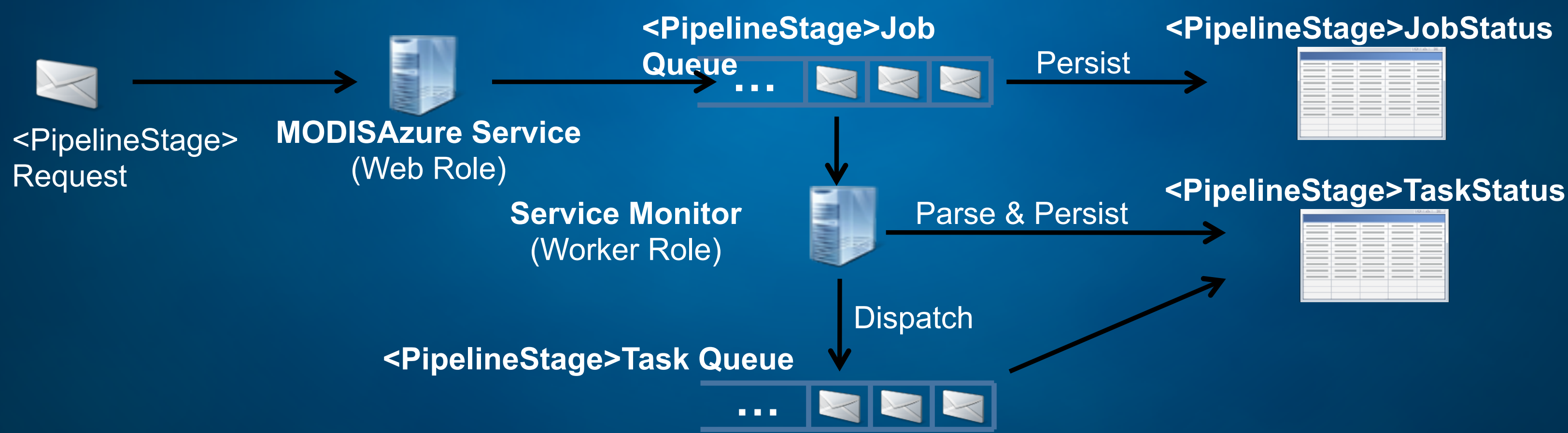
- First stage visible to scientist
- Computes ET in our initial use

Analysis reduction stage

- Optional second visible stage
- Enables production of science analysis artifacts such as maps



MODIS Azure: Architectural Overview



<u>ModisAzure Service</u> is the Web Role front door

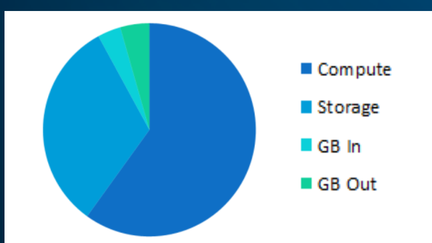
- Receives all user requests
- Queues request to appropriate Download, Reprojection, or Reduction Job Queue

<u>Service Monitor</u> is a dedicated Worker Role

- Parses all job requests into tasks – recoverable units of work
- Execution status of all jobs and tasks persisted in Tables

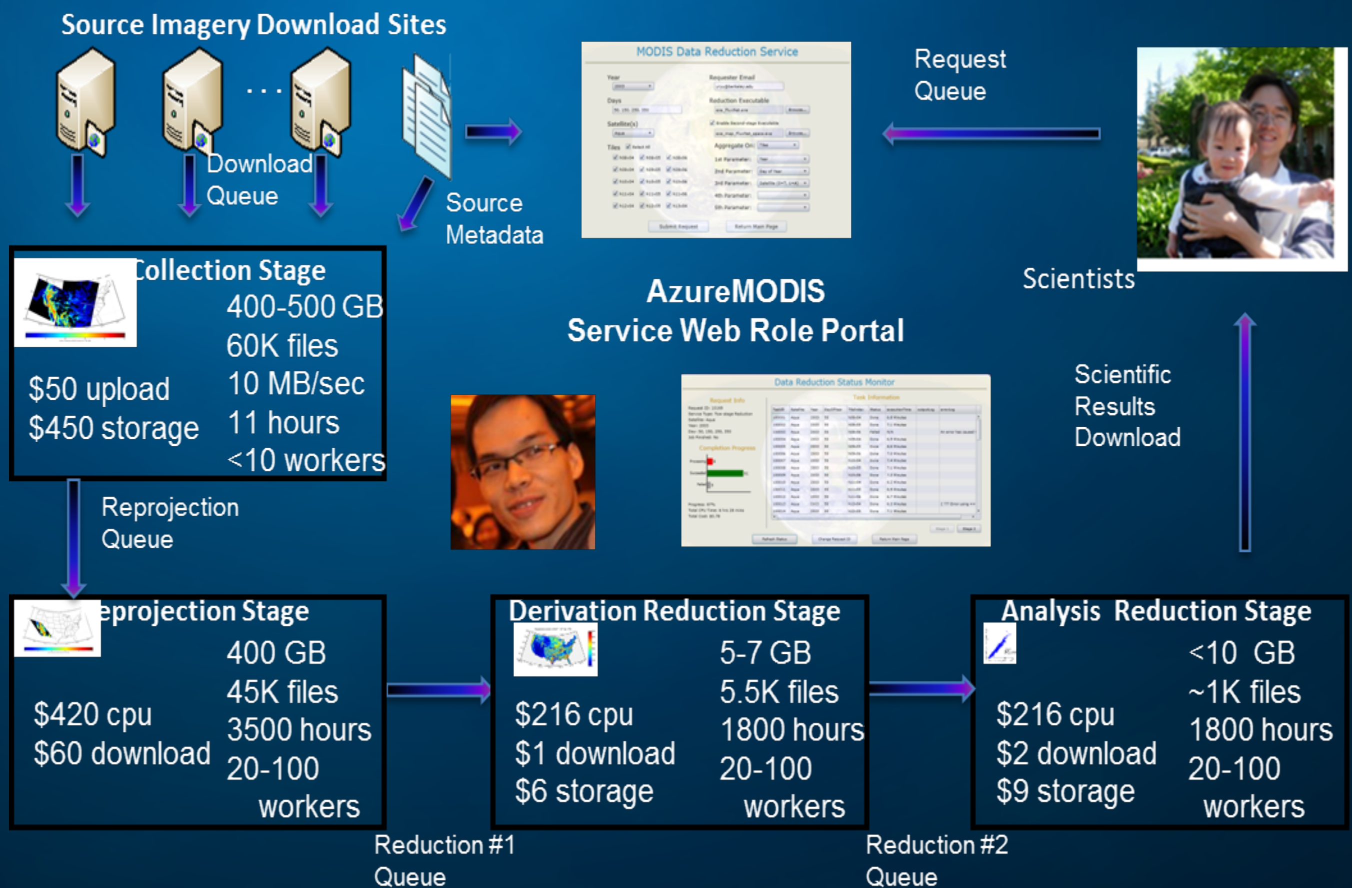
Computing a one US Year ET Computation

- Computational costs driven by data scale and need to run reduction multiple times
- Storage costs driven by data scale and 12 month project duration



Total: $1420

- Small with respect to the people costs even at graduate student rates !



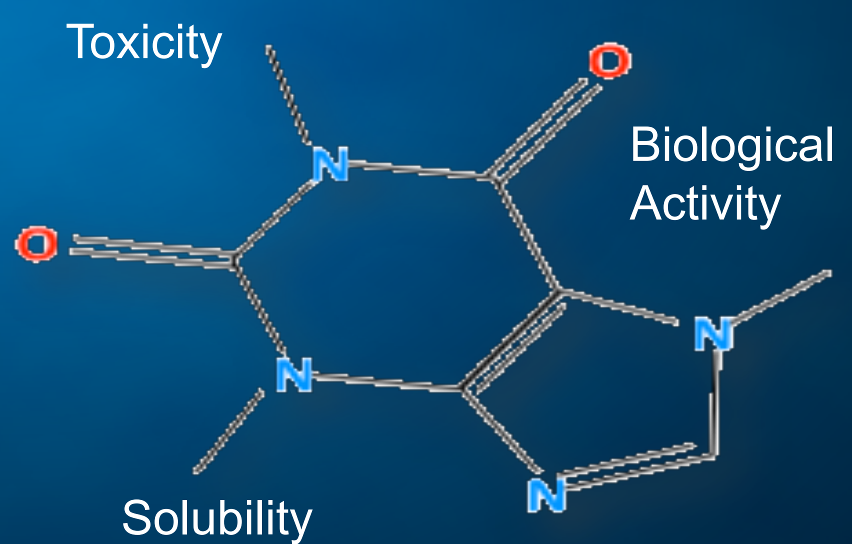
Project Junior

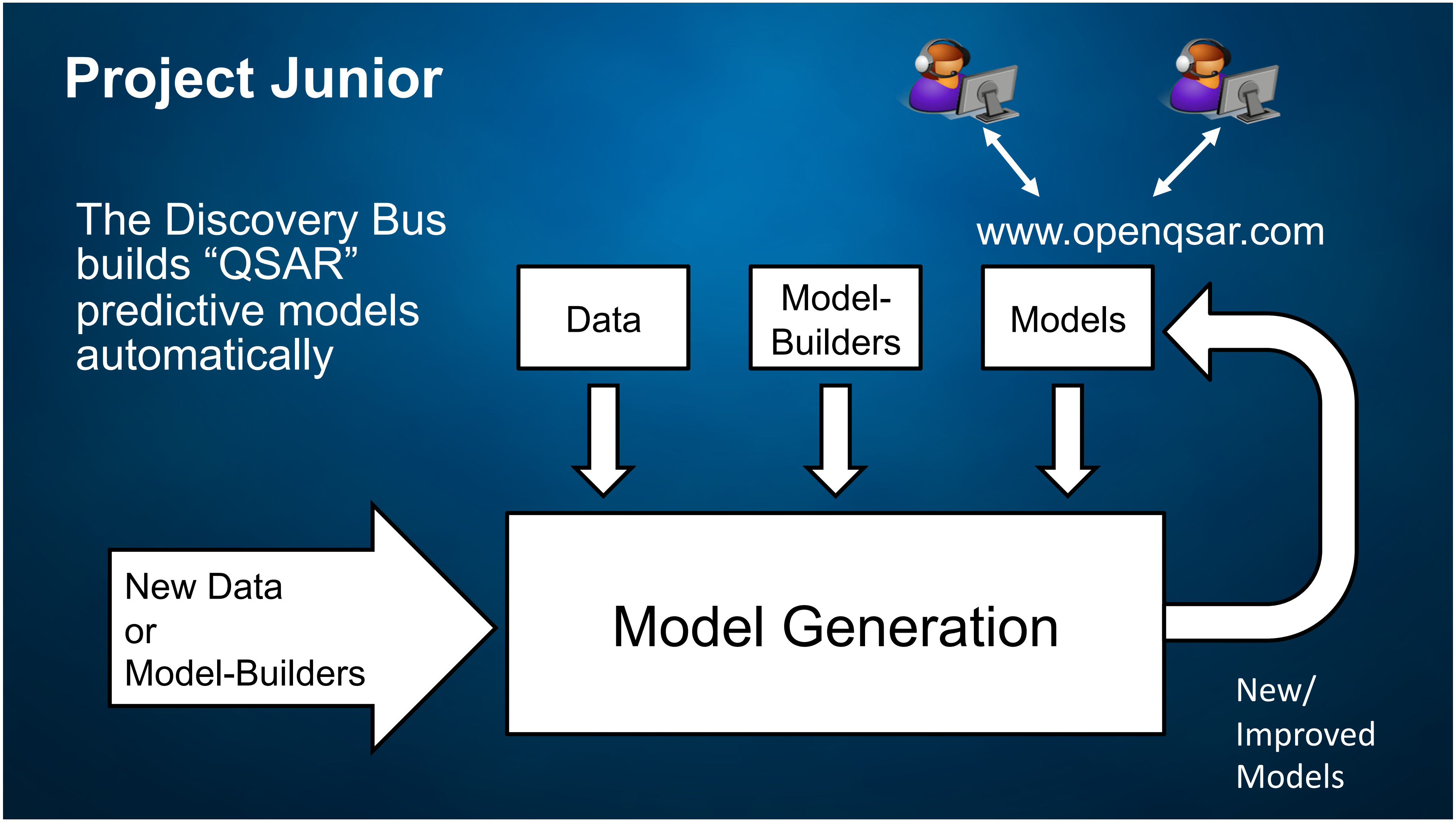
Chemists need to know:

What are the properties of a molecule?

What molecule would have aqueous solubility of 0.1 µg/mL*?*

How can this be done without expensive,time-consuming experimentation?




Project Junior
The Discovery Bus builds “QSAR” predictive models automatically
www.openqsar.com
Data
Model-Builders
Models
New Data
or
Model-Builders
Model Generation
New/
Improved
Models

Project Junior

Increasing amounts of data for model building...

CHEMBL :	data on	622,824 compounds,
	collected from	33,956 publications
WOMBAT :	data on	251,560 structures,
	for over	1,966 targets
WOMBAT-PK:	data on	1,230 compounds,
	for over	13,000 clinical measurements

All contain structure information & numerical activity data

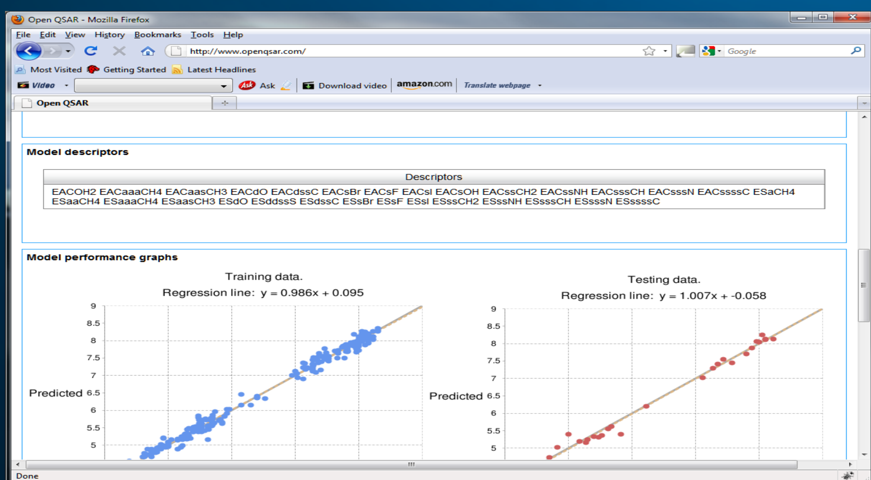
- ✓ More models
- ✓ Better models

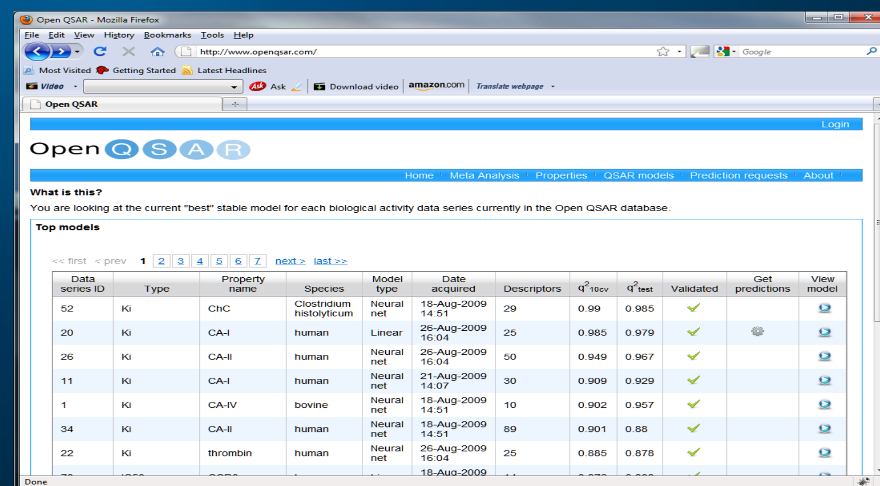
- × Computationally expensive:
  5 years for new datasets on existing Discovery Bus server

Project Junior

Used Windows Azure to generate models in parallel

- 100 workers for 3 weeks (not 5 years!)
- 750K new models available on www.openqsar.com
  (50x more than previously available)





Project Junior - Overview

Chemical Property Prediction on Azure

- QSAR predicts molecular properties
  - e.g. toxicity, solubility
  - reduces time and cost c.f. experimentation
- Vast amounts of new data are now available to build predictive models
  - est. 5 years to process on existing single-server solution
- 100 Azure workers reduced 5 years to 3 weeks
  - used competitive workflow algorithm
  - 10,000 data sets → 750,000 models

    (50x more than before)

VENUS-C

- ***V**irtual multidisciplinary **E**nviro**N**ments **US**ing **C**loud infrastructures*
- EU will fund the project with 4.5 M€ over the first 2 years (1/6/2010-30/5/2012)
- Microsoft will invest up to 3 M€ in Azure resources and research manpower in Redmond, Cambridge/UK, EMIC in Germany and MIC GR in Greece
- This is part of the XCG Cloud Initiative for Research in Europe which includes also direct collaboration with some of the main national funding agencies

Supports multiple basic research disciplines

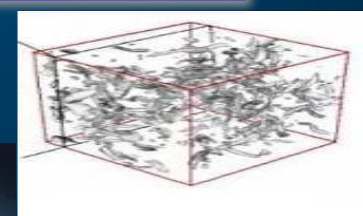
- **Biomedicine:** Integrating widely used tools for Bioinformatics (UPV), System Biology (CosBI) and Drug Discovery (NCL) into the VENUS-C infrastructure
- **Civil Protection and Emergency:** Early fire risk detection (AEG), through an application that will run models on the VENUS-C infrastructure, based on multiple data sources
- **Civil Engineering:** Support complex computing tasks on Building Information Management for green constructions (provided by COLB) and dynamic building structure analysis (provided by UPV)
- **D4Science:** Integrating computing through VENUS-C on data repositories (CNR). In particular focus will be on Marine Biodiversity through Aquamaps

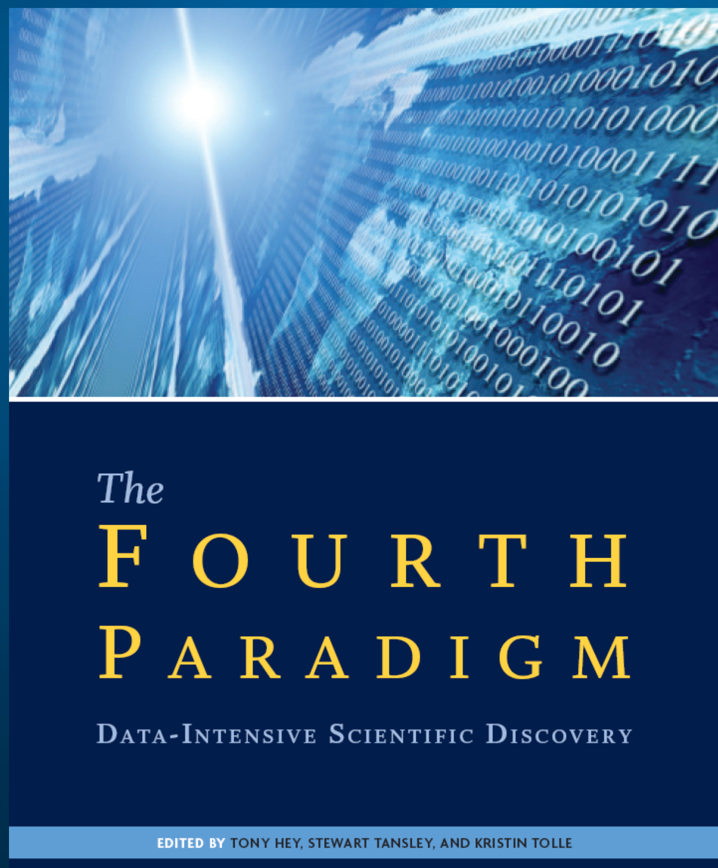
Emergence of a Fourth Research Paradigm

1. Thousand years ago – **Experimental Science**
   – Description of natural phenomena
2. Last few hundred years – **Theoretical Science**
   – Newton's Laws, Maxwell's Equations…
3. Last few decades – **Computational Science**
   – Simulation of complex phenomena
4. Today – **Data-Intensive Science**
   – Scientists overwhelmed with data sets from many different sources
      • Data captured by instruments
      • Data generated by simulations
      • Data generated by sensor networks

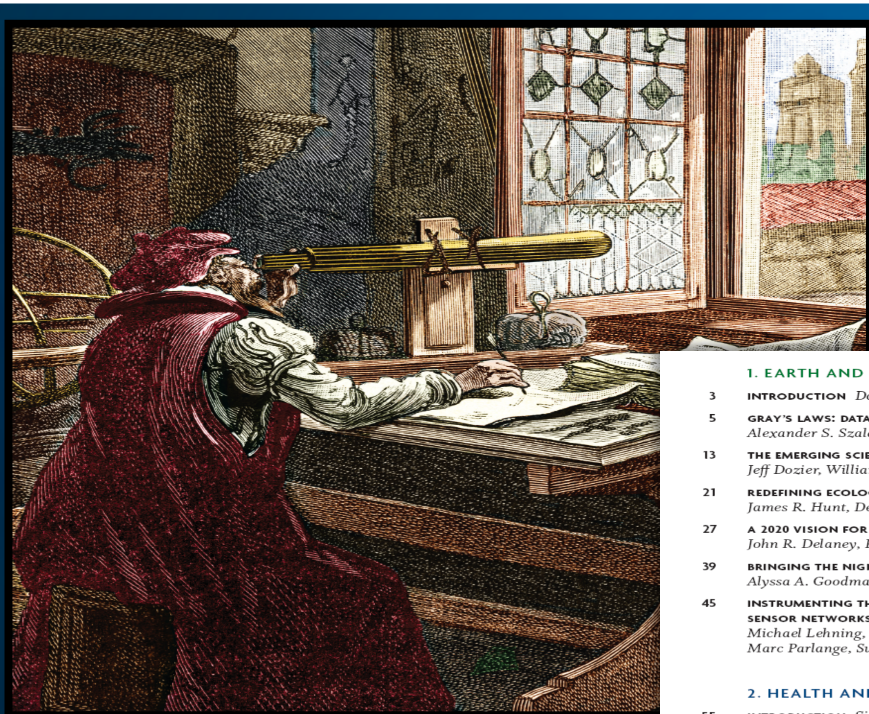
➢ eScience is the set of tools and technologies to support data federation and collaboration
   • For analysis and data mining
   • For data visualization and exploration
   • For scholarly communication and dissemination

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$




The
FOURTH
PARADIGM
DATA-INTENSIVE SCIENTIFIC DISCOVERY
EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE

An edited collection of 26 short technical essays, divided into 4 sections



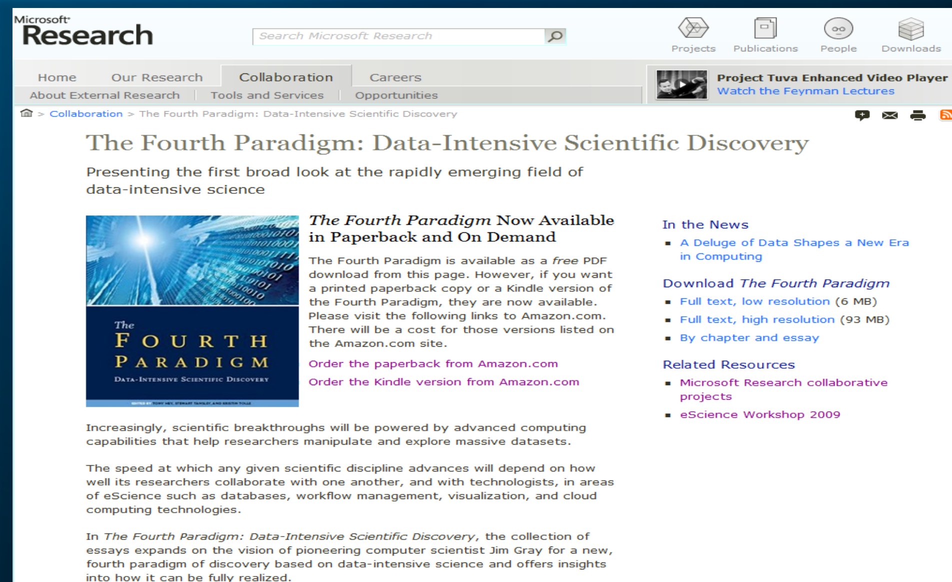
Free PDF Download
Amazon Kindle version; Paperback print on demand

http://research.microsoft.com/fourthparadigm/

- “The impact of Jim Gray’s thinking is continuing to get people to think in a new way about how data and software are redefining what it means to do science."
- — **Bill Gates,** Chairman, Microsoft Corporation

- “One of the greatest challenges for 21st-century science is how we respond to this new era of data-intensive science. This is recognized as a new paradigm beyond experimental and theoretical research and computer simulations of natural phenomena—one that requires new tools, techniques, and ways of working.”
- — **Douglas Kell**, University of Manchester

- “The contributing authors in this volume have done an extraordinary job of helping to refine an understanding of this new paradigm from a variety of disciplinary perspectives.”
- — **Gordon Bell**, Microsoft Research



Future Cyberinfrastructure for Research

Mixture of
Client + Cloud resources

National Science Foundation
WHERE DISCOVERIES BEGIN

Office of Cyberinfrastructure (OCI)

Data Task Force - Co-Chairs:

Dan Atkins, University of Michigan
Tony Hey, Microsoft Research

Open Workshop on Data Management and Data Visualization Needs and Priorities for 21st Century CyberInfrastructure
Berkeley, CA
Oct 10, 2010

For more information email: hedstrom@umich.edu
jimpi@microsoft.com