



# A Real Use Case Demonstrating Interoperability

Paul Watson

School of Computing Science

Newcastle University, UK

[Paul.Watson@ncl.ac.uk](mailto:Paul.Watson@ncl.ac.uk)

The team: David Leahy, Jacek Cala, Hugo Hiden,  
Dominic Searson, Vladimir Sykora, Martyn Taylor, Simon Woodman

With thanks to Microsoft External Research  
for their financial support for Project Junior

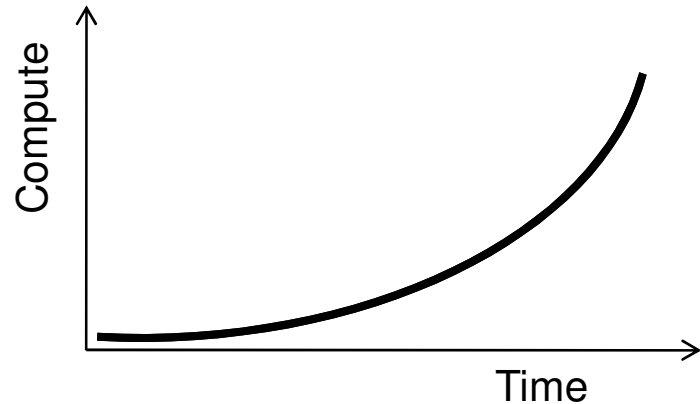
# Why Cloud Computing for Science?

- illusion of Infinite computing resources On Demand
- Pay for use of resources on a short-term basis as needed

(from “Above the Clouds: A Berkeley View of Cloud Computing”)

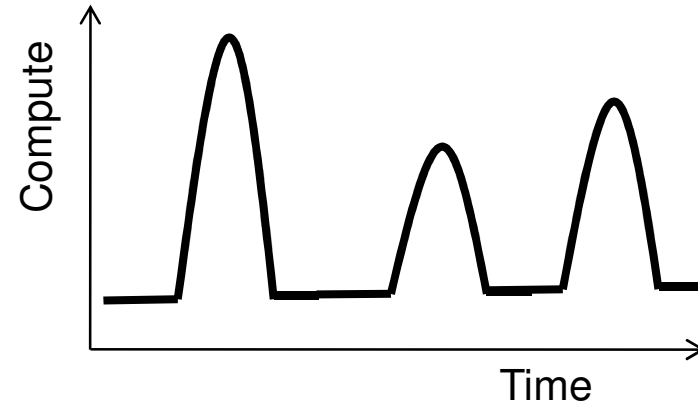
# Good Workload Patterns for Clouds

(with acknowledgements to Dianne O'Brien)



## Fast Growth

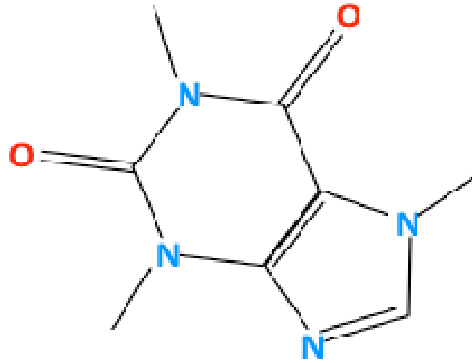
- datasets / applications with rapidly growing popularity



## Bursting

- new data processed
- new algorithm runs
- unpredictable event triggers computation
  - (e.g. earthquake)

# Cloud Science Example - Drug Discovery



*What is the Activity of this molecule?  
- toxicity/ solubility/ ...*

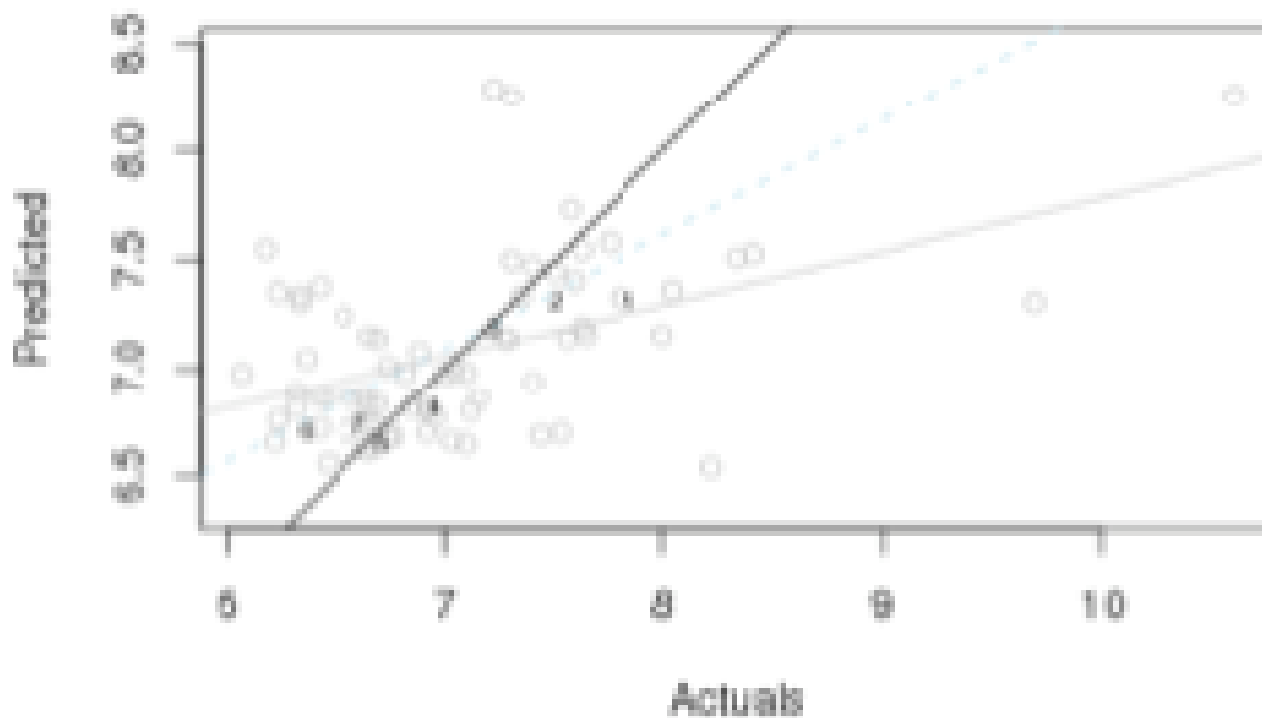
*What molecule would have an  
aqueous solubility of 0.1  $\mu\text{g}/\text{mL}$ ?*

How to answer: run experiments?  
or can we predict?

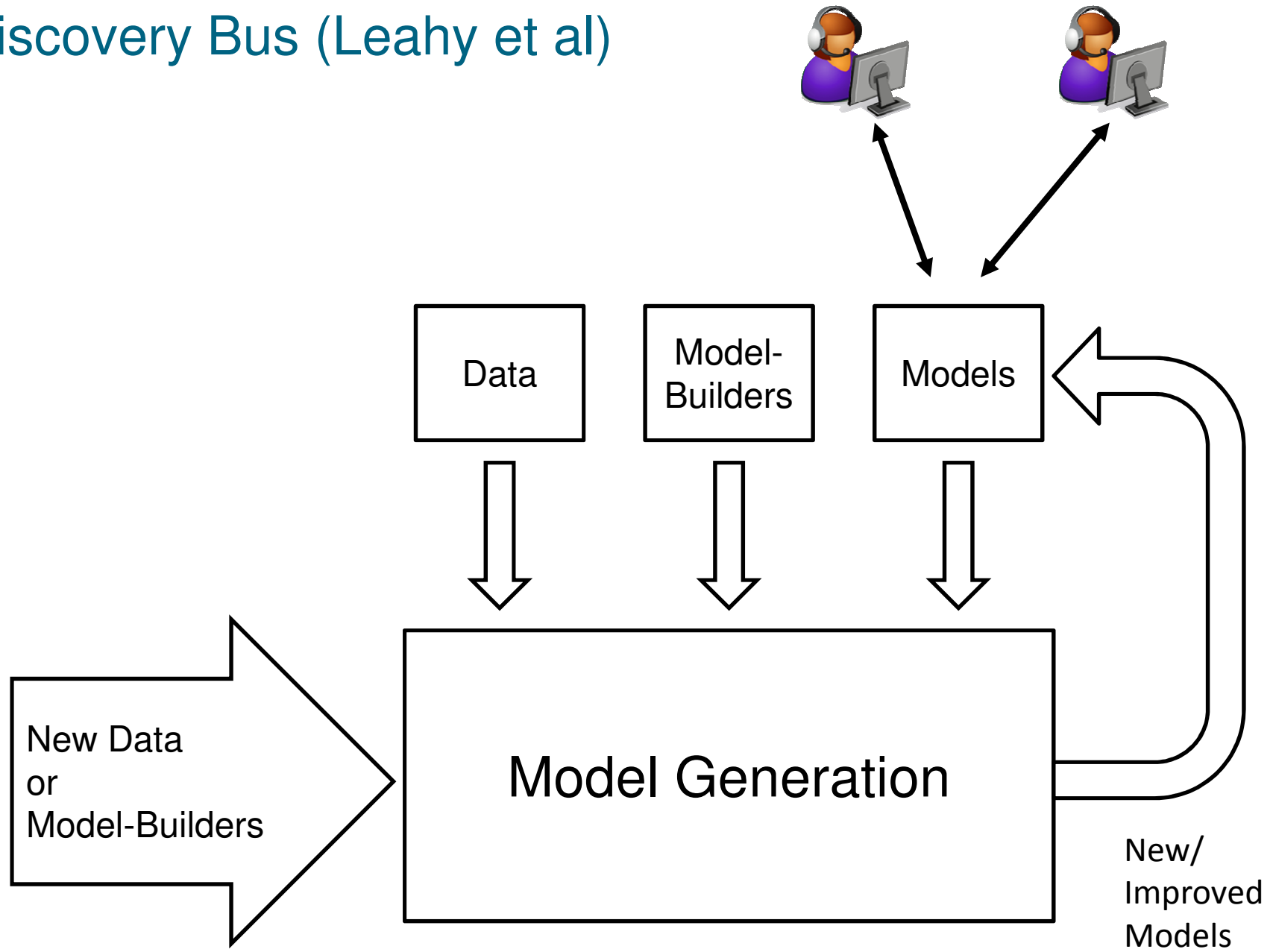
# Quantitative Structure-Activity Relationship (QSAR)

- Chemists use QSAR for prediction
  - based on analysis of existing data

**Predicted Versus Actuals**



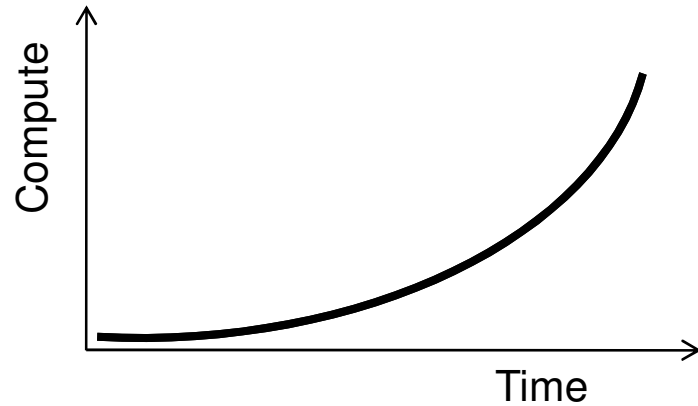
# Discovery Bus (Leahy et al)



# Good Workload Patterns for Clouds

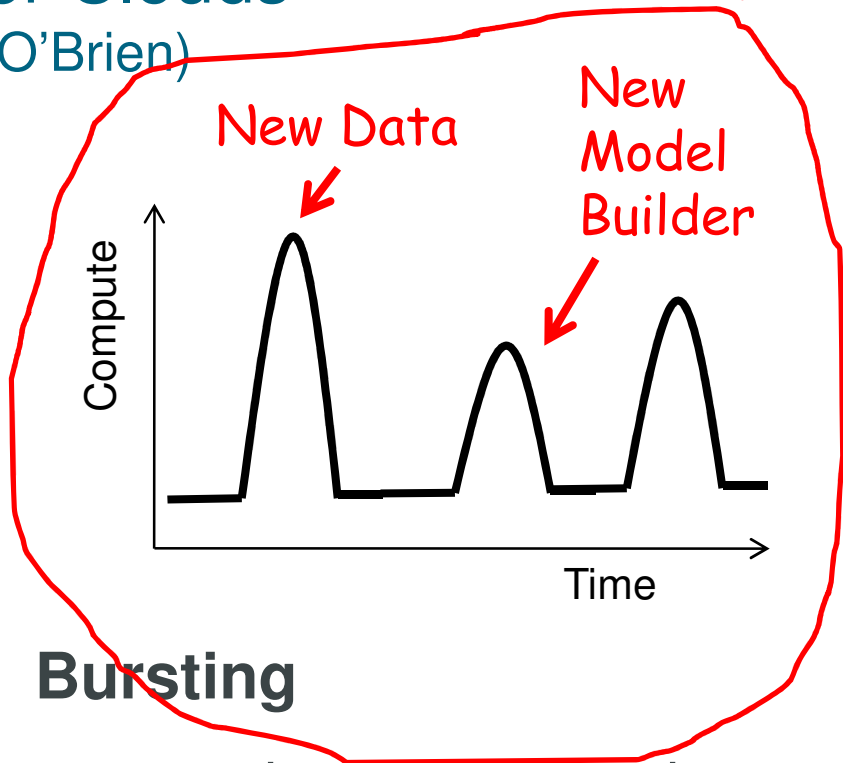
(with acknowledgements to Dianne O'Brien)

Discovery Bus



## Fast Growth

- datasets / applications with rapidly growing popularity

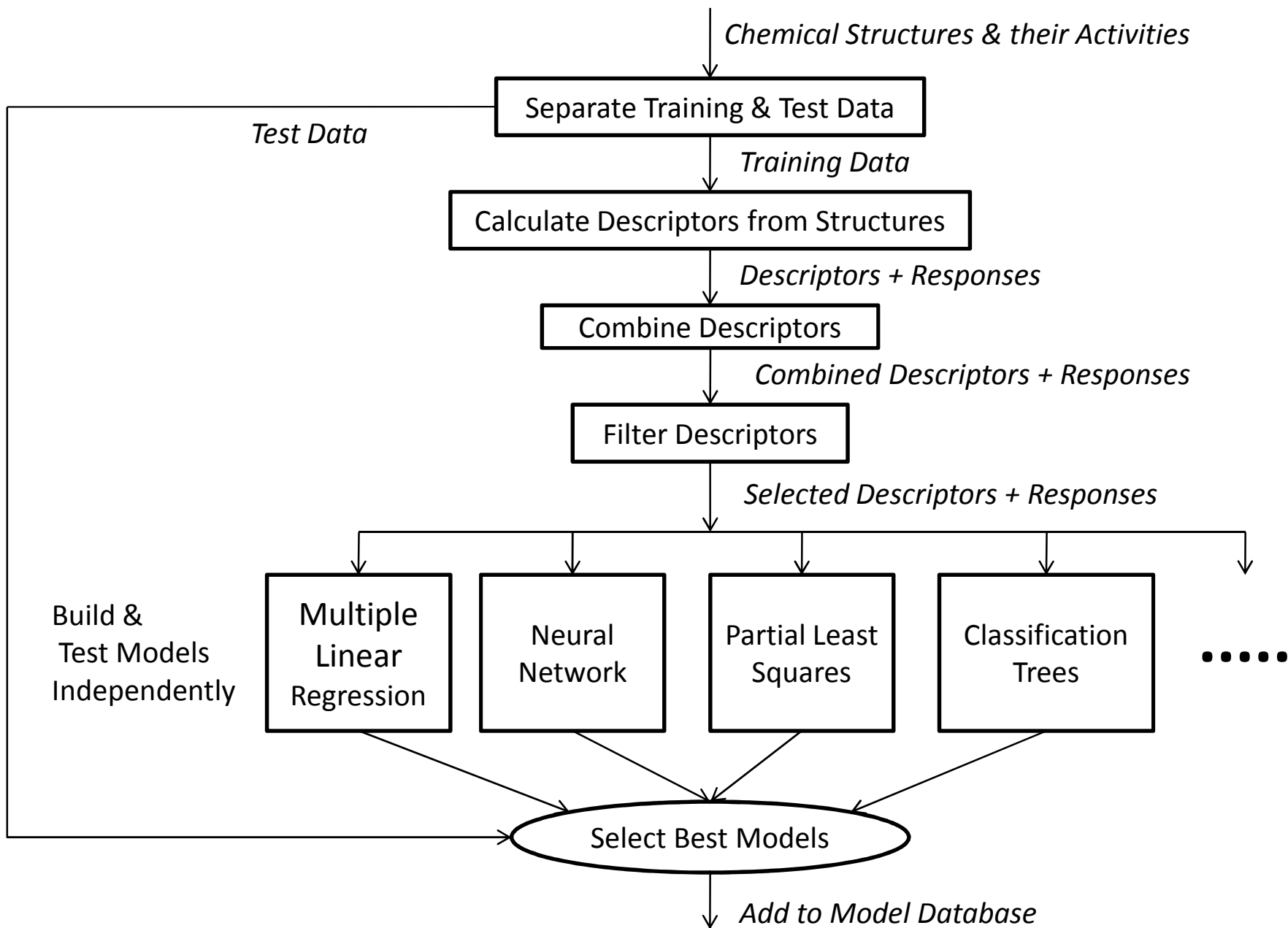


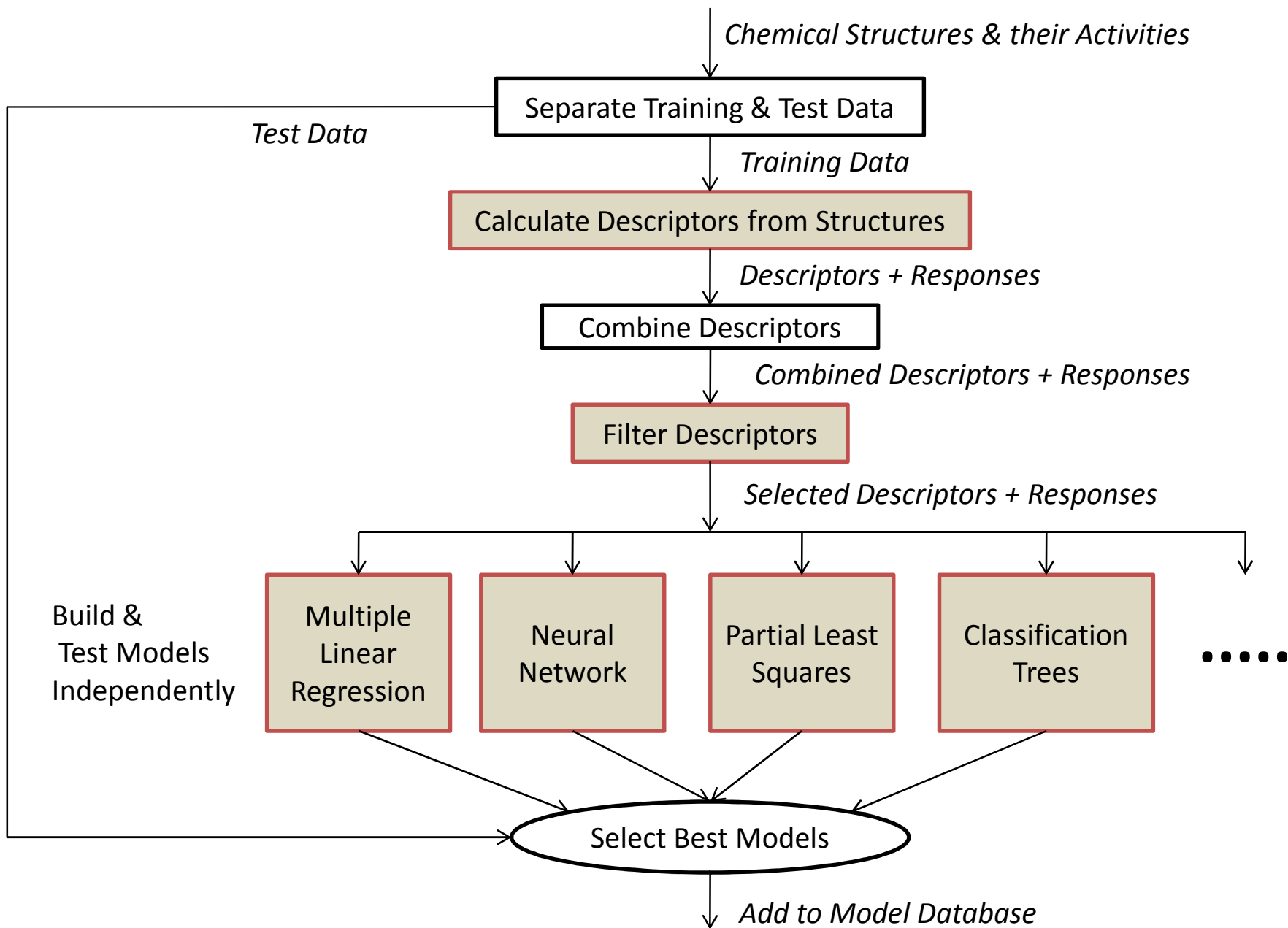
## Bursting

- new data processed
- new algorithm runs
- unpredictable event triggers computation
  - (e.g. earthquake)

# Problem

- Bursty
- Computationally expensive
  - predicted 5 years for 1 big, new dataset on existing server
- Solution:
  - Design Scalable Solution using Azure Cloud





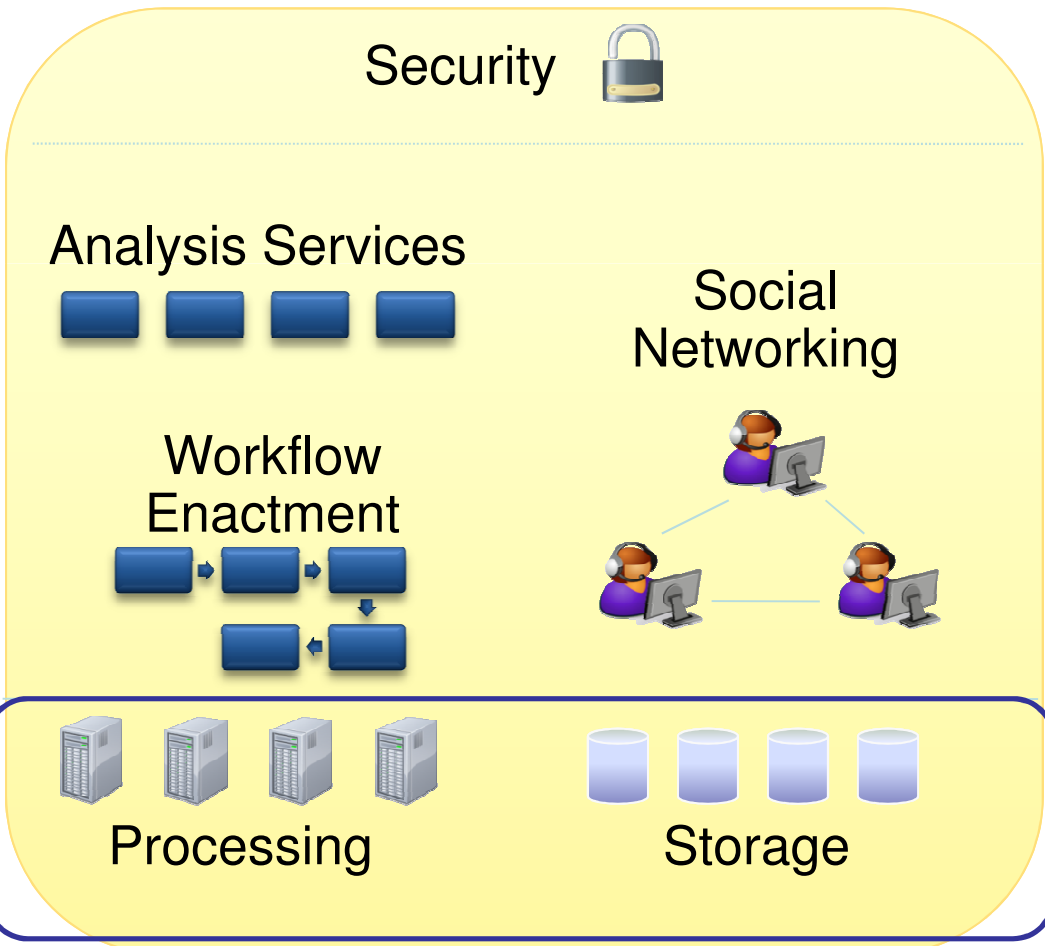
# Approach

- Avoid rewriting all existing Discovery Bus software
- Move existing Discovery Bus to Amazon Cloud
  - without parallelisation
- Move critical tasks to run in parallel on Azure

Discovery Bus  
Planner

*Amazon*

App API



e-Science Central

*Azure*

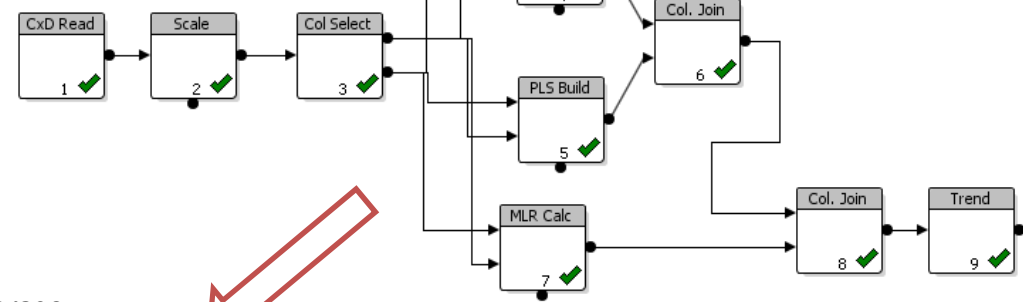
1

Discovery Bus invokes e-Science Central Workflow via API



2

Workflow decomposed to Message Plan



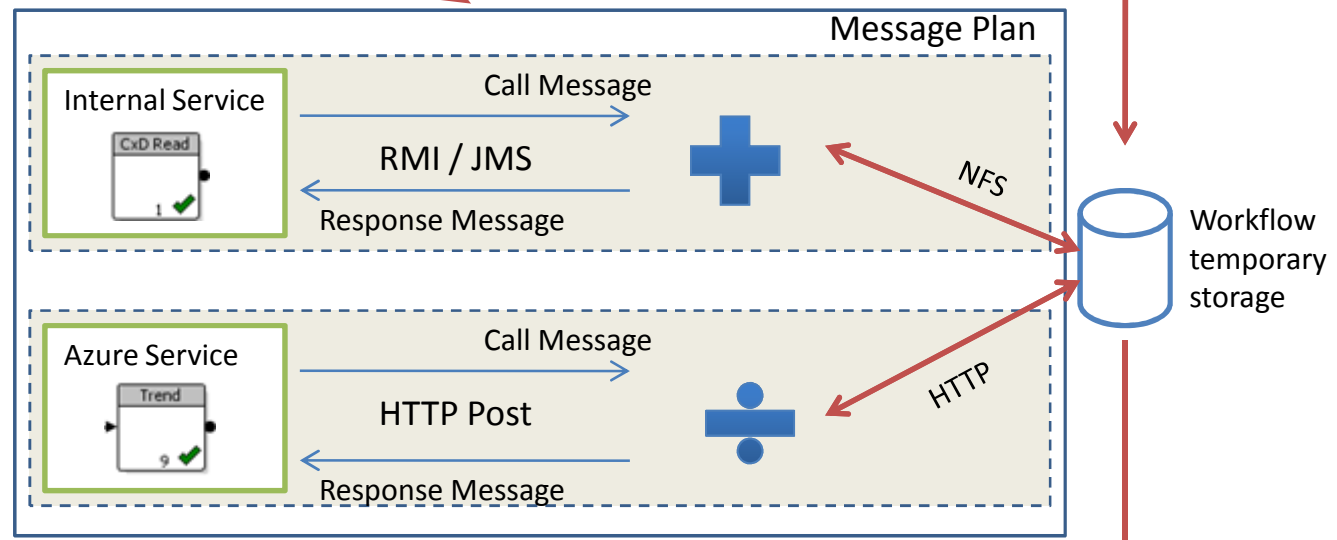
Temporary workflow storage assigned, Message Plan queued for execution.

3



4

Messages sent in sequence



5

Workflow Execution Completes



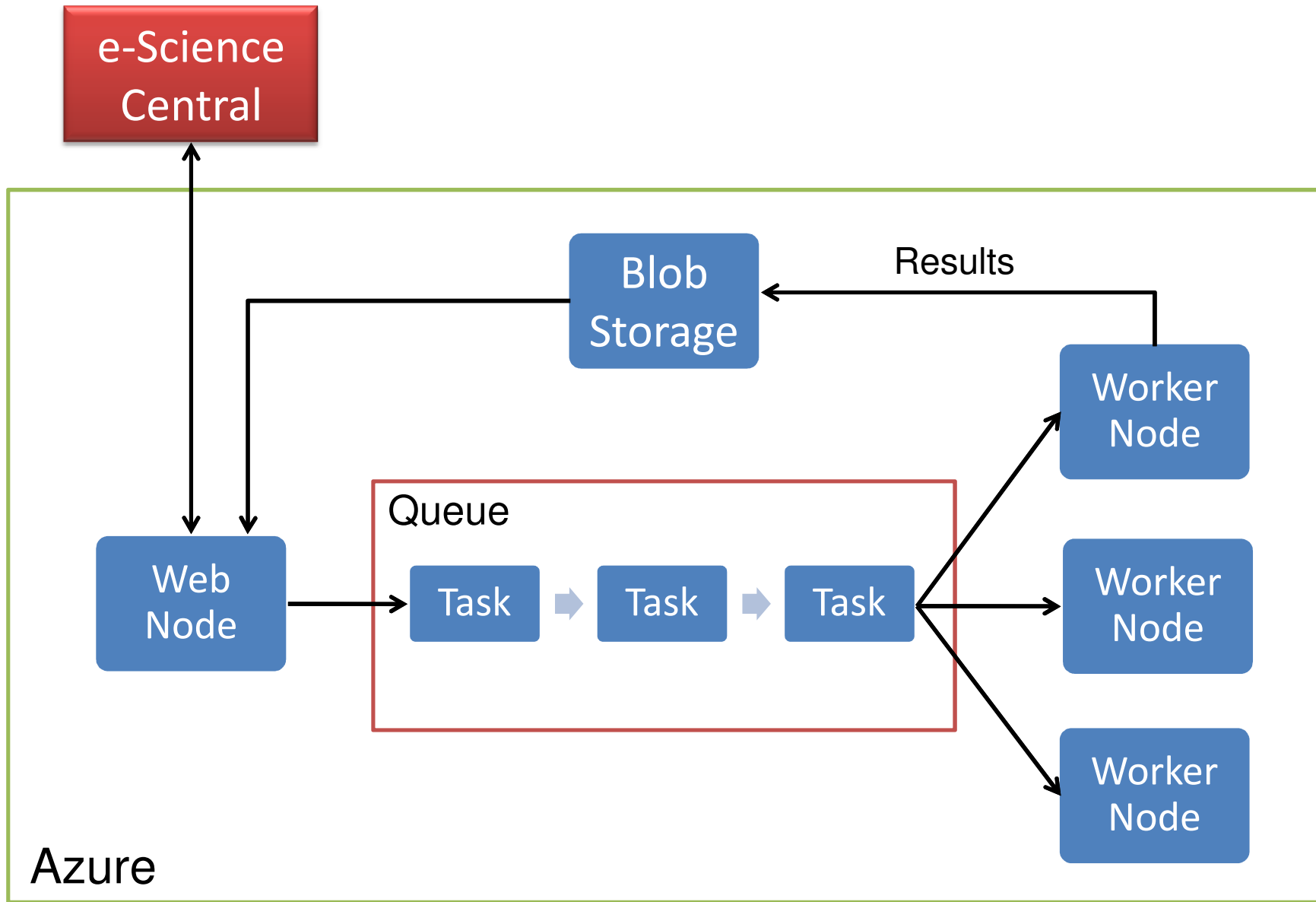
Results data stored in e-Science Central folder



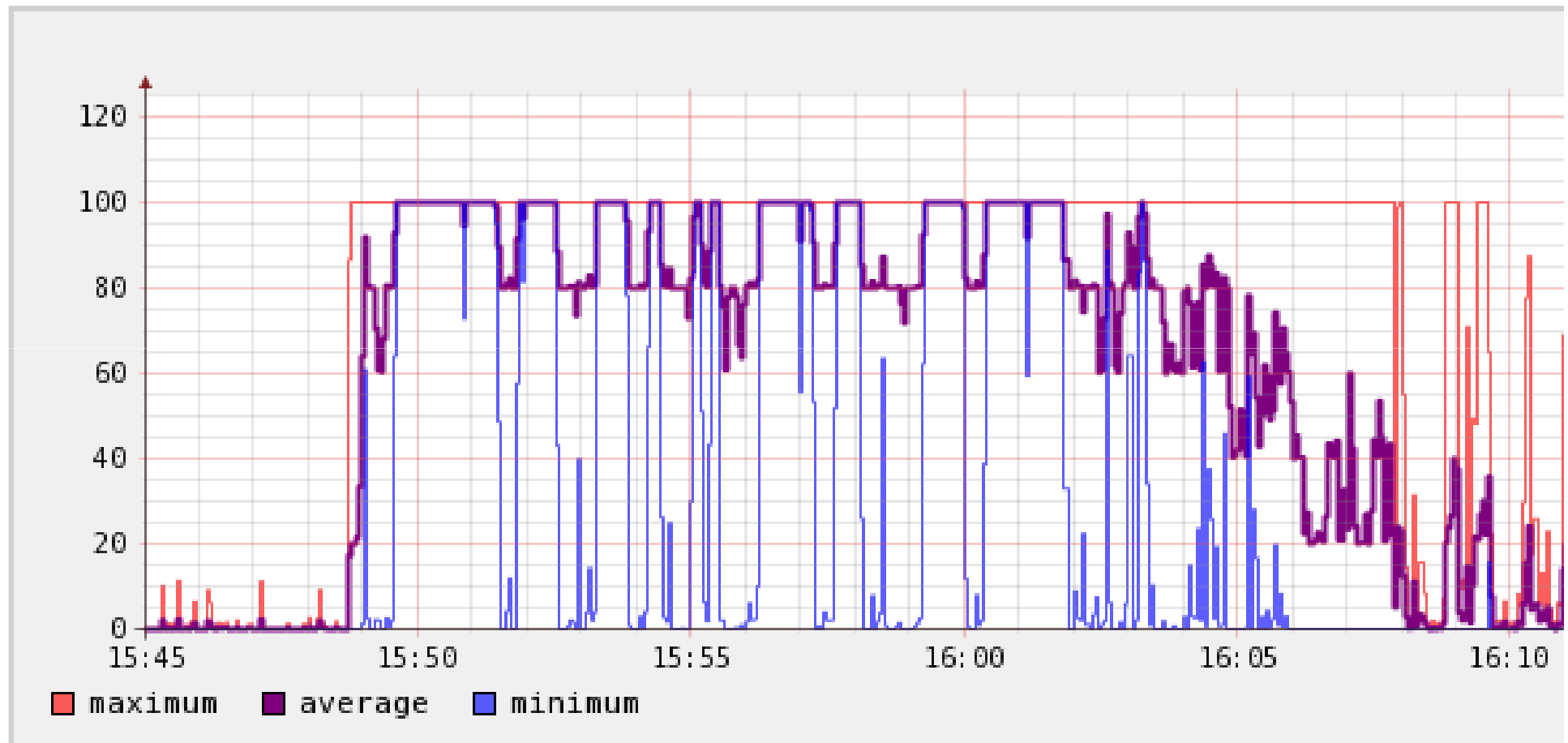
5

Discovery Bus notified with results





# CPU Utilization on 5 Azure Nodes



# Summary

- Clouds can revolutionise e-science
  - sustainable, scalable infrastructure
  - reduce time from idea to realisation
- Discovery Bus exemplifies a good Cloud pattern
  - large, but bursty requirements
- Clouds do NOT make it easier to build complex, scalable, dependable distributed systems
  - we need higher-level “Science Cloud Platforms”
  - (see tomorrow’s talk on “e-Science Central”)