



Digital Institute

Practical Insights of HPC/HTC on the Cloud

Stephen McGough

Vassilis Glenis, Chris Kilsby, Vedrana Kutija,
Simon Wodman

Workshop 40

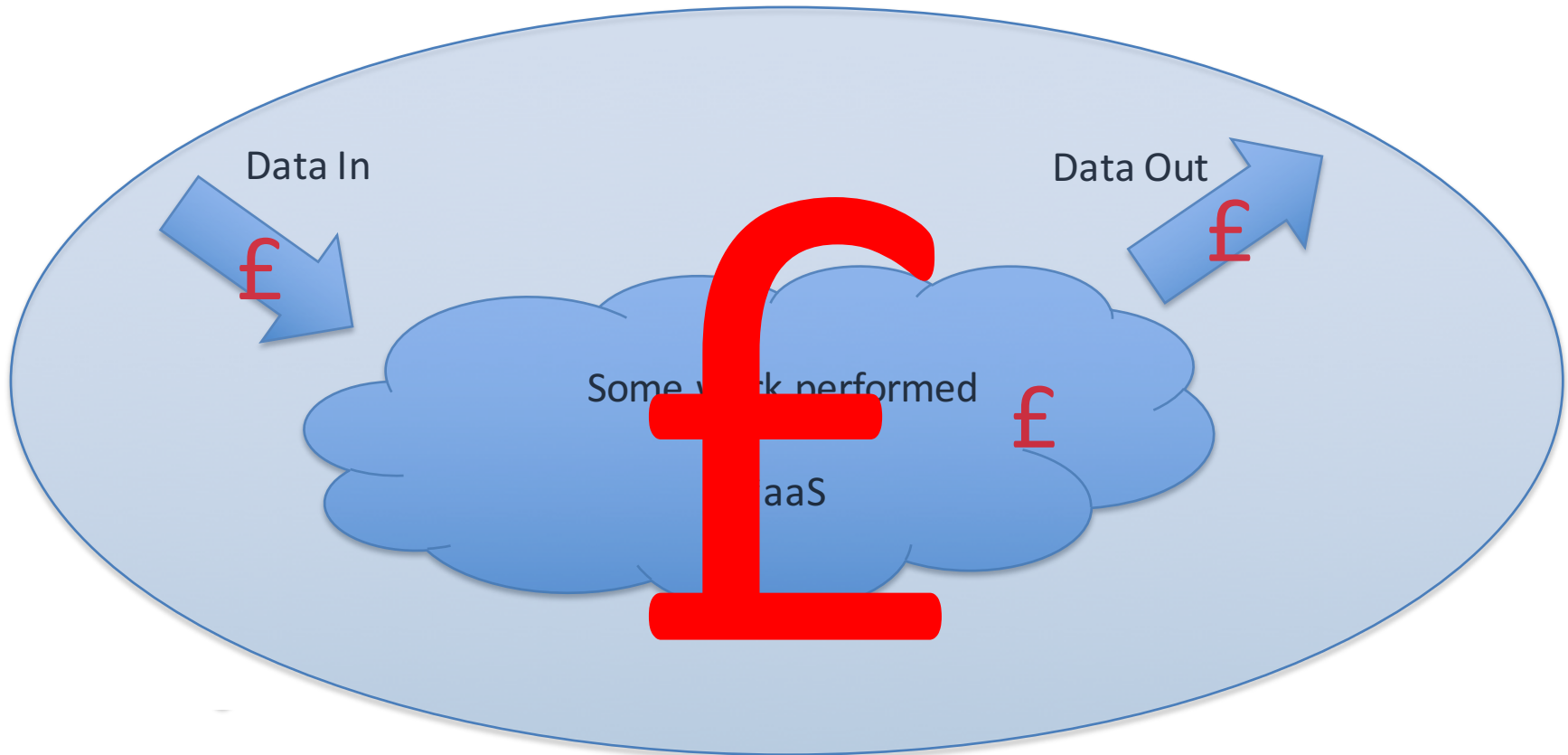
April 4th 2012



Digital Institute

CLOUD COMPUTING

The Cloud model



- SaaS – Software as a Service, you use programs provided by someone else
- PaaS – Platform as a Service, you provide the program and above
- IaaS – Infrastructure as a Service, you provide the OS and above

What is the Cloud Good for?

- Highly democratic access to resources
 - Anyone with a credit card can use
- Rapid provisioning
 - Minutes as opposed to days
- Only pay for the time you use
 - Operational cost rather than capital cost
- Very good for small players
 - Can access more resources than they could buy

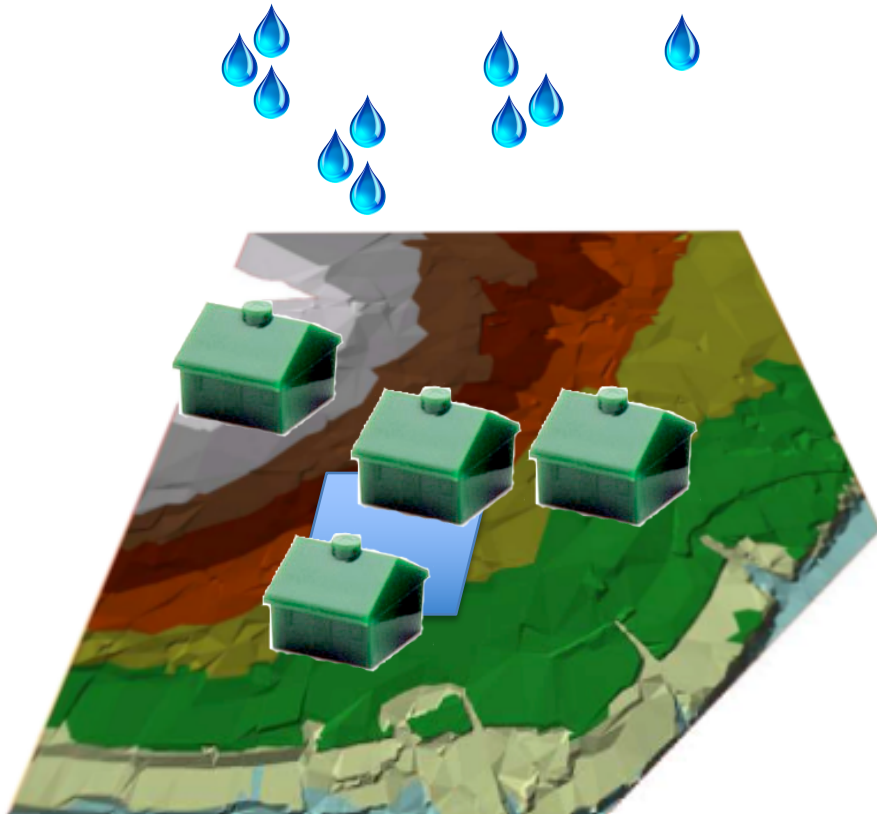
- General problems
 - Data (program) security / legality
 - Can I send my data (program) there?
 - Interaction with program
 - GUI interfaces
 - Interactive use
 - Data volumes
 - How long to transfer data?
 - data to program vs program to data
- Cloud specific problems
 - Data volumes
 - Cost of transferring
 - Execution time
 - Execution time (cost) vs ownership cost



Digital Institute

THE PROBLEM

Flood Modelling



- Landscape topology
- Add buildings
- Simulate rain storms
- Flow rainfall over landscape
- Determine which areas are prone to flooding

- Available resources
 - Simulate a city suburb (small area low resolution) requires ~1.5GB memory
 - Really want to simulate an entire city (large area) and/or at higher resolution
 - Requiring much more memory (up to 48GB)
- Running multiple ensembles
 - Different rain storms
 - Once every 2/10/20/50/100/200 year storm event
 - Different durations
 - 15 minutes, 30 minutes, 1 hour, 2 hours, 3 hours and 6 hours
 - Use high throughput computing
- Run time
 - Simulations last 1 – 13 days

Problem Size

Problem	X Small	Small	Medium	Large	Thames
Simulation size	2km by 2km	2km by 2km	2km by 2km	10km by 12km	200km by 55km
Sim resolution	2m by 2m	1m by 1m	0.5m by 0.5m	4m by 4m	15m by 15m
Cells	1 million	4 million	16 million	7.5 million	5 million
Memory	1.5GB	6GB	48GB	24GB	24 GB
Disk Space	100GB	100GB	200GB	200GB	200GB
Runtime	5 – 24 hours	1 – 3 days	7 – 10 days	7 – 11 days	8 – 13 days



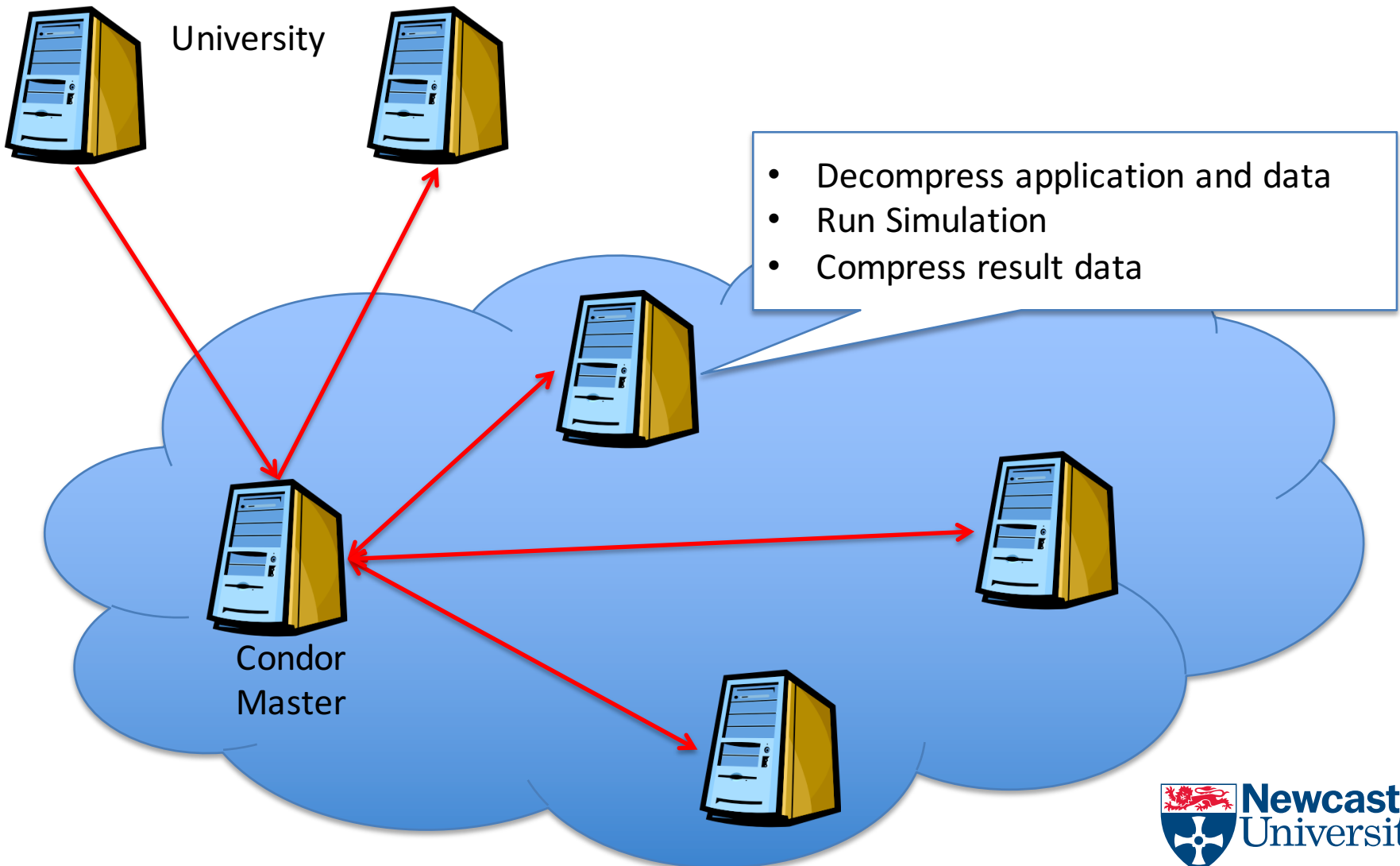
Digital Institute

THE PLAN

The Plan

- Allocated ~£20,000 to run work on the Cloud
- Using IaaS
- Deploy Linux Based Virtual Machines
- Run Condor over virtual Cluster
 - Cluster completely in the Cloud
 - Well understood system
- Instances can be added and removed as required
- Data Staged back to computer in the university on completion of work

Design





Digital Institute

ISSUES

- Software
 - Driven through GUI interface
 - Originally command line but evolved into GUI application
 - Written in Delphi
 - Limits use to Windows based Instances
 - Higher cost for using Windows
- Solution:
 - Removal of GUI interface code – large effort
 - Recompiled with open source Delphi compiler for Linux (Lazarus)

Legal Issues: Licensing

- Landscape topology data free to use under any circumstance
- Building data under license from external company
 - Licensed to use for non commercial use within the University
 - Cloud not in the University
- Solution:
 - Needed to draft up new collaboration agreement with company

- £20,000
 - Can't do this by stealth
- University decided that this should be spent in-house not go to external company
 - Can't we do this in-house?
 - No appropriate resources
 - Can't we buy the resources?
 - Not enough money, £20,000 would buy 2-3 servers
 - ~ 3 years work
 - European Tendering laws
 - Producing a price comparison report
 - Can't we do this in-house – we have resources?
 - Resources are running key university services
- No infrastructure in place to deal with Cloud

Pricing Comparison

Amazon EC2

This information was collected from:
<http://aws.amazon.com/ec2/pricing/>
<http://aws.amazon.com/ec2/instance-types/>

Small Jobs

Instance Name	Number of cores/ Memory	Number of simulations which could run concurrently	Cost per hour	Cost per simulation per hour
Small Instance	1/1.7GB	1	\$0.085	\$0.085
Large	4/7.5GB	4	\$0.34	\$0.085
Extra Large	8/15GB	8	\$0.68	\$0.085
High-Memory Extra Large	6.5/17.1GB	6	\$0.50	\$0.083
Double Extra Large	13/34.2GB	13	\$1.00	\$0.077
Quadruple Extra Large	26/68.4GB	26	\$2.00	\$0.077
High-CPU Medium	5/1.7GB	1	\$0.17	\$0.17
High-CPU Extra Large	20/7GB	7	\$0.68	\$0.097
Cluster Compute	33.5/23GB	23	\$1.60	\$0.07

Cost per simulation per hour

\$0.12

\$0.12

\$0.12

\$0.12

Large Jobs

Instance Name	Number of cores/ Memory	Number of simulations which could run concurrently	Cost per hour	Cost per simulation per hour
Quadruple Extra Large	26/68.4GB	1	\$2.00	\$2.00

Payment Issue

- Most cloud providers 'insist' on payment by credit card
- We wanted to do £20,000 worth of work in one month
 - Can't get a university credit card with this sort of limit
 - Special arrangements made to allow head of finance to have this limit
 - Doesn't work as he wants to pay for an item not have an open card payment
 - Raised existing card limit to £20,000 and split work over payment months
- Discovered too late that Cloud providers can take payment by invoice (not advertised)

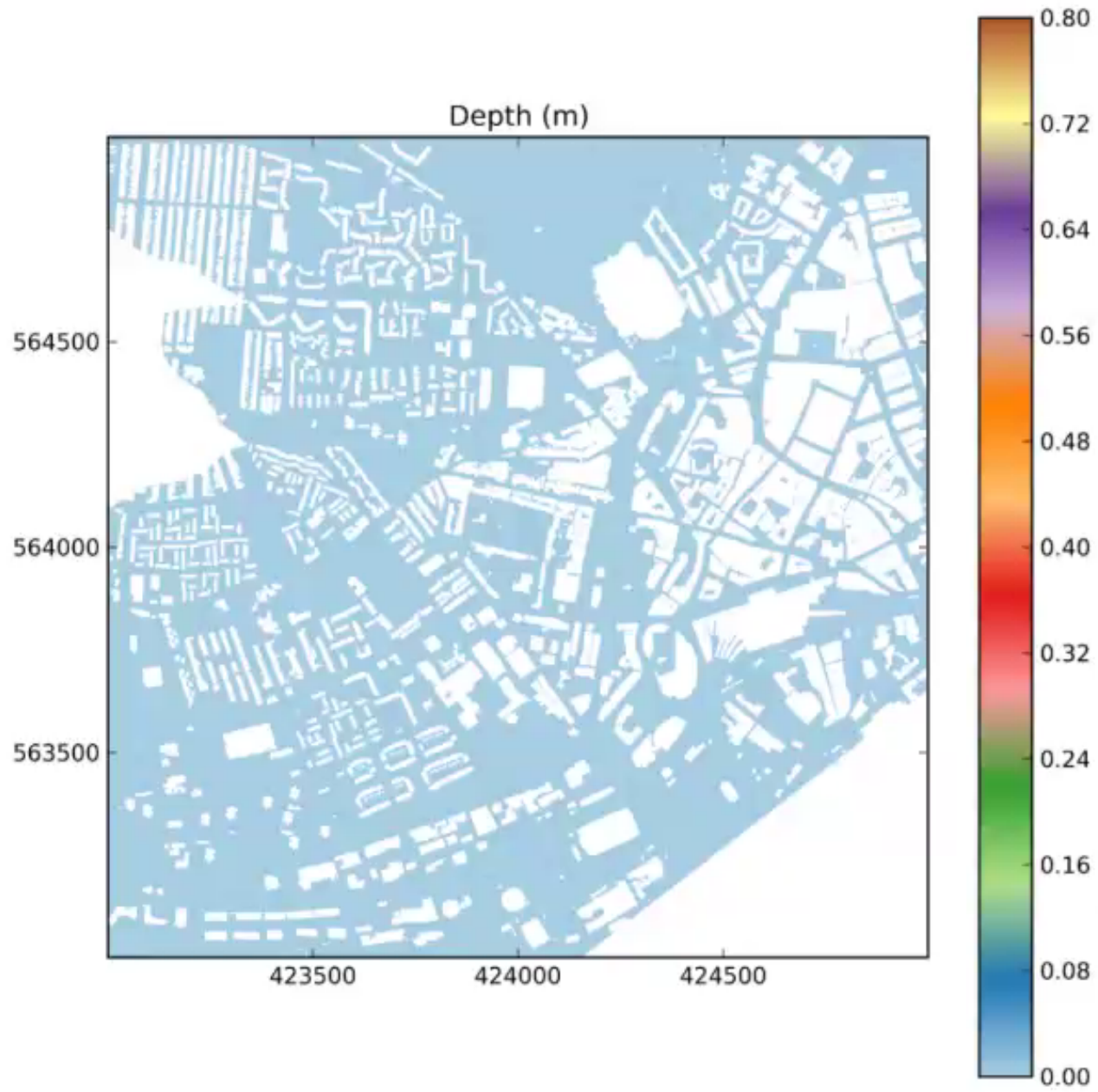
- Upload files:
 - 3 – 25MB depending on simulation run
- Download files:
 - 1.4 – 9GB
- With compression we can bring this down to
 - 125 – 1024MB
- Uploads free downloads not
- Data transfer speeds
 - Dependent on time of day
 - Varied from 200-600KB/s (worst when US awake)
 - Keeping instances alive



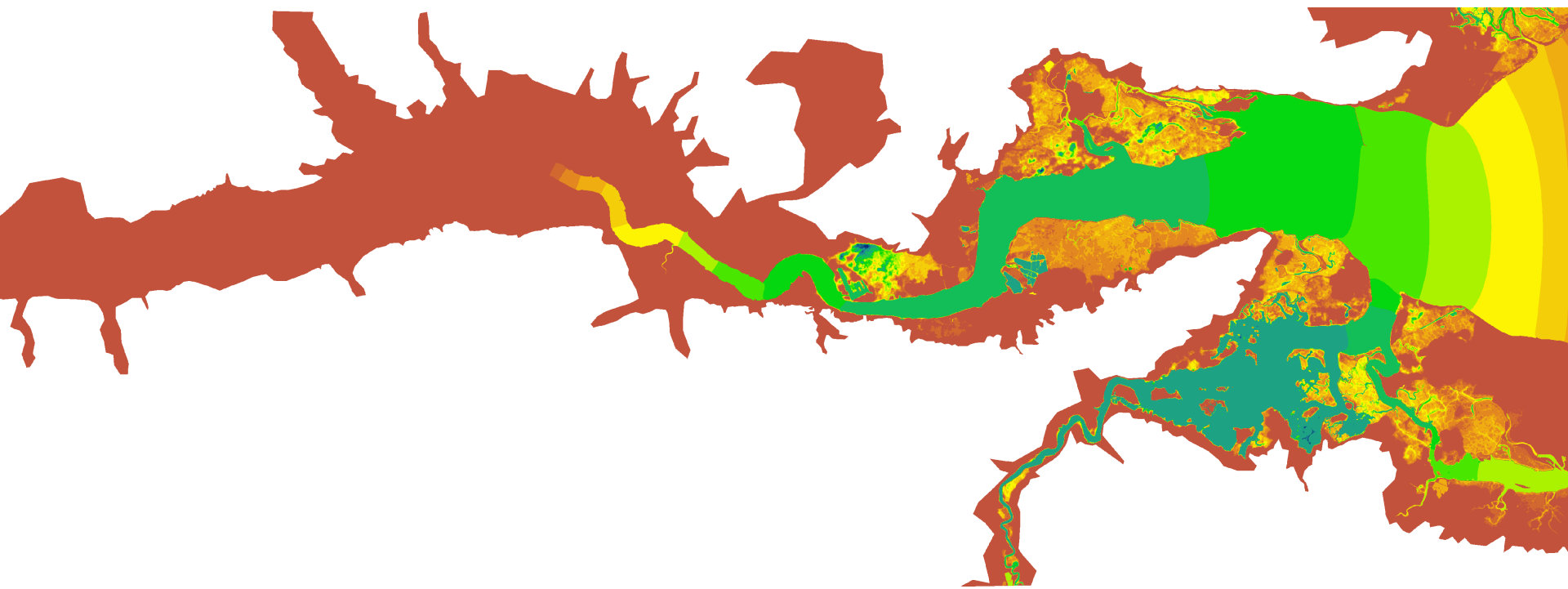
Digital Institute

SOME PRELIMINARY RESULTS









Conclusions

- Technical issues on using the Cloud are relatively easy to solve
- Political issues are more difficult to solve
 - Breaks the model of how University does work
 - Lots of people to convince this is a valid way to do things
- Hopefully next time will be easier!
- Need more (Cloud) time to process results



Digital Institute

stephen.mcgough@ncl.ac.uk

Thank you

QUESTIONS