

Big Data

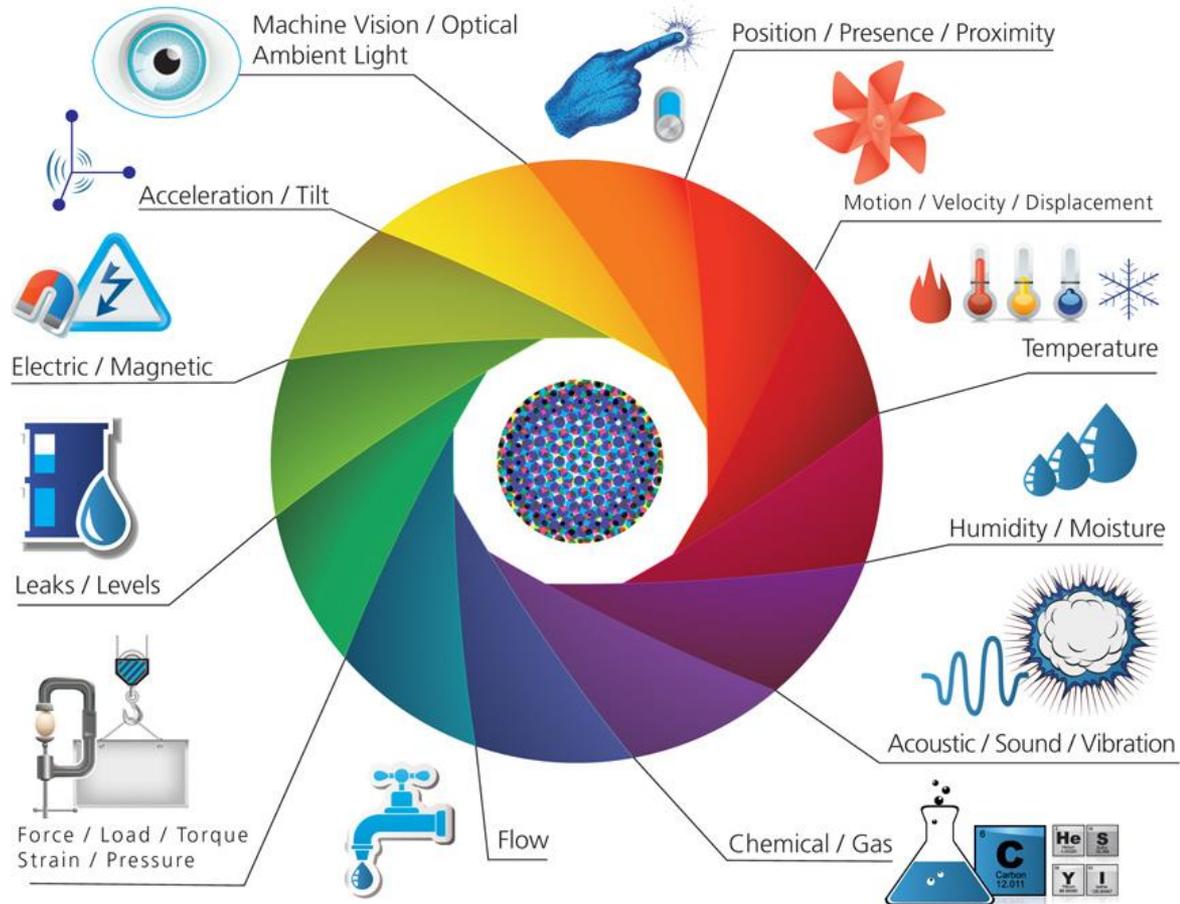
Paul Watson, Hugo Hiden, Simon Woodman,
Mike Simpson
Digital Institute
Newcastle University
UK

ICL Goldrush MegaServer 1995



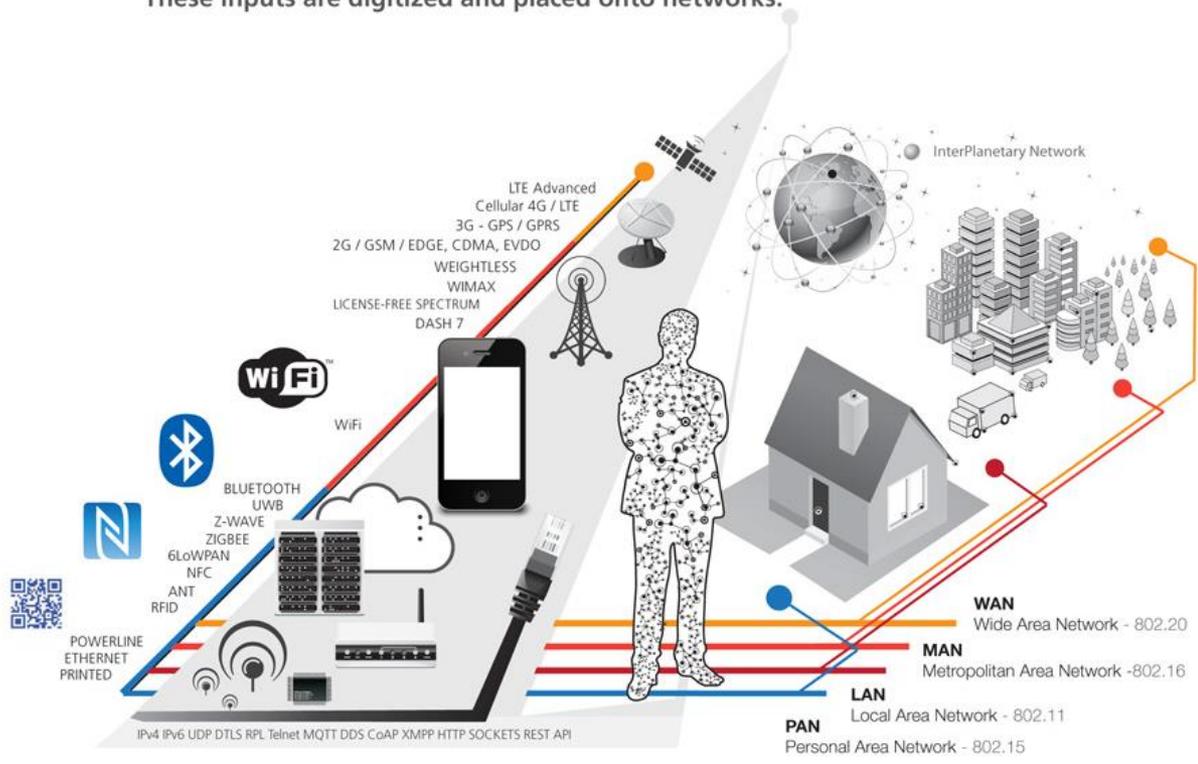
1 SENSORS & ACTUATORS

We are giving our world a digital nervous system. Location data using GPS sensors. Eyes and ears using cameras and microphones, along with sensory organs that can measure everything from temperature to pressure changes.



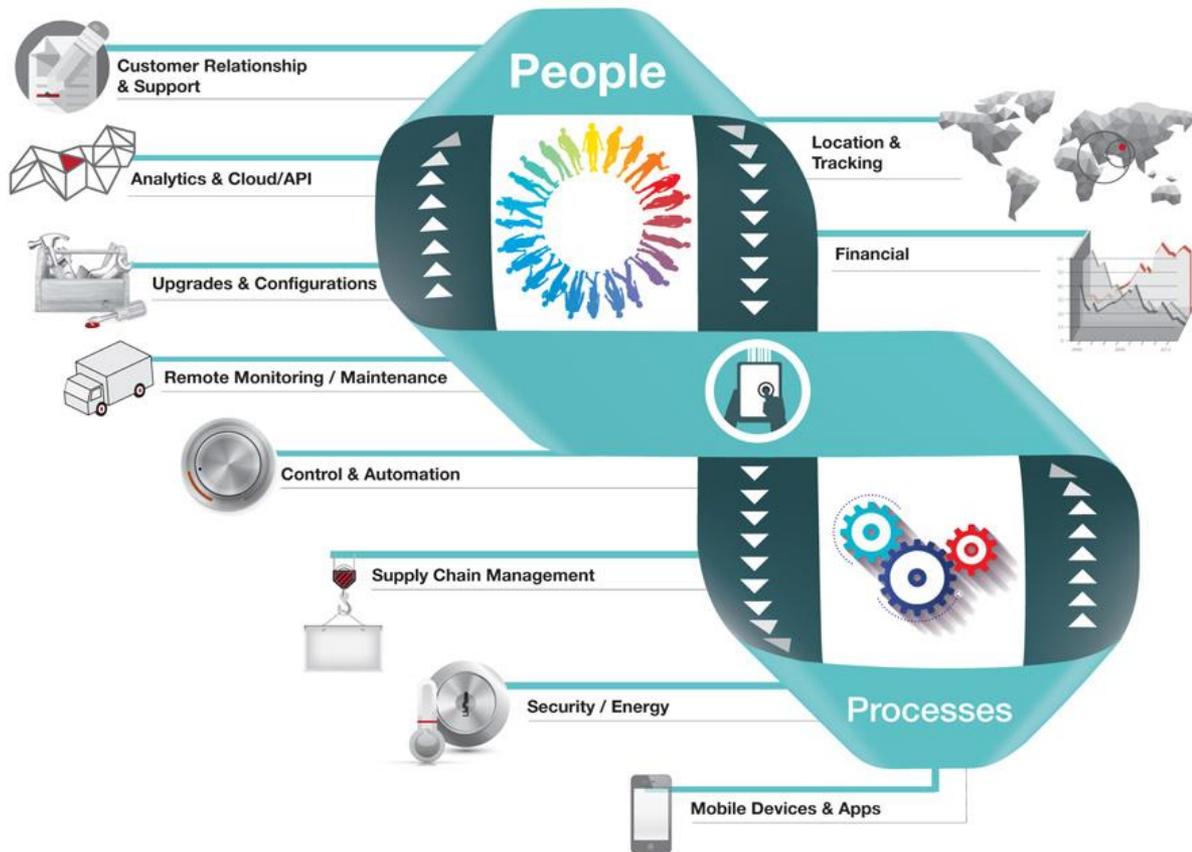
2 CONNECTIVITY

These inputs are digitized and placed onto networks.



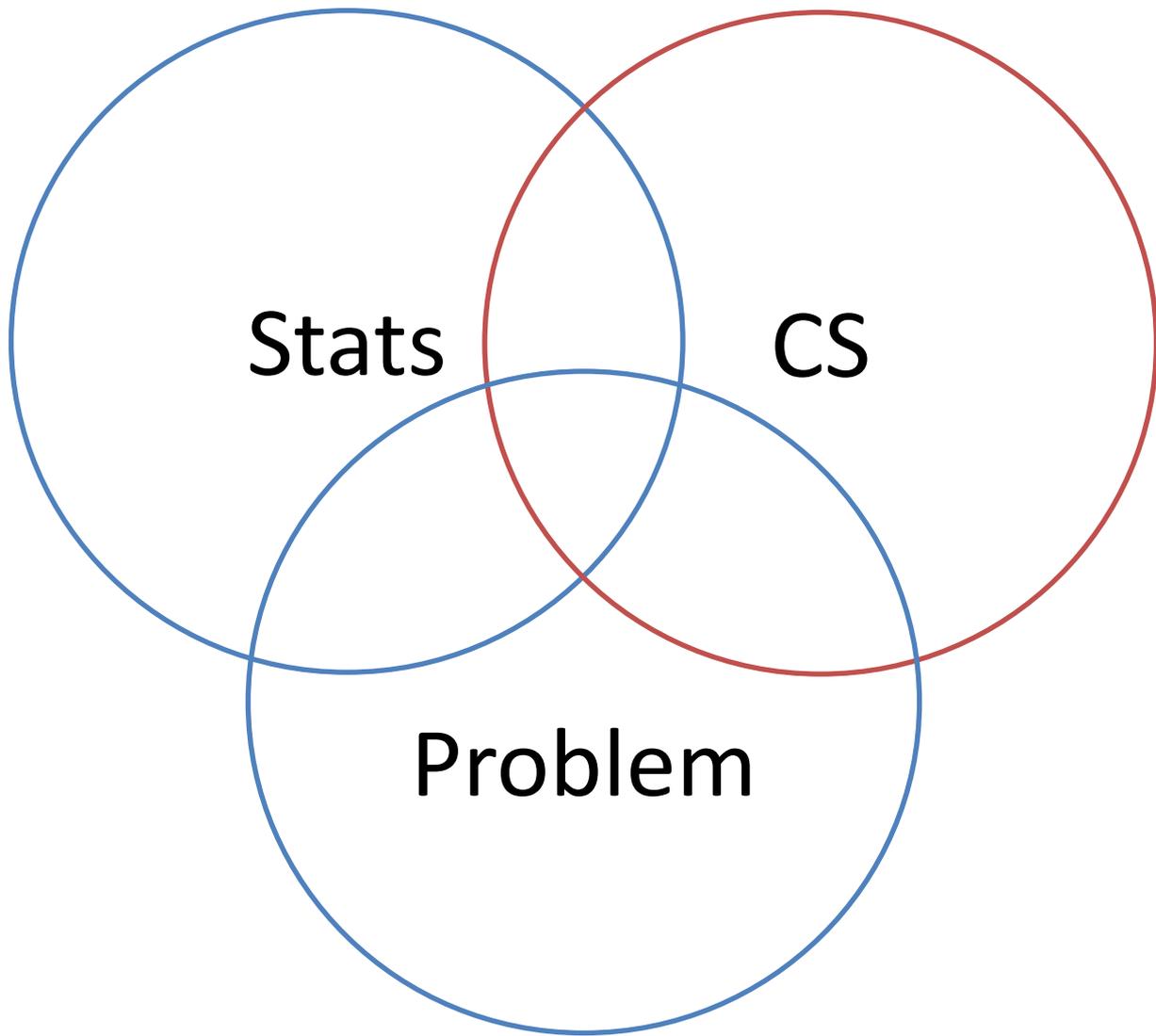
3 PEOPLE & PROCESSES

These networked inputs can then be combined into bi-directional systems that integrate data, people, processes and systems for better decision making.



Challenge

Data $\overset{?}{\Rightarrow}$ Value

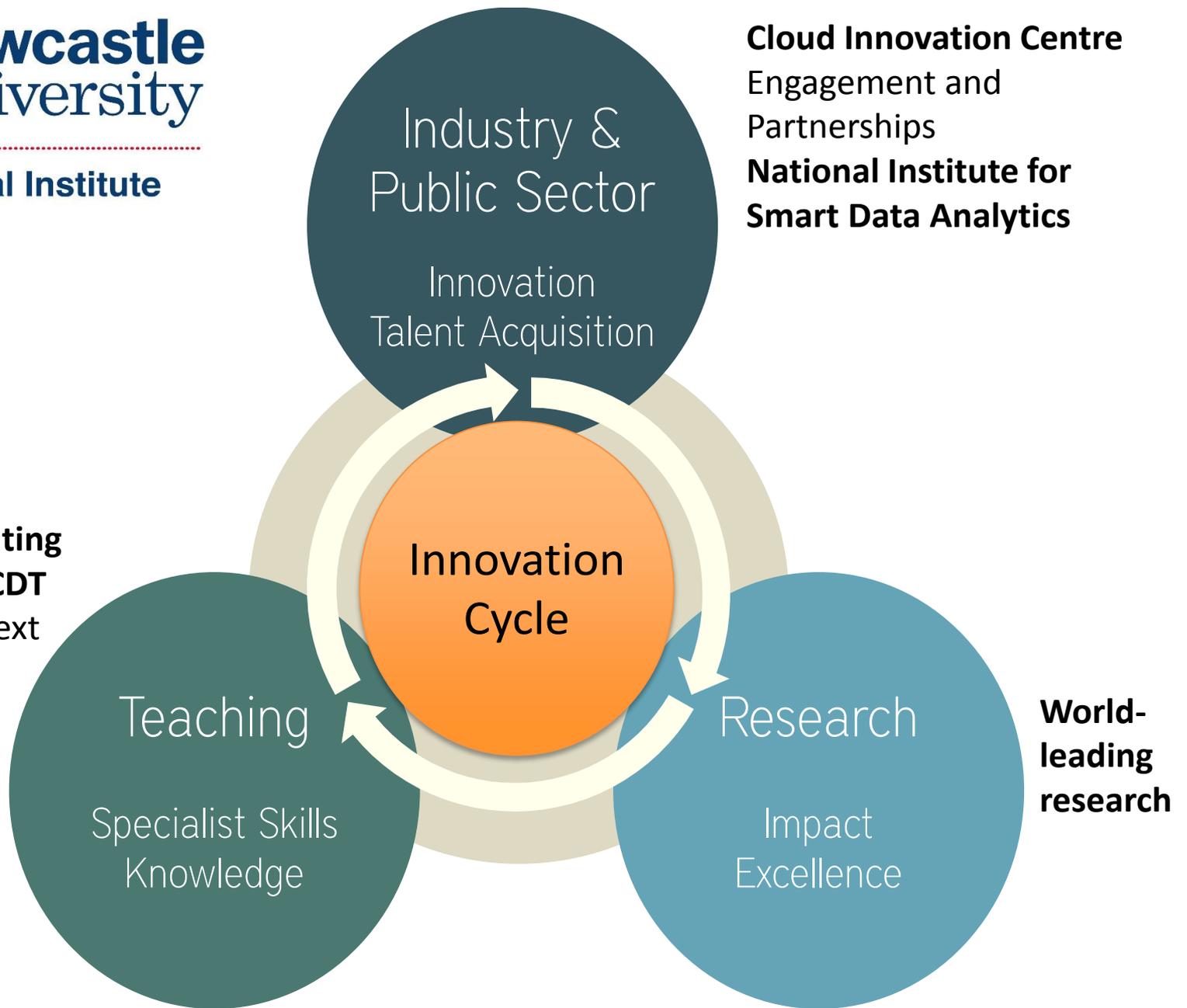


Stats

CS

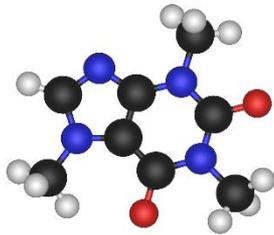
Problem

Cloud Computing for Big Data CDT
Developing next generation of leaders

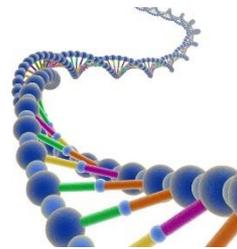


Example Application Areas

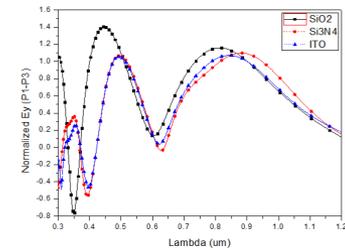
Chemical Informatics



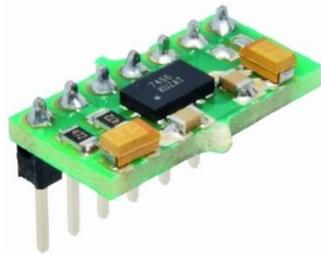
Bioinformatics



Materials



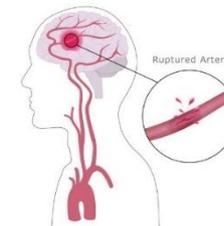
Dementia



Diabetes



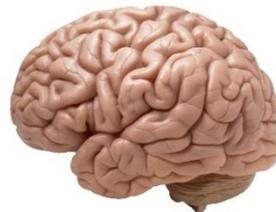
Stroke



Transport



Neuroscience



Architecture



How can we transform rehabilitation?



16 million people a
year suffer a stroke

limbsalive.escapp.net/circus-challenge/clinicians/patients/current/

Limbs Alive | Clinician Portal

Search...

Dashboard

Patients

My Patients

Add Patient

Transfer Patient

Discharge Patient

Goals

Messages

Settings

My Patients

Search

Patient ID	Average Play Time	Last Played	Total Play Time	Total Assessments	Current CAHAI
CL	00:01:010	2013-10-04 17:01:51	03:46:19	2	59.6
P006	00:02:23	2013-11-19 13:06:13	00:07:10	2	49.7
P007	00:00:00		00:00:00	0	0.0
P1	00:00:34	2013-11-19 13:06:13	00:05:06	3	49.7
P2	00:00:00		00:00:00	1	49.7
P3	00:06:02	2013-11-20 10:54:03	00:06:02	1	49.7
P4	00:00:00		00:00:00	0	0.0
SW001	00:00:00		00:00:00	0	0.0

Showing 1 to 8 of 8 entries

Previous 1 Next

Limbs Alive | Clinician Portal

Search...

Dashboard

Patients

My Patients

Add Patient

Transfer Patient

Discharge Patient

Goals

Messages

Settings

Patient PA0002

Overview

Patient Code PA0002
 Gender Male
 Age 60.0
 Stroke Date 01-Jan-2010 00:00:00
 Stroke Side Left
 Pre Stroke Dominant... Left

CAHAI

Date	CAHAI Score
02/01/2014	35
04/01/2014	45
10/01/2014	52.5
12/01/2014	55
14/01/2014	58
18/01/2014	60

App 1

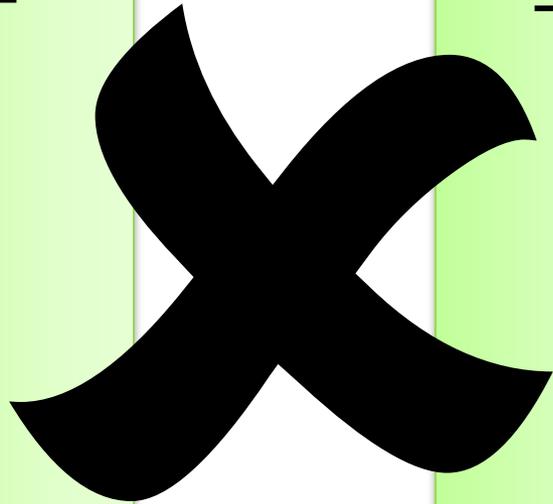
App n

...

Infrastructure:
Storage & Compute

App 1

App n



Infrastructure:
Storage & Compute

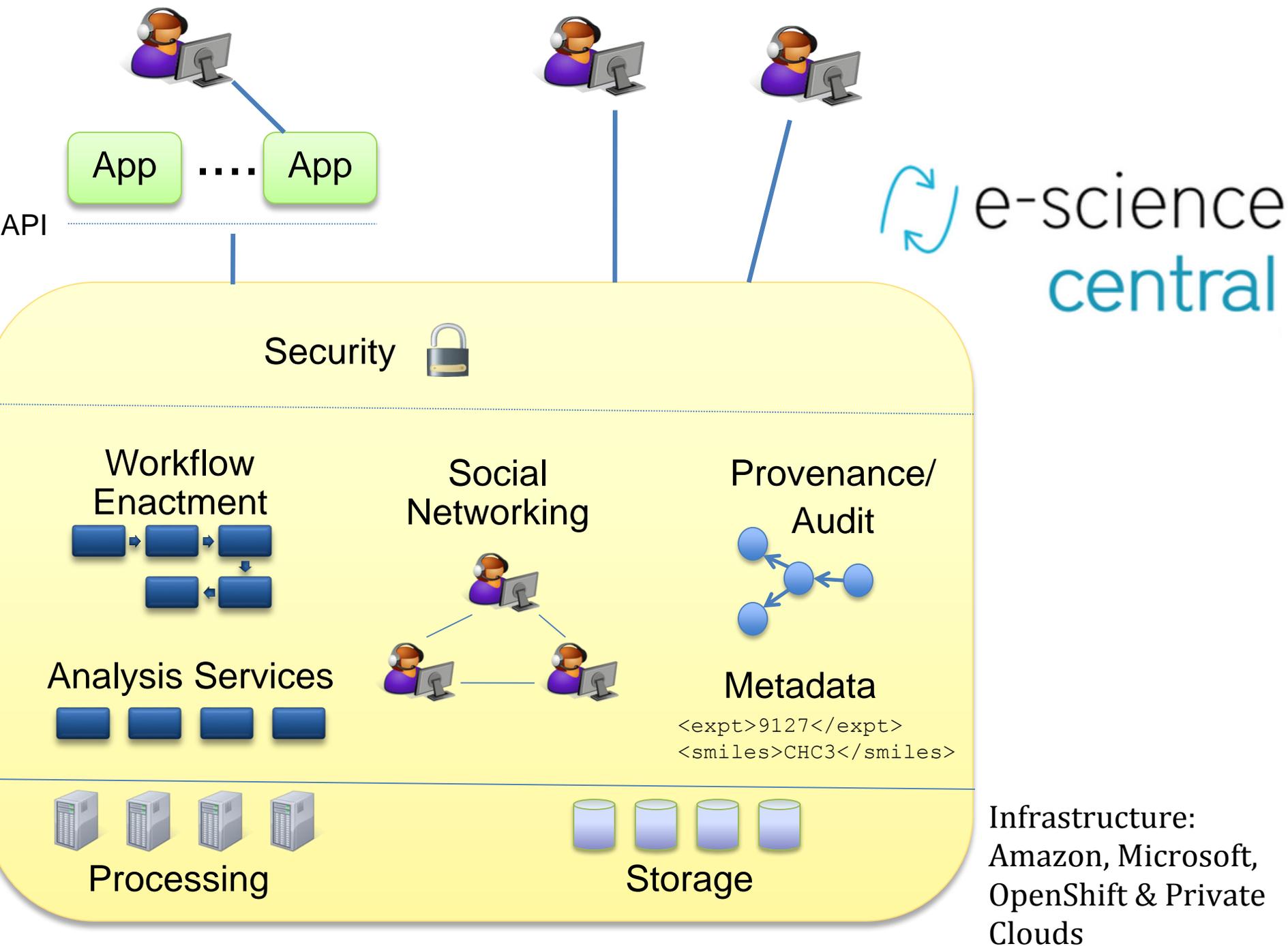
App 1

....

App n

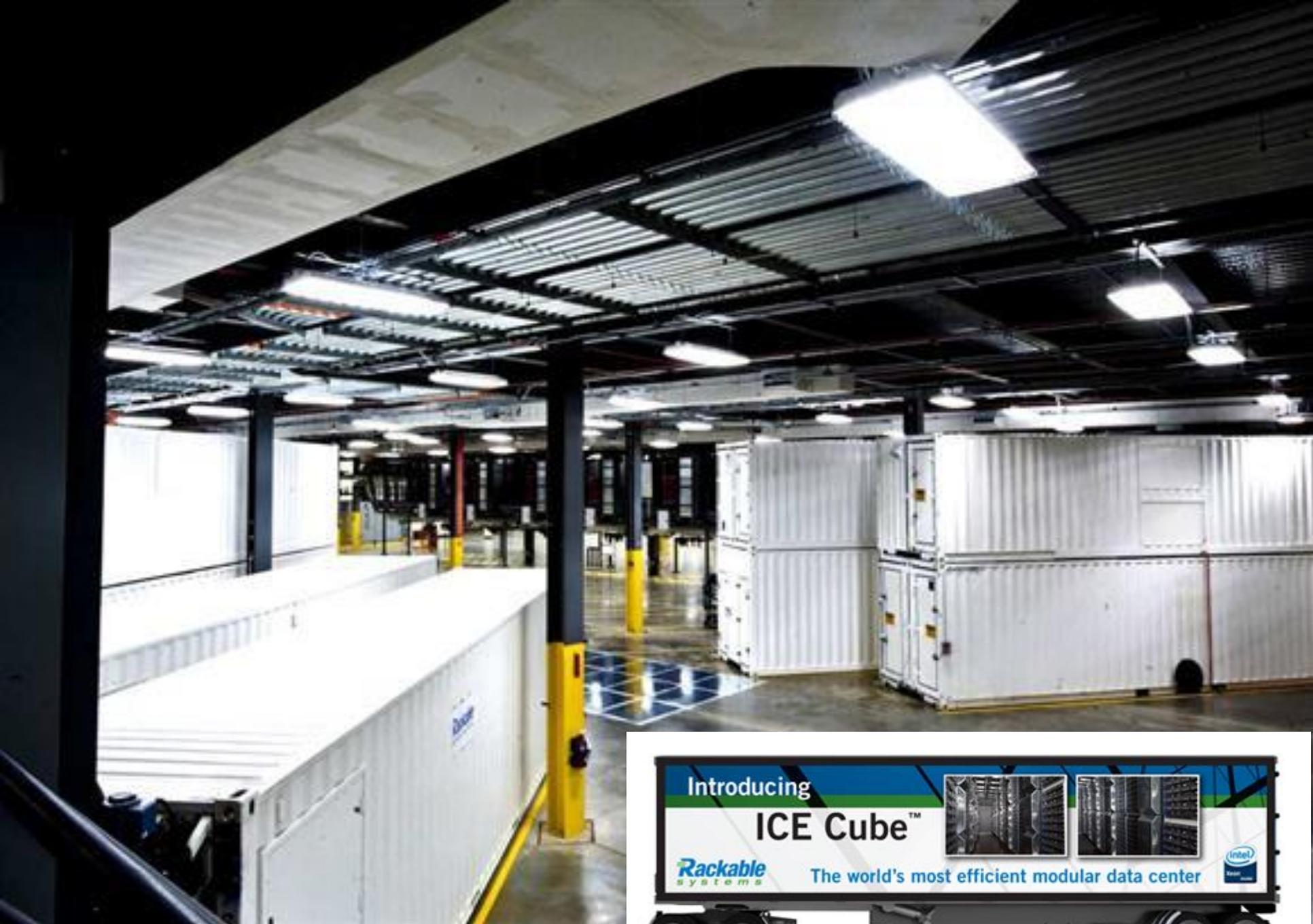
Science Platform:
Store, Share, Analyse

Infrastructure:
Storage & Compute



Demo





Introducing
ICE Cube™

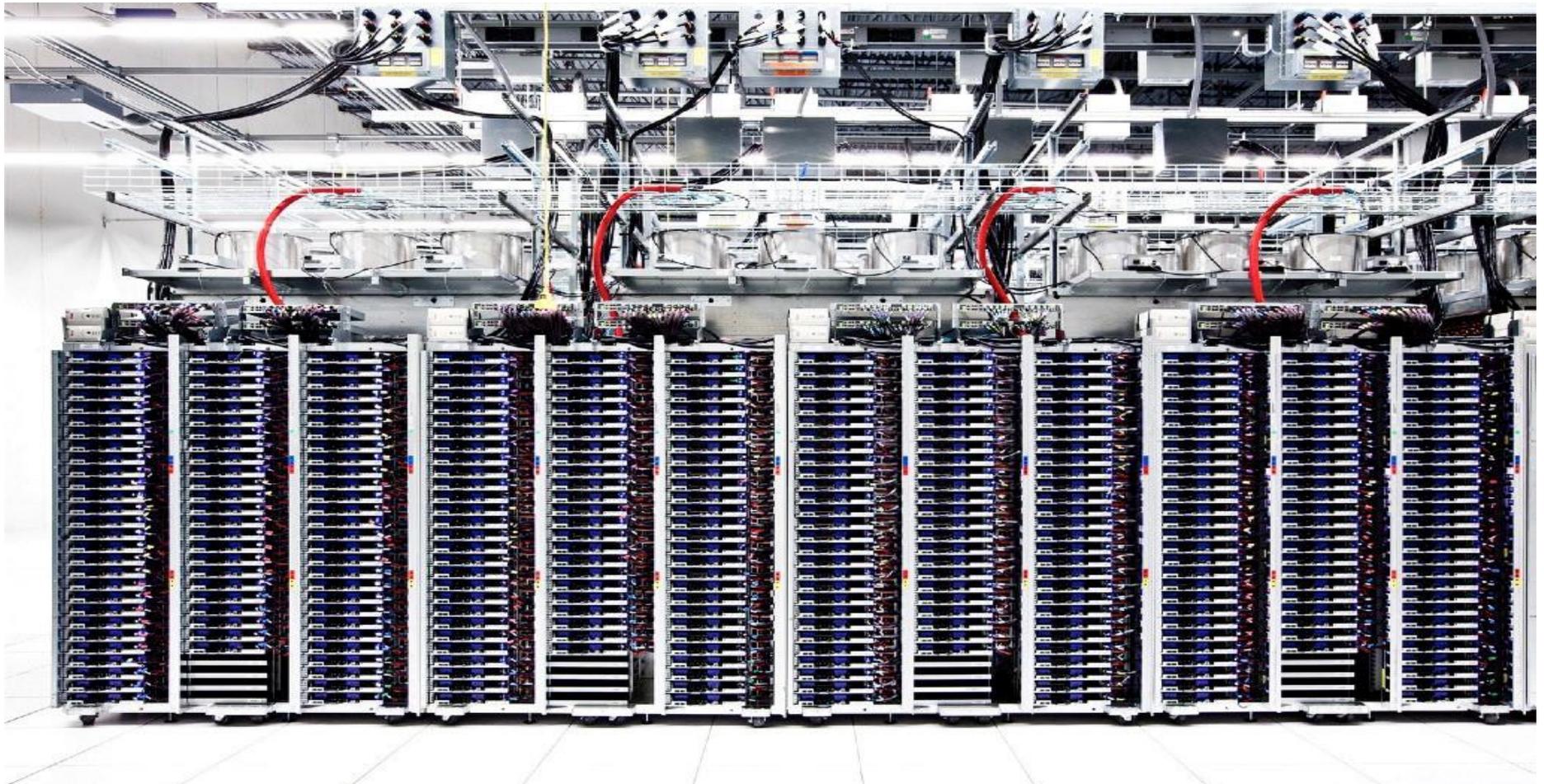
Rackable systems

The world's most efficient modular data center





Racks in a Google Datacentre



What's New?

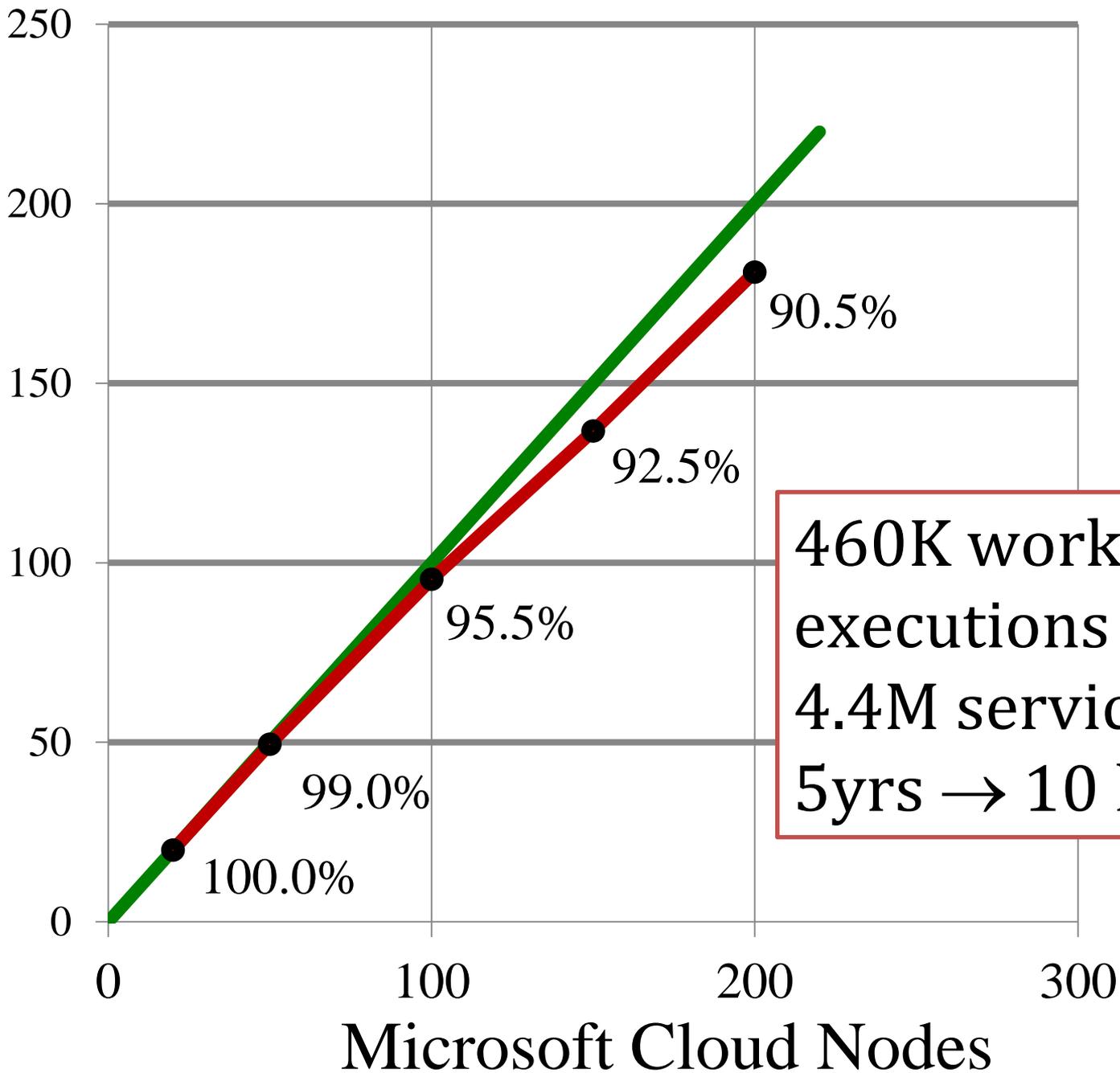
- illusion of infinite computing resources on-demand
- no up-front commitment by users
- Pay for use of resources on a short-term basis as needed

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

The Promise of Clouds

- Agility
- CapEx → OpEx
- Scalability
- Energy Reduction

Relative speed-up



Provenance / Audit

Provenance Report

Document piechart.jpg
Document Id 256712
Document Version Id 256713
Date Tue, 26 Aug 2014 08:01:

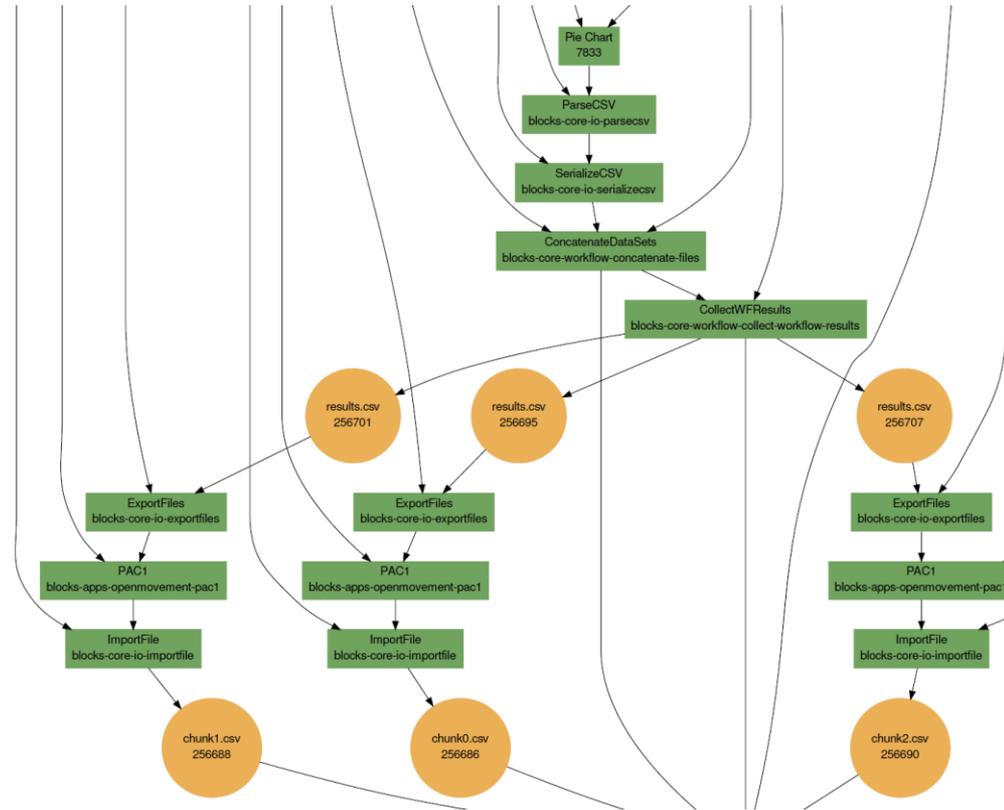
Provenance of piechart.jpg

SerializeCSV (Service Run)

ID blocks-core-io-serializecsv
Version Id 74582
Version Num 5
Start Time Thu Aug 21 13:33:16 UTC 2014
End Time Thu Aug 21 13:33:17 UTC 2014
Block UUID 1C41ACCA-FB8A-C2D2-B09F-A41

ConcatenateDataSets (Service Run)

ID blocks-core-workflow-concatenate-files
Version Id 9708
Version Num 2
Start Time Thu Aug 21 13:33:15 UTC 2014
End Time Thu Aug 21 13:33:16 UTC 2014
Block UUID B146CF0D-D3B1-DD18-9DBC-C1724EEBDD50





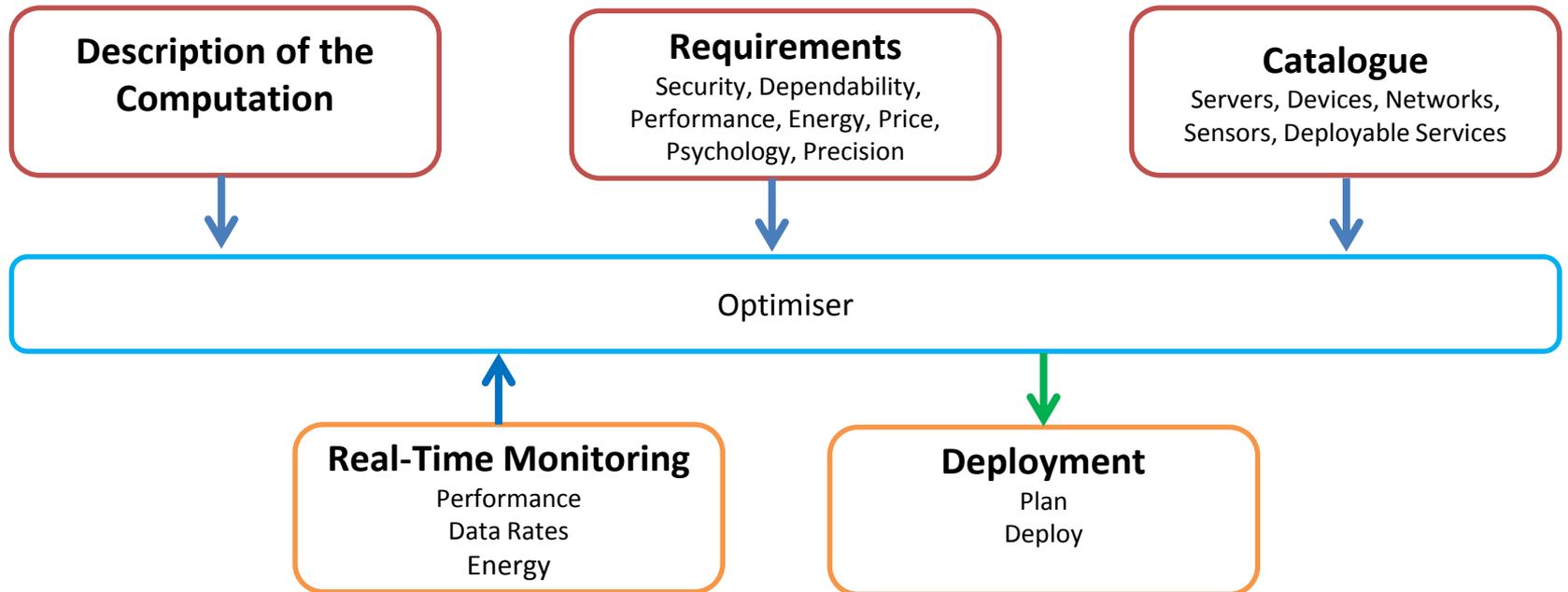
IoT



- High-level description of computation on streaming data – dynamically mapped onto platforms



IoT System Overview



Our CESI Role

- Data and model sharing within & outside the project
 - using the e-Science Central platform
 - Integration with the UKERC energy data centre
- Data analysis
 - high scalability through cloud computing
- Data visualisation and exploration
 - e.g. in Newcastle's 3D Decision Theatres
 - also provides compelling method of engagement

Pathways to Impact for CESI?

Industry Impact

- e-Science Central is open source
 - now used in industry

Impact: Cloud Innovation Centre ⇒ National Institute for Smart Data Innovation



National Institute for Smart Data Innovation

- £30M investment
 - £15M Treasury + £15M Newcastle University
- This will fund
 - new “beacon” building on Science Central (2019)
 - ten years of activities with industry & the public sector

- The National Institute for Smart Data Innovation (NISDI) will enable industry to unleash the huge potential for innovation offered by the explosive growth in digital data.
- Once processed and analysed, digital data can become “Smart Data”, enabling new products and services that greatly benefit the country’s economy and its citizens. However, a major skills gap is preventing the UK from realising the potential of Smart Data.
- NISDI will overcome this barrier by creating a unique new facility which brings together industry, the public sector and universities to create the skills, ideas and resources needed to exploit these opportunities.
- Driven by the needs of industry, NISDI will allow the region, the Northern Powerhouse and the UK to become global leaders in this important sector, estimated to be worth around \$125bn per annum.



NISDI Themes



Health care

Smart
Cities

Automotive

BIM

Manufactur-
ing

Others e.g.
Subsea

Internet of Things

Data Analytics & Visualisation

Cloud Computing

Other technologies – e.g. cyber security

Issues/Opportunities

- Model sharing
- Data preservation
- Meeting Data Security requirements
- “Cloud based controlled energy management systems”
- “Blockchain as a tool for energy retail exchange”
- National Institute for Smart Data Innovation
 - energy theme