



Supercomputing and the cloud – the next big paradigm shift?

Martin Hamilton

Computed by Olivier H. Beauchesne and SCImago Lab, data by Elsevier Scopus

1. Re-use and reproducibility
 2. How do cloud technologies help?
 3. A new paradigm – next steps in open science
-

About Jisc

About Jisc:

- › Registered charity championing the use of digital technologies in research, education and skills
 - › Shared services, national deals, advice & guidance:
 - Janet: reaches 18m in education & public sector + eduroam wireless roaming
 - Digital capabilities programme, supporting next generation of digital leaders
 - Embracing student ideas and using them to build new digital solutions
 - Helping institutions to achieve efficiency gains through digital technology, e.g. sector deals for Microsoft Office365 and Azure cloud services
 - › My role:
 - Generating and channelling new ideas
 - Building partnerships to bring them to fruition
-

1. Re-use and reproducibility

SECTIONS LATEST POPULAR SEARCH THE NEW YORKER

ANNALS OF SCIENCE DECEMBER 13, 2010 ISSUE

THE TRUTH WEARS OFF

Is there something wrong with the scientific method?

By Jonah Lehrer

On September 18, 2007, a few dozen neuroscientists, psychiatrists, and drug-company executives gathered in a hotel conference room in Brussels to hear some startling news. It had to do with a class of drugs known as atypical or second-generation antipsychotics, which came on the market in the early nineties. The drugs, sold under brand names such as Abilify, Seroquel, and Zyprexa, had been tested on schizophrenics in several large clinical trials, all of which had demonstrated a dramatic decrease in the subjects' psychiatric symptoms. As a result, second-generation antipsychotics had become one of the fastest-growing and most profitable pharmaceutical classes. By 2001, Eli Lilly's Zyprexa was generating more revenue than Prozac. It remains the company's top-selling drug.



Many results that are rigorously proved and accepted start shrinking in later studies.

ILLUSTRATION BY LAURENT CILLUFFO

<http://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off>

SECTIONS LATEST POPULAR SEARCH THE NEW YORKER

ANNALS OF SCIENCE DECEMBER 13, 2010 ISSUE

THE TRUTH WEARS OFF

Is there something wrong with the scientific method?

By Jonah Lehrer

On September 18, 2007, a few dozen neuroscientists, psychiatrists, and drug-company executives gathered in a hotel conference room in Brussels to hear some startling news. It had to do with a class of drugs known as atypical or second-generation antipsychotics, which came on the market in the early nineties. The drugs, sold under brand names such as Abilify, Seroquel, and Zyprexa, had been tested on schizophrenics in several large clinical trials, all of which had demonstrated a dramatic decrease in the subjects' psychiatric symptoms. As a result, second-generation antipsychotics had become one of the fastest-growing and most profitable pharmaceutical classes. By 2001, Eli Lilly's Zyprexa was generating more revenue than Prozac. It remains the company's top-selling drug.



Many results that are rigorously proved and accepted start shrinking in later studies.
ILLUSTRATION BY LAURENT CILLUFFO

<http://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off>

<http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>

nature International weekly journal of science

Home News & Comment Research Careers & Jobs Current Issue Archive Audi

Archive Volume 533 Issue 7604 News Feature Article

NATURE | NEWS FEATURE

1,500 scientists lift the lid on reproducibility


Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016 | Corrected: 28 July 2016

PDF Rights & Permissions

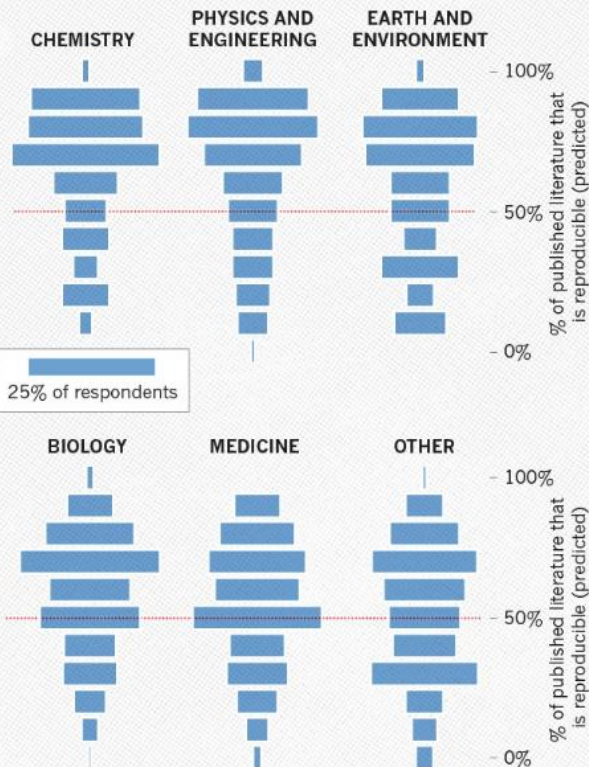
Is there a reproducibility crisis in science?



00:00 02:03

HOW MUCH PUBLISHED WORK IN YOUR FIELD IS REPRODUCIBLE?

Physicists and chemists were most confident in the literature.



Number of respondents from each discipline:
 Biology 703, Chemistry 106, Earth and environmental 95,
 Medicine 203, Physics and engineering 236, Other 233



Open-use and reproducibility

<http://www.nature.com/news/1-500->

[-lid-on-](#)

[19970](#)

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio

Archive | Volume 533 | Issue 7604 | News Feature | Article

NATURE | NEWS FEATURE

1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016 | Corrected: 28 July 2016

PDF | Rights & Permissions

Is there a reproducibility crisis in science?

00:00 | 02:03

ANNALS OF SCIENCE DECEMBER 13, 2010 ISSUE

THE TRUTH WEARS OF

Is there something wrong with the scientific method?

By Jonah Lehrer

On September 18, 2007, a few dozen neuroscientists, psychiatrists, and drug-company executives gathered in a hotel conference room in Brussels to hear some startling news. It had to do with a class of drugs known as atypical or second-generation antipsychotics, which came on the market in the early nineties. The drugs, sold under brand names such as Abilify, Seroquel, and Zyprexa, had been tested on schizophrenics in several large clinical trials, all of which had demonstrated a dramatic decrease in the subjects' psychiatric symptoms. As a result, second-generation drugs became one of the fastest-growing and most profitable products in the pharmaceutical industry. Eli Lilly's Zyprexa was generating more revenue than the company's top-selling drug.

Many have accepted this as the truth. ILLUSTRATION

<http://www.newyorker.com>

SECTIONS LATEST POPULAR SEARCH

ANNALS OF SCIENCE DECEMBER 13, 2010 ISSUE

THE TRUTH WEARS OFF

Is there something wrong with the scientific method?

By Jonah Lehrer

On September 18, 2007, a few dozen neuroscientists, psychiatrists, and drug-company executives gathered in a hotel conference room in Brussels to hear some startling news. It had to do with a class of drugs known as atypical or second-generation antipsychotics, which came on the market in the early nineties. The drugs, sold under brand names such as Abilify, Seroquel, and Zyprexa, had been tested on schizophrenics in several large clinical trials, all of which had demonstrated a dramatic decrease in the subjects' psychiatric symptoms. As a result, second-generation drugs became one of the fastest-growing and most profitable products in the pharmaceutical industry. Eli Lilly's Zyprexa was generating more revenue than any other drug in the company's top-selling drug.

Many have accepted this as the truth. ILLUSTRATION

<http://www.newyorker.com>

Validating key experimental results via independent replication

Learn more »

As seen in



Major projects



Reproducibility Initiative



Reproducibility Project: Cancer Biology



The screenshot shows the Research Councils UK website. At the top left is the logo and name 'RESEARCH COUNCILS UK'. To the right is a search bar with the text 'search' and a 'Search' button. Below the logo is a navigation menu with items: Home, Funding, Research, Innovation, Skills, Public Engagement, News, Events and Publications, and About Us. Below the navigation menu is a breadcrumb trail: Home / Research / RCUK Common Principles on Data Policy. The main heading is 'RCUK Common Principles on Data Policy'. Below the heading is a paragraph: 'Making research data available to users is a core part of the Research Councils' remit and is undertaken in a variety of ways. We are committed to transparency and to a coherent approach across the research base. These RCUK common principles on data policy provide an overarching framework for individual Research Council policies on data policy.' Below this is a section titled 'Principles' followed by a bulleted list of seven principles.

RESEARCH COUNCILS UK

Accessibility | Media Enquiries

search Search

Home Funding Research Innovation Skills Public Engagement News, Events and Publications About Us

Home / Research / RCUK Common Principles on Data Policy

RCUK Common Principles on Data Policy

Making research data available to users is a core part of the Research Councils' remit and is undertaken in a variety of ways. We are committed to transparency and to a coherent approach across the research base. These RCUK common principles on data policy provide an overarching framework for individual Research Council policies on data policy.

Principles

- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner.
- Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.
- RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.
- To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.
- In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

<http://www.rcuk.ac.uk/research/datapolicy/>

RESEARCH COUNCILS UK

Accessibility | Media Enquiries

search Search

Home Funding Research Innovation Skills Public Engagement News

Home / Research / RCUK Common Principles on Data Policy

RCUK Common Principles on Data Policy

Making research data available to users is a core part of the Research Councils' remit and is undertaken to transparency and to a coherent approach across the research base. These RCUK common principles framework for individual Research Council policies on data policy.

Principles

- Publicly funded research data are a public good, produced in the public interest, which should be made available with as few restrictions as possible in a timely and responsible manner.
- Institutional and project specific data management policies and plans should be in accordance with good practice. Data with acknowledged long-term value should be preserved and remain accessible.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be available to enable other researchers to understand the research and re-use potential of the data and information on how to access the supporting data.
- RCUK recognises that there are legal, ethical and commercial constraints on release of research data and that the research process should not be damaged by inappropriate release of data, research organisation policies and practices should be developed to address these constraints.
- To ensure that research teams get appropriate recognition for the effort involved in collecting and managing research data, Research Council funded work may be entitled to a limited period of privileged use of the data to publish the results of their research. The length of this period varies by research discipline and, where appropriate, the published policies of individual Research Councils.
- In order to recognise the intellectual contributions of researchers who generate, preserve and share research data should acknowledge the sources of their data and abide by the terms and conditions of their data.
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

<http://www.rcuk.ac.uk/research/datapolicy/>

EPSRC

Engineering and Physical Sciences
Research Council

Home FUNDING RESEARCH INNOVATION SKILLS NEWS, EVENTS AND PUBLICATIONS

About us

Service standards and policies

EPSRC policy framework on research data

Principles

Scope and benefits

Exploitation of research results and collaborative research

Impact, timescales and support

Responsibility for costs

EPSRC policy framework on research data

This policy framework sets out EPSRC's expectations concerning the management and provision of access to EPSRC-funded research data. EPSRC recognises that a range of institutional policies and practices can satisfy these expectations, and encourages research organisations to develop specific approaches which, while aligned with EPSRC's expectations, are appropriate to their own structures and cultures.

The expectations arise from seven core principles which align with the core RCUK principles on data sharing. Two of the principles are of particular importance: firstly, that publicly funded research data should generally be made as widely and freely available as possible in a timely and responsible manner; and, secondly, that the research process should not be damaged by the inappropriate release of such data.

The framework was endorsed by the EPSRC Council in March 2011 and implemented from 01 May 2011. It was developed with the benefit of advice from university administrators, from academics, and from research collaborators based in industry.

<https://www.epsrc.ac.uk/about/standards/researchdata/>

<http://www.rcuk.ac.uk/research/opendata/>

Concordat on Open Research Data

The Concordat on Open Research Data has been developed by a UK multi-stakeholder group. This concordat will help to ensure that the research data gathered and generated by members of the UK research community is made openly available for use by others wherever possible in a manner consistent with relevant legal, ethical, disciplinary and regulatory frameworks and norms, and with due regard to the costs involved.



Published 28th July 2016

Concordat on Open Research Data

The Concordat on Open Research Data has been developed by a UK multi-stakeholder group. This concordat will help to ensure that the research data gathered and generated by members of the UK research community is made openly available for use by others wherever possible in a manner consistent with relevant legal, ethical, disciplinary and regulatory frameworks and norms, and with due regard to the costs involved.



Published 28th July 2016

<http://www.rcuk.ac.uk/research/opendata/>

The UK is on course to make all taxpayer-funded research publications available in an open access format. Open research data is the next step in achieving the UK's open science ambitions. I see open access to research data as a fundamental good: combining research publications with their data will help drive transparency, improve co-operation and strengthen the UK's position as a global science leader.

The *Concordat*, for the first time, proposes a series of clear and practical principles for working with research data that cover the many roles needed to support the research process. It is not a rulebook, but a set of expectations of best practice developed by the research community itself.

This is not a Government owned document, nor should it be. The research community has worked hard to arrive at the consensus delivered in this report and I would like to thank the members of the UK Open Research Data Forum for their valuable contributions. I would also like to thank Professors Nick Wright, Rick Rylance and Duncan Wingham for their leadership.

Rt. Hon Jo Johnson MP

Minister of State for Universities and Science

<https://researchdata.jiscinvolve.org/wp/>

Jisc R&D Email Feedback

Research Data Discovery Service (Alpha)

Datasets Organisations About Reports FAQ

Home > Datasets

Filter by location Clear

Search datasets... Search

Order by Relevance

Advanced Search

14,628 datasets found

Hourly automated weather station (AWS) data from Climoor fieldsite in Clocaen...

This dataset contains hourly automated weather station (AWS) data from the Climoor field site in Clocaenog forest, NE Wales. It runs from 10/6/2008 until 31/12/2013, and...

NERC Biogeochemical Ocean Flux Study (BOFS) data in the Southern Ocean (1992)

Map data © OpenStreetMap contributors
Tiles by MapQuest

Organisations

- UK Data Service (6946)
- NERC (3507)
- University of Edinb... (1304)

Jisc R&D Email Feedback

Research Data Discovery Service (Alpha)

<https://researchdata.jiscinvolve.org/wp/>

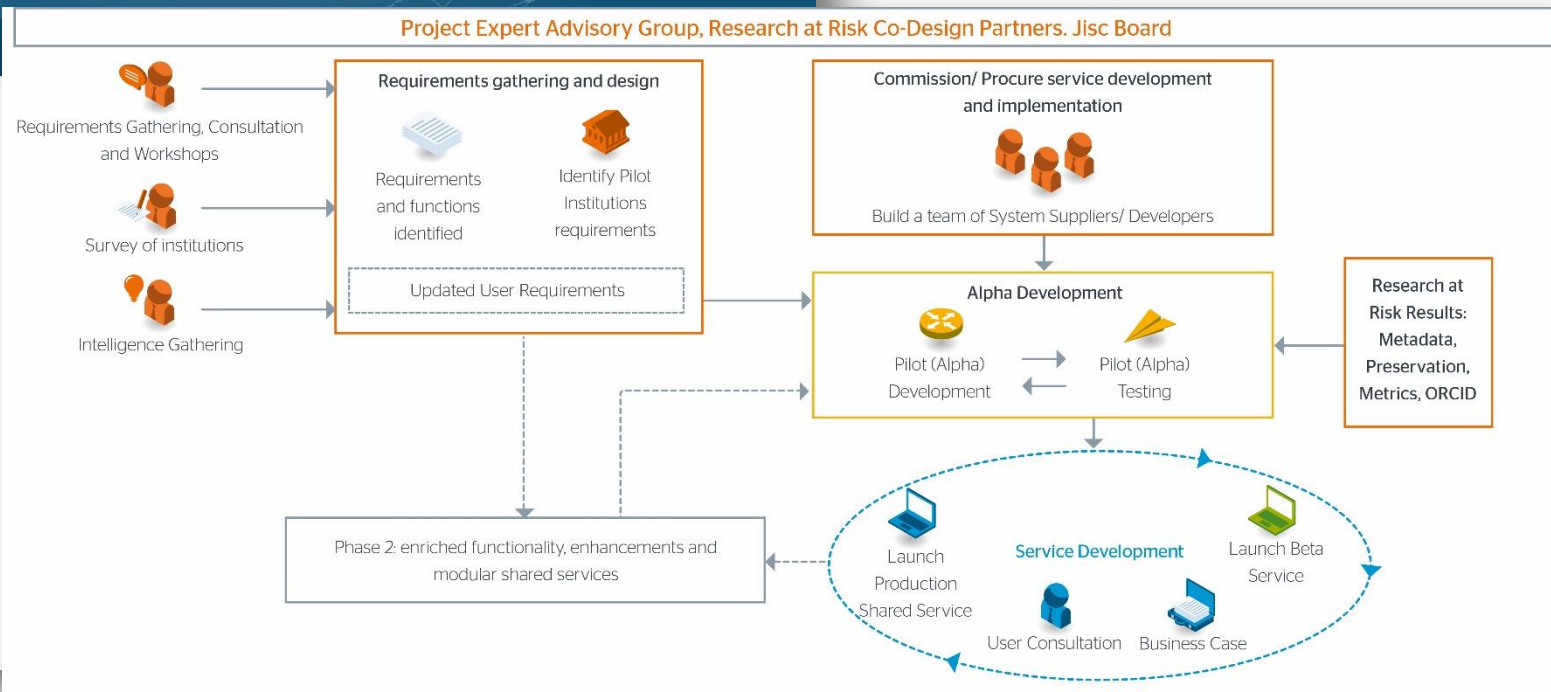
Home > Datasets

Filter by location Clear

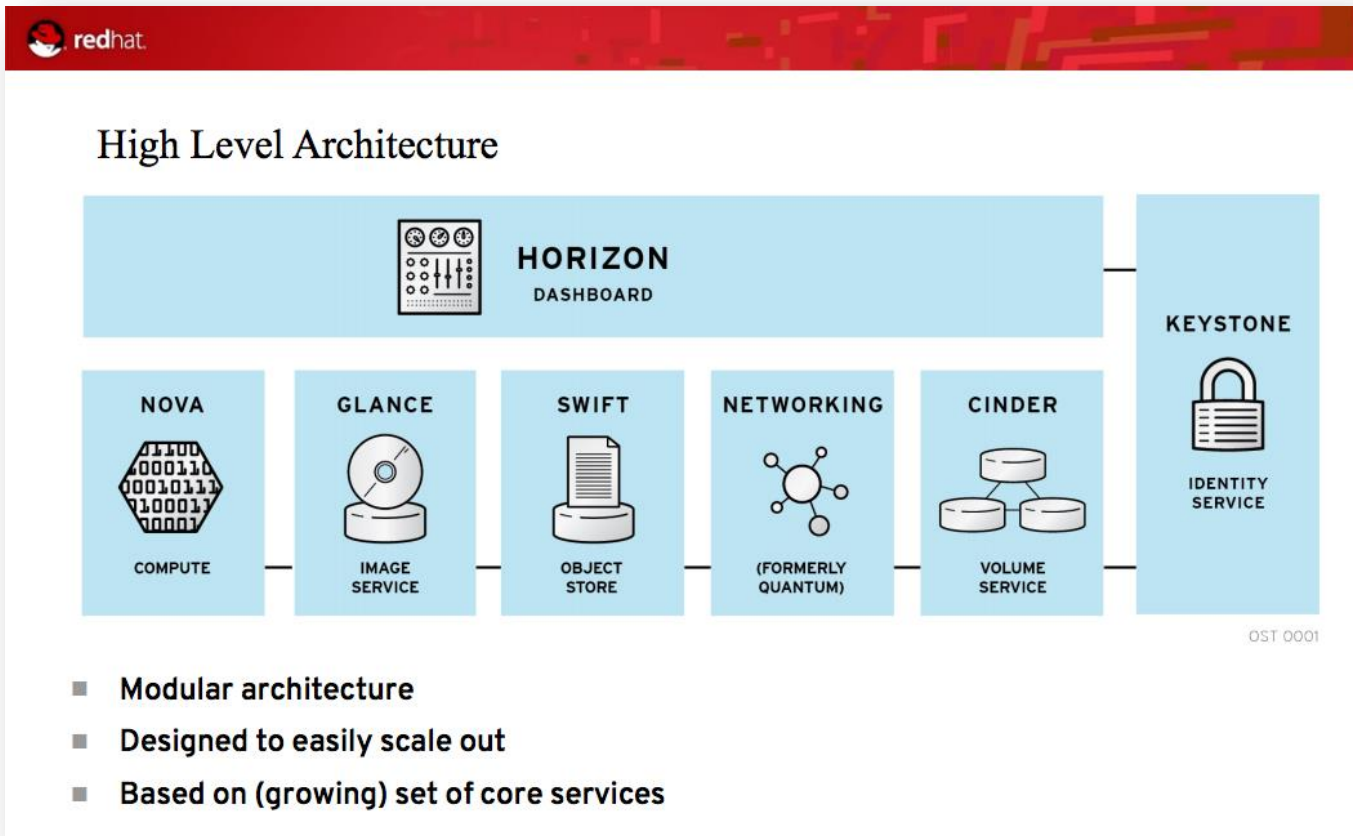
Map data © OpenStreetMap contributors
Tiles by MapQuest

Organisations

- UK Data Service (6946)
- NERC (3507)
- University of Edinb... (1304)



2. How do cloud technologies help?



BIG DATA, CLOUD COMPUTING, EUROPE, FEATURED

CERN's OpenStack Cloud to Reach 150,000 Cores by 2015

BY YEVGENIY SVERDLIK ON NOVEMBER 7, 2014

ADD YOUR COMMENTS



The team overseen by Tim Bell, infrastructure manager at CERN (pictured), is preparing to upgrade its OpenStack cloud from 115,000 to 150,000 compute cores to handle an increase in data that will be generated by the Large Hadron Collider when its energy is doubled in 2015.

<http://www.datacenterknowledge.com/archives/2014/11/07/cerns-openstack-cloud-to-reach-150000-cores-by-2015/>

The amount of data that CERN has to accommodate as they design their new center is virtually unfathomable. As the researchers collide particles into each other, 7,000 ton instruments that can be thought of as 100 megapixel digital cameras the size of the Notre Dame Cathedral capture 40 million pictures a second. That's a petabyte of data generated every second that needs to be processed and analyzed. Bell says that while they do have server farms to knock those numbers down to reasonable levels, in the end, they're still dealing with 35 petabytes a year that they have to record, and with the upgrades coming, they're expecting that to double.

http://www.datanami.com/2013/09/23/cern_turns_to_google_for_datacenter_direction/

<http://www.datacenterknowledge.com/archives/2014/11/07/cerns-openstack-cloud-to-reach-150000-cores-by-2015/>

BIG DATA, CLOUD COMPUTING, EUROPE, FEATURED

CERN's OpenStack Cloud to Reach 150,000 Cores by 2015

BY YEVGENIY SVERDLIK ON NOVEMBER 7, 2014

[ADD YOUR COMMENTS](#)



The team overseen by Tim Bell, infrastructure manager at CERN (pictured), is preparing to upgrade its OpenStack cloud from 115,000 to 150,000 compute cores to handle an increase in data that will be generated by the Large Hadron Collider when its energy is doubled in 2015.

VM Depot
by Microsoft Open Technologies

BROWSE PUBLISH SHARE MY ACCOUNT HELP

Search for your virtual machine image...

What

A community managed repository of Linux and FreeBSD virtual machine images for easy deployment to Microsoft Azure.

Why

Discovering and running open source software on Microsoft Azure makes sense, VM Depot makes it easy.

How

Publishers can share Virtual Machine Images for free. Users discover and deploy to Microsoft Azure with a Free Trial.

CiviCRM 4.6.4-0 (Ubuntu 14.04)

CiviCRM powered by Bitnami is a pre-configured, ready to run image for running CiviCRM on Microsoft Azure. CiviCRM is an open source and freely downloadable constituent relationship management solution. CiviCRM is web-based, open source, internationalized, and designed specifically to meet the needs of advocacy, non-profit and non-governmental groups. Bitnami has integrated CiviCRM with the Drupal content management ...

Publisher: Bitnami **Platform:** ubuntu

Create Virtual Machine Deployment Script Deployment Tutorial

Sort By:

- Featured
- Date Added
- Name
- Platform
- Rating

aws marketplace Amazon Web Services Home

Sign in or [Create a new account](#) [Your Account](#) | [Help](#) | [Sell on AWS Marketplace](#)

Shop All Categories ▾ [Your Software](#)

Categories

[All Categories](#)

[Business Software](#)

[High Performance Computing](#)

Filters

Operating System

All Linux/Unix

Software Pricing Plans

Free (8)

Hourly (24)

Monthly (5)

Annual (6)

Bring Your Own License (2)

Software Free Trial

Free Trial (9)

Delivery Method

Amazon Machine Image (39)

SaaS (9)

Average Rating

★★★★★ & up (3)

★★★★☆ & up (3)

★★★★☆ & up (4)


★★★★☆ & up (6)

Architecture

64-bit (39)

Business Software >

High Performance Computing (48 results) showing 1 - 10 1 2 3 4 5 ▶

 **ADAPA Decision Engine**


★★★★★ (1) | Version 4.1.2 | Sold by [Zementis](#)

Free Trial

\$0.99 to \$7.99/hr for software + AWS usage fees

ADAPA is a predictive analytics decision engine based on the PMML (Predictive Model Markup Language) standard. With ADAPA, deploy one or many predictive models from data ...

Linux/Unix, Ubuntu 14.04 x64 | 64-bit Amazon Machine Image (AMI)


 **Bayesian Machine Learning on the Limit Order Book**

★★★★★ (1) | Sold by [BMLL Technologies](#)

From \$10 per user per hour + (AWS and data and 3rd party software charges.)

Access to the historical limit order book for all global liquid securities, at full L2 or L3 resolution. This AMI contains the interface to this AWS hosted 10PB dataset, ...

Software as a Service


 **Queue-it**

★★★★★ (2) | Sold by [Queue-it ApS](#)

Free trial and several subscription packages available

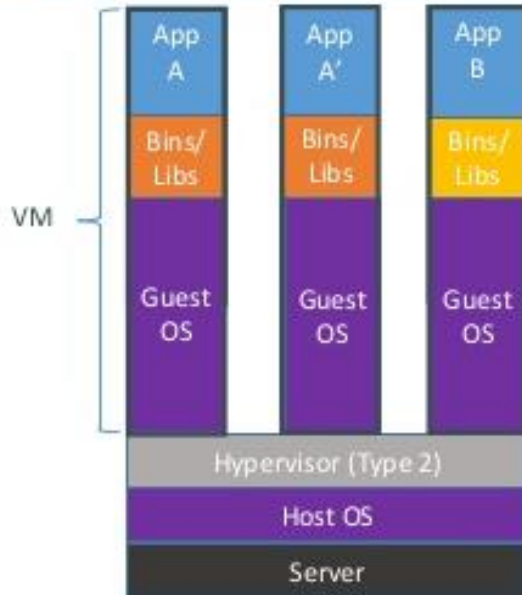
Queue-it is an online queueing system for managing website overload during extreme user peaks in relation to popular events and activities online. The solution maintains ...

Software as a Service

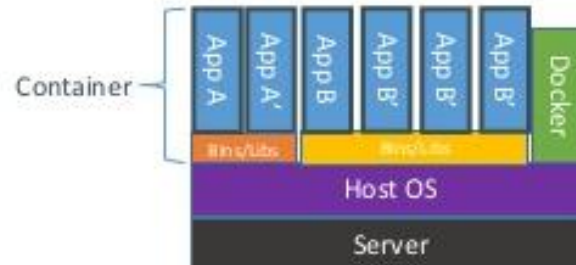
 **CoreOS Linux**

★★★★★ (2) | Version CoreOS | Sold by [CoreOS](#)

Containers vs. VMs



Containers are isolated, but share OS and, where appropriate, bins/libraries



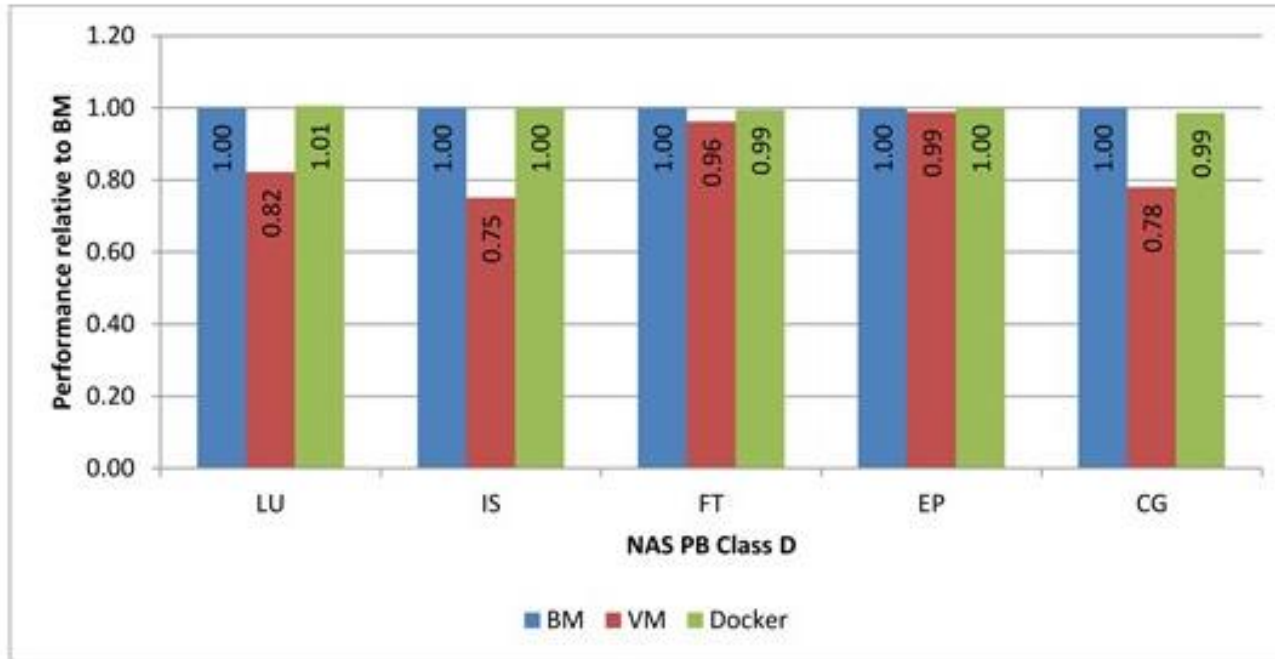


Figure 4 Performance of NPB on containers and VMs relative to BMs

OPEN CONTAINER INITIATIVE

ABOUT COMMUNITY FAQ JOIN NEWS CONTACT US

The Open Container Initiative

is an open governance structure for the express purpose of creating open industry standards around container formats and runtime.

Supporting Companies

amazon web services | anchore | APCERA | apprenda | aqua | at&t

CISCO | ClusterHQ | ContainerShip | CoreOS | DATERA | DELL

docker | EasyStack | EMC² | facebook | FUJITSU | Goldman Sachs

Google | Hewlett Packard Enterprise | HUAWEI | IBM | Infoblox | intel

Joyent | Kyup | MESOSPHERE | Microsoft | midokura | NUTANIX

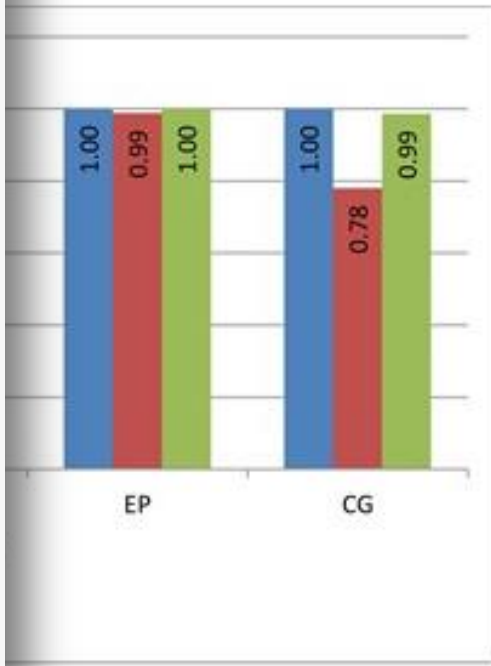
ORACLE | Pivotal | POLYVERSE | portworx | RANCHER | redhat

REPLICATED | resin.io | ROBIN | SUSE | sysdig | Twistlock

Twitter | UNIVA | Verizon Labs | Virtuozzo | vmware | weaveworks

wercker | Western Digital

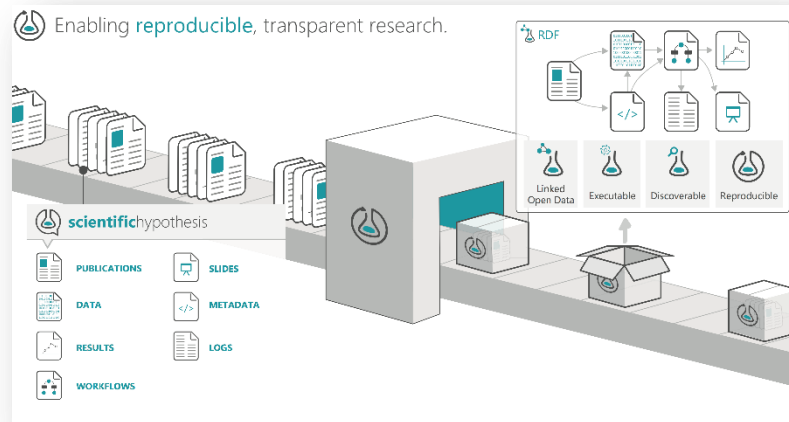
<https://www.opencontainers.org/>



3. A new paradigm?

Next steps in open science?

- › Identifiers – from Researcher IDs (ORCID) to Digital Object Identifiers (DOIs) + organisation IDs
- › Equipment / calibrations used in an experiment – potentially building on the national equipment.data.ac.uk database
- › Moving “beyond the PDF” as canonical research output to individually addressable components – tables, figures, underlying data and code
- › Making it easier to share whole workflows / pipelines
- › Bringing in expert Research Software Engineers to work with the researcher – see <http://www.rse.ac.uk>



<http://www.researchobject.org/>

Martin Hamilton
Futurist, Jisc, London

[@martin_hamilton](#)
martin.hamilton@jisc.ac.uk



Except where otherwise noted, this work is licensed under CC-BY