HPC Cluster Computing from 64 to 156,000 cores

Jason Stowe, Founder

CYCLECOMPUTING

You'll learn about:

How continuous/instant delivery helps invention/discovery

CYCLECOMPUTING

You'll learn about:

Climate, Organic Semiconductors,
Lean manufacturing
Technical Computing

About Cycle Computing

We believe utility access to technical computing accelerates discovery & invention

Our mission:

Easy HPC applications
for anyone,
on any resources,
at any scale

About real use cases:

NOVARTIS

Life Sciences

- **39.5 years of drug compound computations in 9 hours**, at a total cost of **$4,372**
- **10,000 server cluster** seamlessly spanned regions
- Advanced 3 new otherwise unknown compounds in wetlab

AEROSPACE
Assuring Space Mission Success

Rocket Design & Simulation

- Moving HPC workloads to cloud for burst capacity
- Amazon GovCloud, Security, and key management
- Up to 1000s of cores for burst / agility

SCHRÖDINGER.

Life Sciences

- **50,000-core utility supercomputer** in the Cloud
- Analyzed 21 million compounds in 3 hours
- **12.5 processor years for < $15,000**

PACIFIC LIFE

Asset & Liability Modeling (Insurance / Risk Mgmt)

- Completes monthly/quarterly runs 5-10x faster
- Use 1600 cores in the cloud to shorten elapsed run time from ~ 10 days to ~ 1-2 days

Nuclear engineering **Utilities / Energy**

- Approximately 600-core MPI workloads run in Cloud
- Ran workloads in months rather than years
- Introduction to production in 3 weeks

Johnson & Johnson

Manufacturing & Design

- Enable end user on-demand access to 1000s of cores
- Avoid cost of buying new servers
- Accelerated science and CAD/CAM process

About me:

I’m a geek from the 90s

90s geek??

Whilst I like hacker news, reddit, etc.

Slashdot is my homepage

UN Report: Climate Changes Overwhelming

Posted by **samzenpus** on Monday March 31, 2014 @03:23PM
from the it's-getting-hot-in-here dept.

iONiUM (530420) writes

> "'The impacts of global warming are likely to be "severe, pervasive and irreversible", a major report by the UN has warned.' A document was released by the IPCC outlining the current affects on climate change, and they are not good. For specific effects on humans: 'Food security is highlighted as an area of significant concern. Crop yields for maize, rice and wheat are all hit in the period up to 2050, with around a tenth of projections showing losses over 25%.'"

Read the **987** comments

x obvious x science x climate x earth x un story

CYCLECOMPUTING

Study Rules Out Global Warming Being a Natural Fluctuation With 99% Certainty

Posted by **Soulskill** on Saturday April 12, 2014 @11:39AM
from the let's-blame-the-dinosaurs dept.

An anonymous reader writes

> "A study out of McGill University sought to examine historical temperature data going back 500 years in order to determine the likelihood that global warming was caused by natural fluctuations in the earth's climate. The study concluded there was less than a 1% chance the warming could be attributed to simple fluctuations. 'The climate reconstructions take into account a variety of gauges found in nature, such as tree rings, ice cores, and lake sediments. And the fluctuation-analysis techniques make it possible to understand the temperature variations over wide ranges of time scales. For the industrial era, Lovejoy's analysis uses carbon-dioxide from the burning of fossil fuels as a proxy for all man-made climate influences – a simplification justified by the tight relationship between global economic activity and the emission of greenhouse gases and particulate pollution, he says. ... His study [also] predicts, with 95% confidence, that a doubling of carbon-dioxide levels in the atmosphere would cause the climate to warm by between 2.5 and 4.2 degrees Celsius. That range is more precise than – but in line with — the IPCC's prediction that temperatures would rise by 1.5 to 4.5 degrees Celsius if CO2 concentrations double.'"

Read the **849** comments

x news x statistics x collapse x cuethedeniers x climatechange story

CYCLECOMPUTING

BBC News | Sport | Weather | Capital | Future | Shop

NEWS SCIENCE & ENVIRONMENT

Home | US & Canada | Latin America | UK | Africa | Asia | Europe | Mid-East | Business | Health | Sci/Environm

31 March 2014 Last updated at 07:35 ET

Share

Climate impacts 'overwhelming' - UN

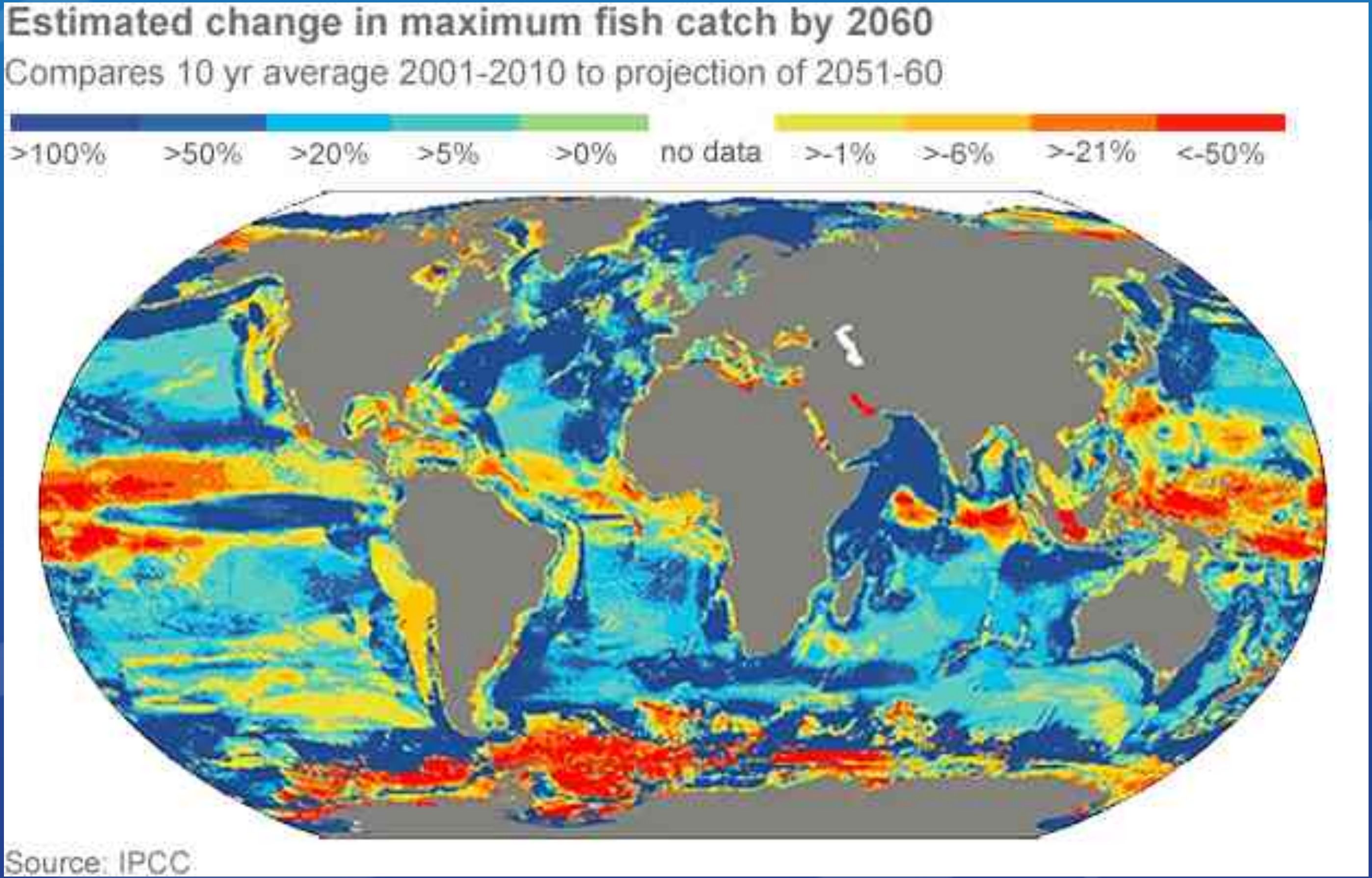
COMMENTS (1872)

By Matt McGrath
Environment correspondent, BBC News, Yokohama, Japan

It will affect everyone

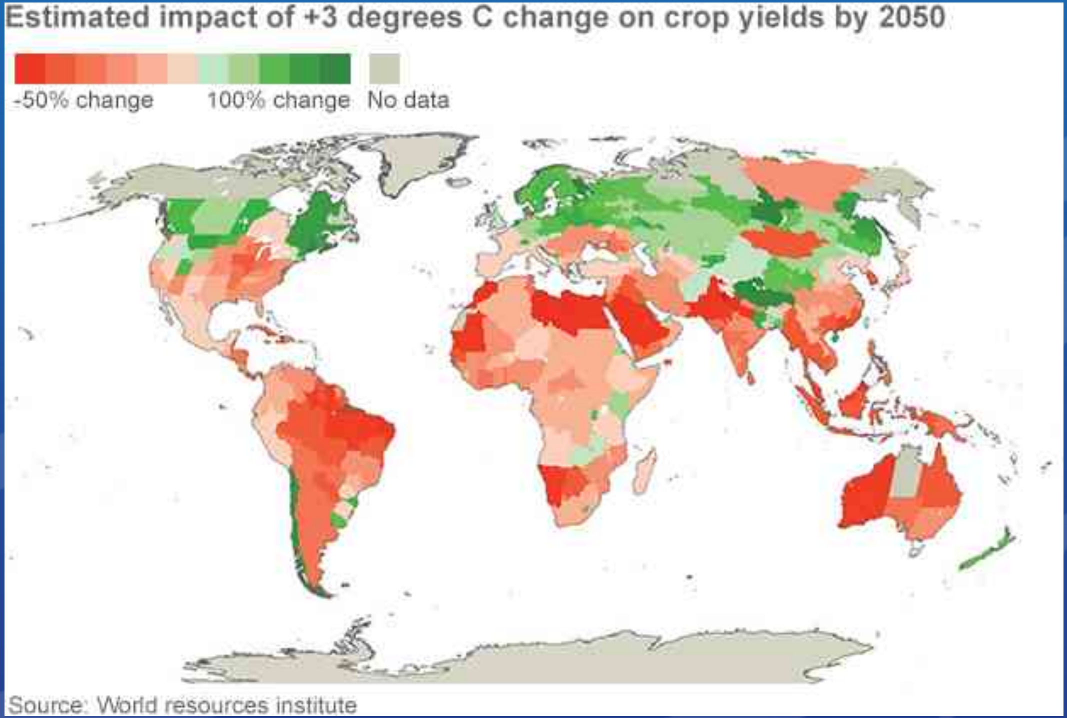
CYCLECOMPUTING

CYCLECOMPUTING



CYCLECOMPUTING

What do you think?



CYCLECOMPUTING

Buy land in Canada?

CYCLECOMPUTING

Maybe, but hopefully you're getting:
this is a big deal

Just in case…

Chairman Rajendra Pachauri

Intergovernmental Panel on Climate Change

"Nobody on this planet is going to be untouched by the impacts of climate change"

US Secretary of State John Kerry:

"Unless we act dramatically and quickly, science tells us our climate and our way of life are literally in jeopardy.

Denial of the science is malpractice."

Climate Scientist: Climate Engineering Might Be the Answer To Warming

Posted by **samzenpus** on Monday April 14, 2014 @05:27PM
from the warm-up-the-cloud-gun dept.

Lasrick (2629253) writes

> "Tom Wigley is one of the world's top climate scientists, and in this interview he explains his outspoken support for both nuclear energy and research into climate engineering. Wigley was one of the first scientists to break the taboo on public discussion of climate engineering as a possible response to global warming; in a 2006 paper in the journal Science, he proposed a combined geoengineering-mitigation strategy that would address the problem of increasing ocean acidity, as well as the problem of climate change. In this interview, he argues that renewable energy alone will not be sufficient to address the climate challenge, because it cannot be scaled up quickly and cheaply enough, and that opposition to nuclear power 'threatens humanity's ability to avoid dangerous climate change.'"

Read the **338** comments

x science x earth x technocopian x whatcouldpossiblygowrong story

UN: Renewables, Nuclear Must Triple To Save Climate

Posted by **samzenpus** on Sunday April 13, 2014 @05:15PM
from the its-getting-hot-in-here dept.

An anonymous reader writes

> "On the heels of a study that concluded there was less than a 1% chance that current global warming could be simple fluctuations, U.N. scientists say energy from renewables, nuclear reactors and power plants that use emissions-capture technology needs to triple in order keep climate change within safe limits. From The Washington Post: 'During a news conference Sunday, another co-chair, Rajendra K. Pachauri of India, said the goal of limiting a rise in global temperatures "cannot be achieved without cooperation." He added, "What comes out very clearly from this report is that the high-speed mitigation train needs to leave the station soon, and all of global society needs to get on board."'"

Read the **426** comments

x science x nuclear x renewables x earth x wealthtransfer story

Climate Engineering?

CYCLECOMPUTING

Solar Energy

Wind power

Nuclear Fission energy

GeoThermal

Possibilities?

Climate Engineering

Nuclear Fusion

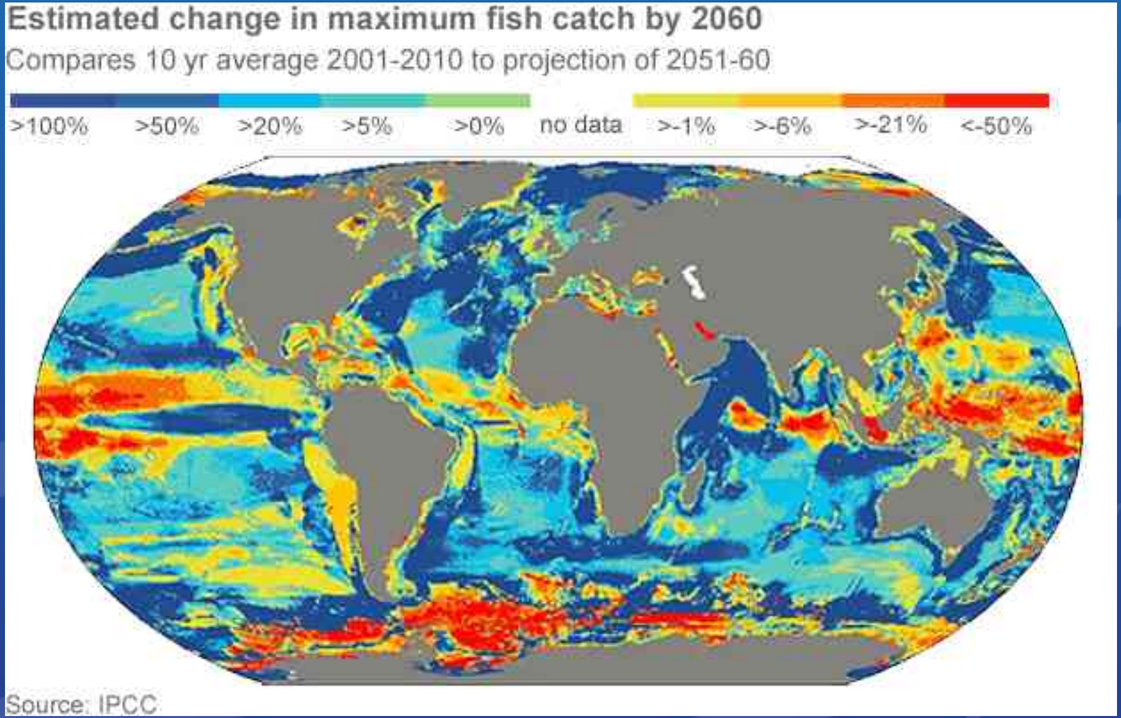
BioFuels

CYCLECOMPUTING

How do we simulate/engineer a solution?

CYCLECOMPUTING

How do they create these?



CYCLECOMPUTING

They use a supercomputer…

CYCLECOMPUTING

Supercomputers = great

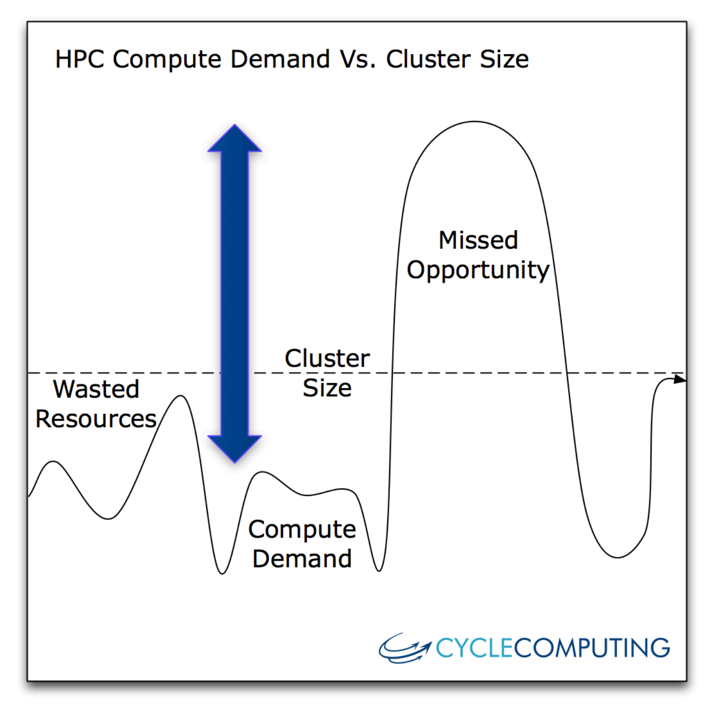
- Supercomputers have special features, like fast interconnects
- So good they should never run R, genomics, monte carlo, etc. that don't use *Fast Interconnect*
- During peak usage, our users have wanted:
  - Latency sensitive
  - Large scale/grid MPI

  Running on bare metal, put other jobs elsewhere

Flickr: LouisRix

CYCLECOMPUTING

Limitations of fixed infrastructure



Too small when needed most,
Too large every other time…

- Upfront CapEx anchors you to aging servers
- Costly administration
- Miss opportunities to do better risk management, product design, science, engineering

CYCLECOMPUTING

Lets talk about materials science

CYCLECOMPUTING

Specifically…

Organic semiconductor

From Wikipedia, the free encyclopedia

This section **does not cite any references or sources**. Please help improve this section by adding citations to reliable sources. Unsourced material may be challenged and removed. *(July 2013)*

STM image of self-assembled supramolecular chains of the organic semiconductor Quinacridone on Graphite.

An **organic semiconductor** is an organic material with semiconductor properties, that is, with an electrical conductivity between that of insulators and that of metals. Single molecules, oligomers, and organic polymers can be semiconductive. Semiconducting small molecules (aromatic hydrocarbons) include the polycyclic aromatic compounds pentacene, anthracene, and rubrene.[*not verified in body*] Polymeric organic semiconductors include poly(3-hexylthiophene), poly(p-phenylene vinylene), as well as polyacetylene and its derivatives.[*not verified in body*]

There are two major overlapping classes of organic semiconductors. These are organic charge-transfer complexes and various linear-backbone conductive polymers derived from polyacetylene. Linear backbone organic semiconductors include polyacetylene itself and its derivatives polypyrrole, and polyaniline.[*not verified in body*] At least locally, charge-transfer complexes often exhibit similar conduction mechanisms to inorganic semiconductors. Such mechanisms arise from the presence of hole and electron conduction layers separated by a band gap.[*not verified in body*] Although such classic mechanisms are important locally, as with inorganic amorphous semiconductors, tunnelling, localized states, mobility gaps, and phonon-assisted hopping also significantly contribute to conduction, particularly in polyacetylenes.[*not verified in body*] Like inorganic semiconductors, organic semiconductors can be doped. Organic semiconductors susceptible to doping such as polyaniline (Ormecon) and PEDOT:PSS are also known as *organic metals*.[*not verified in body*]

Designing Better Solar Materials

The challenge is efficiency - turning photons to electricity

The number of possible materials is limitless:

- Need to separate the right compounds from the useless ones
- Researcher Mark Thompson, PhD:

  "If the 20th century was the century of silicon, the 21st will be all organic. Problem is, how do we find the right material without spending the entire 21st century looking for it?"

Needle in a Haystack Challenge:

205,000 compound families totaling 2,312,959 core-hours or 264 core-years

Supercomputers = Fast & costly

Flickr: LouisRix

CYCLECOMPUTING

Another tool in the box: Cloud

- At peak load, use up to massive scale to run throughput workloads on cloud:
  - Needle in a Haystack
  - Whole Sample set
  - BigData/Throughput i/o
  - Integration, monte-carlo paths
  - Interactive, Service oriented workloads (master worker)
- Workload run-time rather than individual job run-time

Flickr: MattDavis

CYCLECOMPUTING

Most of Cycle's users use
40 - 4000 cores,

Mark Thompson's problem
wasn't that problem

CYCLECOMPUTING

Needle in a Haystack Challenge:

205,000 compounds totaling 2,312,959 core-hours or 264 core-years

CYCLECOMPUTING

205,000 molecules
264 years of computing

16,788 Spot Instances,
156,314 cores

8-Region Deployment

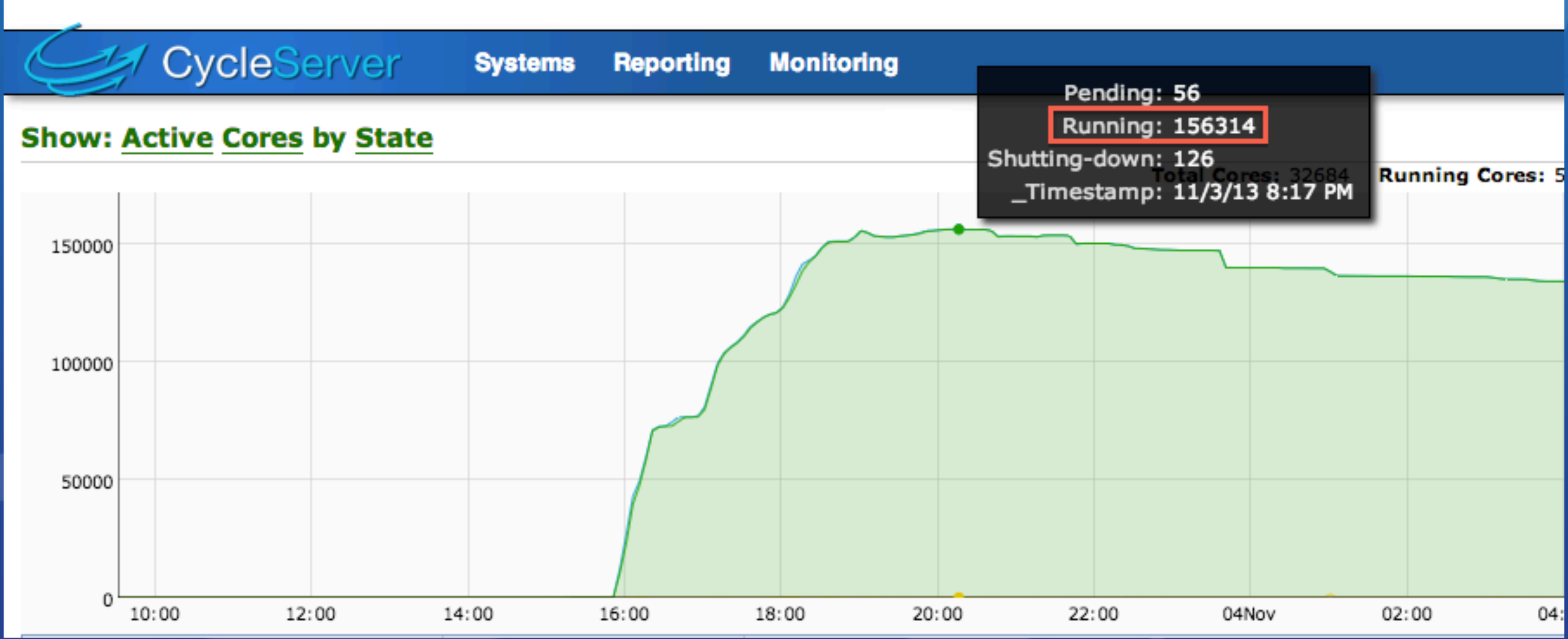
CYCLECOMPUTING

205,000 molecules
264 years of computing

156,314 cores =
1.21 PetaFLOPS (~Rpeak)

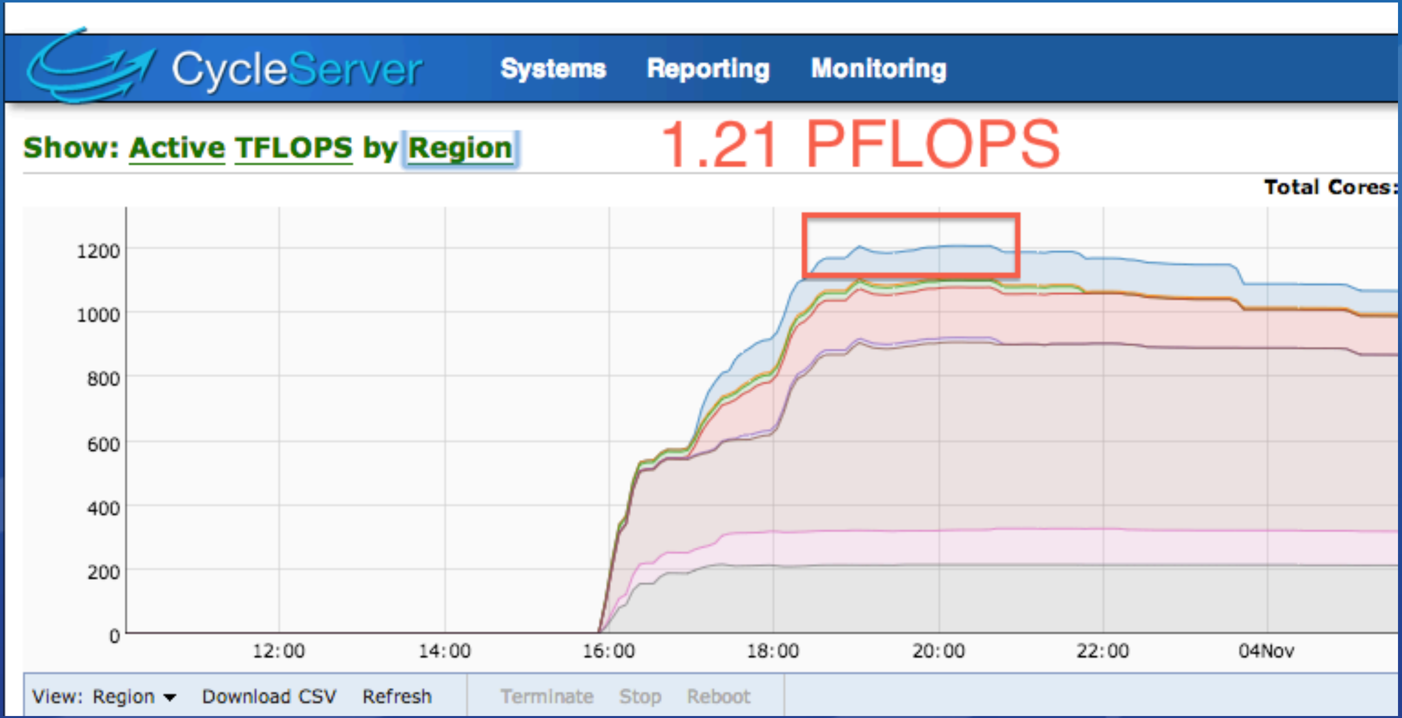
CYCLECOMPUTING

1.21 PetaFLOPS (Rmax+Rpeak)/2, 156,314 cores



CYCLECOMPUTING

Benchmark individual machines



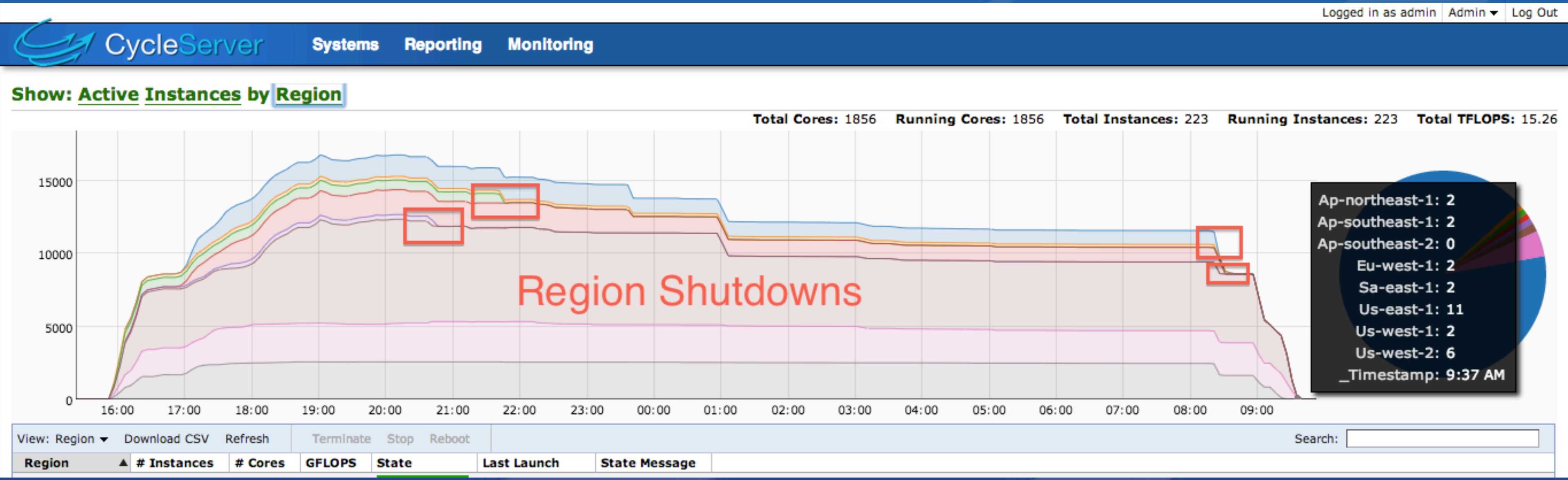
CYCLECOMPUTING

205,000 molecules
264 years of computing

Done in 18 hours
Access to $68M system
for $33k

CYCLECOMPUTING

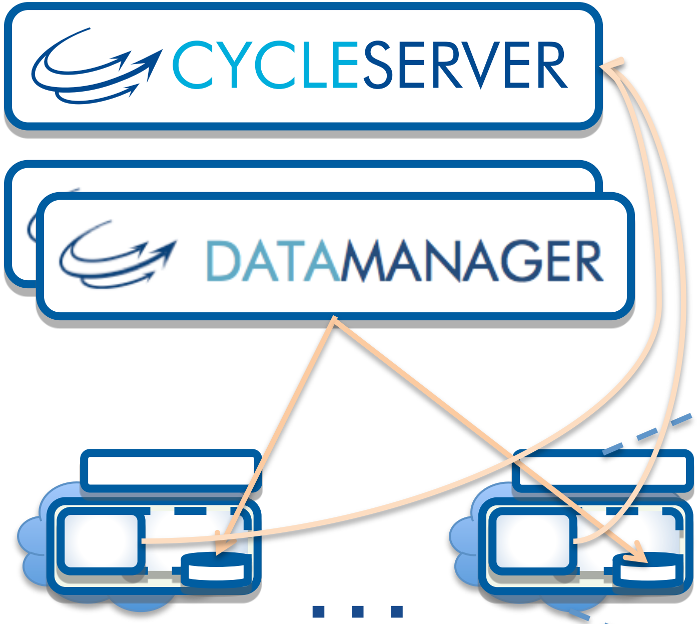
Resilient Workload Scheduling



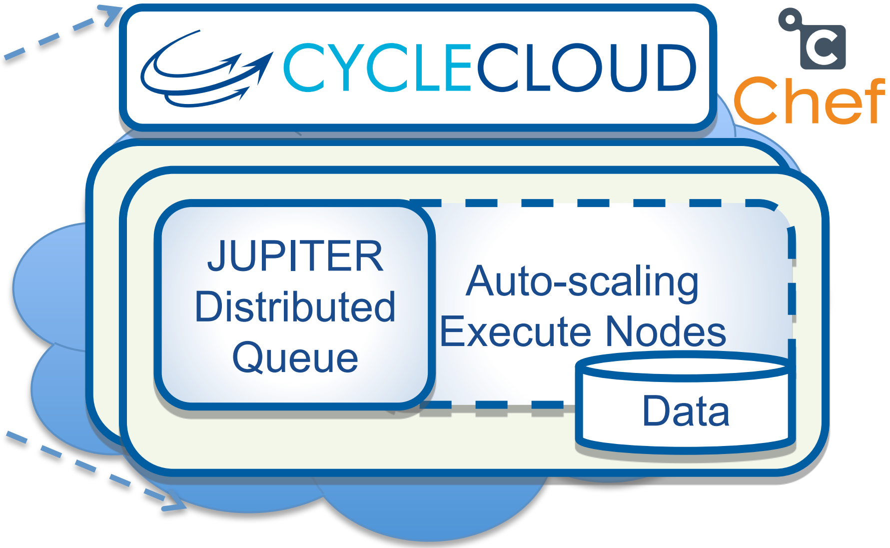
CYCLECOMPUTING

How did we do this?

Automated in 8 Cloud Regions, 5 continents, double resiliency



14 nodes controlling 16,788



Continuous Software Deployment/Configuration

Chef

Now Dr. Mark Thompson

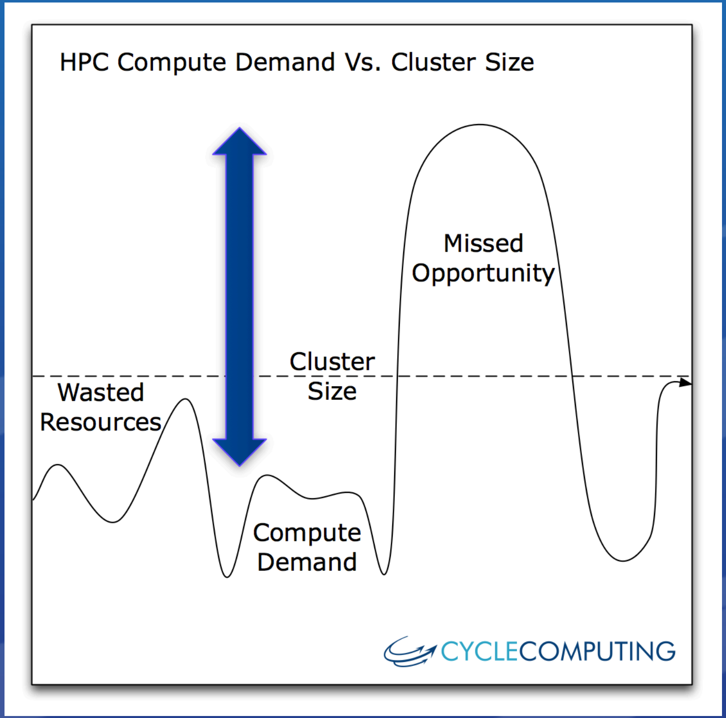
Is 264 compute years closer to making efficient solar a reality using organic semiconductors…

CYCLECOMPUTING

Important?

CYCLECOMPUTING

Solving this problem…



CYCLECOMPUTING

History repeats itself

Lean Manufacturing (Japanese Manufacturing to turn raw materials to inventory)

Continuous Delivery
(Modern software process to turn builds to production applications)

Users want to decrease
Cycle time
=
(prep time + queue time +
run time)

CYCLECOMPUTING

The Scientific Method

Ask a Question → Hypothesiz
e → Predict → Experiment / Test → Analyze → Final Results

Test and Analyze stages require the most time, compute, and data

What’s holding us (Engineering, Science, Risk) back?

Utility HPC Access = Continuous HPC capacity delivery

CYCLECLOUD

Chef

CYCLECOMPUTING

The Innovation Bottleneck:

Scientists/Engineers forced to size expectations to the infrastructure they have access to

CYCLECOMPUTING

How does cloud help?

CYCLECOMPUTING

Put the important,
fast interconnect jobs on
fast interconnect machines

Put everything else on
cloud capacity, when needed

Workloads are actually helping

Fixed Cluster	Serial or MPI job (may need 1 core or many servers)	Serial or MPI job	Serial or MPI job

Workload can benefit from dynamic cluster

- Serial or MPI job (may need 1 core or many servers)
- Serial or MPI job (may need 1 core or many servers)
- Serial or MPI job (may need 1 core or many servers)

For these workloads:
$$$/science and throughput

****are as important as****

individual task performance

RUN TIME

CYCLECOMPUTING

Utility Access is Revolutionizing clients

NOVARTIS **Life Sciences**

- **39.5 years of drug compound computations in 9 hours**, at a total cost of **$4,372**
- **10,000 server cluster** seamlessly spanned US/EU regions
- Advanced 3 new otherwise unknown compounds in wet lab

AEROSPACE Assuring Space Mission Success **Rocket Design & Simulation**

- Moving HPC workloads to cloud for burst capacity
- Amazon GovCloud, Security, and key management
- Up to 1000s of cores for burst / agility

SCHRÖDINGER. **Life Sciences**

- **156,000-core utility supercomputer** in the Cloud
- Used $68M in servers for 18 hours for $33,000
- Simulated 205,000 materials **(264 years) in 18 hours**

PACIFIC LIFE **Asset & Liability Modeling (Insurance / Risk Mgmt)**

- Completes monthly/quarterly runs 5-10x faster
- Use 1600 cores in the cloud to shorten elapsed run time from ~10 days to ~ 1-2 days

Nuclear engineering **Utilities / Energy**

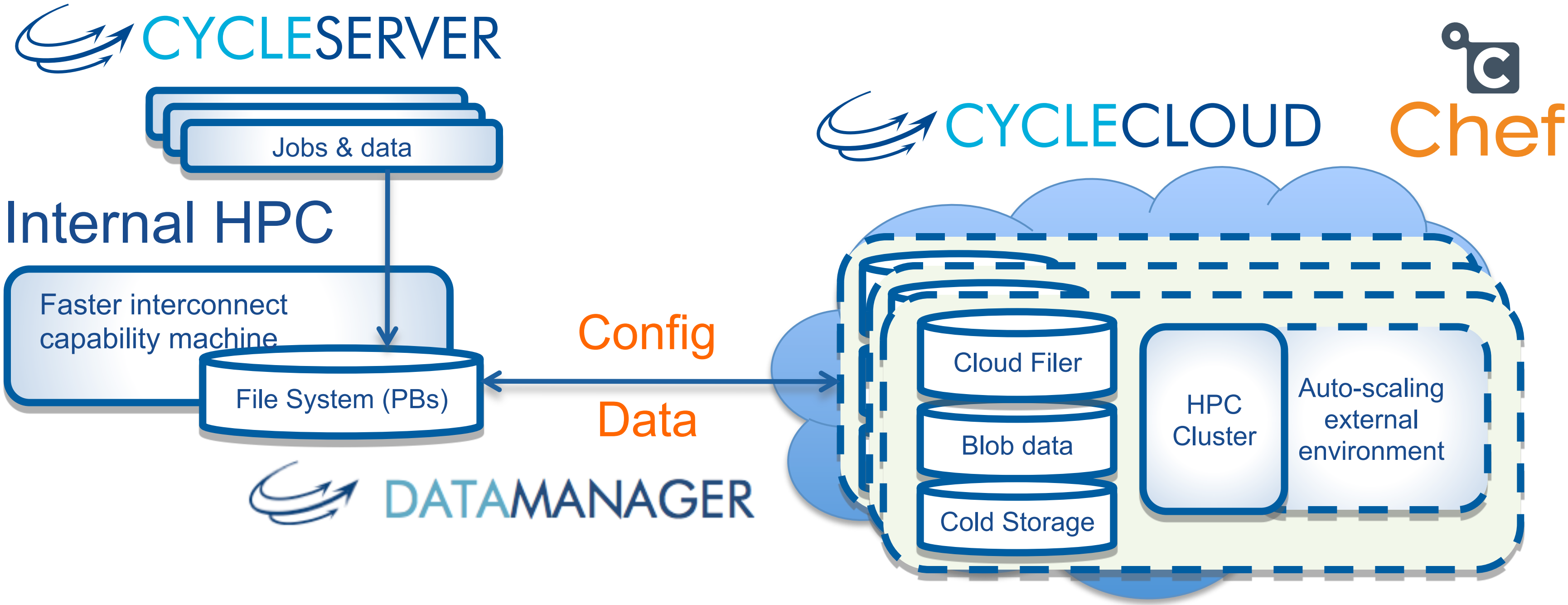
- Approximately 600-core MPI workloads run in Cloud
- Ran workloads in months rather than years
- <u>Introduction to production in 3 weeks</u>

Johnson & Johnson **Manufacturing & Design**

- Enable end user on-demand access to 1000s of cores
- Avoid cost of buying new servers
- Accelerated science and CAD/CAM process

Cycle HPC Deployment Architecture

Internal & External



Example Use Cases

Cycle software tools securely orchestrate scientific, engineering, and finance workloads across AWS

10,600 instance

Top 10 Pharma

Molecular Modeling for Cancer Targets

160 core

MPI/Engineering

Nuclear Reactor, Hard Disk, & Rocket

1 Million hours

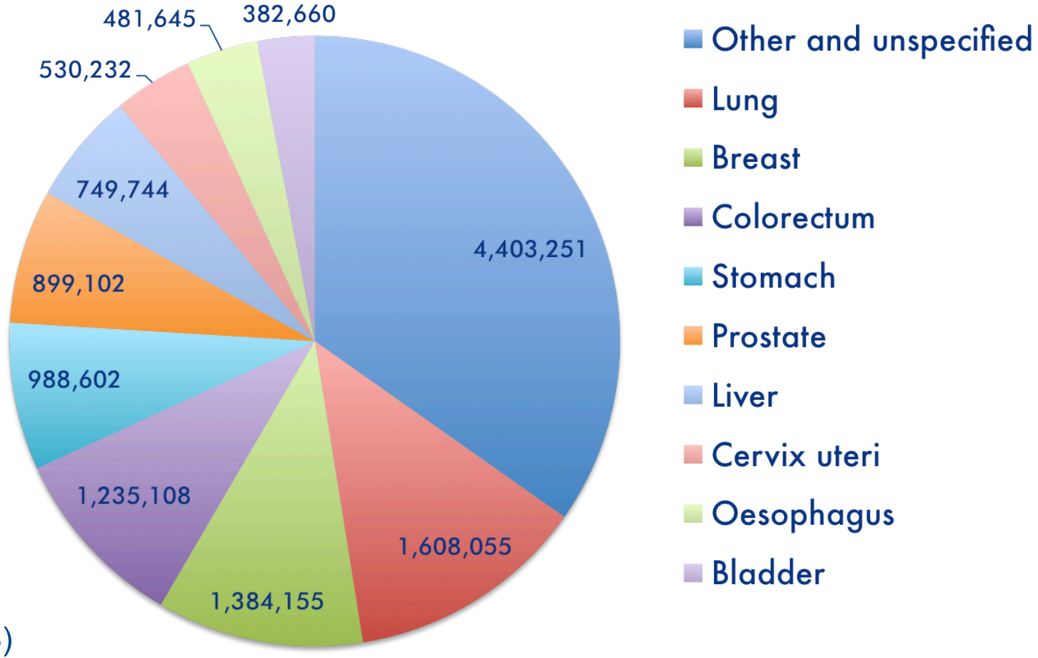
RNA Analysis

Genomics/Next-Gen Sequencing

Survey of Use Cases

- ☑ **Drug Design**
- ☑ CAE
- ☑ Genomics

New Cases of Cancer per Year: 12.66 Million

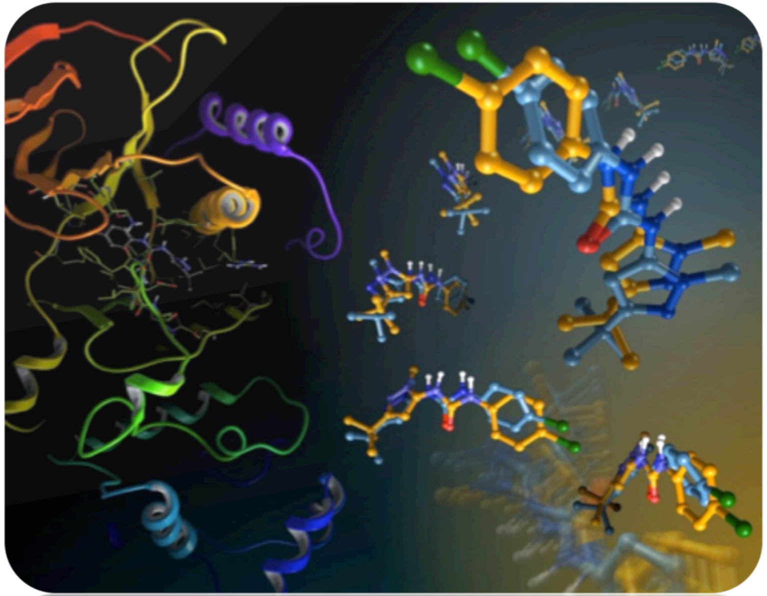


(W.H.O./Globocan 2008)

CYCLECOMPUTING

Every day is crucial and costly

Challenge: Novartis run 341,700 hours of docking against a cancer target (impossible to do in-house)



Novartis Instance Benchmarking

The Cloud: Flexible Science on Flexible Infrastructure

InstanceType	Count	Avg Run Time	Avg Spot Price/1K Jobs ▲	Avg On Demand Price/1K Jobs
cc2.8xlarge	5120	32s	$0.08	$0.67
c1.xlarge	1280	54s	$0.13	$1.09
c1.medium	560	55s	$0.14	$1.11
m1.xlarge	640	55s	$0.20	$1.83
m1.medium	120	57s	$0.20	$1.89
m3.2xlarge	1280	57s	$0.23	$1.97
m3.xlarge	640	58s	$0.23	$2.02
m1.small	91	1m 59s	$0.23	$1.99
m1.large	320	1m 7s	$0.24	$2.25

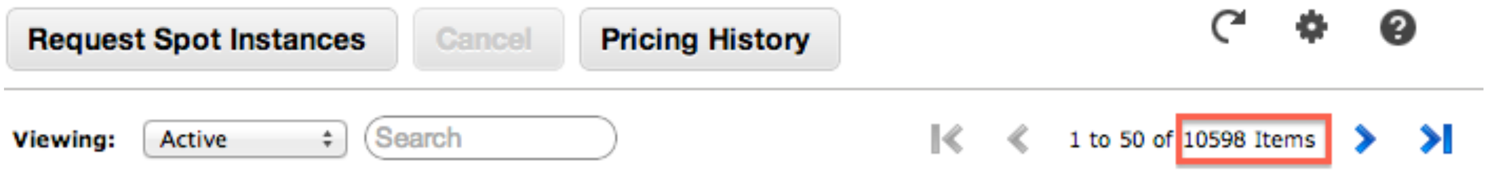
Engineering the right infrastructure for a workload:

- Software runs the same job many times across instance types
- Measures the throughput and determines the $ per job
- Use the instances that provide the best scientific ROI
- CC2 instance (Intel Xeon® 'Sandy Bridge') ran best for this

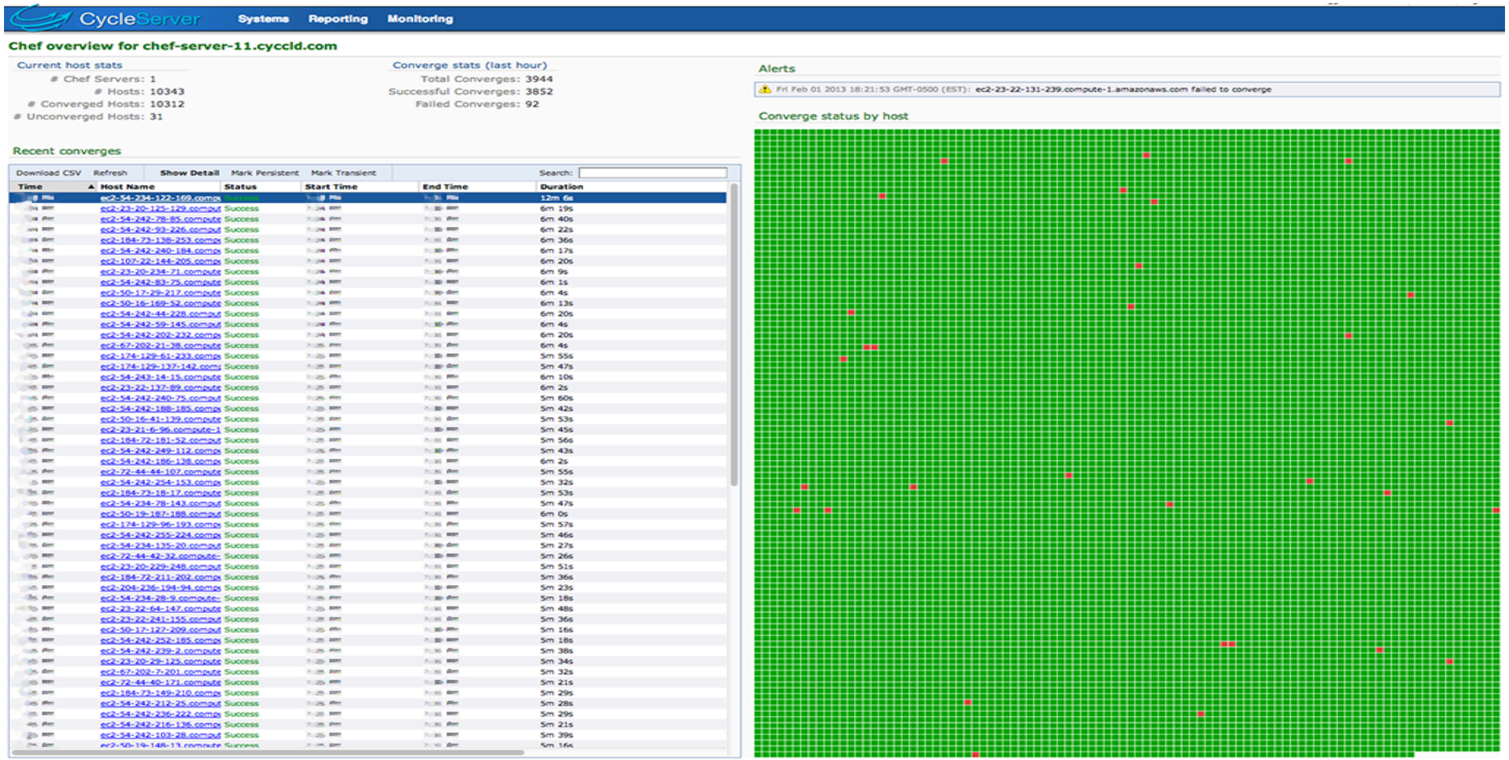
Utility technical computing server count:

```
$ knife node list | wc -l
   10599
$ 
```

AWS Console view:



Cycle Computing's cluster view:



CYCLECOMPUTING

Metric	Count
Compute Hours of Science	341,700 hours
Compute Days of Science	14,238 days
Compute Years of Science	39 years
AWS Instance Count	10,600 instances

CycleCloud & Cloud finished impossible run, $44 Million in servers, in…

39 years of drug design in 11 hours on 10,600 servers for < $4,372

CYCLECOMPUTING

Does this lead to new drugs?

Novartis announced <u>3 new compounds</u> going into screening based upon this 1 run

Survey of Use Cases

- ☑ Drug Design
- ☑ **CAE**
- ☑ Genomics

…

CYCLECOMPUTING

The Aerospace Corporation

Aerospace Corp @ AWS re:Invent

High Performance Computing @ Aerospace

- Allow engineers and scientists to focus on their discipline and research
- Reduce and eliminate complexity in using High Performance Computing (HPC) resources
- Supply and support centralized and networked HPC resources

AWS re:Invent

Reached out to Cycle Computing in Summer 2013

Aerospace Corp @ AWS re:Invent

What makes this work?

- AWS GovCloud
  - GovCloud is FedRAMP compliant
- Secure transport to and from Aerospace
  - VPC provides an additional layer of security while data is in transit
- Cyclecomputing
  - Cycle provides cluster auto-scaling

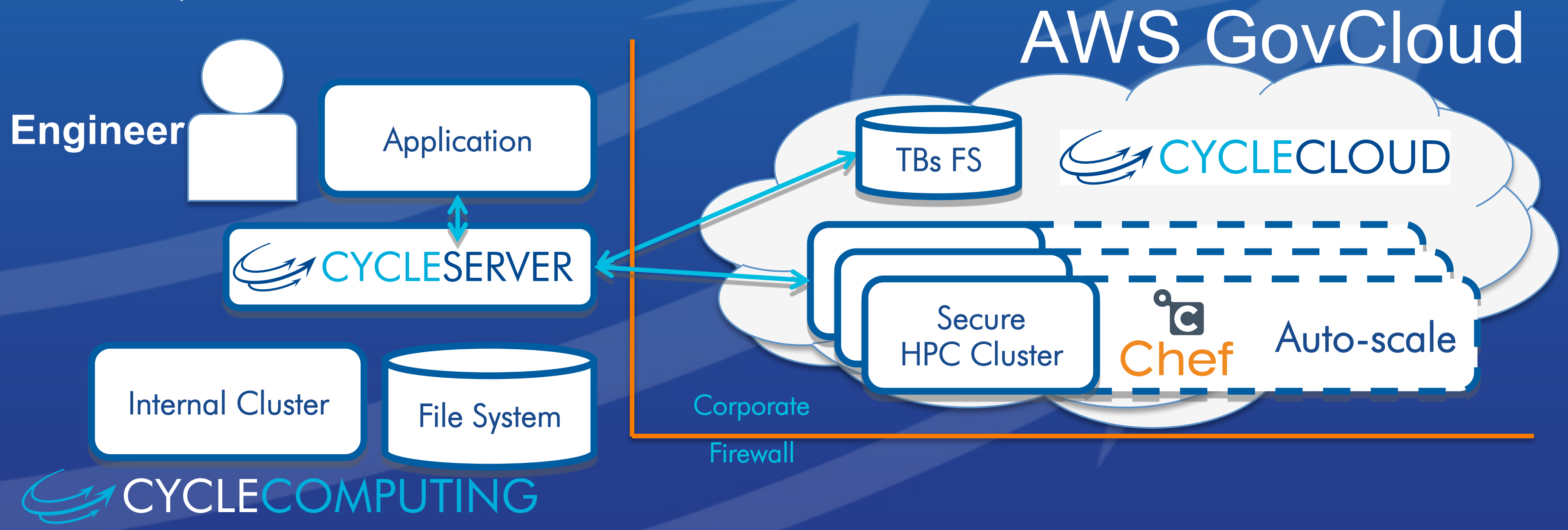
AWS re:Invent

CYCLECOMPUTING

Rocket Design on AWS/Cycle

Remove caps on scale, speed time to result

No QueueWait for users



HGST, A Western Digital Company

CYCLECOMPUTING

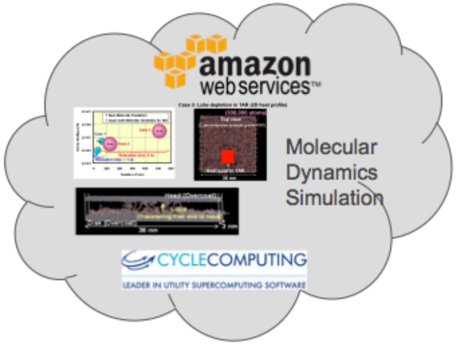
HGST, A Western Digital Company

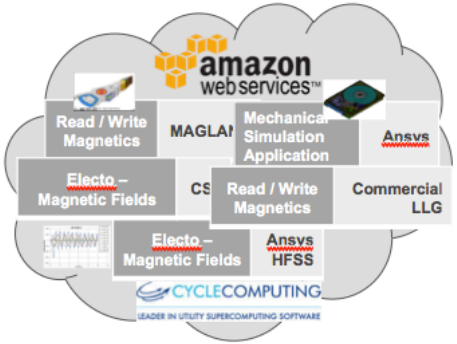
HGST
a Western Digital company

- Founded in 2003 through the combination of the hard drive businesses of IBM, the inventor of the hard drive, and Hitachi, Ltd ("Hitachi")
- Acquired by Western Digital in 2012
- More than 4,200 active worldwide patents
- Headquartered in San Jose, California
- Approximately 41,000 employees worldwide
- Develops innovative, advanced hard disk drives, enterprise-class solid state drives, external storage solutions and services
- Delivers intelligent storage devices that tightly integrate hardware and software to maximize solution performance

Burst for TC workloads

HGST's Amazon HPC Platform





Base HPC Platform

- Scalable to thousands of instances to support numerous simultaneous simulations

Pre- and Post-Processing Server Farms

- New G2 Instances Add Visualization Capabilities

Results

- Gave 10 minutes to get 64-core to 512-core environments!
- Parallel scaling greatly improved time-to-results
- Simplified submission interface removed file transfer and remote computing complexity
- Auto-scaling environments decreased costs while increasing science completed.

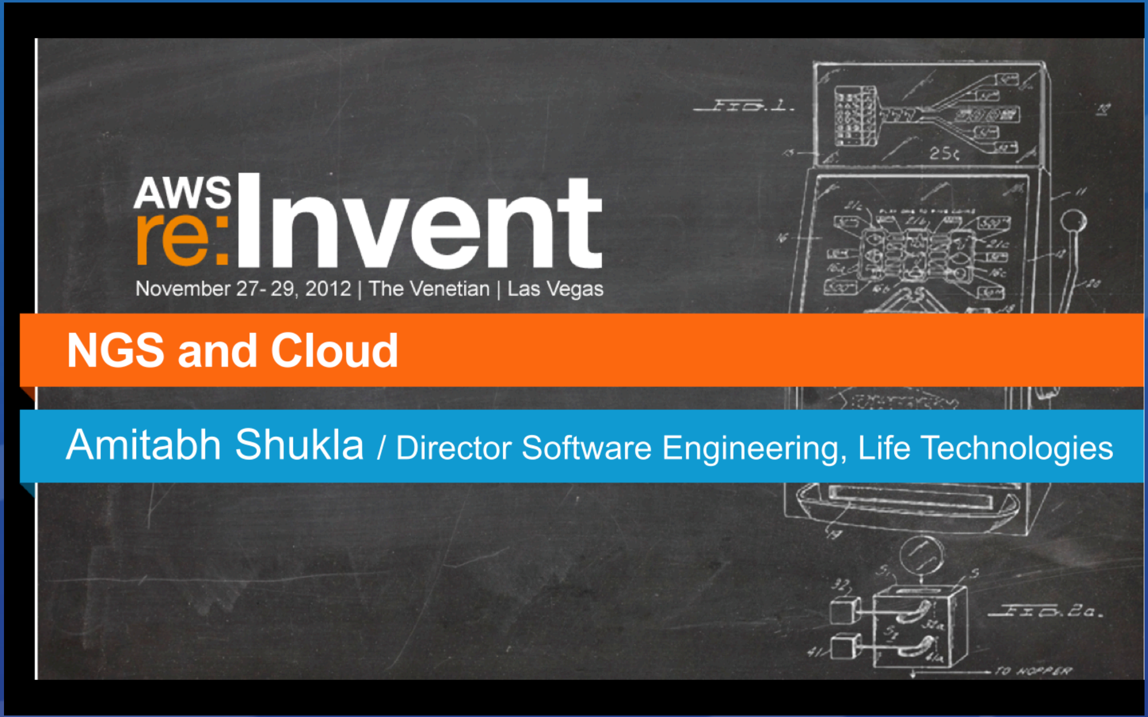
Survey of Use Cases

☑ **Drug Design**
☑ **CAD/CAM**
☑ **Genomics**
…

IonTorrent on Cloud

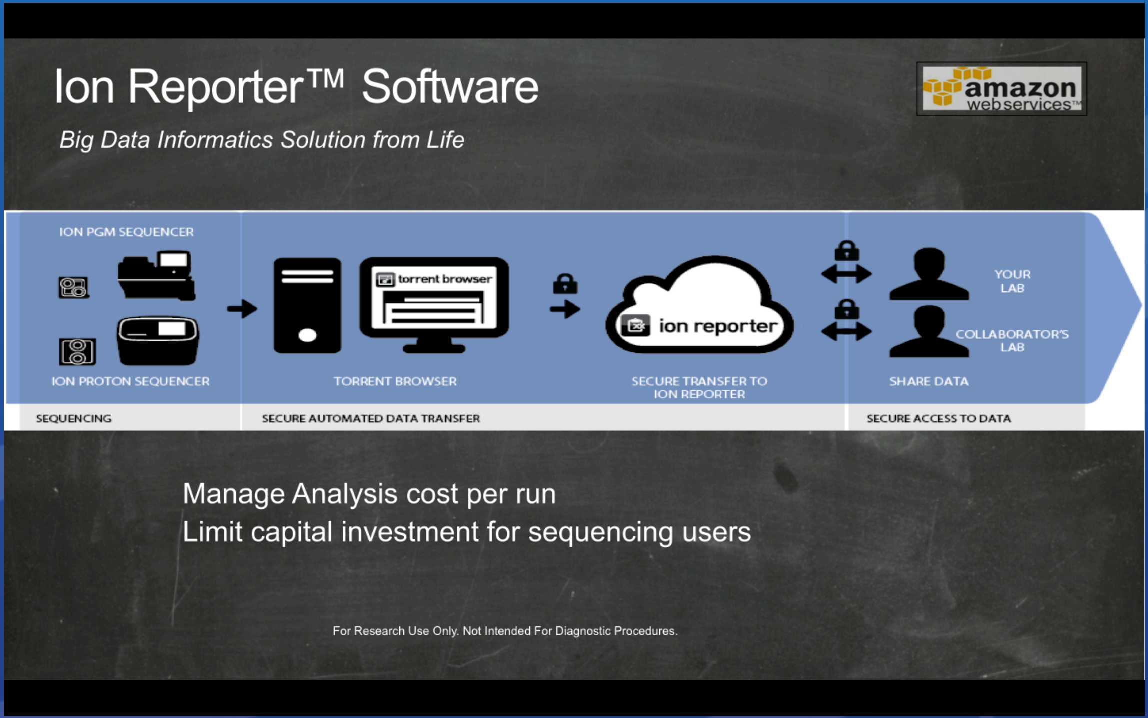
CYCLECOMPUTING

IonTorrent on CycleCloud



CYCLECOMPUTING

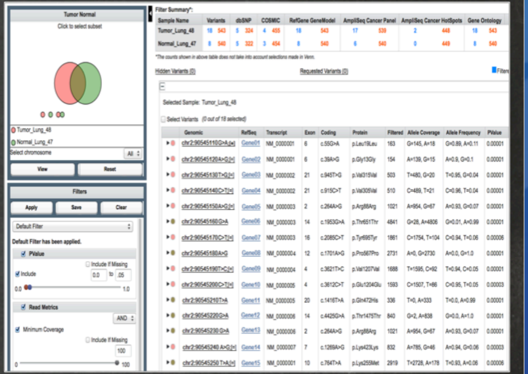
IonTorrent on CycleCloud



IonTorrent on CycleCloud

IR Data Analyses – Using AWS

- Fixed
  - Application servers
  - Data upload servers
  - NoSQL database
  - RDS-MySql



- Elastic
  - On demand CC2 instances: CycleCloud
    - Hadoop: Bring data to compute nodes.
  - Storage
    - EBS for primary storage
    - S3 to Archive data
      - Future release: Glacier for Archival solution

CYCLECOMPUTING

Results

- Continuously deliver up to 8,000 core of compute power to support online genomic analysis
- Complicated application w/Web, NoSQL, TC clusters
- Analysis as a Service for all IonTorrent users

Survey of Use Cases

- ☑ **Drug Design**
- ☑ **CAD/CAM**
- ☑ **Genomics**

…

Now hopefully…

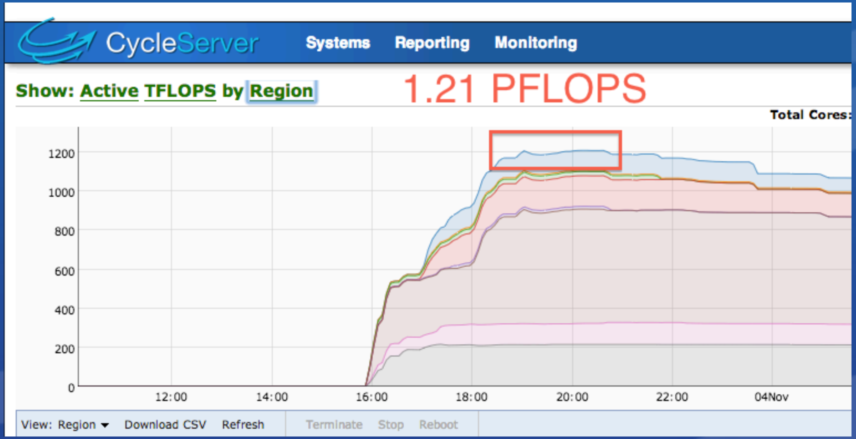
CYCLECOMPUTING

You believe
utility access to technical computing
accelerates invention

Supercomputers = great

Flickr: LouisRix

And so is Cloud…



Flickr: MattDavis

CYCLECOMPUTING

Computing can revolutionize materials science

Organic semiconductor

From Wikipedia, the free encyclopedia

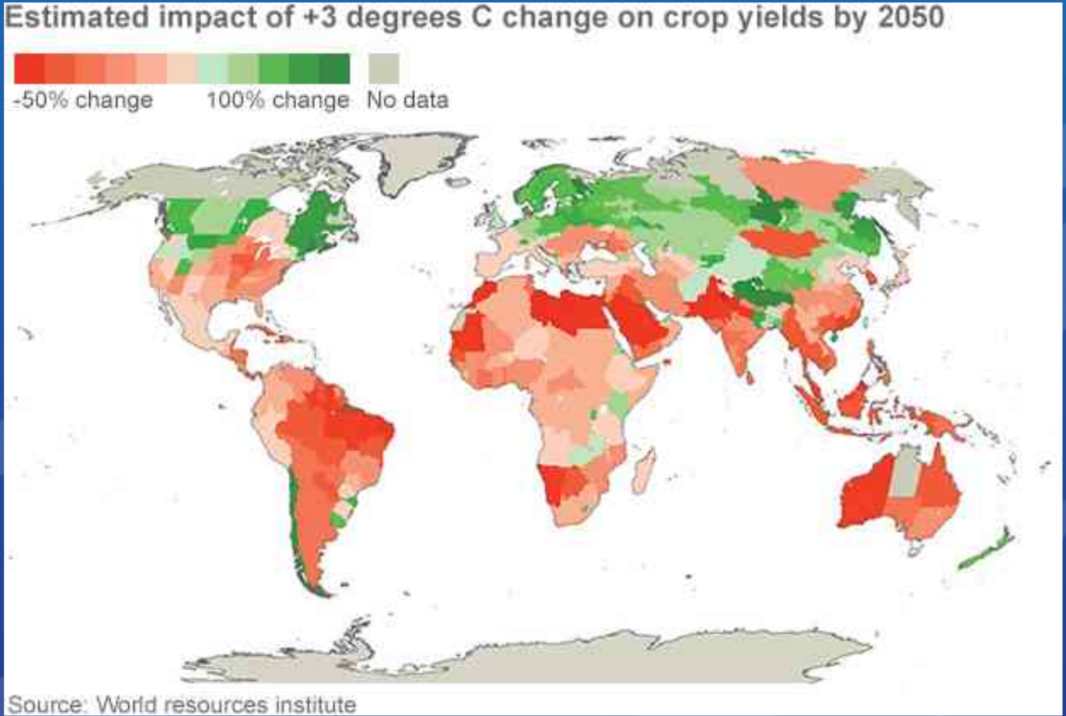
This section **does not cite any references or sources**. Please help improve this section by adding citations to reliable sources. Unsourced material may be challenged and removed. *(July 2013)*

STM image of self-assembled supramolecular chains of the organic semiconductor Quinacridone on Graphite.

An **organic semiconductor** is an organic material with semiconductor properties, that is, with an electrical conductivity between that of insulators and that of metals. Single molecules, oligomers, and organic polymers can be semiconductive. Semiconducting small molecules (aromatic hydrocarbons) include the polycyclic aromatic compounds pentacene, anthracene, and rubrene.[*not verified in body*] Polymeric organic semiconductors include poly(3-hexylthiophene), poly(p-phenylene vinylene), as well as polyacetylene and its derivatives.[*not verified in body*]

There are two major overlapping classes of organic semiconductors. These are organic charge-transfer complexes and various linear-backbone conductive polymers derived from polyacetylene. Linear backbone organic semiconductors include polyacetylene itself and its derivatives polypyrrole, and polyaniline.[*not verified in body*] At least locally, charge-transfer complexes often exhibit similar conduction mechanisms to inorganic semiconductors. Such mechanisms arise from the presence of hole and electron conduction layers separated by a band gap.[*not verified in body*] Although such classic mechanisms are important locally, as with inorganic amorphous semiconductors, tunnelling, localized states, mobility gaps, and phonon-assisted hopping also significantly contribute to conduction, particularly in polyacetylenes.[*not verified in body*] Like inorganic semiconductors, organic semiconductors can be doped. Organic semiconductors susceptible to doping such as polyaniline (Ormecon) and PEDOT:PSS are also known as *organic metals*.[*not verified in body*]

We need to engineer solutions for climate, and lower our footprint

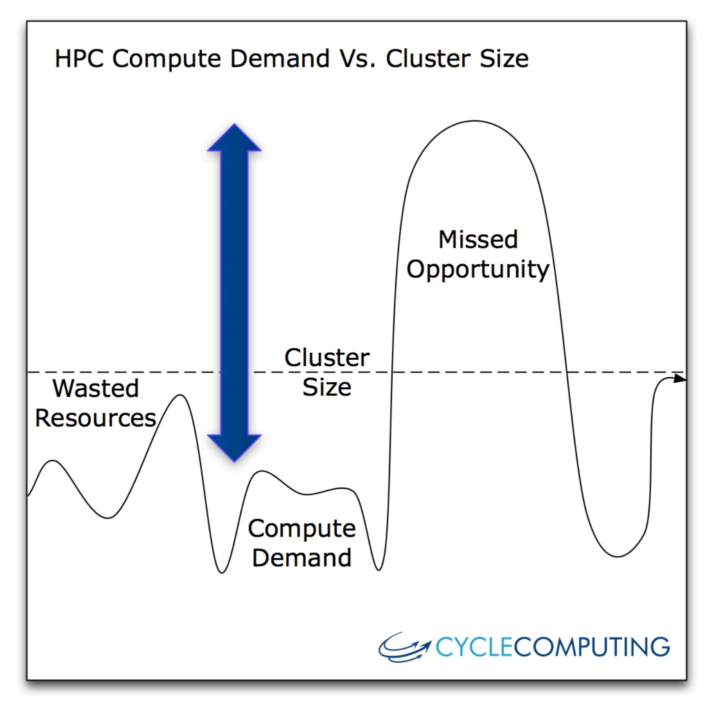


CYCLECOMPUTING

Buy land in Canada?

CYCLECOMPUTING

We can solve this problem



CYCLECOMPUTING

Across many science, engineering, and risk

NOVARTIS — **Life Sciences**

- **39.5 years of drug compound computations in 9 hours**, at a total cost of **$4,372**
- **10,000 server cluster** seamlessly spanned US/EU regions
- Advanced 3 new otherwise unknown compounds in wet lab

AEROSPACE — Assuring Space Mission Success — **Rocket Design & Simulation**

- Moving HPC workloads to cloud for burst capacity
- Amazon GovCloud, Security, and key management
- Up to 1000s of cores for burst / agility

SCHRÖDINGER. — **Life Sciences**

- **156,000-core utility supercomputer** in the Cloud
- Used $68M in servers for 18 hours for $33,000
- Simulated 205,000 materials **(264 years) in 18 hours**

PACIFIC LIFE — **Asset & Liability Modeling (Insurance / Risk Mgmt)**

- Completes monthly/quarterly runs 5-10x faster
- Use 1600 cores in the cloud to shorten elapsed run time from ~10 days to ~ 1-2 days

Nuclear engineering — **Utilities / Energy**

- Approximately 600-core MPI workloads run in Cloud
- Ran workloads in months rather than years
- <u>Introduction to production in 3 weeks</u>

Johnson & Johnson — **Manufacturing & Design**

- Enable end user on-demand access to 1000s of cores
- Avoid cost of buying new servers
- Accelerated science and CAD/CAM process

To continuously deliver HPC to accelerate the human mind…

CYCLECLOUD

Chef

CYCLECOMPUTING

Does this resonate with you?

We're hiring like crazy:
Software developers,
HPC engineers,
devops, sales, etc.

jobs@
cyclecomputing.com

CYCLECOMPUTING