Trends in Computational and Data-enabled Science and Need for Coordinated e-Infrastructure

Edward Seidel

Director, National Center for Supercomputing Applications Founder Professor of Physics, Professor of Astronomy University of Illinois at Urbana-Champaign



PHILOSOPHIÆ NATURALIS PRINCIPIA MATHEMATICA MATHEMATICA Mather Langlann, & Societatis Regalis Sodali. IMPRIMATUR: & PEPY S, Reg. Sor. PRÆSES Juli 5, 1656.

LONDINJ, Julfa Sectoratis Regie ac Typis Jofepli Streater. Profiat apud plures Bibliopolas. does MDCLXXXVII. 5



Convergence



Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond

National Center for Supercomputing Applications University of Illinois at Urbana–Champaign



Collaborations Grow as Problems Become More Complex; Systems and Data Sizes, too! TRENDS IN SCIENCE





Three tech trends made this possible:

Instrumentation

• HPC

Data Science
 They are intricately
 connected!

Precisely 100 years ago, Einstein predicted gravitational waves. Feb 11, 2016 their historic discovery was announced!



2 Black Hole Collision: Changing Culture of Science



Figure 17. The collision of two Black Holes. The event horizons ∂B_1 and ∂B_2 merge to form the event horizon ∂B_3 . The apparent horizons ∂T_2 do not merge but are enveloped by a new apparent horizon ∂T_3 .

<text><text><text><text>



1600-1972: 1994: 10 people, 1998: 15 people, Calileo - Hawking. NCSA Cray Y-MP, NCSA Origin, 1 person, no 50MB 50GB

Community Einstein Toolkit

"Einstein Toolkit : open software for astrophysics to enable new science,





54

LIGO Scientific Collaboration Go-WA

- 1004 Authors on the breakthrough paper...
- How do they collaborate?
- They have a data sharing agreement and services to suppor them



LIGO-L



Gravitational Wave Astrophysics Interplay between Simulation and F^{*} priment odels & Simulation





Obs

News Flash! Sept 14, 2016, Announced Last Month: Gravitational waves and black holes discovered!

oscillati

waves, ...



...that simple chirp, which rose to the note of middle C before abruptly stopping, seems destined to take its place among the great sound bites of science, ranking with Alexander Graham Bell's "Mr. Watson - come

here" and Sputnik's first beens from orbit." NY Times

Stakes in Neva **Rise for Hillary** Clinton as Caucuses Near

By ADAM NAGOURNEY

Ξ

After his defeat of Mrs. Clinto a chance to prove himself in a state as recipile. state as racially diverse as Io and New Hampshire are not.

Bernie Sanders Intris a South Carolina Tov That Loves Hillary Clinton

By RICHARD FAUSSET

ago.

In Orangeburg, S.C., some vo who said in January that they



SCIENCE & ENVIRONMENT

542 ld undoubtedly vote fo 256frequency 12369 'Ripples' from blac 0:30 For the first time, scientists have d space and time - gravitational wav



SUBSCRIBI

CLASSIFI



BREAKTHROUGH HPC SYSTEMS SUPPORT BREAKTHROUGH SCIENCE

But HPC is only part of greater ecosystem!

Cray XE6/XK7 - 276 Cabinets

XE6 Compute Nodes - 5,688 Blades - 22,640 Nodes -362,240 FP (bulldozer) Cores - 724,480 Integer Cores 4 GB per FP core XK7 GPU Nodes 1,056 Blades – 4,224 Nodes 33,792 FP Cores - 11,354,112 cuda cores – 4,224 K20X GPUs, 4 GB per FP core

Blue Waters Computing System







100M atoms! AIDS Virus, Chromatophore Schulten, UIUC

UNDERSTANDING MATERIALS & DESIGNING NEW ONES



Starry Night *Turbulence in Fluids*



Vincent van Gogh, 1889

"When I die and go to heaven there are two matters on which I hope for enlightenment. One is quantum electrodynamics, and the other is the turbulent motion of fluids. And about the former I am rather optimistic." Horace Lamb, 1932...

- Turbulent flow via direct numerical simulation (DNS) just possible on modern Petascale systems
 - 100PF+ would allow complete engine simulation with turbulence
 - 40 Exallop would allow complete airplane simulation!





Half of the Blue Waters Projects Require Connection to Real-World Data...

Real World Science and Engineering Require Integration of Compute and Data FUTURE DIRECTIONS

The Growth of Data



The FOURTH PARADIGM DATA- INTENSIVE SCIENTIFIC DISCOM "I'm still here..."

But I'm your new baby big brother...



With millions of processors...

Data-enabled Transformation of Science





Astronomy 1500- 2000:

- Single scientist looks through telescope
- Record KB of data in notebook
- Require reproducibility

Sloan Digital Sky Survey 2000+

- Record data for decade (40TB)
- Serve to entire world
- Thousands of scientists
 work "together"



DES (now)

- 200GB/night
- PB in decade
- LSST (6 years)
 - Record data for decade
 - SDSS/night!
 - 200 PB/decade



Integrating it all: Multi-Messenger Astronomy...

Astronomy c. 2020!

- New era: seeing events
 as they occur
- Here now
 - ALMA, EVLA
 - Ice Cube neutrinos
- On horizon
 - 20-30m optical?
 - LSST
 - aLIGO, Indigo
- SKA = exabytes
 - Simulations integrate physics

Communities communicate by sharing data, software... knowledge...

Scenarios like this in all fields...















Los Angeles







National Academy Reports...

"*Convergence* [is] a culture shift for academic organizations that are traditionally organized around discipline-based departments...."

"...most difficult problems do not respect disciplinary boundaries...convergent science, [integrates] insights and approaches from many fields...barriers to convergent science...



We need e-Infrasturctures that better support convergence....



Modern Research & Education Ecosystem

Blue Waters



Key challenge will be to create a deeply integrated compute and data environment that supports complex problem solving for academia and industry...

Campus

XSEDE

Campus

En*10Gb

Campus

CANARIE n*10Gb

Salt Lake City

Angel

San Diego Phoen

ic Wave Sunnyvale

ENIC n*10Gb Tijuana

TransLight 1000

Campus

Campus

ARIE n*10Gb

Campus

Software

Campus

Campus

Education: I need all of this to start to solve my problem!









Fundamental Problems with this

- It is piecemeal...
- It is not integrated...
- It's chaotic: Pieces come and go...
- Big instrument projects be
 - Can't depend on any
 - Therefore they "ro'
- Its expensive: nee

What is needed are long term planning horizons before construction! Can this be done?

- But with current mod
 - We are (presumation wasting money...
 - It is harder to support the way science is moving, towards convergence...



Science-driven Cyberinfrastructure

Long-term planning & investment in national computing and data CI by the National Science Foundation

Home Current Drafts and Contributions

Get Involved

Participants

Participants

Steering Committee

The efforts of developing this concept are guided by a steering committee consisting of:

Dan Atkins

W.K. Kellogg Professor of Community Information Professor of Electrical Engineering and Computer Science College of Engineering, University of Michigan

Alan Blatecky





Interdisciplinary Research and Education DATA AS ENABLER FOR COLLABORATIVE WORK

Materials Innovation Grand Challenge Communities

- Combining approaches in a digital world
 - Theory and computation
 - Instrumentation
 - Data and informatics
- Cyberinfrastructure
 - Software centers
 - Data services + Instruments
 - Computing
- Policy
 - Open data will accelerate discovery, enhance interdisciplinarity, speed innovation, commercialization



"Long Tail" vs Big Data: Advanced Photon Source Upgrade

Highly integrated computing/data services at ANL



Materials Data Facility *Chicago, NIST, NCSA, ...*

We aim to make it simple for materials datasets and resources to be ...





Building Grand Challenge Communities around Data NSF BIG DATA HUBS









The National DATA SERVICE

National Data Service Workshop October 19-21, 2015

The National Data Service Consortium fourth plenary meeting will be in San Diego, October 19-21 and **limited space is still available**!

Researchers, educators, students communicate by sharing data...this is central to enabling everything above! Services needed to make it work!

Summary

- Major paradigm and culture shifts: research is changing dramatically
 - Complex problems require collaborations at new scales
 - Computing and data capabilities growing unprecedented rate
- Major infrastructure projects are highly computing & data intensive
 - Data services are needed to support
 - A comprehensive, integrated approach would better serve science, reduce costs

Data sharing supports int'l cooperation









