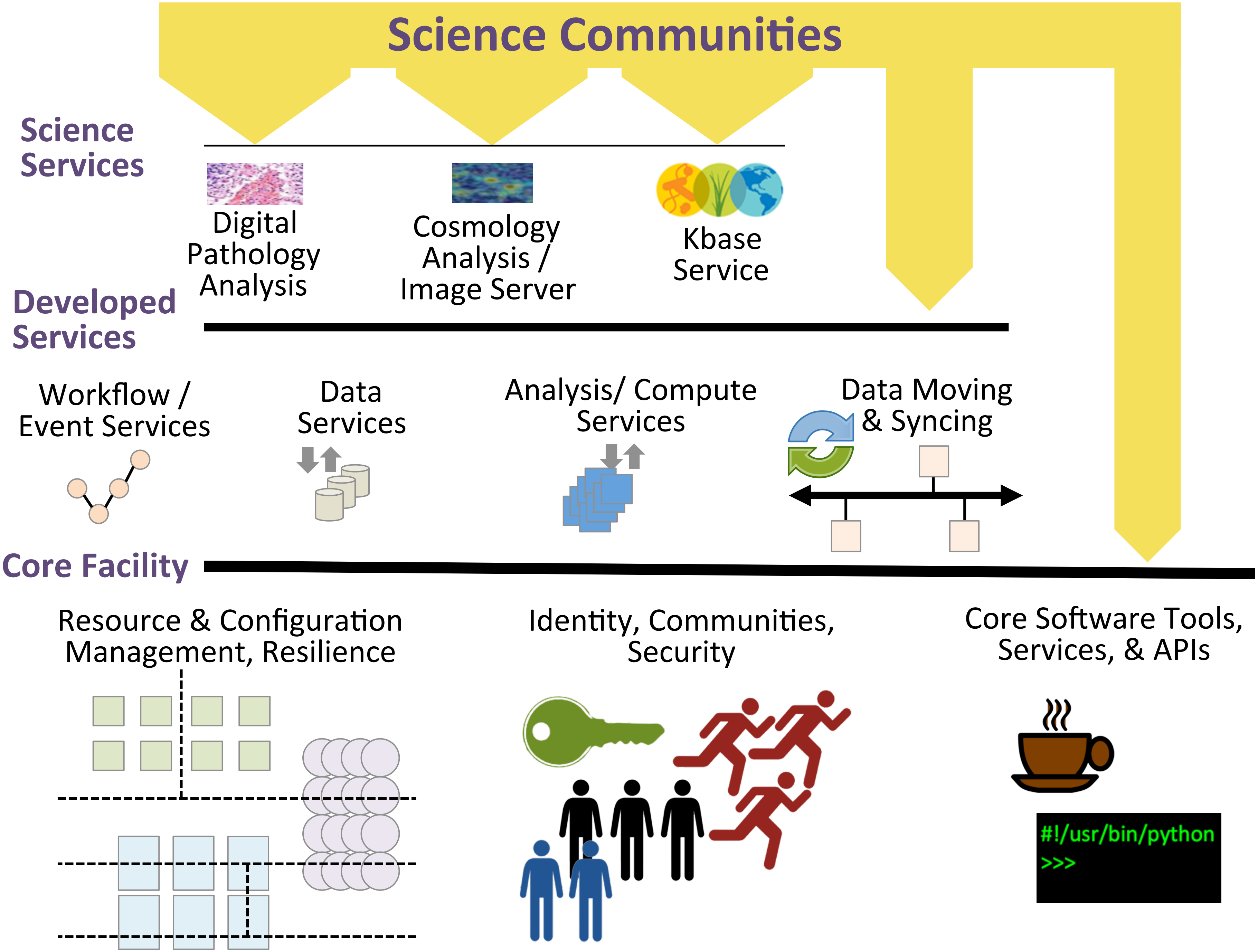
10^18 BDEC

Exec Committee

Pete Beckman
Jean-Yves Berthou
Jack Dongarra
Yutaka Ishikawa
Satoshi Matsuoka
Philippe Ricoux

BIG DATA AND EXTREME-SCALE COMPUTING


Science Communities
Science Services
Digital Pathology Analysis
Cosmology Analysis / Image Server
Kbase Service
Developed Services
Workflow / Event Services
Data Services
Analysis/ Compute Services
Data Moving & Syncing
Core Facility
Resource & Configuration Management, Resilience
Identity, Communities, Security
Core Software Tools, Services, & APIs
#!/usr/bin/python
>>>

Summary: New Ideas (EC must change or be BD irrelevant)

- Resource Management
  - Elastic, Interactive, Reservations/co-scheduling
  - Economic models/cost, new incentive models, QoS
- System Complexity
  - Virtualization / BareMetal (encapsulation, elasticity)
  - Management tools
  - Machine learning & automation in operations
- Identity management & sharing / analysis
  - must move toward dropbox models
- Data sync/move/stream must be part of system

ToDo (need more exploration):

- Workflow / execution model / programming
  - Still quite diverse, must create a few examples for discussion
- Archetypes for platforms, with quantitative measures & benchmarks
- Testbeds: Understand what we are missing, and how we can provide it
- Mini-apps: We need some real examples (with workflows) that the community can measure, re-code, explore
- Future tech that could help address BDEC
  - 100Gb networks, virtualization,

Expectations

- ***Long term*** archival storage of science data is out of scope
- Looking at intersection of BD and EC

Data Services
(2nd afternoon session)

James Ahrens, Chaitan Baru

- Beckman report
  - Report from database community about big data challenges
  - Recommending separation between storage layer and runtime layer.
- Propose a BDEC Workshop at VLDB 2015, Hawaii

- Session 1: 90 Minutes
  - Describe what is needed in Common/Basic data services for BDEC system
    - What are the most accepted (deployed) services?
    - What are the largest gaps for BDEC science communities?
    - What kinds of optimizations / specializations are required for science communities?
  - Report out during Plenary
- Session 2: 120 Minutes
  - Analyze and discuss the initial work of the other breakouts
  - Describe:
    - What is needed to support Federation, Provenance, and Curation?
    - What are the programming models needed?
    - What basic services should be included in BDEC facilities first?
  - Report out during Plenary

Federation

- Federation: Needed to access data from different sources
- Need catalog services
- Load up large amounts of observational, remote sensing data, simulation data, etc into one place
  - Be able to access data from the cloud
- First problem: identify management
- Examples:
  - ESA Geohazard project: Combine ESA data with WHO data for malaria
  - AIST: GeoGrid project—remote sensing, in situ data.
  - Many agencies store their own data. Requires federated access.
  - NCI: Cancer Cloud initiative, 3PB cancer genomic data
- Need higher level analysis services with data
  - Should be able to run a toolbox in the cloud

Provenance

- Provenance is a subset of the reproducibility problem, e.g. where did the data come from; who accessed it;
- Having good provenance information can help with reuse of data
  - Avoid re-computation; avoid replication (deduplication)
- If the data is in the cloud and we can monitor usage
  - Then easier to provide provenance on data
- Characterize/understand "Provenance preservation" between consecutive steps.
  - Provenance preservation is property of each transformation

Curation

- Maintaining data for the long-term such that it is available for reuse and preservation
- Need to provide ready <u>access</u> to curated data
  - Need an architecture that allows you to easily access stored data
- How long do we keep the data?
  - Sensor data is precious—natural, observed phenomena; things that cannot be reproduced
- Unsure of how to sustain curation of data
  - Can industry assist: curation as a service?
  - Could they charge for the CPU for using the data?

Programming Models

- Current models are not adequate for interacting with data
  - Declarative approaches are good
  - Need more abstractions; interact well with the runtime
  - Need an “explain” capability (debuggability)
  - Many users would like to have DSLs
- Support for resilience (tolerate faults)

Basic services to include in BDEC first?

- Ability to specify and orchestrate workflows
  – On-demand access to compute and storage
- Virtualization (with hardware support)
- Analytics libraries at scale
  – Need to understand current performance and scaling characteristic of some of these packages, including GIS…
- Some runtime support, e.g. for data movement/caching, …

BDEC February 2014 Applications Breakout

Session 2

The Desire for a Mini-App

- The complexity is in the data
  - Can you get a mini-data
  - Why isn't it so important in commercial analytics?

Why is data movement so important

- Google moves the data once and keeps it
- Sensible designs do not require repeating the data movement over and over and over
- Is virtualization a need or is it an artifact of an assumption of how to run/allocate the system
- Domain-specific computing is an alternative that is likely to be more cost effective

Work Flow Description

- Components
  - Algorithm
  - Compute pattern
  - Data size
  - Data access pattern
- Links – data transferred
- Data sources and sinks
- Coupling between compute and data
- Synchronous vs. asynchronous

Spatio-temporal Sensor Integration, Analysis, Classification

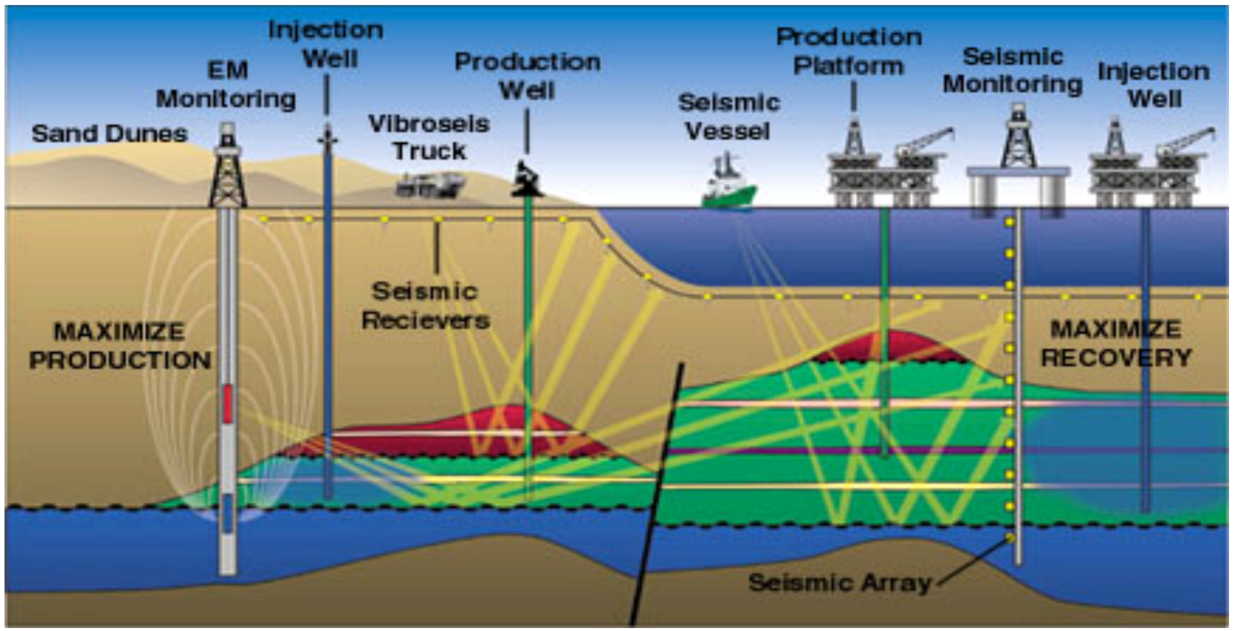
- Multi-scale material/tissue structural, molecular, functional characterization. Design of materials with specific structural, energy storage properties, brain, regenerative medicine, cancer
- Integrative multi-scale analyses of the earth, oceans, atmosphere, cities, vegetation etc – cameras and sensors on satellites, aircraft, drones, land vehicles, stationary cameras
- Digital astronomy
- Hydrocarbon exploration, exploitation, pollution remediation
- Aerospace – wind tunnels, acquisition of data during flight
- Solid printing integrative data analyses
- Autonomous vehicles, e.g. self driving cars
- Data generated by numerical simulation codes – PDEs, particle methods
- Fit model with data

Typical Computational/Analysis Tasks

Spatio-temporal Sensor Integration, Analysis, Classification

- Data Cleaning and Low Level Transformations
- Data Subsetting, Filtering, Subsampling
- Spatio-temporal Mapping and Registration
- Object Segmentation
- Feature Extraction
- Object/Region/Feature Classification
- Spatio-temporal Aggregation
- Diffeomorphism type mapping methods (e.g. optimal mass transport)
- Particle filtering/prediction
- Change Detection, Comparison, and Quantification

Coupled data acquisition, data analysis, modeling, prediction and correction – data assimilation, particle filtering etc.



Detect and track changes in data during production
Invert data for reservoir properties
Detect and track reservoir changes

Assimilate data & reservoir properties into the evolving reservoir model
Use simulation and optimization to guide future production

Soft real time and streaming Sensor Data Analysis, Event Detection, Decision Support

- Integrated analyses of patient data – physiological streams, labs, mediations, notes, Radiology, Pathology images, mobile health data feeds
- High frequency trading, arbitrage
- Real time monitoring earthquakes, control of oilfields
- Control of industrial plants, aircraft engines
- Fusion – data capture, control, prediction of disruptions
- Internet of things
- Twitter feeds
- Intensive care alarms

Typical Computational Analysis Tasks

Streaming Sensor Data Analysis, Event Detection, Decision Support

- Prediction algorithms – Kalman, particle filtering
- Machine learning algorithms on aggregated data to develop model, use of model on streaming data for decision support
- Searching for rare events
- Statistical algorithms to distinguish signal from noise
- On the fly integration of multiple complementary data streams

“omics”

- Sequence assembly
- Metagenomics – identification/characterization of populations of organisms based on DNA/RNA sequencing
- Phylogenetics, genetic based population biology, cancer mutation landscaping
- Pathway modeling using integrated sequence, expression, epigenetics, protein, glycans
- Genetic/genomic based design of organisms with specific properties

Typical Computational/Analysis Tasks
"omics"

- Discrete algorithms – hashing, searching, sorting, comparisons, dynamic programming, indexing, similarity search
- Compression
- Statistical algorithms to distinguish biological signal from experimental artifacts/noise
- Graph construction, traversal, partial/sub graph matching, graph partitioning
- Statistical methods on graphs e.g. Bayesian networks

Population and Social Network Analyses

- Aggregated electronic health data to predict likelihood of disease onset, treatment response, likelihood of re-hospitalization etc
- Predict demand for products, target advertising, store shelf placement
- Characterize influence in social networks

Typical Computational/Analysis Tasks

Population and Social Network Analyses

- Structural properties of graphs – PageRank, diameter, radius, connected component
- Graph spectral analysis
- Graph mining
- Machine learning, cluster analysis
- Statistical modeling and analyses
- Natural language processing

Approach

- Detailed example workflows led by application experts
- Key cases involve interplay between simulation and data acquisition – "data assimilation"
- Scenarios involving current and future state associated with origin and movement of data between workflow stages
- Definition of workflow components

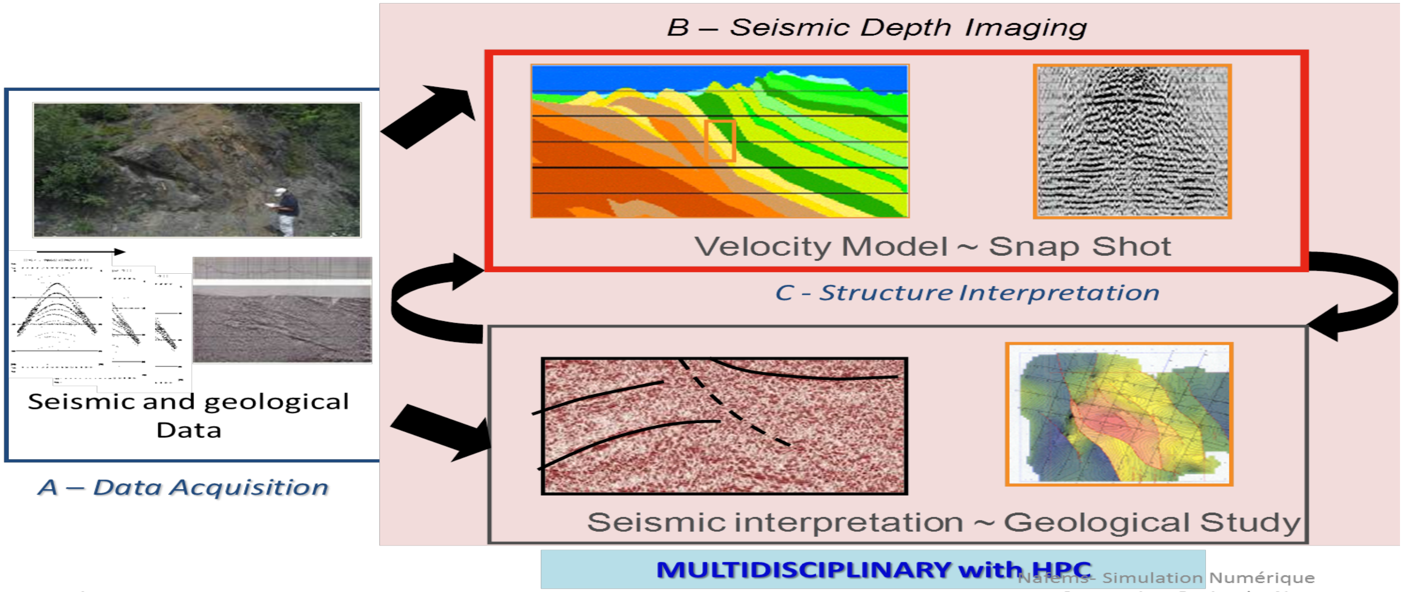
Follow up

- Yutaka Akiyama– Metagenomics
- Geoffrey Fox – Components
- Jean-Claude Andre – Climate
- Philippe Ricoux – Oil exploration/reservoir management
- Joel Saltz – Medical imaging

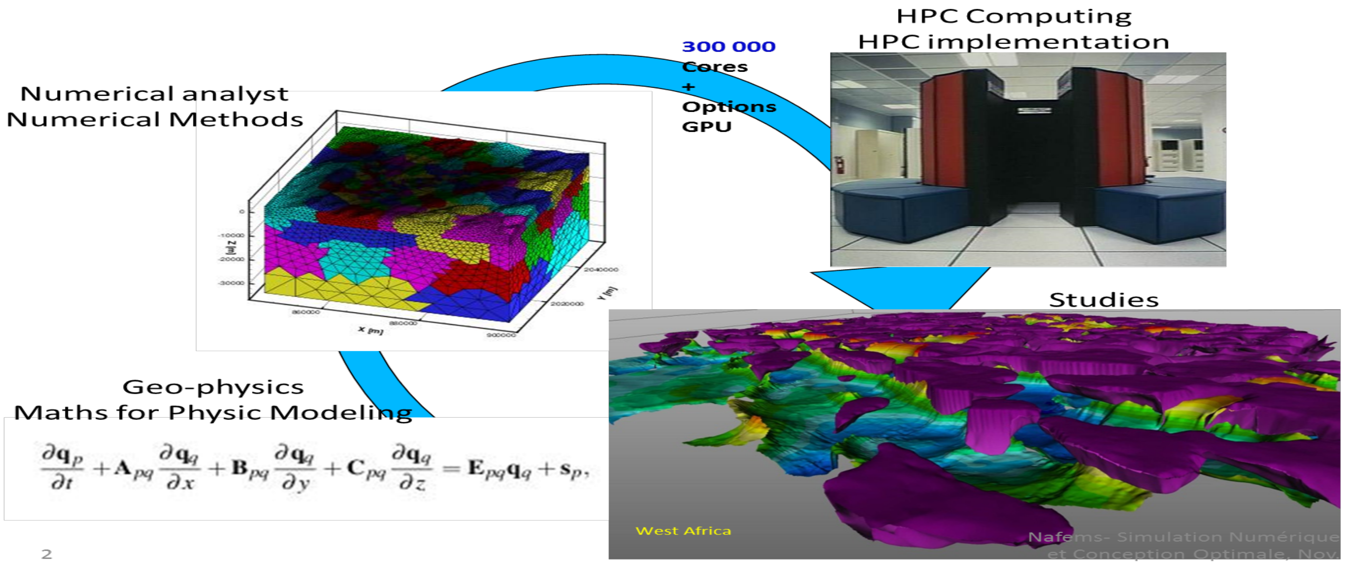
Oil Exploration

- Seismic data acquisition – 20PB
- Transport by plane
- Extract 400TB, load to HPC center
- Obtain complementary datasets – EM, gravimetry and conductivity by network
- Use model to analyze at HPC center
- Generate 3D image (3TB)
- Analysts examine, modify model, modify extraction

Seismic Imaging and geological structures

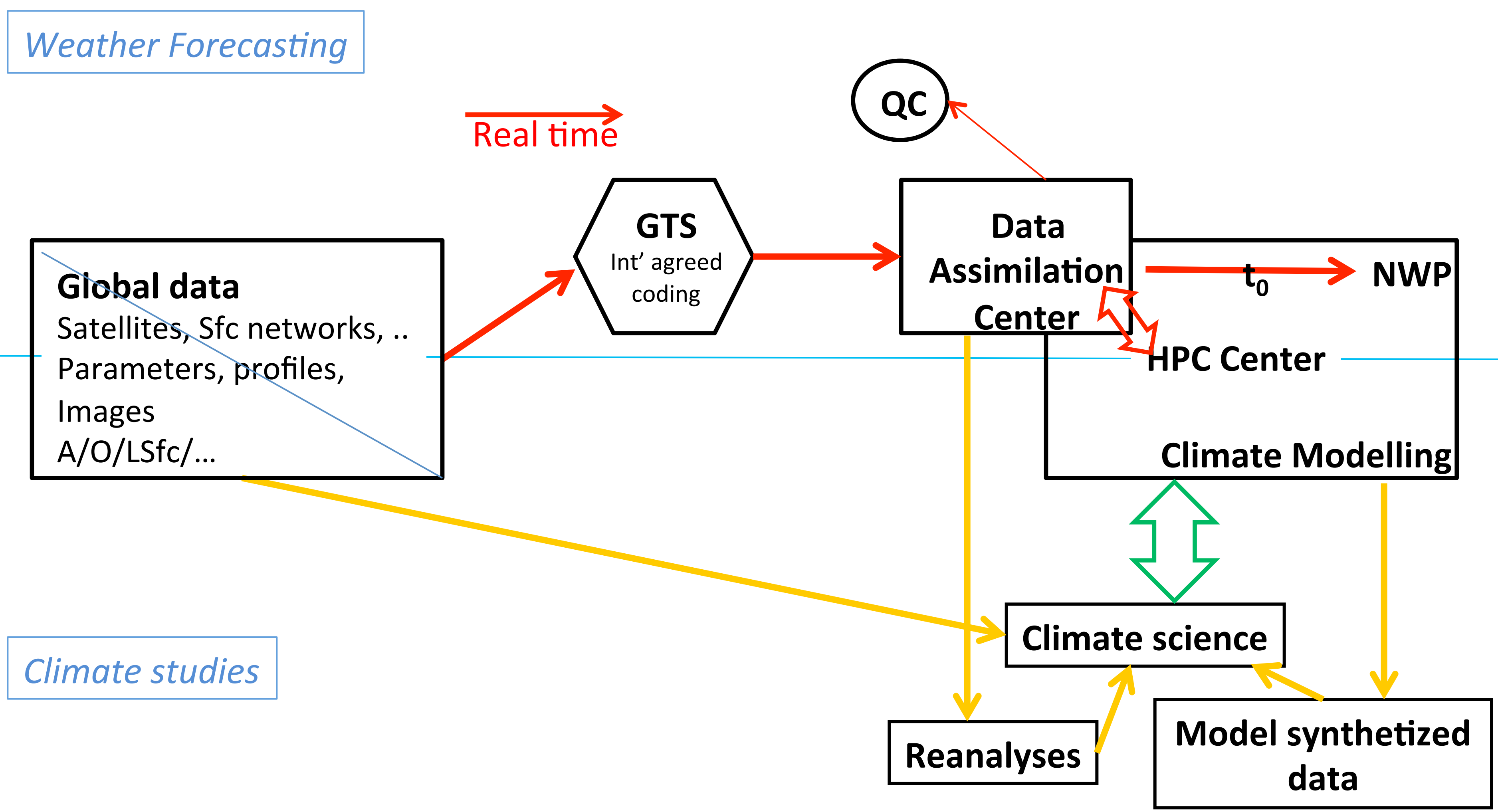


HPC for Depth Imaging : 3 fundamental steps



$$\frac{\partial \mathbf{q}_p}{\partial t} + \mathbf{A}_{pq}\frac{\partial \mathbf{q}_q}{\partial x} + \mathbf{B}_{pq}\frac{\partial \mathbf{q}_q}{\partial y} + \mathbf{C}_{pq}\frac{\partial \mathbf{q}_q}{\partial z} = \mathbf{E}_{pq}\mathbf{q}_q + \mathbf{s}_p,$$

A simplified (?) view of workflows for weather and climate

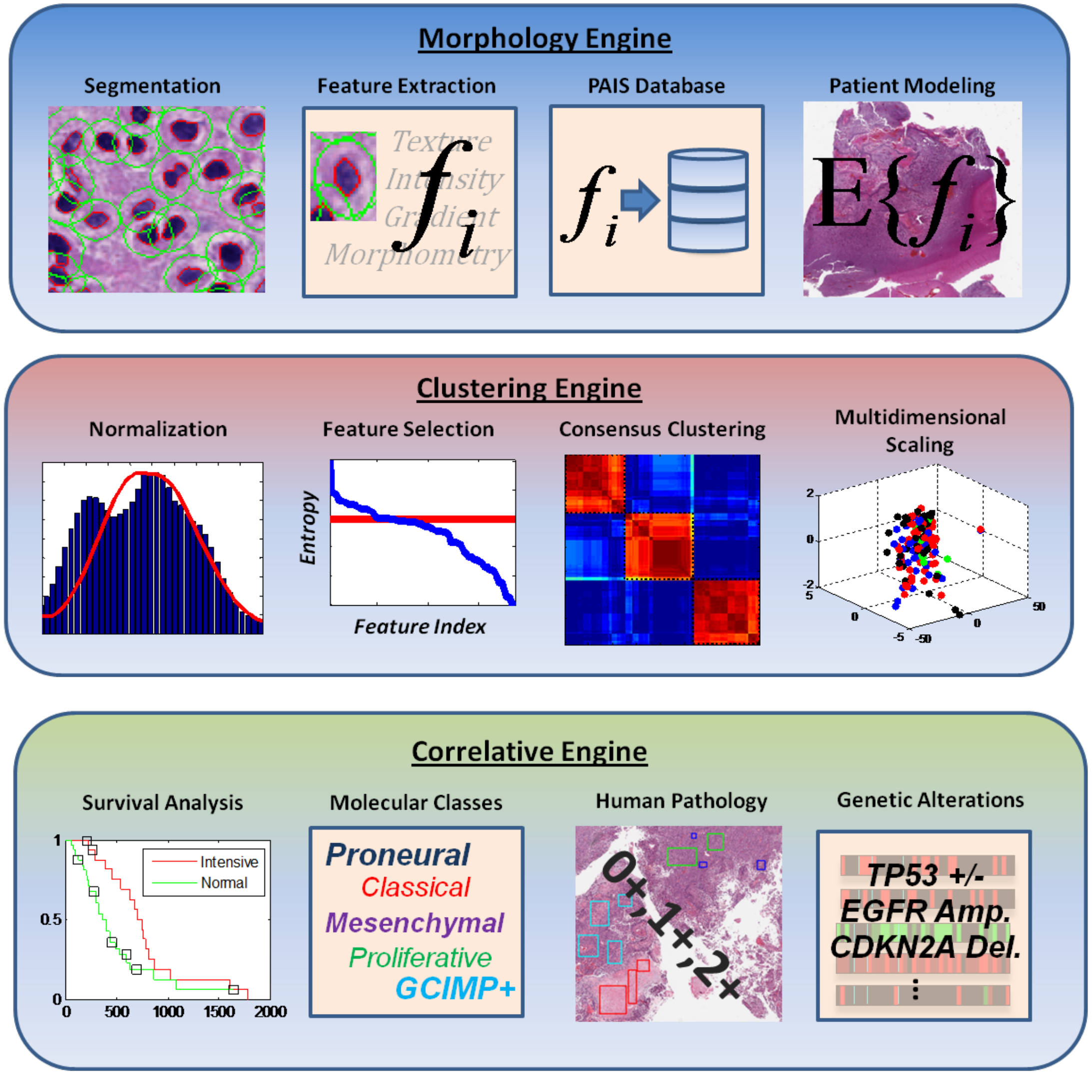


J.C. André, Feb. 28, 2014

Bottleneck

- Long term storage for simulation data - 10-30 PB/year
- Access to facilities to analyze

Direct Study of Relationship Between Image Features vs Clinical Outcome, Response to Treatment, Molecular Information



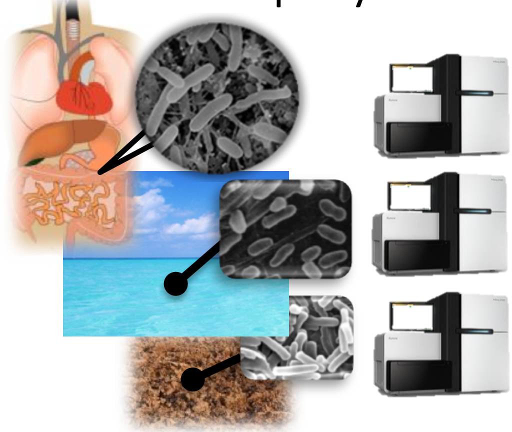
Future State

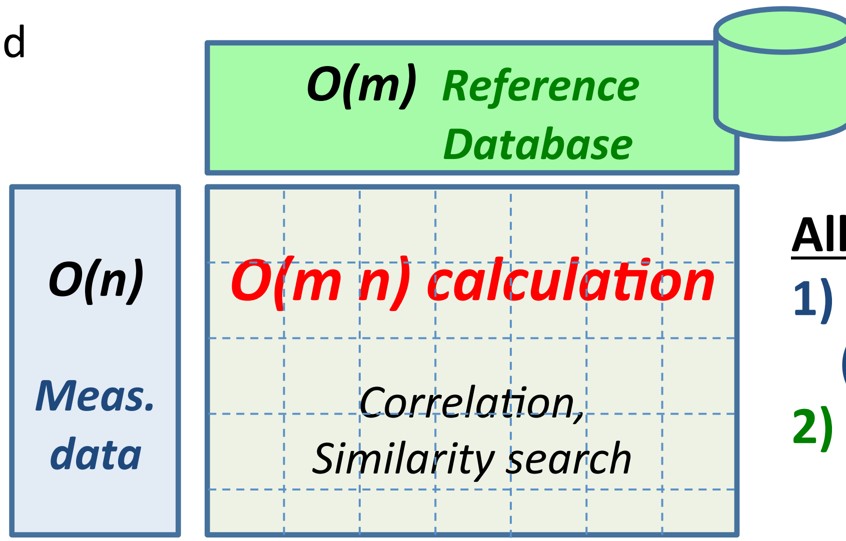
- 100K – 1M pathology slides/hospital/year
- 2GB compressed per slide
- 1-10 slides used for Pathologist computer aided diagnosis
- 100-10K slides used in hospital Quality control
- Groups of 100K+ slides used for clinical research studies -- Combined with molecular, outcome data

Metagenome Analysis

Direct sequencing of uncultured microbiomes in a target environment

1TB/run at Year 2012 and
300% increase per year





Allocation on nodes:
1) Measured data (transient)
2) References DB (resident)

Thousands of sequencers will work every day on **Year 2015**

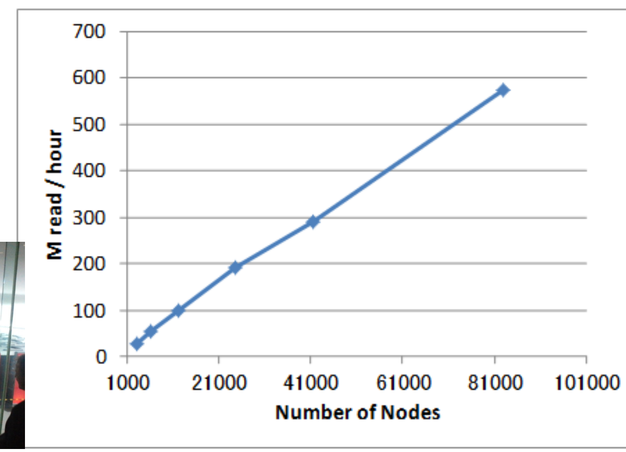
Year 2010

0.18 M Reads / hour
144core Xeon Cluster

3000-fold speed-up achieved

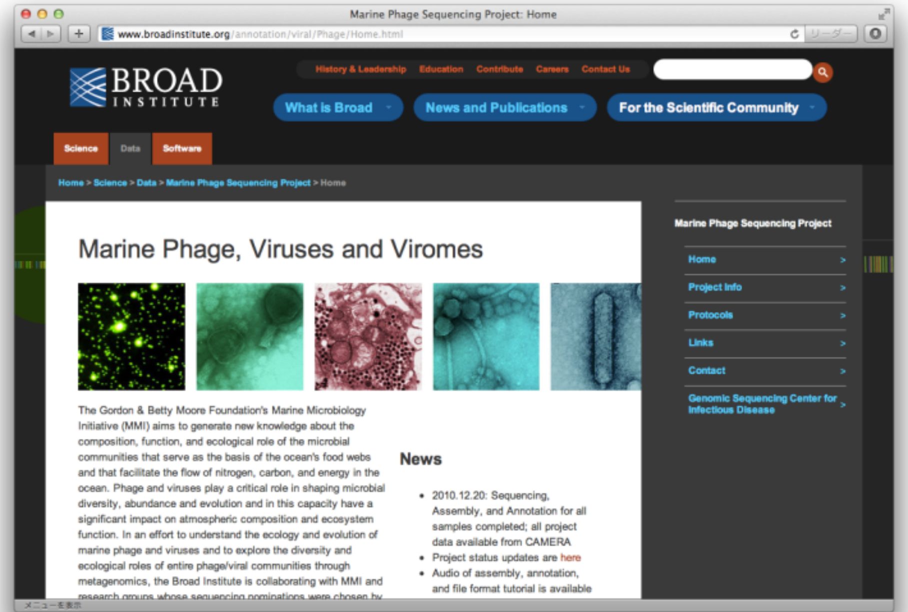
GHOST-MP
OpenMP / MPI

Year 2012

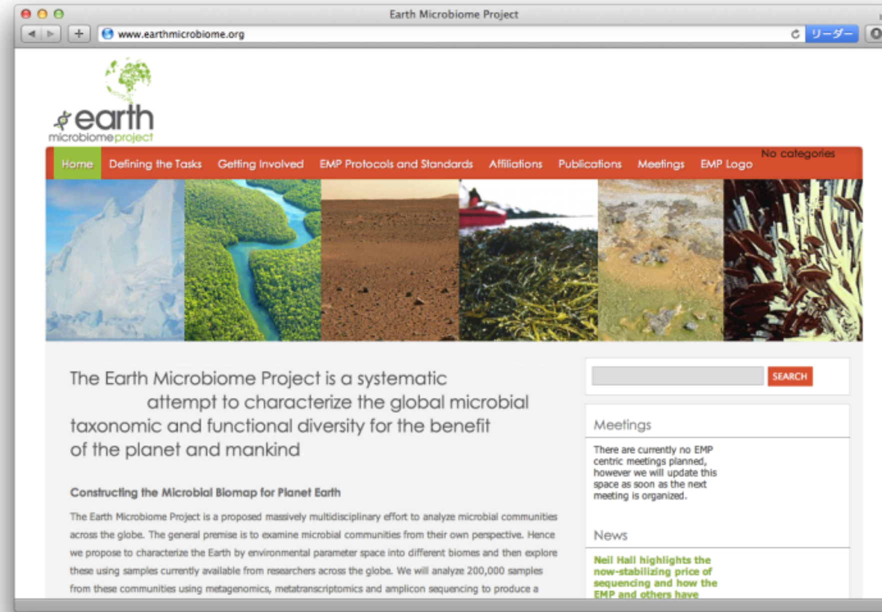


573 M Reads / hour with
82944 node **K-computer**

Marine Phage Sequencing Project

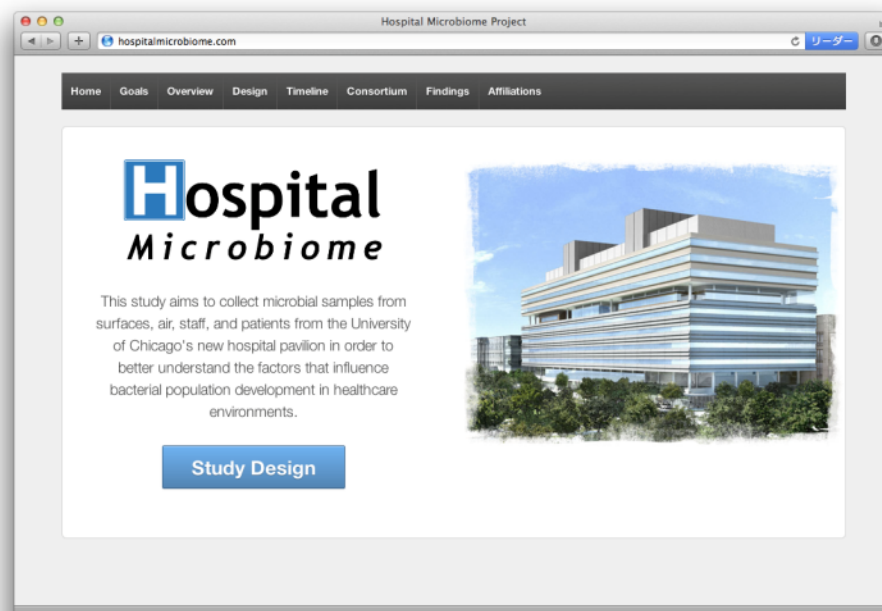


Earth Microbiome Project



Home Microbiome Study

Hospital Microbiome Project



General Structure/Features/Components

- **Data Gathering**
  - Blocked or Real Time (Streaming)
  - Interval = Month(Seismic, Remote sensing) -- Day (genomic) – seconds
- **Storage**
  - Type SQL NoSQL Files etc
  - Style: Dedicated or Permanent or Transient
  - Possible static databases of other data for comparison
- **Process** (Analytics, Visualization)
  - Analytics algorithm: Pattern recognition, Clustering, Blast, Collaborative Filtering, learning network, outlier detection ….
  - Pleasingly parallel or not
  - Flops per I/O byte
  - Communication/Interconnection needs
- **Produce output**
  - Often output much lower size
- Then possibly ***iterate process*** by further analysis of produced output
- Possibly ***share data*** from different sources