

SAND2013-10490P

Science Application Big Data Analytics and the Relations to Other Big Data Applications

Jay Lofstead

Scalable System Software

Sandia National Laboratories

Albuquerque, NM, USA

gflofst@sandia.gov

Texas Tech

December 10, 2013

Sandia National Laboratories

Exceptional service in the national interest

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

What is “Big Data”?

Popular current definition revolves around the three Vs:

- Velocity
  - Quick insights
- Variety
  - Web page hits and purchases on an e-commerce site
- Volume
  - Fast moving stream of tweets or Facebook posts
- and sometimes others like Value
  - Does this data stream offer potential useful data points if it can be correlated?

What is “Big Data”?

Typical characteristics:

- Small data items
  - Formats may vary depending on source, but each item small (< 1 KB)
- Sorting and correlating data to find patterns
  - “Map” to tag items and “Reduce” to sort them into buckets
- Timely relevance
  - “trending” news stories based on what people are saying or tagging
  - What items to promote to people viewing this item?

What is “Big Data”?

- What about data sets of 10s of TB of discrete items?
  - Data warehouses of transactional data
- What about a single data item of 1 PB or more?
  - World climate model at 1 km scale including multiple layers of atmosphere, land, ocean, and ice sheets
- What about years (decades) worth of data?
  - Long-term health studies
  - Climate data
  - Census data

Big Data of Old

Pre-21st century two extremes:

- Data Warehousing – LOTS of small items to correlate
  - Healthcare
  - Commerce (offline world)
- Computational Science – a small number of BIG data items
  - Simulations
  - Data analytics for experimental or observational data (or both simultaneously)

Data Warehousing

- Lots of data, but…

- Data fed from transactional sources to larger storage offline
  - Avoid impact on transactional system
- Data "normalized" into expected relational format
  - Format data into pre-defined fields for a range of queries

- Examples
  - For a particular diagnosis, what treatment regimens yield what outcomes at what costs, order by cost, success rate (drive treatment recommendations and reimbursement amounts)
  - What people buy particular products (direct marketing expenditures for direct mail campaign for a different product)

Computational Science

BIG codes that compute heavily for every output. Relatively small number of immense data "items"

- Climate simulations
  - Land, ocean, land ice, sea ice, and atmosphere models all interact to represent the global climate
- Fusion reactors
  - Model the magnetic containment field and plasma for generating clean energy on demand
- Materials
  - Shock physics for explosions
- Oil exploration
  - Simulation of ground + observational data to determine where to drill for best yield at lowest cost

Emerging "Big Data"

- Science data analysis sharing characteristics
  - Lots of small items
  - Independent evaluation

  - Biology-applications (e.g., genomics searches)
    - Partially "old" science application characteristics (large single data set)
    - Partially "new" big data (independent analysis of individual items)

Spectrum of Big Data

Big data items
Smaller number of items

Small individual items
Large numbers of items

Science Simulations

Biology-Style Science

Current "Big Data"

Data Warehousing

Science Analytics

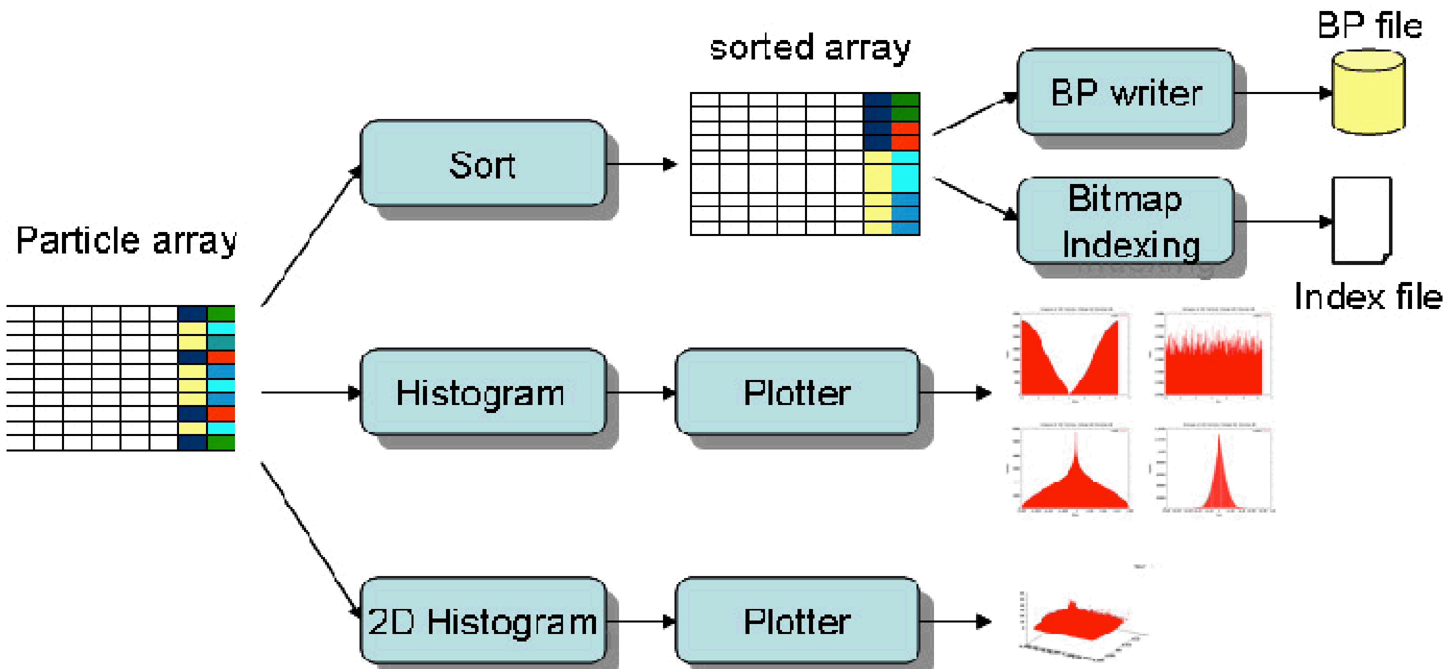
- Visualization
  - What are data ranges?
  - What is the rendering space compared to the simulation space?
  - What angle to view from?
  - How to render each data element?
  - What data elements to render on the same image?

Science Analytics

- Particle, fragment, or “energy” tracking
  - For particles that escape magnetic containment, how did they generate the velocity?
  - When a material explodes, how does it fragment and where do these fragments go and break apart over the simulation time?
  - Identify an atmospheric energy burst (storm) and where is it predicted to go
  - Trending for a value over time (when does the temperature in a combustion simulation reach the ignition temperature/pressure?)

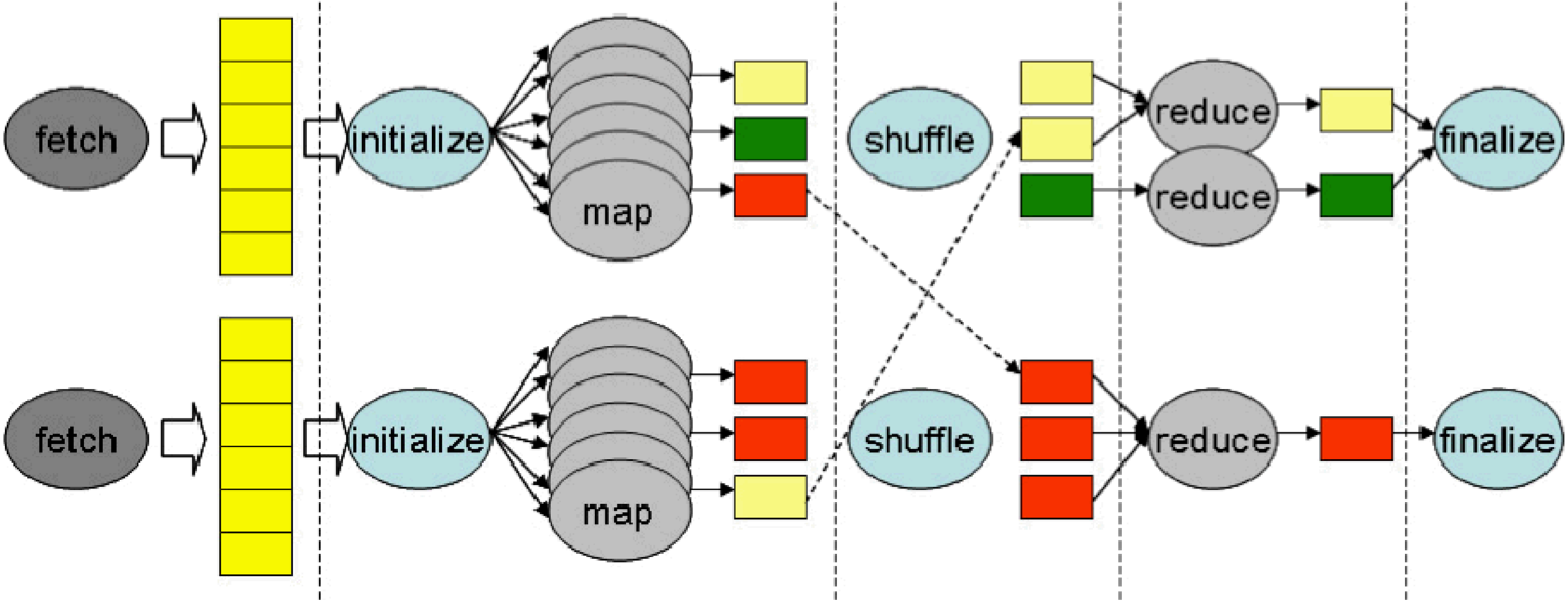
Science Analytics like Big Data

- What kinds of analysis can be done using big data tools?
  - Histograms
  - Merge sort of a data set
  - Bounding box data selection
  - Data indexing (bucket-based ranges, e.g., FastBit)



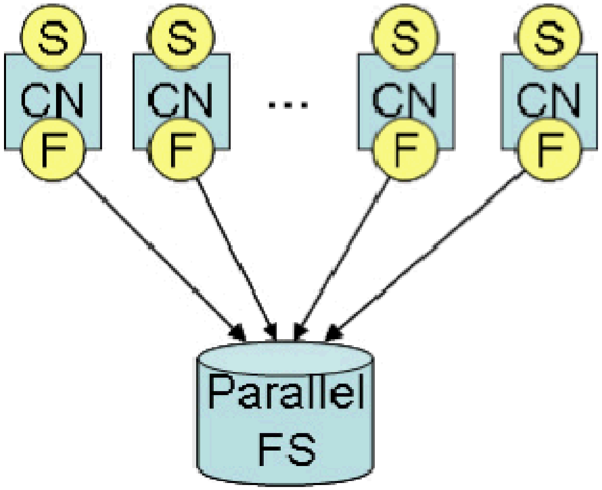
Workflow Structure

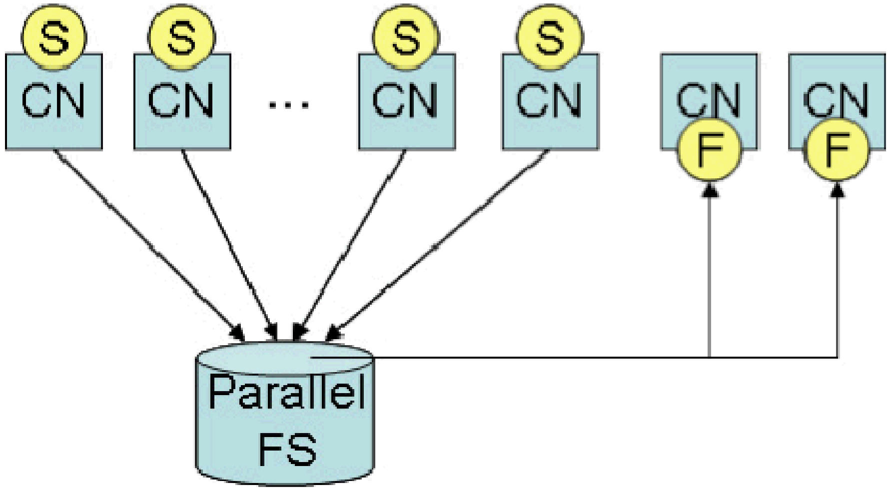
- Similar to Map-Reduce, but with constraints



When and Where to do Analytics?

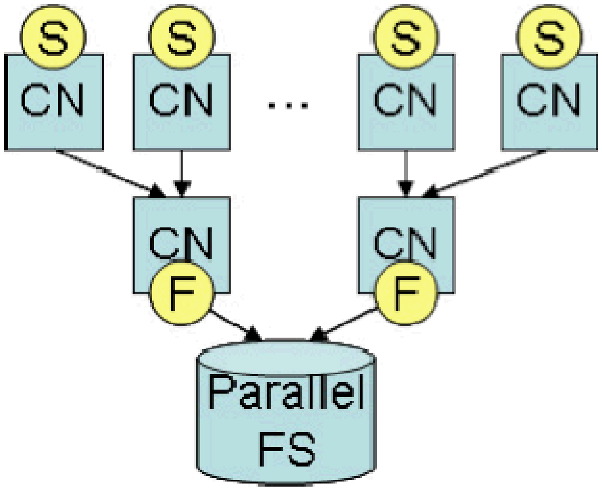
- In-Situ – in same process space as application
- Offline – based on data stored on persistent storage





When and Where to do Analytics?

- In-Transit – use resources on IO path to transform or annotate data as it moves



Fusion Science Today

- Physical experimental equipment: Joint European Torus
  - About 80% efficient
  - LOTS of diagnostic data
    - Visible and infrared cameras
    - Magnetic field and current measurement
    - X-ray monitors
    - Several others

- ITER (next generation)
  - 40 TB data per shot (400 seconds long) every 1600 seconds

- Gather data to validate and drive simulation to build next generation (may need to change simulation approach)

Fusion Science Today

- How do we capture 40 TB in 400 seconds?
- Can we process it in 1600 seconds to make adjustments or determine if maintenance is required?

Shock Physics

- Explosion modeling for anything from a pipe bomb to a nuclear warhead
- Model is a collection of atoms at particular distances from each other
- Simulation progresses as energy (or ignition) added to see how material behaves

- Ideal situation: See how fragments are formed and track their movement through life of simulation
- Problem: IO bandwidth woefully inadequate today

Shock Physics

- In-Situ problematic
  - 30 MB for just CTH, 300 MB CTH + ParaView
  - CTH scales fine to 100K processes, ParaView, not so much

- Offline impossible
  - I/O time would exceed computation time just to output sufficient data
  - Does not include reading time, analysis time, or writing time for fragments

Shock Physics

- In Transit
  - Decouple scalability issues
  - Absorb data at network rather than storage speeds
  - Process data down into fragments and write fragments (MUCH smaller data)
  - Enables fragment tracking!

What is Wrong with In Transit?

- In Situ – fast, clear transfer/communication of data.
  - If simulation fails, analysis fails (or vice versa)
- Offline – disk offers persistence and semantics for data processing, but slow
  - Failures just require restarting failed component

- Must address limitations of both to get In Transit acceptable:
  1. What is equivalent to a file when doing a network transfer (MxN)?
  2. How to control access to data prematurely?
  3. How to evolve data reliably?
  4. What should the interfaces be?

Storage Semantics for In Transit

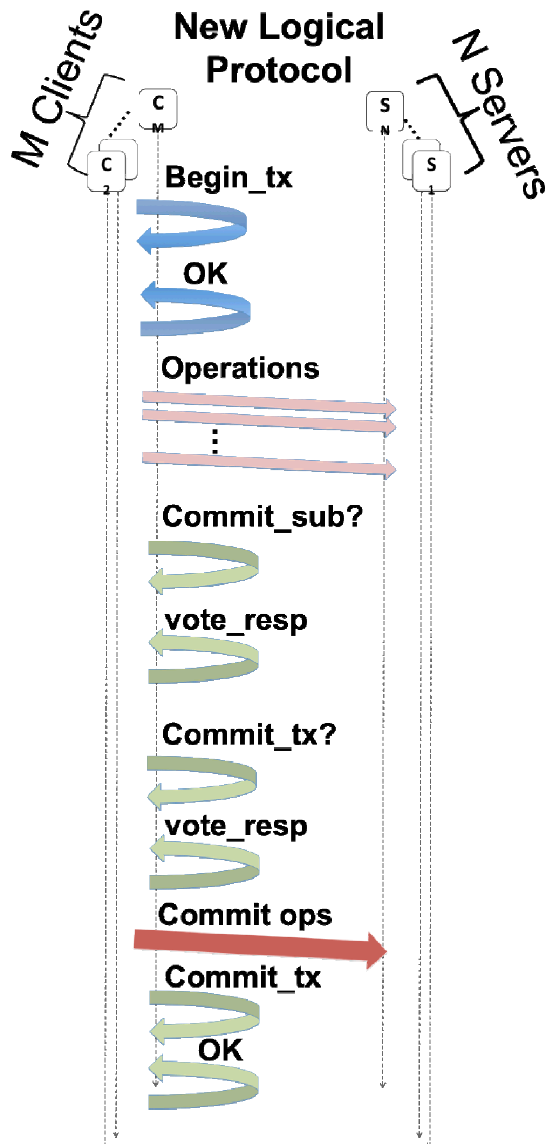
- Storage Semantics worth replicating
  - “blocking” of data into a data set
  - Controlled access to a data set
  - Persistence
- Storage Semantics worth avoiding
  - Delays in IO stack preventing knowledge of “real” completion of IO task

Transactions

- Yes, just like from databases—but much harder
- Problems:
  - We have parallel clients and multiple servers (MxN)
  - Must incorporate black box servers
  - We have several data items in a single set, all of which are distributed

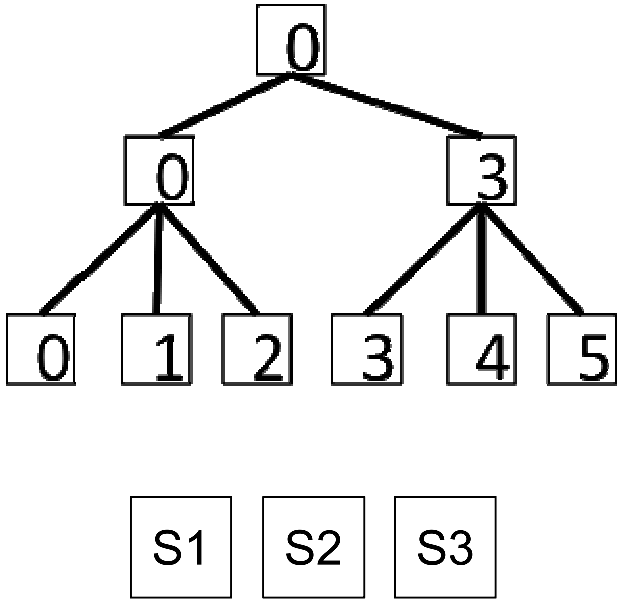
D^{2}T – Doubly Distributed Txns

- Revised protocol
- Sub transactions for each item (variable)
- Overall transaction for overall blocking
- Low requirements for servers to participate
- Client-side only coordination



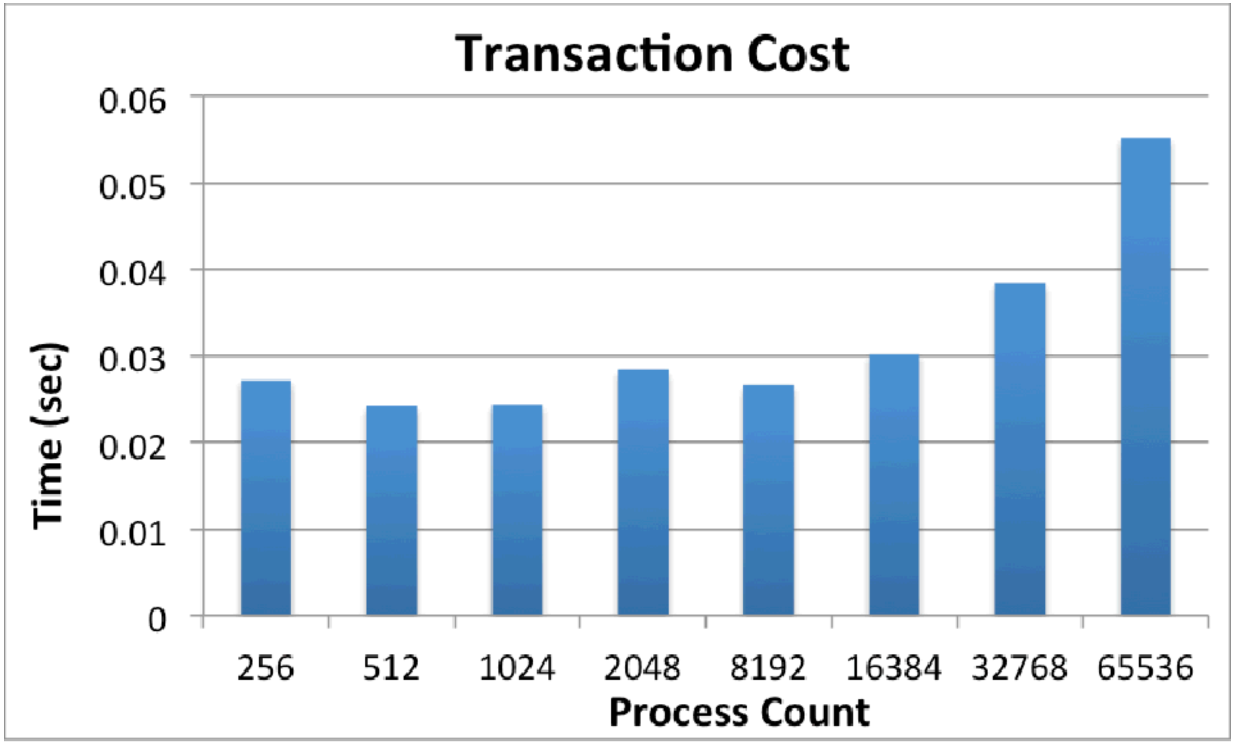
D^2T Structure

- Multiple roles for some processes
- 0 is coordinator, sub-coordinator, and subordinate
- 3 is sub-coordinator and subordinate
- 1, 2, 4, 5 are all just subordinates
- S1, S2, S3 are servers



Transaction Performance

- Notes:
  - Always used at least 2 sub-coordinators to slow it down
  - Added a sub-coordinator when subordinate count exceeded 256
  - 64K processes = 256 sub-coordinators with 256 subordinates each
  - Overhead only for complete set of transaction calls (no op. costs)



What About Persistence?

- Persistence for data staging is an open problem
- What role will new memory types play?

Harder Problems

- Coupling the physics
  - What are the time and space scales for simulations and analysis?
  - Is the simulation written to accept a partner computation?
  - What about only in one direction?
  - How often can data be exchanged?
  - How often should data be exchanged?
  - How to handle the interface zones where meshes (or other simulation data organizations) meet?
  - How does the physics change in the presence of the new data?

Big Data Tools and Science Analytics

- Science simulation analytics may work with some big data tools (Sci-Hadoop)
- Some scale out techniques useful for scale up
  - process everything possible in parallel isolation and only coordinate in a tightly integrated, small area
  - Incorporate resilience as a fundamental feature of any workflow

Questions?