National Aeronautics and Space Administration

Applying Apache Hadoop to NASA's Big Climate Data

Use Cases and Lessons Learned

Glenn Tamkin (NASA/CSC)

Team: John Schnase (NASA/PI), Dan Duffy (NASA/CO), Hoot Thompson (PTP), Denis Nadeau (CSC), Scott Sinno (PTP), Savannah Strong (CSC)

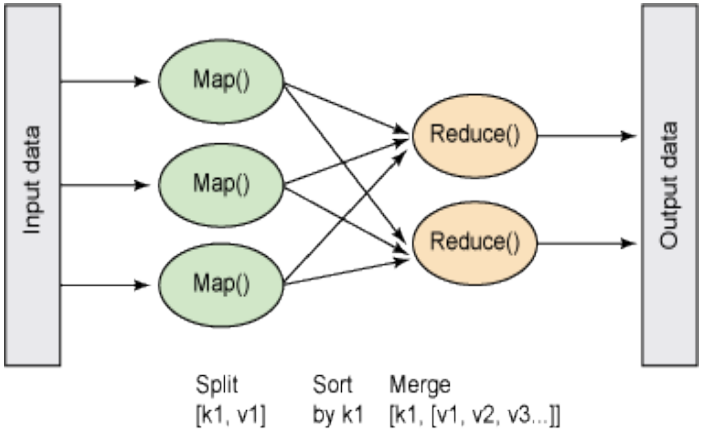
Overview

- The NASA Center for Climate Simulation (NCCS) is using Apache Hadoop for high-performance analytics because it optimizes computer clusters and combines distributed storage of large data sets with parallel computation.
- We have built a platform for developing new climate analysis capabilities with Hadoop.

The Apache Software Foundation

Solution

- Hadoop is well known for text-based problems. Our scenario involves binary data. So, we created custom Java applications to read/write data during the MapReduce process.
- Our solution is different because it: a) uses a custom composite key design for fast data access, and b) utilizes the Hadoop Bloom filter, a data structure designed to identify rapidly and memory-efficiently whether an element is present.



Why HDFS and MapReduce ?

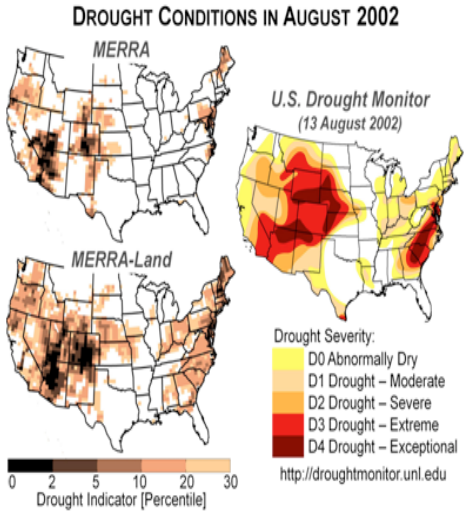
- Software framework to store large amounts of data in parallel across a cluster of nodes
  - Provides fault tolerance, load balancing, and parallelization by replicating data across nodes
  - Co-locates the stored data with computational capability to act on the data (storage nodes and compute nodes are the same – typically)
  - A MapReduce job takes the requested operation and maps it to the appropriate nodes for computation using specified keys

Who uses this technology?
- Google
- Yahoo
- Facebook

Many PBs and probably even EBs of data.

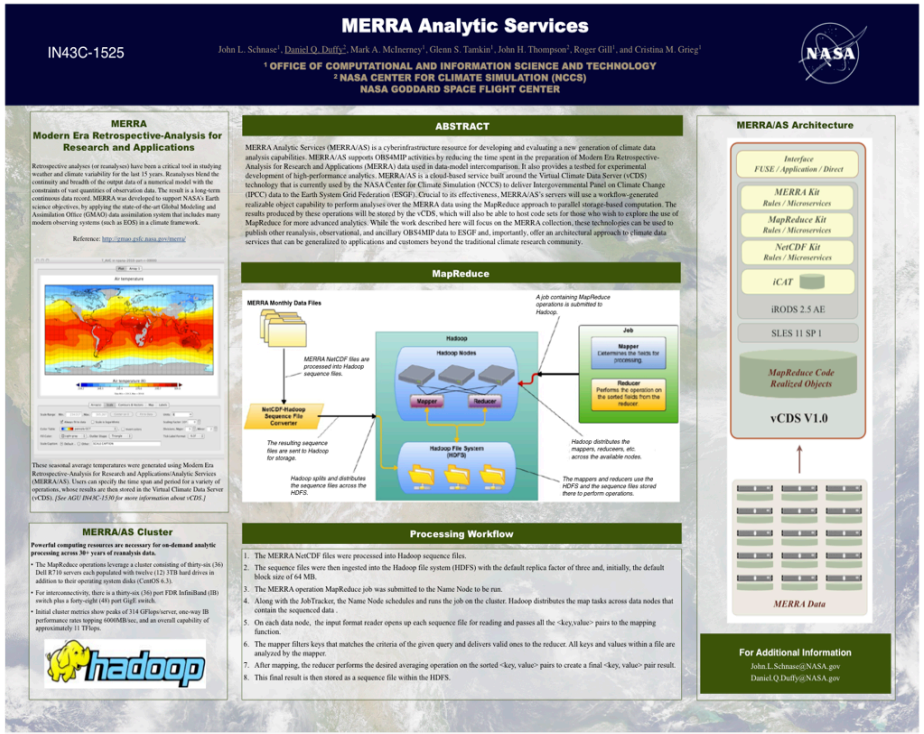
Background

- Scientific data services are a critical aspect of the NASA Center for Climate Simulation's mission (NCCS). Modern Era Retrospective-Analysis for Research and Applications Analytic Services (MERRA/AS) …
  - Is a cyber-infrastructure resource for developing and evaluating a next generation of climate data analysis capabilities
  - A service that reduces the time spent in the preparation of MERRA data used in data-model inter-comparison



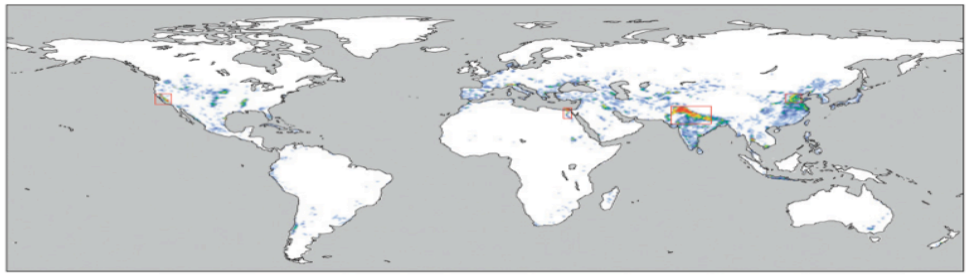
Vision

- Provide a test-bed for experimental development of high-performance analytics
- Offer an architectural approach to climate data services that can be generalized to applications and customers beyond the traditional climate research community



Example Use Case - WEI Experiment

- Wei team used MERRA data to study four intensively irrigated regions: northern India/Pakistan, the North China Plain, the California Central Valley, and the Nile Valley.
- Seasonal rates of evapotranspiration with and without irrigation over the studied areas were then compared to assess the impact of irrigation.
- The data required for these calculations include average daily precipitation, evapotranspiration, temperature, humidity, and wind at different tropospheric levels at six-hourly time steps from 1979 to 2002.
- This early-stage data reduction—average values for environmental variables over specific spatiotemporal extents—is the type of data assembly that historically has been performed on the scientist's workstation after transfers from public archives of large blocks of data.



THE UNIVERSITY OF TEXAS AT AUSTIN

JACKSON SCHOOL OF GEOSCIENCES

FEBRUARY 2013 WEI ET AL.

Where Does the Irrigation Water Go? An Estimate of the Contribution of Irrigation to Precipitation Using MERRA

JIANGFENG WEI*
Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland

PAUL A. DIRMEYER
Department of Atmospheric, Oceanic and Earth Sciences, George Mason University, Fairfax, Virginia, and Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland

DOMINIK WISSER
Department of Physical Geography, Utrecht University, Utrecht, Netherlands

MICHAEL G. BOSILOVICH
Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

DAVID M. MOCKO
SAIC and Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

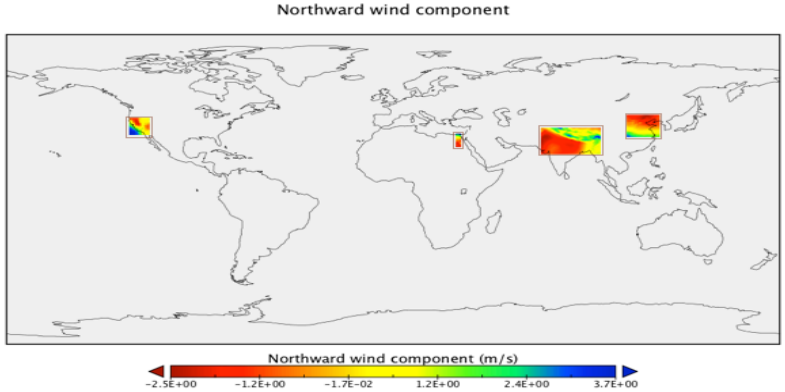
(Manuscript received 24 May 2012, in final form 21 September 2012)

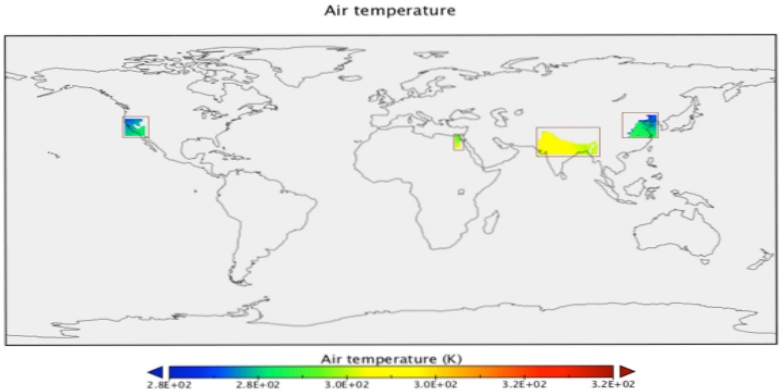
ABSTRACT

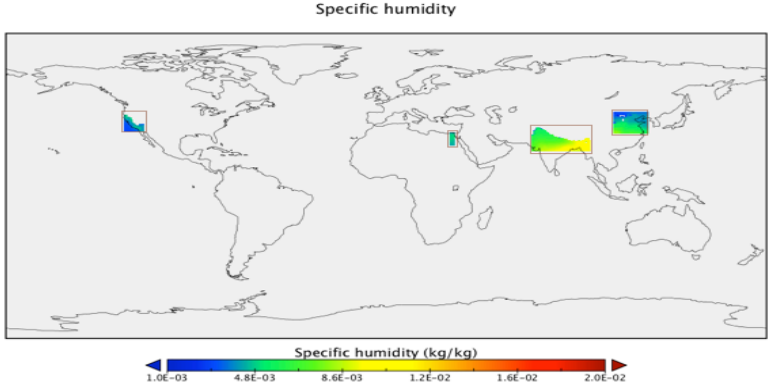
Irrigation is an important human activity that may impact local and regional climate, but current climate model simulations and data assimilation systems generally do not explicitly include it. The European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim) shows more irrigation signal in surface evapotranspiration (ET) than the Modern-Era Retrospective Analysis for Research and Applications (MERRA) because ERA-Interim adjusts soil moisture according to the observed surface temperature and humidity while MERRA has no explicit consideration of irrigation at the surface. But, when compared with the results from a hydrological model with detailed considerations of agriculture, the ET from both reanalyses show large deficiencies in capturing the impact of irrigation. Here, a back-trajectory method is used to estimate the contribution of irrigation to precipitation over local and surrounding regions, using MERRA with observation-based corrections and added irrigation-caused ET increase from the hydrological model. Results show substantial contributions of irrigation to precipitation over heavily irrigated regions in Asia, but the precipitation increase is much less than the ET increase over most areas, indicating that irrigation could lead to water deficits over these regions. For the same increase in ET, precipitation increases are larger over wetter areas where convection is more easily triggered, but the percentage increase in precipitation is similar for different areas. There are substantial regional differences in the patterns of irrigation impact, but, for all the studied regions, the highest percentage contribution to precipitation is over local land.

Wei, J., Dirmeyer, P. A., Wisser, D., Bosilovich, M. G., & Mocko, D. M. (2013). Where does irrigation water go? An estimate of the contribution of irrigation to precipitation using MERRA. *Journal of Hydrometeorology*, 14(2), 271–289.

Example Use Case - WEI Experiment







Wei, et al.

- ~8.4 TB transferred from archive to local workstation (weeks)
- Clipping, averaging performed by Fortran program on local workstation (days)

MERRA/AS

- Clipping, averaging performed by MERRA/AS (~28 hrs)
- Only ~35 GB final product transferred to local workstation (minutes)

- Significant time savings in data wrangling,
- rapid screening over monthly means files takes minutes, and
- there's a possibility of folding Dr. Wei's modeling algorithm back into the CDS API …

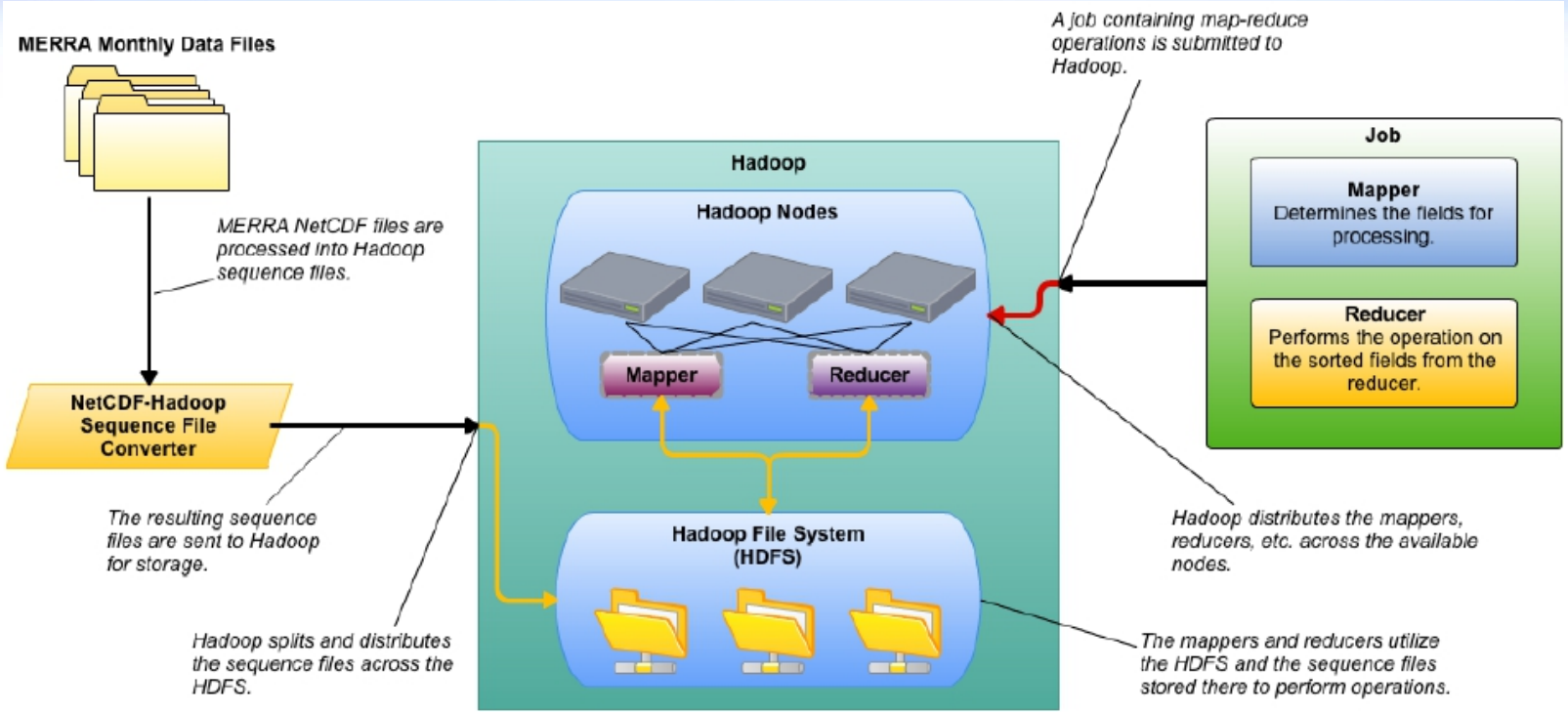
MERRA Data

- The GEOS-5 MERRA products are divided into 25 collections: 18 standard products, 7 chemistry products
- Comprise monthly means files and daily files at six-hour intervals running from 1979 – 2012
- Total size of NetCDF MERRA collection in a standard filesystem is **~80 TB**
- One file per month/day produced with file sizes ranging from ~20 MB to ~1.5 GB

Name	Description	Size Gbytes/day // Tbytes
const_2d_asm_Nx	Constant fields	
inst6_3d_ana_Nv	Analyzed fields on model layers	0.452
inst6_3d_ana_Np	Analyzed fields at pressure levels	0.291
inst3_3d_asm_Cp	Basic assimilated fields from IAU corrector	0.231
tavg3_3d_cld_Cp	Upper-air cloud related diagnostics	0.075
tavg3_3d_mst_Cp	Upper-air diagnostics from moist processes	0.056
tavg3_3d_trb_Cp	Upper-air diagnostics from turbulence	0.147
tavg3_3d_rad_Cp	Upper-air diagnostics from radiation	0.088
tavg3_3d_tdt_Cp	Upper-air temperature tendencies by process	0.191
tavg3_3d_udt_Cp	Upper-air wind tendencies by process	0.224
tavg3_3d_qdt_Cp	Upper-air humidity tendencies by process	0.166
tavg3_3d_odt_Cp	Upper-air ozone tendencies by process	0.083
tavg1_2d_slv_Nx	Single-level atmospheric state variables	0.285
tavg1_2d_flx_Nx	Surface turbulent fluxes and related quantities	0.267
tavg1_2d_rad_Nx	Surface and TOA radiative fluxes	0.189
tavg1_2d_lnd_Nx	Land related surface quantities	0. 146
tavg1_2d_int_Nx	Vertical integrals of tendencies	1.500
inst1_2d_int_Nx	Vertical integrals of quantities	0.115
TOTAL		4.506 // 49.6

Name	Description	Size (Gbytes)
const_2d_chm_Fx	2-D invariants on chemistry grid	
tavg3_3d_chm_Fv	Chemistry related 3-D at model layer centers	0.329
tavg3_3d_chm_Fe	Chemistry related 3-D at model layer edges	0.166
tavg3_2d_chm_Fx	Chemistry related 2-DSingle-level	0.020
tavg3_3d_chm_Nv	Accumulated transport fields at layers	0.915
tavg3_3d_chm_Ne	Accumulated transport fields at edges	0.469
inst3_3d_chm_Ne	Instantaneous fields for off-line transport	0.050
TOTAL CHEM		1.949 // 21.44

Map Reduce Workflow

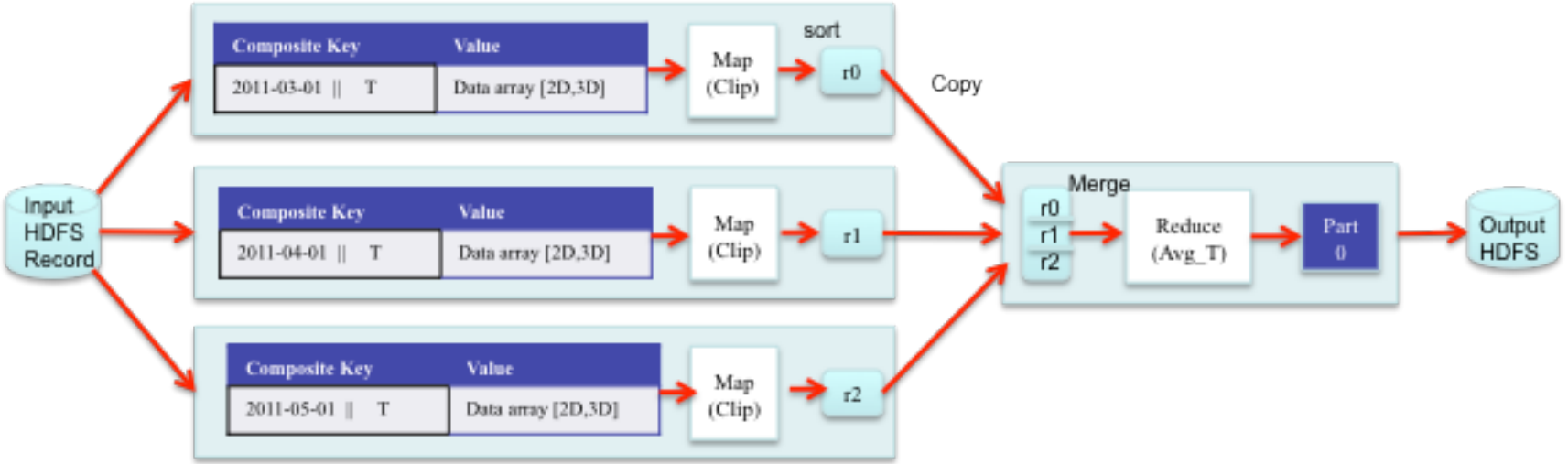


Ingesting MERRA data into HDFS

- Option 1: Put the MERRA data into Hadoop with no changes
  - » Would require us to write a custom mapper to parse
- Option 2: Write a custom NetCDF to Hadoop sequencer and keep the files together
  - » Basically puts indexes into the files so Hadoop can parse by key
  - » Maintains the NetCDF metadata for each file
- Option 3: Write a custom NetCDF to Hadoop sequencer and split the files apart (allows smaller block sizes)
  - » Breaks the connection of the NetCDF metadata to the data
- Chose Option 2

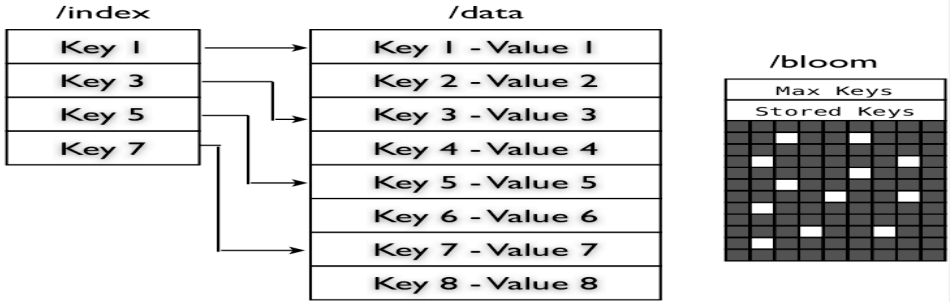
Sequence File Format

- During sequencing, the data is partitioned by time, so that each record in the sequence file contains the timestamp and name of the parameter (e.g. temperature) as the composite key and the value of the parameter (which could have 1 to 3 spatial dimensions)



Bloom Filter

- A Bloom filter, conceived by Burton Howard Bloom in 1970, is a space-efficient probabilistic data structure that is used to test whether an element is a member of a set. False positive retrieval results are possible, but false negatives are not; i.e. a query returns either "inside set (may be wrong)" or "definitely not in set".
- In Hadoop terms, the BloomMapFile can be thought of as an enhanced MapFile because it contains an additional hash table that leverages the existing indexes when seeking data.



Bloom Filter Performance Increase

- The original MapReduce application utilized standard Hadoop Sequence Files. Later they were modified to support three different formats called Sequence, Map, and Bloom.
- Dramatic performance increases were observed with the addition of the Bloom filter (~30-80%).

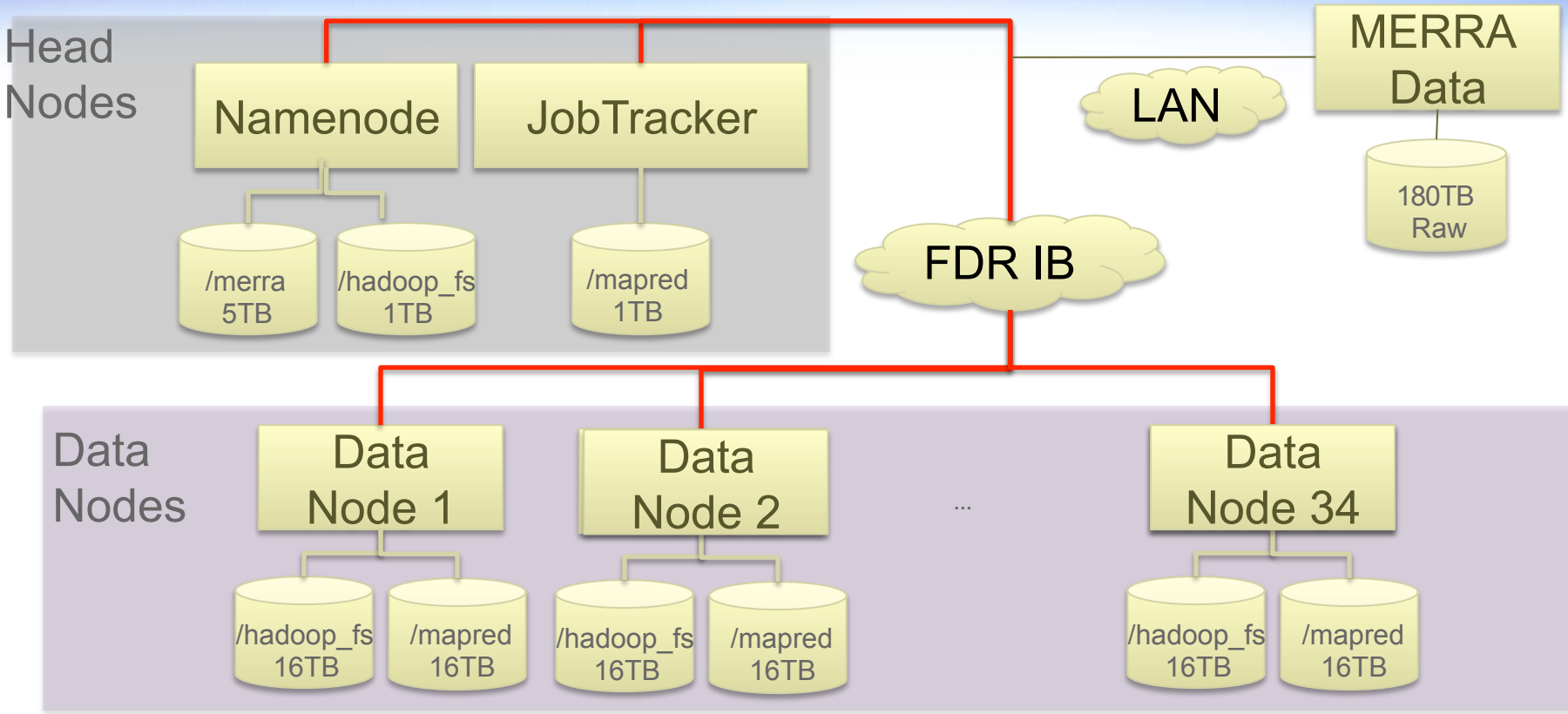
Job Description	Host	Sequence (sec)	Map (sec)	Bloom (sec)	Percent Increase
Read a single parameter (“T”) from a single sequenced monthly means file	Standalone VM	6.1	1.2	1.1	+81.9%
Single MR job across 4 months of data seeking “T” (period = 2)	Standalone VM	204	67	36	+82.3%
Generate sequence file from a single MM file	Standalone VM	39	41	51	-30.7%
Single MR job across 4 months of data seeking “T” (period = 2)	Cluster	31	46	22	+29.0%
Single MR job across 12 months of data seeking “T” (period = 3)	Cluster	49	59	36	+26.5%

Data Set Descriptions

- **Two data sets**
  - MAIMNPANA.5.2.0 (instM_3d_ana_Np) – monthly means
  - MAIMCPASM.5.2.0 (instM_3d_asm_Cp) – monthly means
- **Common characteristics**
  - Spans years 1979 through 2012…..
  - Two files per year (hdf, xml), 396 total files
- **Sizing**

Type	Raw Total (GB)	Sequenced Total (GB)	Raw File (MB)	Sequenced File (MB)	Sequence Time (sec)
MAIMNPANA	84	224	237	565	30
MAIMCPASM	48	119	130	300	15

MERRA Cluster Components

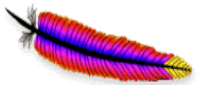


Operational Node Configurations

Configuration	Bare1
Node	Dell R720
Processor Type	Intel Sandy Bridge
Processor Number	E5-2670
Processor Speed	2.60 GHz
Cores per Socket	8
Number of Sockets	2
Cores per Node	16
Main Memory	32 GB
Storage	12 by 3 TB drives = 36 TB RAW
Interconnect	Mellanox MT27500 FDR IB
Operating System	Centos 6.3
Kernel	2.6.32-279.5.1
Hadoop	0.20.2
java-6-sun	1.6.0_24

Other Apache Contributions…

- Avro – a data serialization system
- Maven – a tool for building and managing Java-based projects
- Commons – a project focused on all aspects of reusable Java components
  - Lang – provides methods for manipulation of core Java classes
  - I/O - a library of utilities to assist with developing IO functionality
  - CLI - an API for parsing command line options passed to programs
  - Math - a library of mathematics and statistics components
- Subversion – a version control system
- Log4j - a framework for logging application debugging messages



Apache Commons™
http://commons.apache.org/

SUBVERSION®

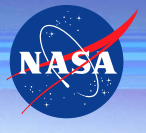
Other Open Source Tools…

- Using Cloudera (CDH), the open source enterprise-ready distribution of Apache Hadoop.
  - Cloudera is integrated with configuration and administration tools and related open source packages, such as Hue, Oozie, Zookeeper, and Impala.
  - Cloudera Manager Free Edition is particularly useful for cluster management, providing centralized administration of CDH.

Next Steps

- Tune the MapReduce Framework
- Try different ways to sequence the files
- Experiment with data accelerators
- Explore real-time querying services on top of the Hadoop file system:
  - Apache Drill
  - Impala (Cloudera)
  - Ceph,
  - MapR…

Conclusions and Lessons Learned



- Design of sequence file format is critical for big binary data
- Configuration is key…change only one parameter at a time for tuning
- Big data is hard, and it takes a long time….
- Expect things to fail – a lot
- Hadoop craves bandwidth
- HDFS installs easy but optimizing is not so easy
- Not as fast as we thought … is there something in Hadoop that we don't understand yet
- Ask the mailing list or your support provider