Using Kerberized Lustre over the WAN for High Energy Physics Data

Josephine Palencia, Robert Budden, Kathy Benninger, Dimitri Bourilkov, Paul Avery, Mengxing Cheng, Yu Fu, Bockjoo Kim, Dave Dykstra, Nirmal Seenu, Jorge Rodriquez, John Dilascio, Donald Shrum, Jorge Wilgenbusch, Drew Oliver, Daniel Majchrzak

Table of Contents

OSG/ExTENCI Background

- ❑ LHC and Experiments
- ❑ CMS Global Data Grid
- ❑ Data Tiers

ExTENCI Lustre WAN

- ❑ Kerberos/Realm, Authenticated Lustre components
- ❑ Packaged images for Kerberized VM Clients
- ❑ Default kerberos flavor
- ❑ Secure, Distributed OSTs/OST pools

High Energy Physics Applications

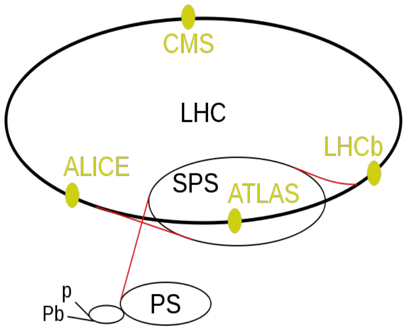
- ❑ CMS (Compact Muon Solenoid), ROOT
- ❑ LQCD
- ❑ ATLAS
- ❑ CVMFS (CERN-VM Filesystem)

USQCD US Lattice Quantum Chromodynamics

Issues/Future Work

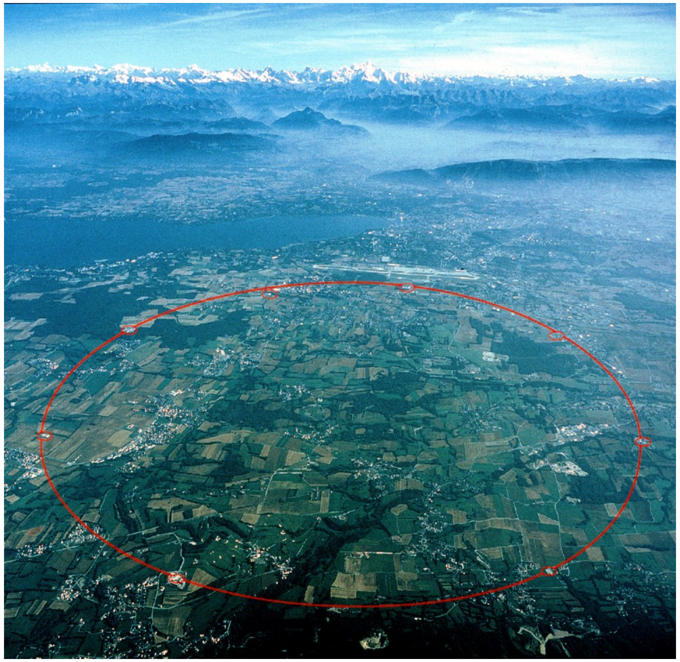
References/Addendum

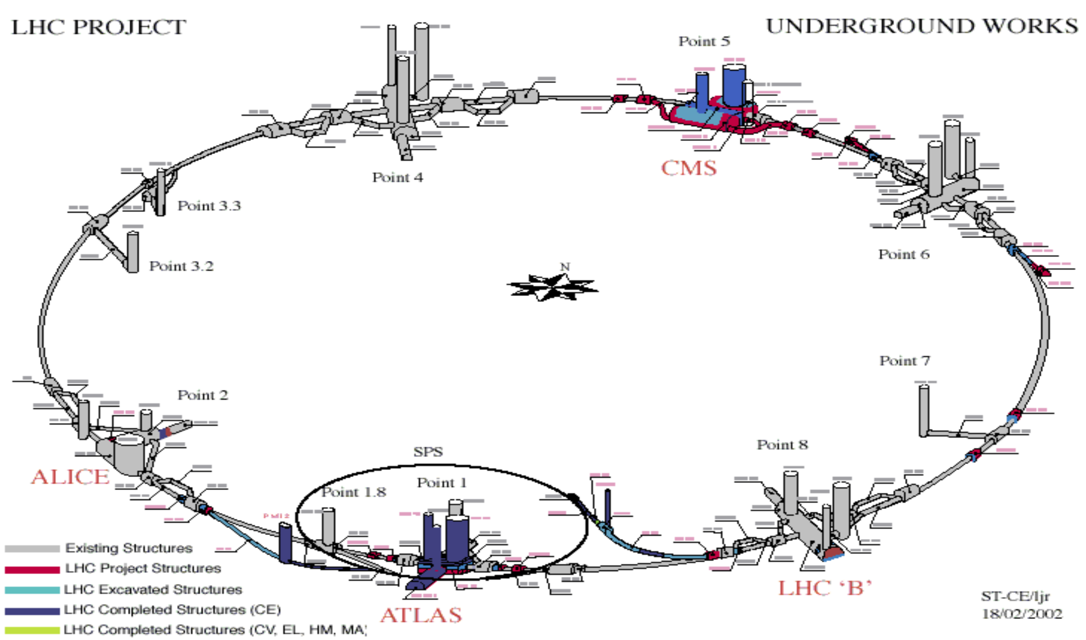
OSG/ExTENCI Background: LHC



Large Hadron Collider (LHC)

- Largest high-energy particle accelerator built (98-08) by CERN near Geneva b/w Switzerland & France : 27km
- 2 beams of hadrons (Pb ions or protons) travel in opposite directions inside to recreate conditions after BB by colliding two beams head-on at very high energy
- Addresses some of the fundamental questions in the laws of high energy/particle physics
- 9000+ physicists, 250+ institutes, 60+ countries

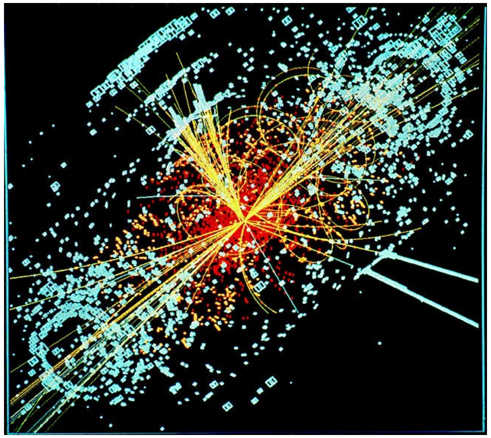




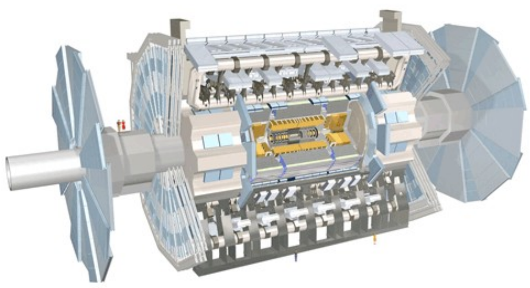
LHC (CMS, ATLAS)

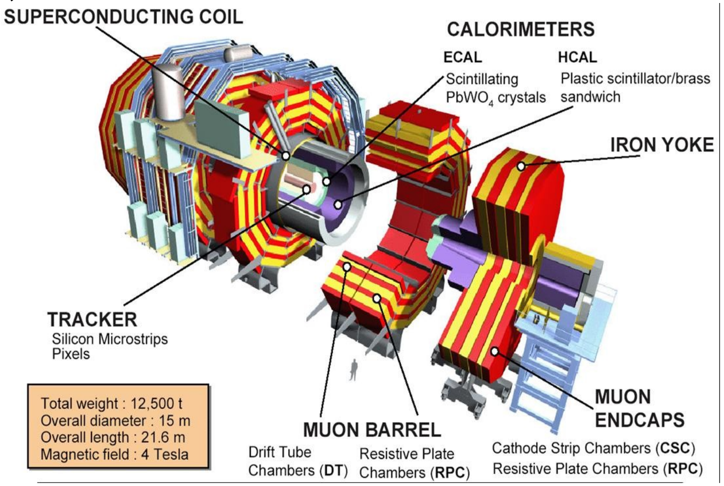
2/6 Experiments

- ❑ **CMS:** uses general purpose detector to investigate **Higgs boson, extra dimensions** and **dark matter particles**
  - Record sets of measurements of **particles** created in **collisions- path, energies** and **identities**
  - 3600 scientists, 38 countries, 183 institutes
- ❑ **ATLAS**: also uses general purpose detector; same goal different method
  - 3000 scientists, 57 countries, 174 institutes



Simulation of how a Higg's boson can appear in the CMS detector at CERN

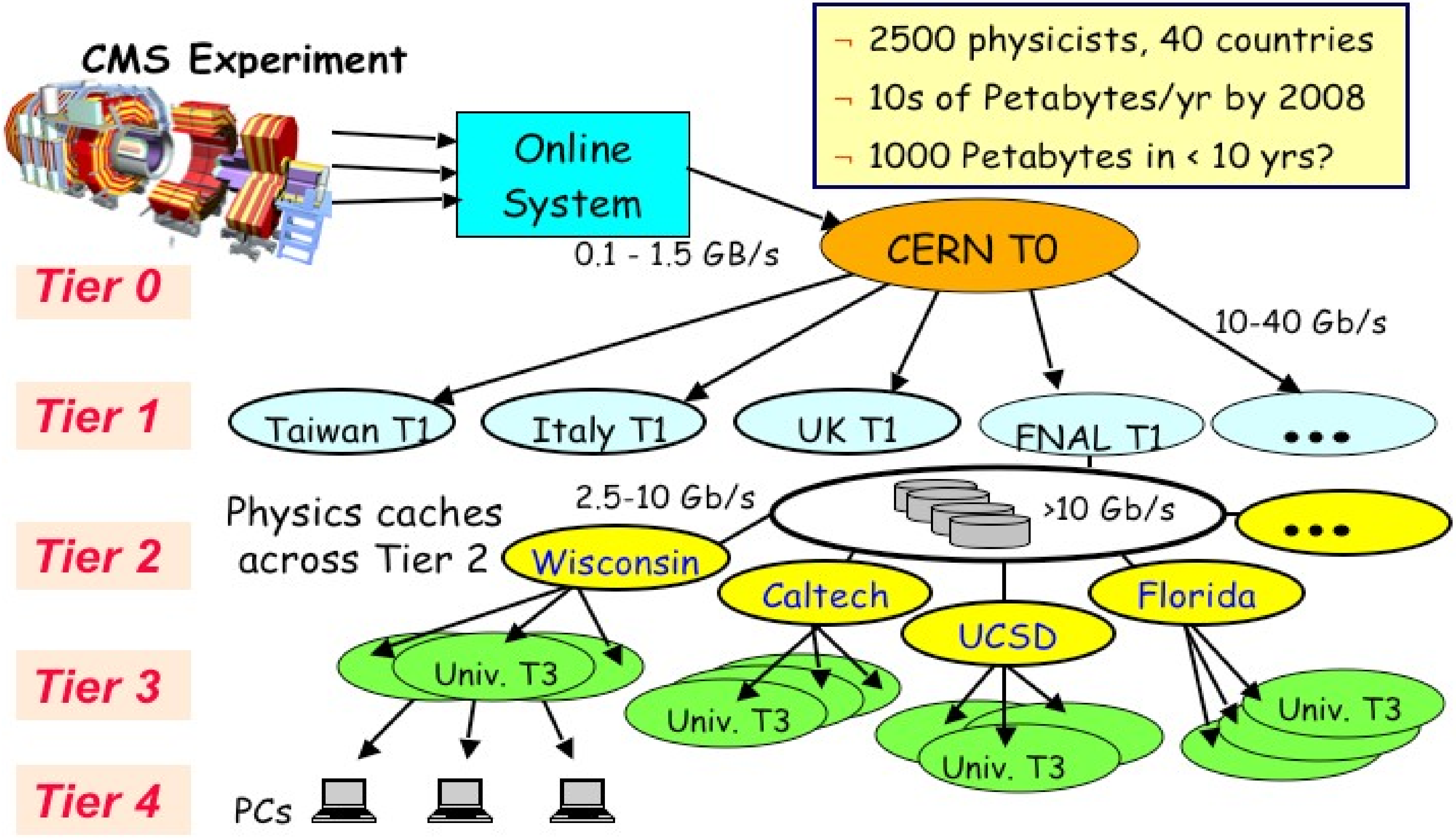




ATLAS DETECTOR
Size: 46 m long, 25 m high and 25 m wide
-**largest volume particle detector constructed.**
Weight: 7000 tonnes
Design: barrel plus end caps
Location: Meyrin, Switzerland

CMS Detector
Size: 21 m long, 15 m wide and 15 m high.
Weight: 12 500 tonnes
Design: barrel plus end caps
Location: Cessy, France

CMS Global Data Grid



Project Partners and Goal

ExTENCI with the OSG: NSF Grant 1007115

Extending Science Through Enhanced National Cyberinfrastructure with the Open Science Grid

Create a secure, distributed filesystem over the WAN that allows access to remote applications data for analysis by scientists at CERN Tier3/4 sites having none or very limited resources and sysadmin personnel.

PITTSBURGH SUPERCOMPUTING CENTER

J. Palencia, R. Budden, K. Benninger

UNIVERSITY of FLORIDA

D. Bourilkov, P. Avery, M. Cheng, Y. Fu, B. Kim

Fermilab

D. Dykstra, N. Seenu

FIU FLORIDA INTERNATIONAL UNIVERSITY

J. Rodriquez, J. Dilascio

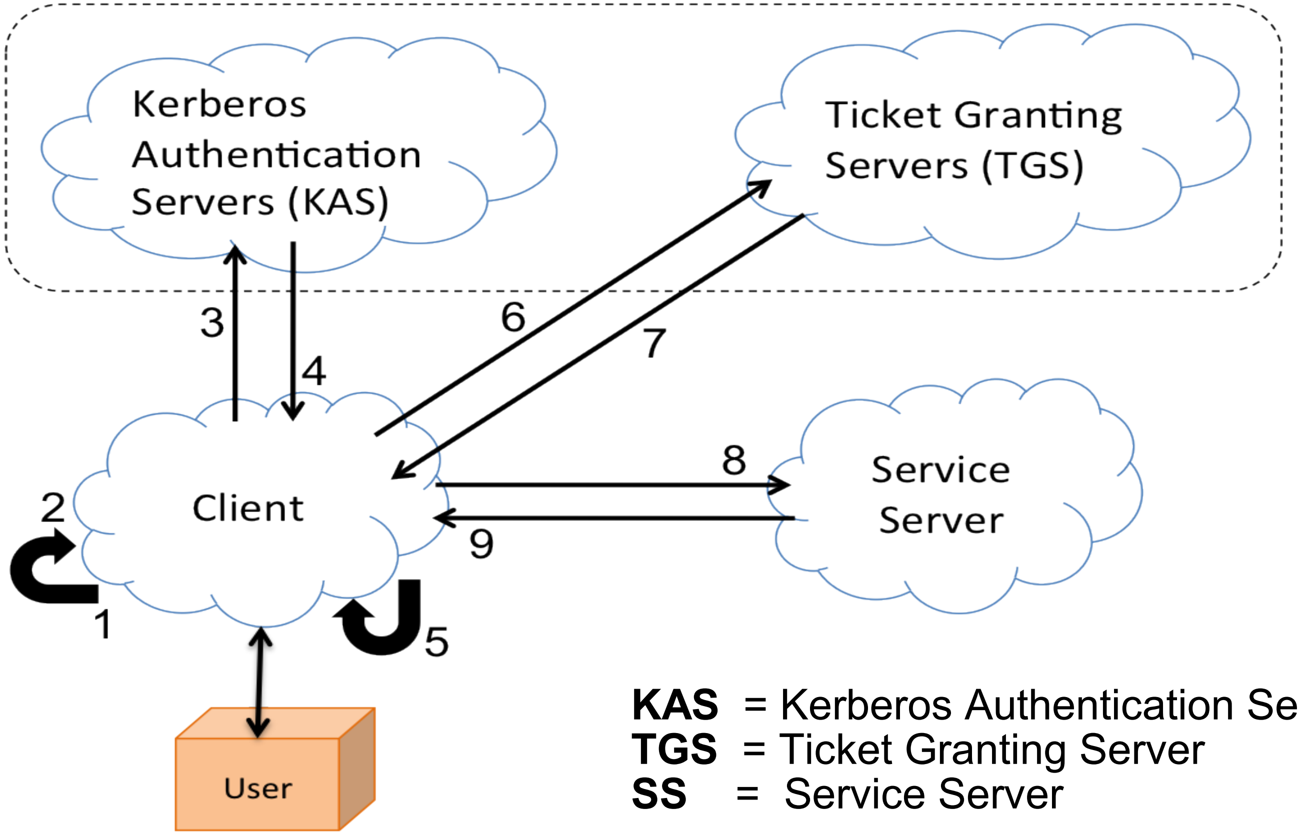
THE FLORIDA STATE UNIVERSITY

D. Shrum, J. Wilgenbusch

USF UNIVERSITY OF SOUTH FLORIDA

D. Oliver, D. Majchrzak

Kerberos Primer

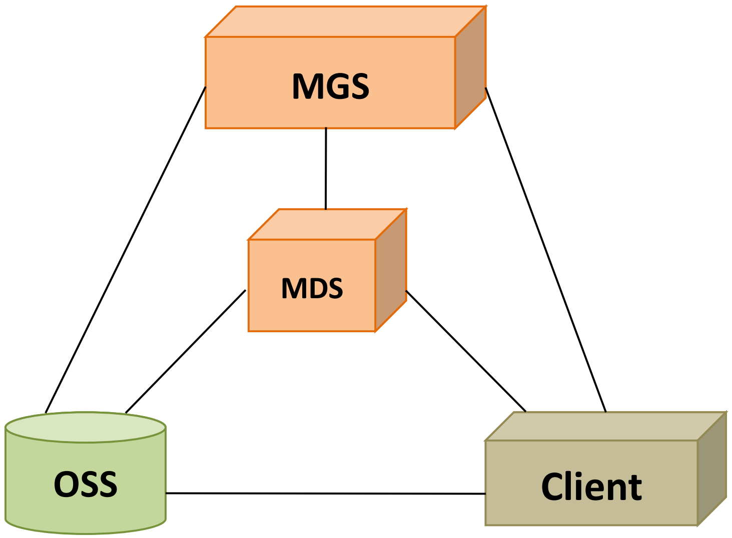


KAS = Kerberos Authentication Server
TGS = Ticket Granting Server
SS = Service Server

CLIENT

- Authenticates itself to KAS
- Demonstrates to TGS that it's authorized to receive a ticket for a service
- Demonstrates to SS that it's been approved to receive service

Authenticated Lustre Components



FLAVOR	AUTH	RPC MESSAGE PROTECTION	BULK DATA PROTECTION
lctl conf_param extenci.srpc.flavor.default = **krb5n**			
null		NULL	NULL
KRB5n	GSS/krb5	NULL	checksum(adler32)
KRB5a	GSS/krb5	PARTLY INTEGRITY	checksum(adler32)
KRB5i	GSS/krb5	INTEGRITY	integrity(sha1)
KRB5p	GSS/krb5	PRIVACY	privacy(sha1/aes128)

lctl conf_param extenci.srpc.flavor.**tcp0**=krb5n
extenci.srpc.flavor.**tcp1**=null
extenci.srpc.flavor.default.**cli2ost**=krb5i
extenci.srpc.flavor.default.**mdt2mdt**=null
extenci.srpc.flavor.default.**mdt2ost**=krb5i
mgs.srpc.flavor.default=krb5p

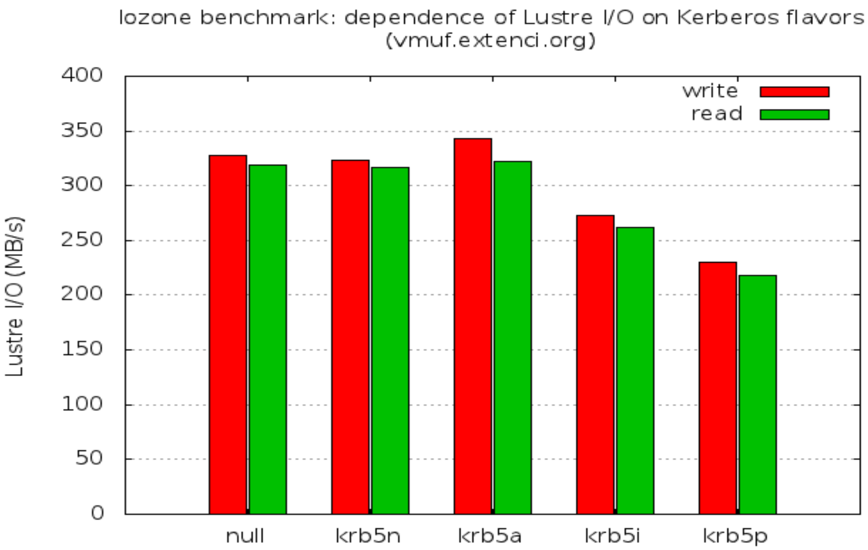
- ❑ Ease in bringing up secure lustre components
- ❑ Kerberos infrastructure is **NOT** required
- ❑ Each system is given a **UNIQUE** keytab
- ❑ Seconded by firewall (becomes **optional**)

Dependence of Lustre IO on Kerberos Flavors

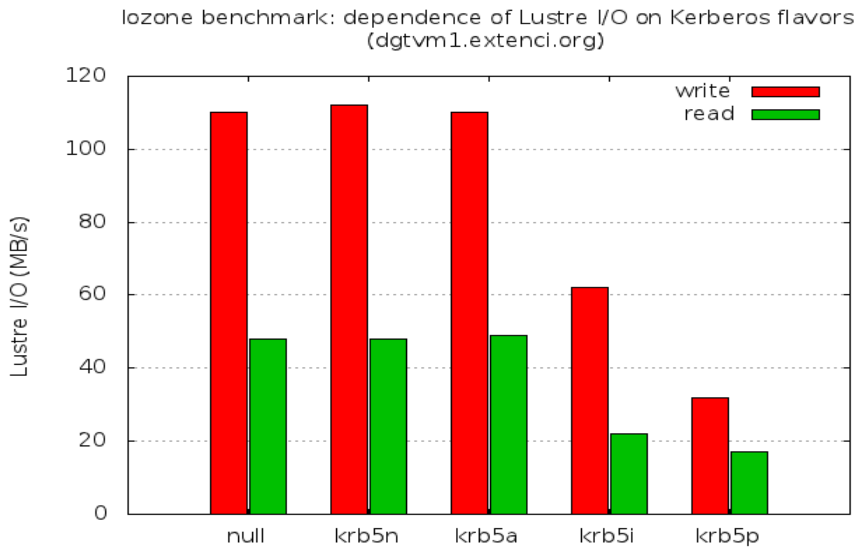
Kerberos flavor	null	**krb5n**	Krb5a	krb5i	krb5p
RPC security	null	null	header integrity	integrity	privacy
Bulk security	null	checksum	checksum	integrity	privacy

Iozone : one instance working on a non-striped file stored in Lustre filesystem.

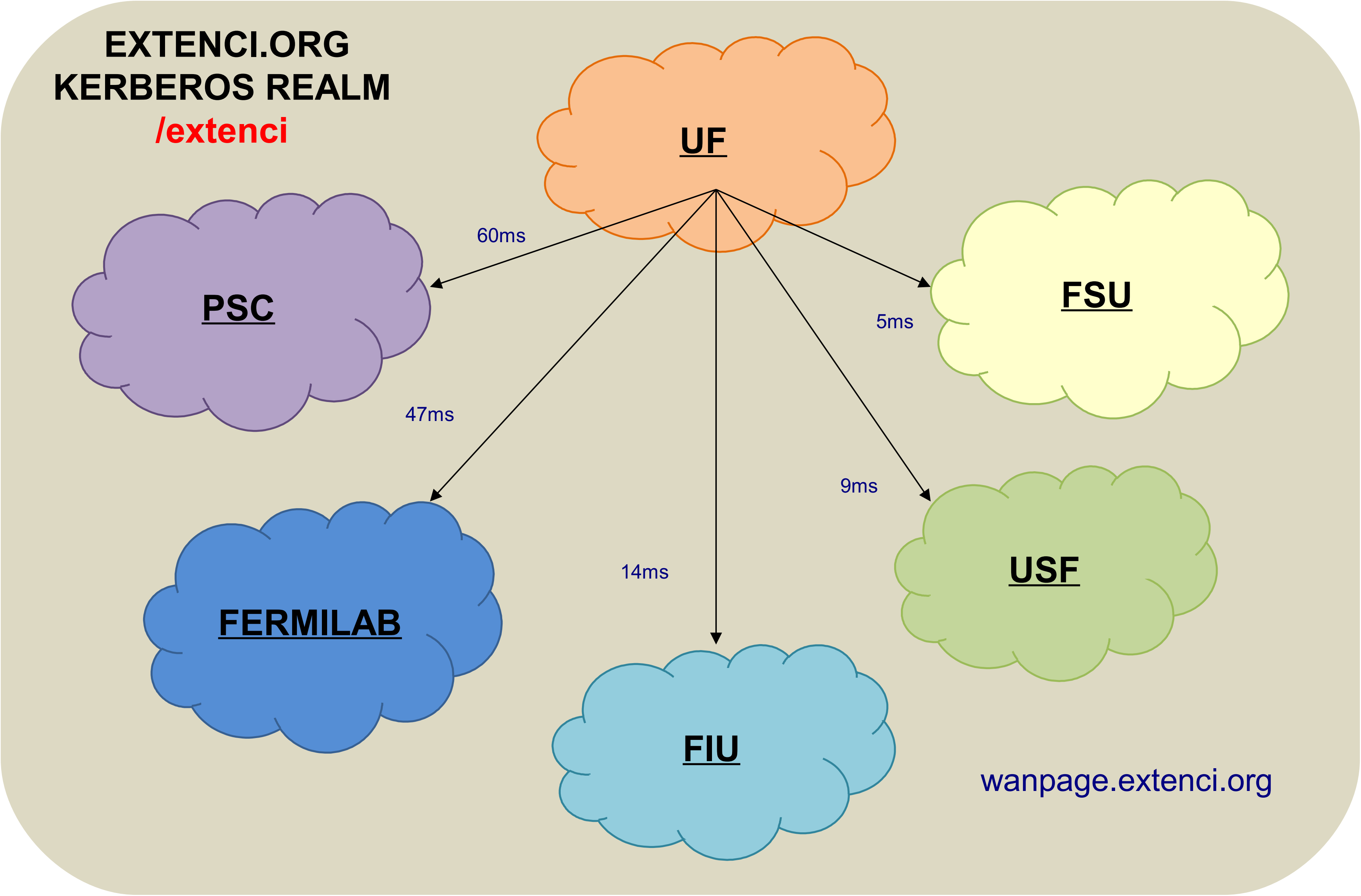
Local client (vmuf, 10GigE)



Remote virtual client (dgtvm1,1GigE)

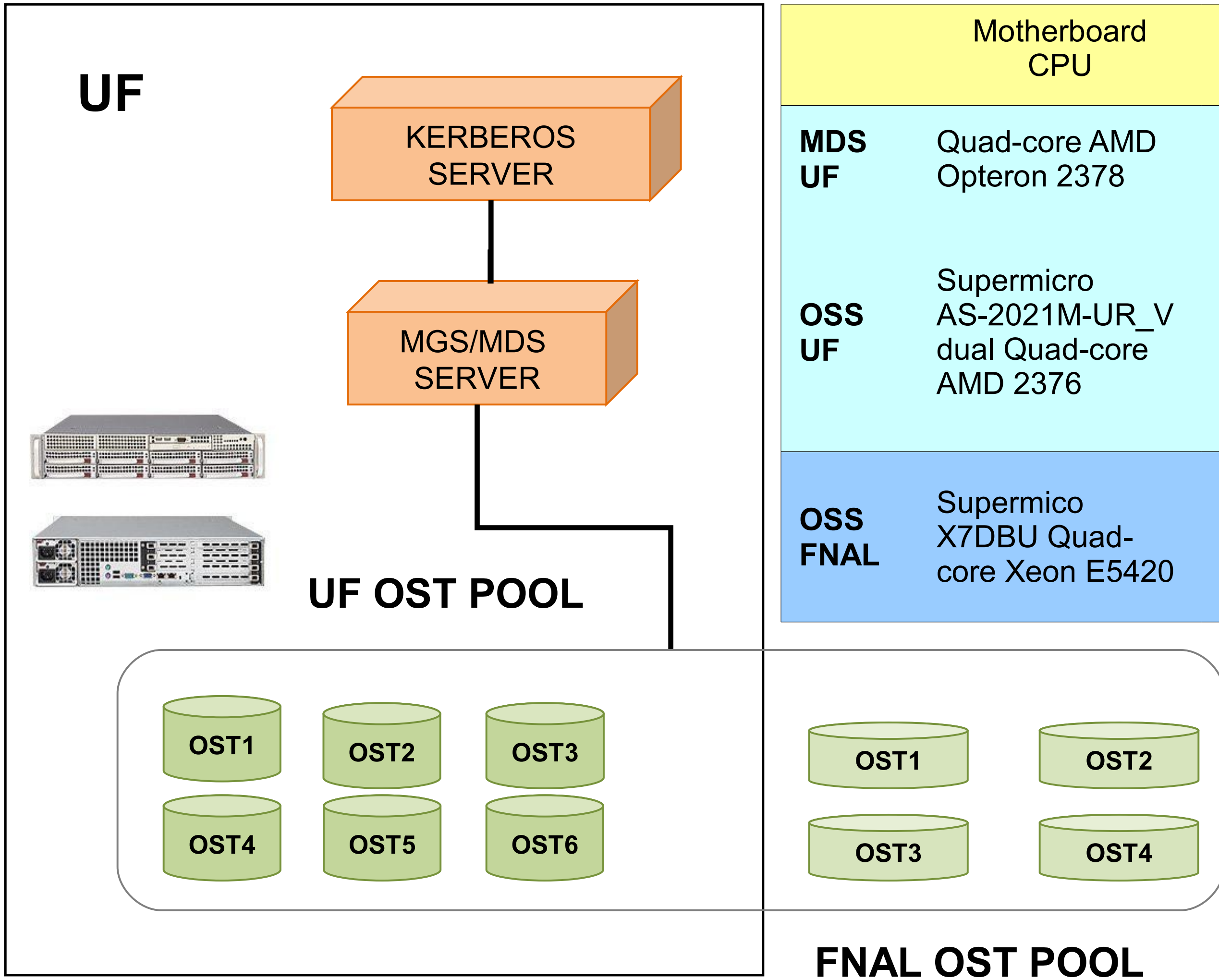


krb5n/krb5a have little impact on Lustre I/O while krb5i/krb5p slow down the fs.


EXTENCI.ORG
KERBEROS REALM
/extenci
UF
PSC
FSU
FERMILAB
USF
FIU
60ms
5ms
47ms
9ms
14ms
wanpage.extenci.org

Hardware at UF and Fermilab

	Motherboard CPU	Mem	Network Interface, Controllers
MDS UF	Quad-core AMD Opteron 2378	8GB	1GigE, 57GB /mds
OSS UF	Supermicro AS-2021M-UR_V dual Quad-core AMD 2376	32GB	10 GigE, 2 Adaptec 51645 raid controllers, 16 2TB drives in in 3 4+1 RAID5 (hot spare)
OSS FNAL	Supermico X7DBU Quad-core Xeon E5420	8GB	1GigE, Nexsan SATABeast 42 500GB, 4x 8/10 RAID6 2GB Qlogic ISP2312



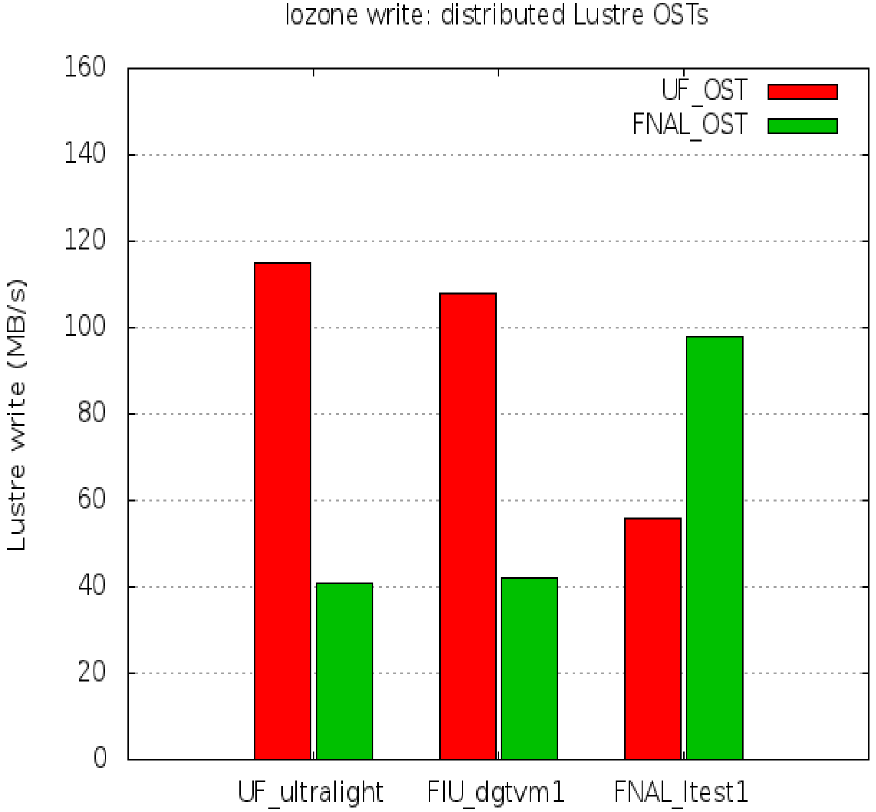
60TB STORAGE

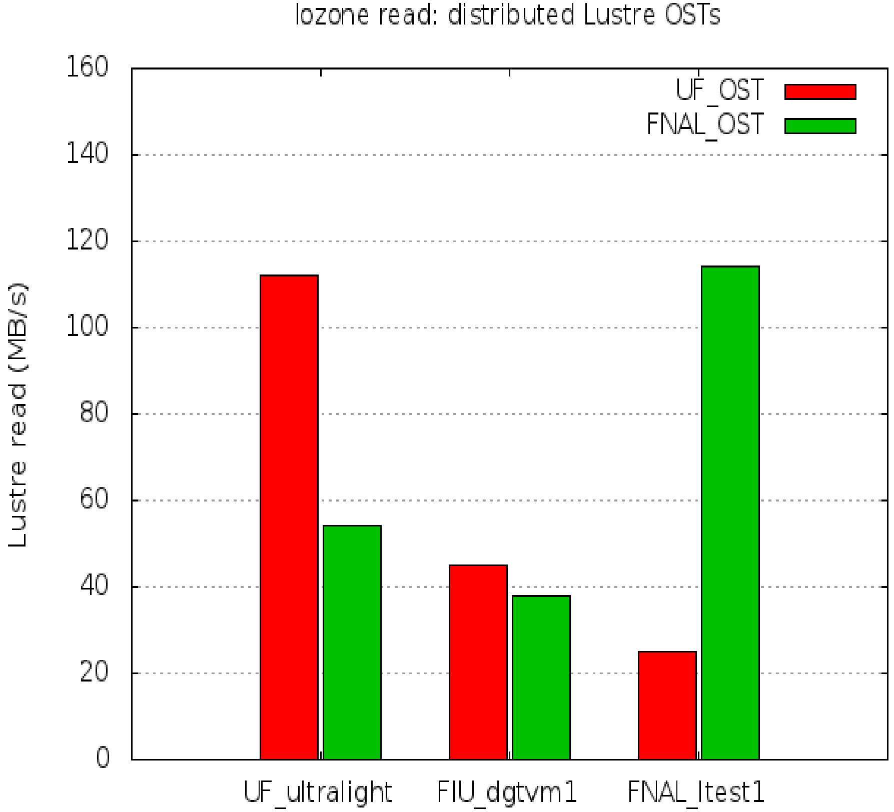
45TB UF OST
15TB FNAL OST

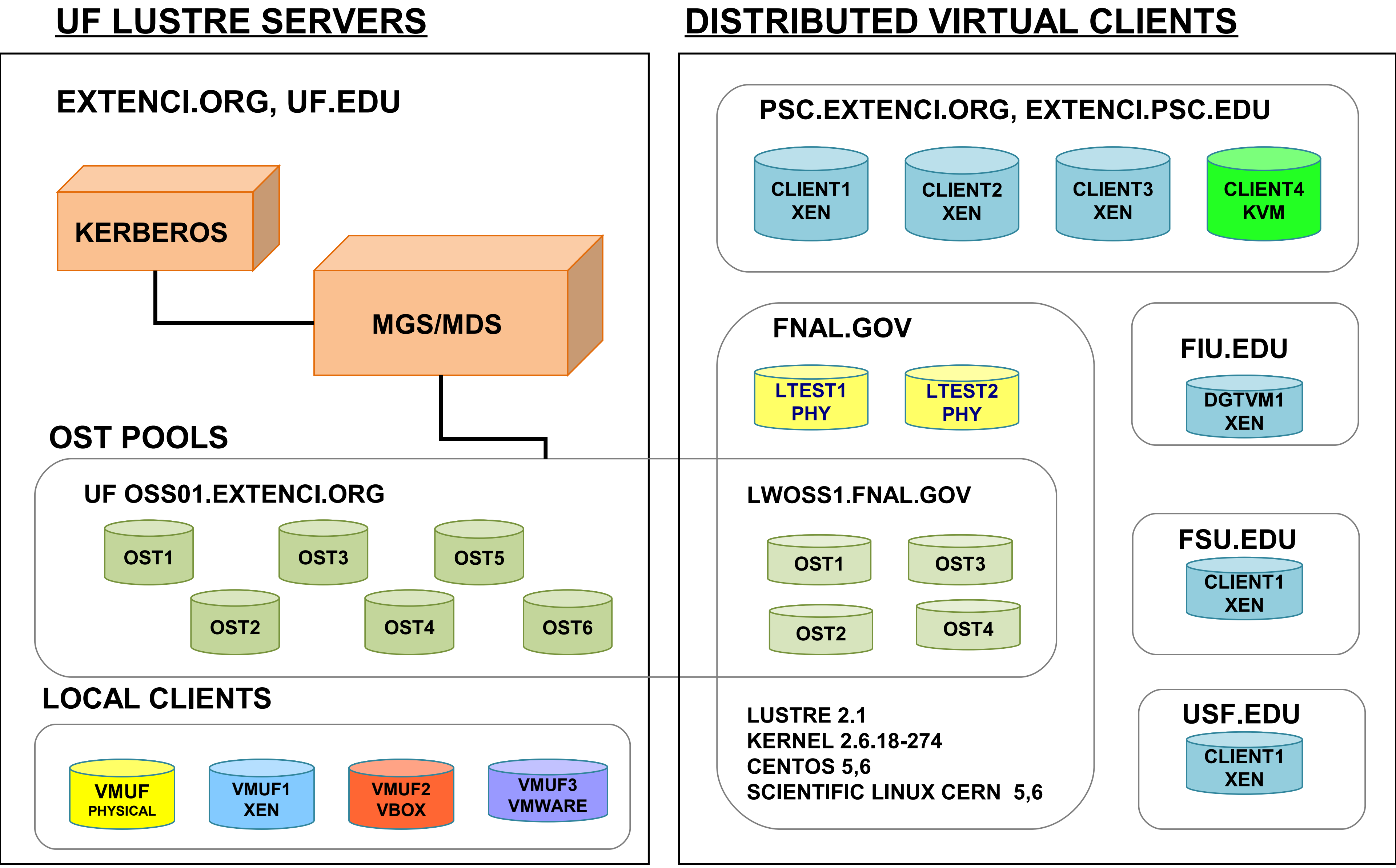
Secure Distributed OSTs

2 OST pools at UF and Fermilab (extenci.uf, extenci.fnal) take advantage of fast IO when local storage is used.

- Ultralight (UF), Dgtvm1 (FIU), Ltest (FNAL)






UF LUSTRE SERVERS
DISTRIBUTED VIRTUAL CLIENTS
EXTENCI.ORG, UF.EDU
KERBEROS
MGS/MDS
OST POOLS
UF OSS01.EXTENCI.ORG
OST1
OST3
OST5
OST2
OST4
OST6
LOCAL CLIENTS
VMUF PHYSICAL
VMUF1 XEN
VMUF2 VBOX
VMUF3 VMWARE
PSC.EXTENCI.ORG, EXTENCI.PSC.EDU
CLIENT1 XEN
CLIENT2 XEN
CLIENT3 XEN
CLIENT4 KVM
FNAL.GOV
LTEST1 PHY
LTEST2 PHY
LWOSS1.FNAL.GOV
OST1
OST3
OST2
OST4
LUSTRE 2.1
KERNEL 2.6.18-274
CENTOS 5,6
SCIENTIFIC LINUX CERN 5,6
FIU.EDU
DGTVM1 XEN
FSU.EDU
CLIENT1 XEN
USF.EDU
CLIENT1 XEN

Packaging Images of Secure Lustre VM Clients

We provide the virtual (VM) images for the local/remote kerberized Lustre clients (and servers) accompanied by complete, detailed **documentation.**

- ❑ **Choice of Virtualization Software**
  - XEN
  - VirtualBox
  - VMware
  - KVM

- ❑ **Software Stack**
  - Kerberos
  - CMS
  - LQCD
  - CVMFS
  - ATLAS
  - ROOT

Network IO of Various Virtualized Systems

- Maximize network IO of various virtual systems
- Study network software interrupts, other network architectures (bonded interfaces)
- Currently XEN is more optimized than VirtualBox and Vmplayer

	CL->OSS (GB/s)	OSS->CL	CL->MDS	MDS->CL	DD lustre Read (MB/s)	Write
VMUF* (PHYSICAL)	9.9	9.9	.95	.95	390	310
VMUF1 (XEN)	9.6	7.2	.95	.95	250	236
VMUF2 (VBOX)	1.9	1.5	.94	.64	125	66
VMUF3 (VMPlayer)	1.5	.78	.94	.625	90	45

PSC
PITTSBURGH SUPERCOMPUTING CENTER

ExTENCI LustreWAN Summary

- **UF** manages the Metadata & OSS storage servers
- Kerberos realm, **EXTENCI.ORG** was established to create secure LNET that only authorized machines and users can access
- Kerberos default security - **krb5n** - provide **adequate security without sacrificing performance** for secure communication among servers/clients as well as user logins and fs access
- **OST pools** at UF and Fermilab allow use of local OSTs for faster IO
- **Prepared system images** (*kerberos+ lustre+ applications*) of Virtual lustre clients make the setup and administration of the systems easier
- Other **CERN Tier 3** sites (FSU, FIU, USF), and PSC can access /extenci; **authorized non-CERN sites** can also easily use the filesystem.
- **Centralized accounts** at UF are autosynced in the clients
- Other kerberos and lustre options enabled to further enhance fs security
  - **Lustre quotas** and **Lustre ACLs**
  - Kerberos **PKINIT** allows use of **X509 certificates** to authenticate in **kerberos framework**

High Energy Physics Applications

- **CMS**
  - CMSS Muon Analysis
  - Root
  - Lustre Scalability of Multiple Root Instances
  - Lustre Scalability of Multiple Clients
- **LQCD**
  - Lustre Scability with Multiple Clients using local OSS Storage
- **ATLAS with CVFS**
  - Benefits of CVMFS
  - CVFS Scalability with Multiple Root Instances

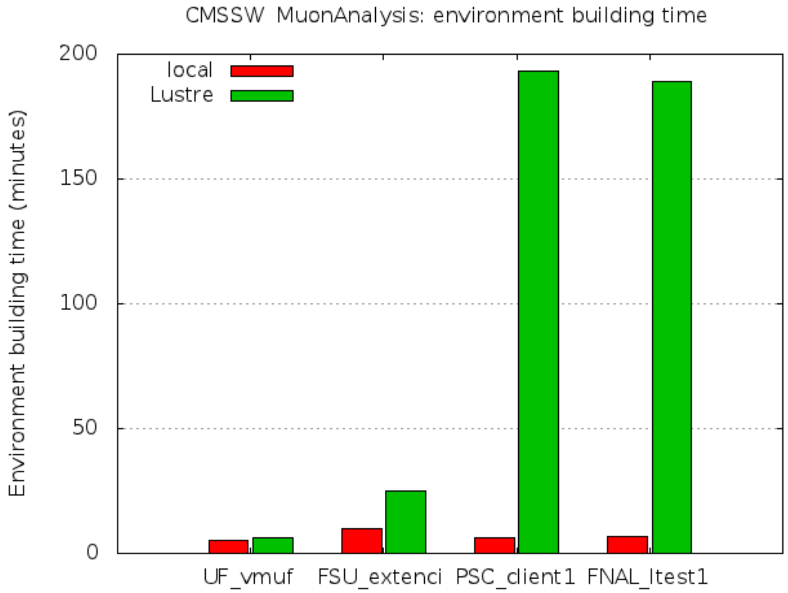
CMSSW MuonAnalysis

We test CMSSW_3_9_7 MuonAnalysis at 4 remote clients

Site	UF	FSU	FNAL	PSC
Latency(ms)	0.1	5	47	60

Environment Building Time (**SCRAM**): compile, link

Local: CMSSW installed on /local
Lustre: CMSSW installed on /extenci lustre fs



- Network latency affects performance (round-trip time for every access back to remote server)
  - Over 85000 file accesses (24k opens, 24k stats, 36k lstats, 1.5k readlinks)
  - 20703 directories, 182913 files ,file size- few range from few KB to several MB (LOSF)

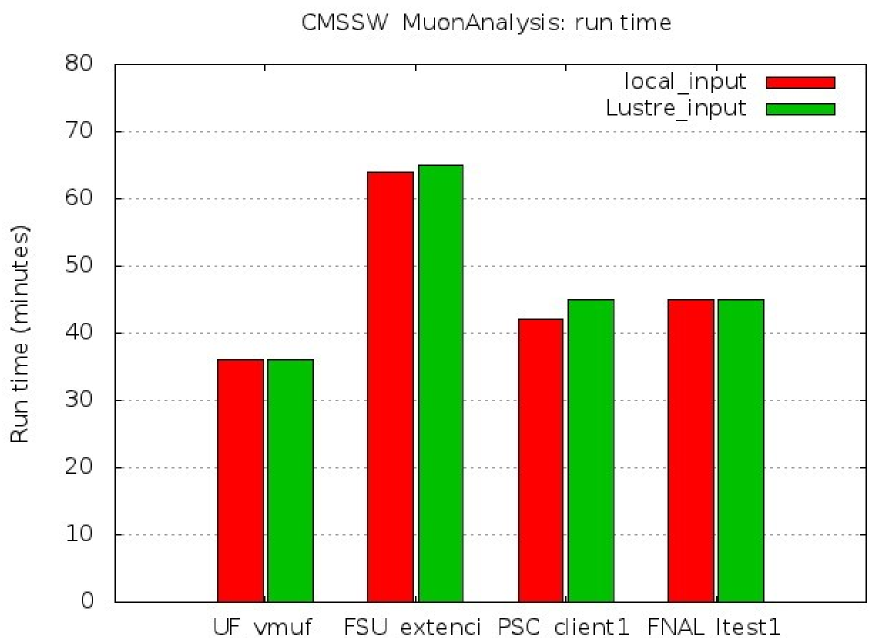
Building time of CMS on Lustre is affected by latency, **slow for remote sites**

CMSSW MuonAnalysis

We test CMSSW_3_9_7 MuonAnalysis at four remote clients

Site	UF	FSU	FNAL	PSC
Latency(ms)	0.1	5	47	60

Run time (local installation)

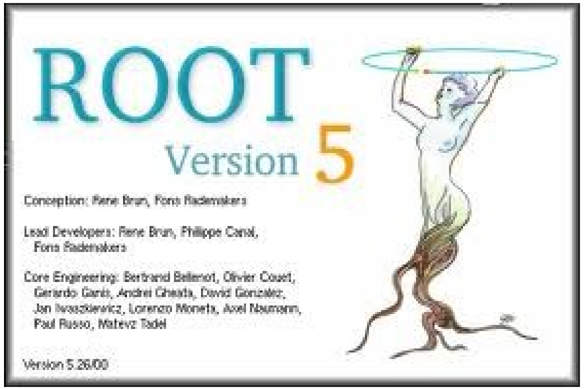


- ❑ CMS software (12GB) made **resident** on the lustre VM client local partition image
  - /local/cmssw
- ❑ CMSSW installed in /local but reads input files locally or from lustre fs
  - local_input, lustre_input
- ❑ Storing input file (2GB data) in lustre doesn't compromise performance

CMS: Lustre Scalability of Multiple ROOT Instances

Background: CERN ROOT

- application responsible for the storage and distribution of CERN LHC data.



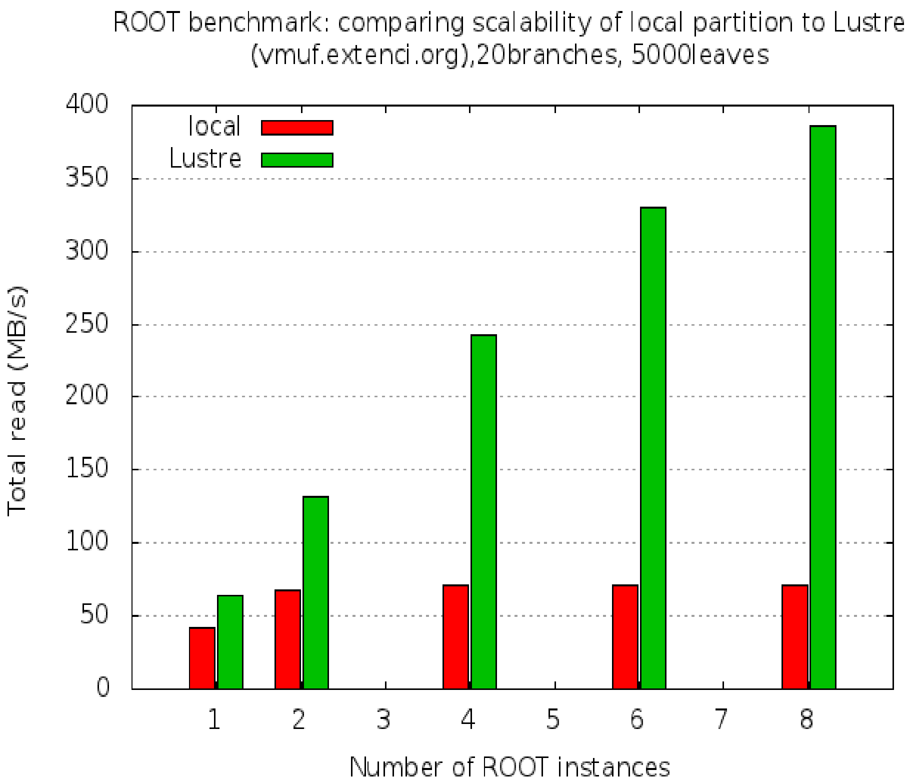
- **Basic, called directly** by all LHC Experiments (CMS, ATLAS, etc) for IO purposes from the huge framework (10^6 lines of code)
- **Optimizable** parameters (e.g. readahead)
- **Zips/unzips data**, stores them in trees and leaves, sorting different but similar events next to each other
- Benefit: **approach is very fast reading a few variables** from each event (but **slow when you read complete** (all leaves) **events**
- Price: for zipping/unzipping - **IO also can become CPU intensive.**

CMS: Lustre Scalability of Multiple ROOT Instances

ROOTv 5.30 reads non-striped file in Lustre local filesystem.

- UF developed ROOT file (10^3 code lines) to test ROOT IO
- Tuned ROOT IO with full control of data format and ease of running different tests
- Tested various data tree structures (number of branches and leaves per branch) from very simple to closely resembling CMS data
- Designed a more IO- (non-CPU) intensive ROOT

Collectl benchmark

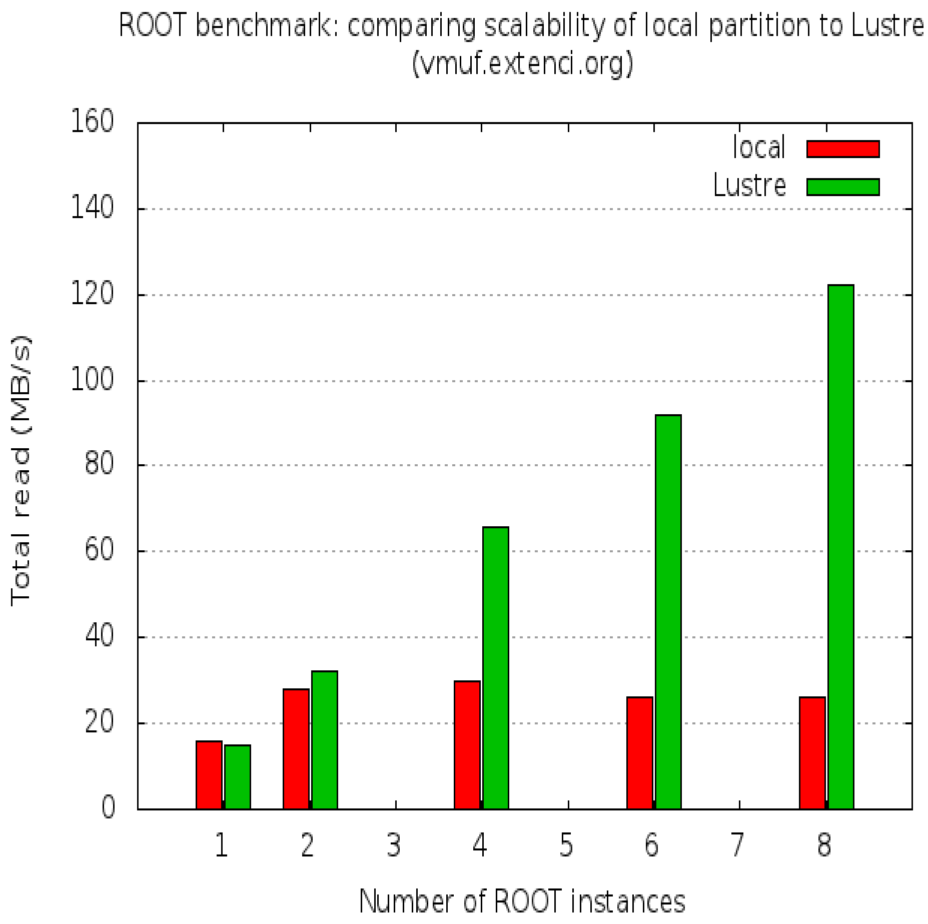


In constrast to local partition, Lustre has very good scability with increasing number of ROOT instances

CMS: Lustre Scalability of Multiple ROOT Instances

- **Tree structure of a file** read by ROOT **determines I/O** speed.

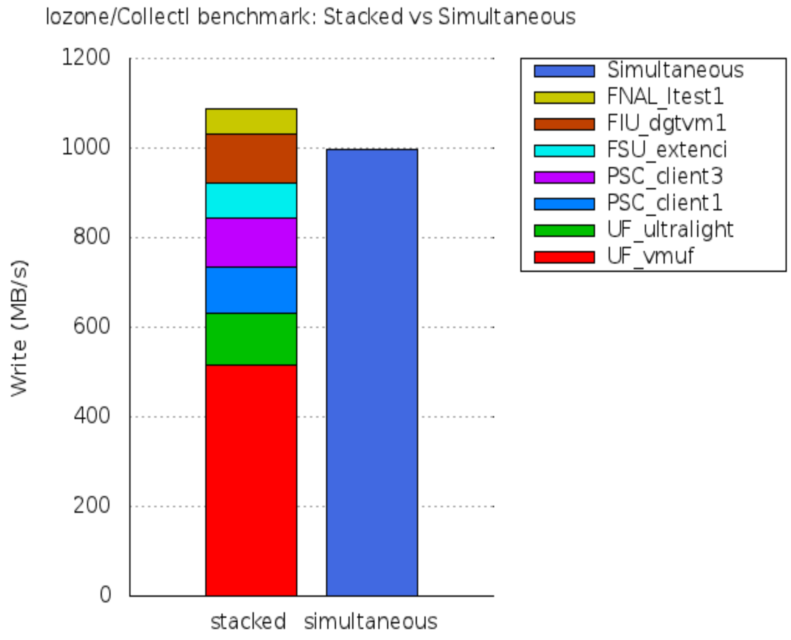
- Graph showing **perfect linear scalability** with lustre on VMUF with increasing ROOT instances used a file with a structure of **2 branches, 2 leaves for each branch, and 20GB random numbers,** giving only 15MB/s for each ROOT instance but gives the linear scalability with increasing ROOT instances.



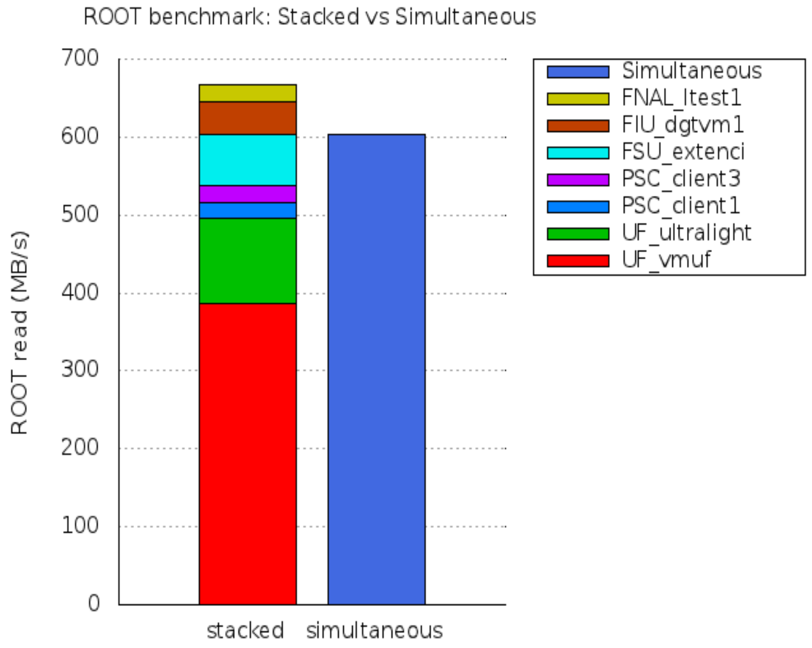
CMS: Lustre Scalability for Multiple Clients

- **Stacked**: Summation over **iozone(write) & ROOT (read)** benchmarks of individual client after sequential runs
- **Simultaneous**: Total I/O throughput on object storage server (OSS) benchmarked by collectl when all clients running in parallel.

Write



ROOT read

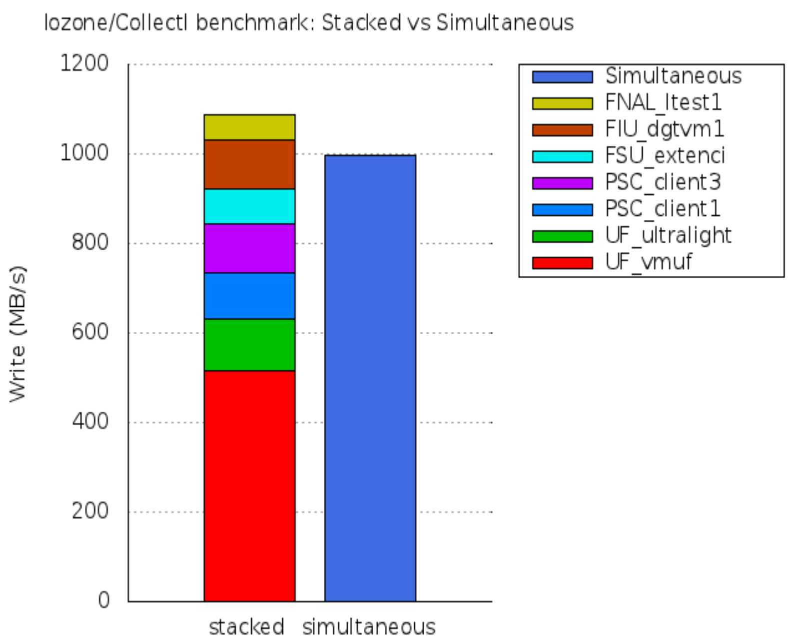


Simultaneous very close to stacked shows Lustre providing scalable total I/O throughputs to multiple clients.

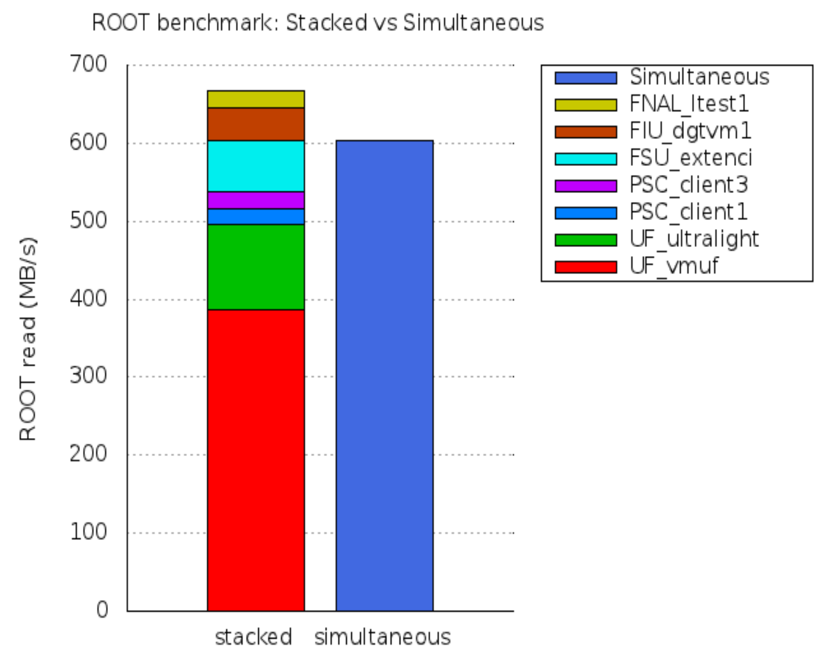
CMS: Lustre Scalability for Multiple Clients

This case showing multi-client scalability used a **Root file with a structure of 20 branches, 5000 leaves for each branch, and 5GB random numbers**. It gives the best (optimized) I/O, say on vmuf 64MB/s for one ROOT instance and 386MB/s for eight ROOT instances, but it doesn't linearly scale as increasing the ROOT instances

Write



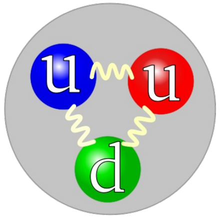
ROOT read



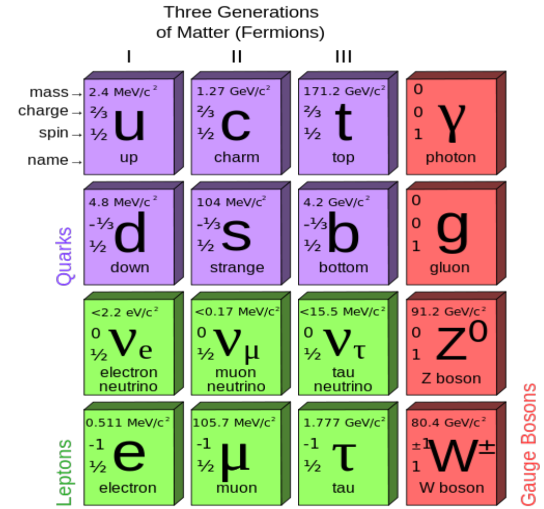
Lattice Quantum Chromodynamics (LQCD)

QCD : theory describing interactions between quarks, gluons

- ❑ **Quark**: fundamental constituent of matter
  - forms hadrons (proton, neutrons)
  - types/flavors: u, d, s, c, b, t
  - intrinsic properties: m, q, spin, color
  - distinguished by their masses and how they decay



Proton: Composed of 2u and 1d quarks

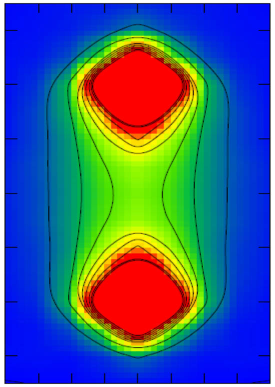


- ❑ **Strong force**: fundamental force that binds quarks together
- ❑ **Gluons**: elementary particles which act as exchange particles (gauge bosons) from strong force between quarks

Scientists **measure quark masses and decay probabilities** as a way to look for new physics beyond standard model (symmetry-violating decays).

LQCD

- ❑ (nonlinear) resort to linear techniques using numerical simulations
- ❑ lattice gauge theory formulation on grid/lattice of points in space-time
- ❑ codes spend much of their time inverting large sparse matrices (MPI)



Meson: LQCD shows meson composed of quark and antiquark

Lattice Quantum Chromodynamics (LQCD)

Software Stack used to build su3_rmd (static)

- ❑ **milc_qcd-7.6.3, mvapich2-1.7-r5225, scidac, scidac-mvapich**
- ❑ milc input files read serially
- ❑ milc output files written using 3 different methods/SciDAC formats
  - save_serial_scidac (1 output file irrespective of number of MPI processes)
  - save_partfile_scidac (1 output file per MPI Process)
  - save_multifile scidac (1 output file per MPI Process)

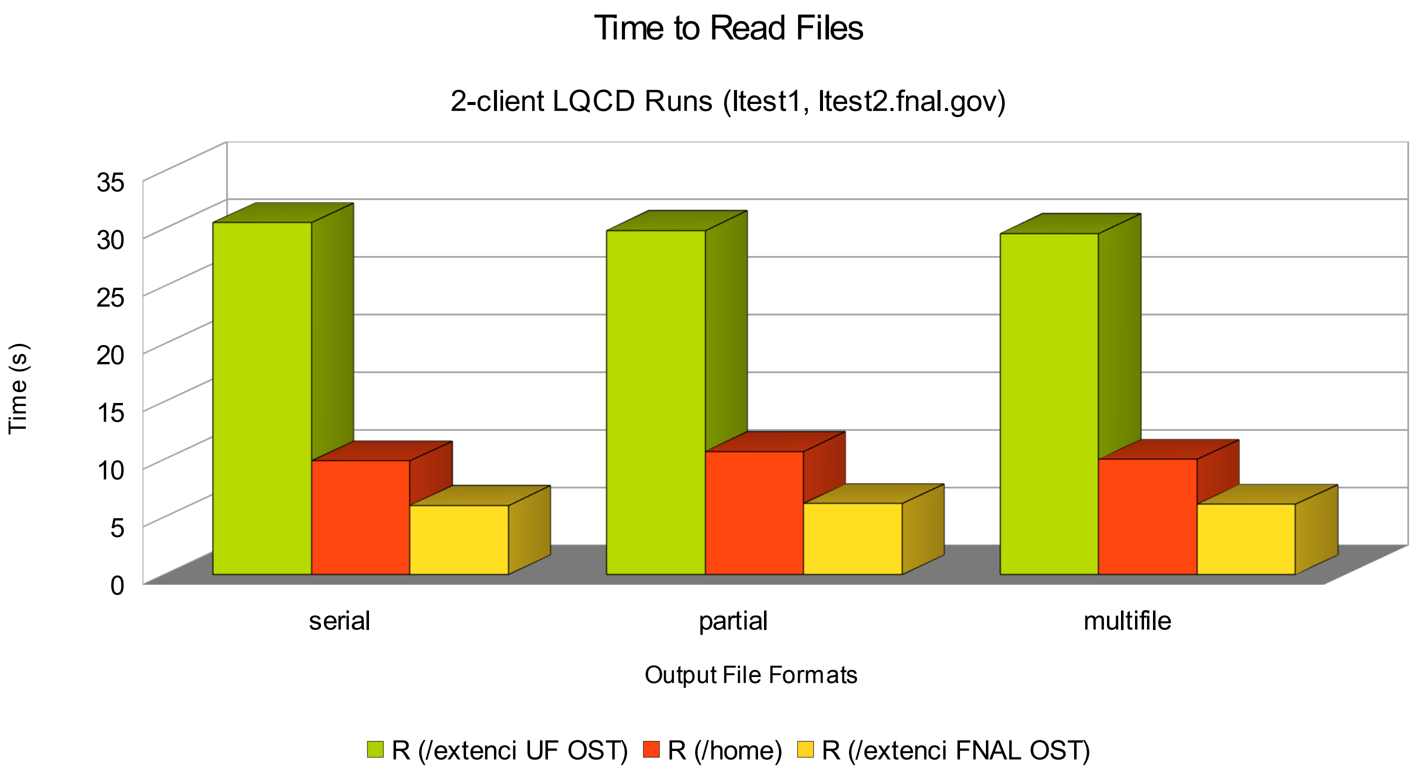
MIMD Lattice Computation (MILC) code

- ❑ **set of codes** developed by MILC collaboration to do simulations of 4D SU(3) lattice gauge theory
- ❑ **code capabilitie**s: molecular dynamics **evolutio**n, fermion action, hadron spectroscopy, matrix elements for leptonic **decay**, dirac matrix eigenvectors and eigenvalues)

SciDAC
Scientific Discovery through Advanced Computing

- ❑ DOE-supported LQCD code adding levels (e.g. QCD **physics** toolbox, QOP optimized in **ASM**, QDP- **data** parallel, QLA-**linear algebra,** QMP-**message passing**)

Lattice Quantum Chromodynamics (LQCD)



Number of Input Files	Number of Output Files	Input File Sizes	Output File Sizes	Output File Formats
2	1	579MB, 495B	579MB	serial
2	16	"	37MB	partial
2	16	"	37MB	multifile

Lattice Quantum Chromodynamics (LQCD)

CERN-VM FileSystem

CMVFS: caching, http-based read-only fs optimized for delivering software to machines

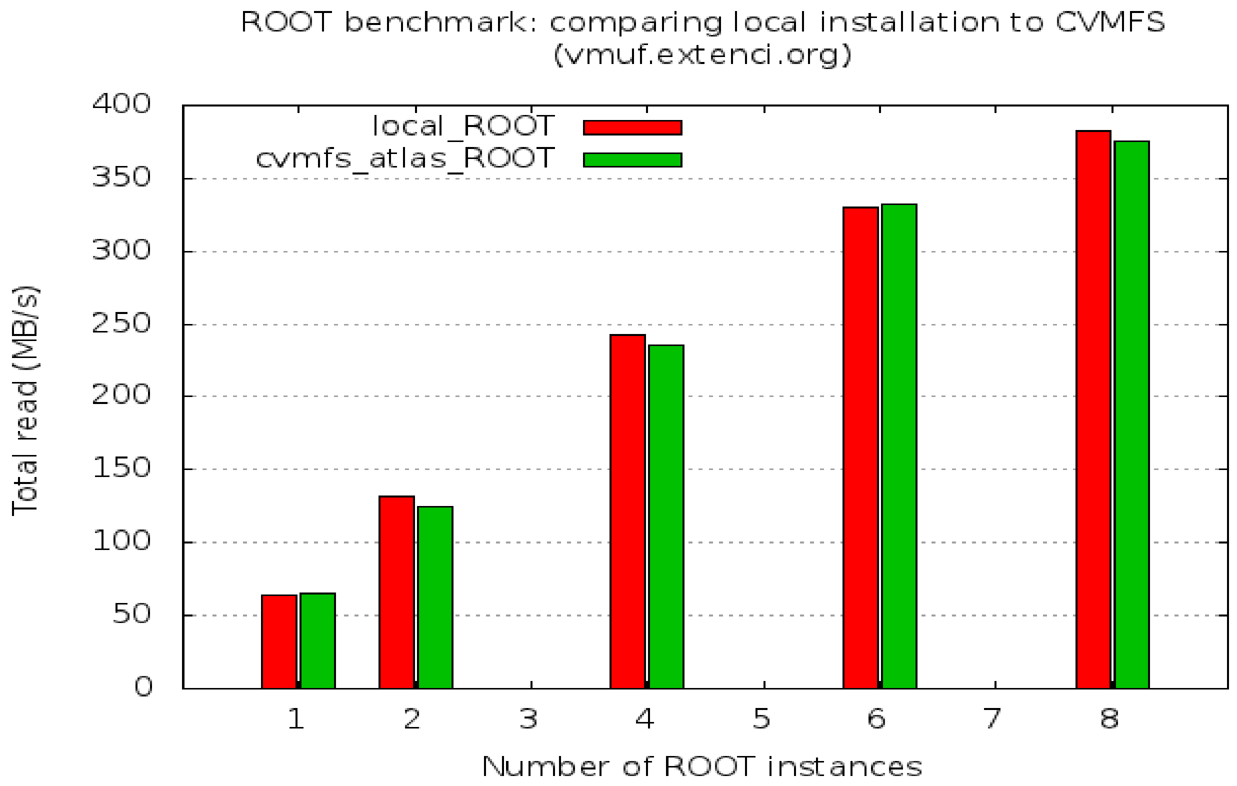
- Removes need for local SW install at every site
- Performance close to a locally installed SW after initial cache population especially for physical machines
- Uses http and fuse to mount a virtual fs
- Scales with additional squid caches
- Verifies file checksums (SHA1) against trusted catalog obtained over https

Benefits

- Network filesystem delivers software on demand
- No hassle of software management and maintenance for user

ATLAS with CVMFS

Compare throughput of local ROOT reading /extenci dbtree3 files to that of CVMFS ROOT reading the same files.



CVMFS with ATLAS ROOT has similar performance as local ROOT, giving 380MB/s for 8 instances

ISSUES/FUTURE WORK

- **Lustre Kerberos (non)support**
  - Clients crashing: lustre kerberos bug
  - Kerberos to be included by NRL into Whamcloud's test suite
- **Application (+ Filesystem) Profiling and Tuning**
  - Filesystems: Lustre, SLASH2, etc
  - Current and other applications
  - Have more clients
- **Better Integration with CERN tools**
- **Integration with XSEDE**
- **Multi-Tier Kerberos**
  - PKINIT, LDAP/PAM, NFS
- **Improve VM Image and its generation**
  - Management Console
  - GUI Automation
    - Users create, boot, launch, specify memory, # CPUs, OS, applications
  - Finetune: sync applications version

PSC
PITTSBURGH SUPERCOMPUTING CENTER

REFERENCES

[1]Palencia, J., Budden, R., and Sullivan, K. 2010. *Kerberized Lustre 2.0 over the WAN*. In Proceedings of the 2010 Teragrid Conference (Pittsburgh, Pennsylvania, August 02 - 05, 2010). TG '10. ACM, New York, NY, 1-5. DOI=http://doi.acm.org/10.1145/1838574.1838589

[2]*MIT Kerberos*: http://web.mit.edu/kerberos/

[3]*Lustre 2.** http://wiki.whamcloud.com/display/PUB/Documentation

[4]*CMS* http://cms.web.cern.ch/content/cms-physics

[5]*LQCD* http://www.physics.utah.edu/~detar/milc/milc_qcd.html

[6]*LQCD Software Stack* http://www.usqcd.org/fnal/qinstall_fermi_bench.html

[7]*ATLAS* http://wiki.whamcloud.com/display/PUB/Documentation

[8]*CVMFS* https://twiki.cern.ch/twiki/bin/view/Atlas/CernVMFS

PSC
PITTSBURGH SUPERCOMPUTING CENTER

Lustre-WAN Metrics

From host: FIU_Client1_VM

To host:	04/01/2012 hi	04/01/2012 lo	04/01/2012 avg	03/31/2012 hi	03/31/2012 lo	03/31/2012 avg	03/30/2012 hi	03/30/2012 lo	03/30/2012 avg	03/29/2012 hi	03/29/2012 lo	03/29/2012 avg	03/28/2012 hi	03/28/2012 lo	03/28/2012 avg	03/27/2012 hi	03/27/2012 lo	03/27/2012 avg	03/26/2012 hi	03/26/2012 lo	03/26/2012 avg
WANLustre	96.2	96.0	96.2	96.6	96.1	96.3	96.7	95.7	96.4	96.3	96.1	96.2	111.1	96.8	106.3	111.2	110.5	110.9	gen	gen	N/A

From host: FSU_Client1

To host:	04/01/2012 hi	04/01/2012 lo	04/01/2012 avg	03/31/2012 hi	03/31/2012 lo	03/31/2012 avg	03/30/2012 hi	03/30/2012 lo	03/30/2012 avg	03/29/2012 hi	03/29/2012 lo	03/29/2012 avg	03/28/2012 hi	03/28/2012 lo	03/28/2012 avg	03/27/2012 hi	03/27/2012 lo	03/27/2012 avg	03/26/2012 hi	03/26/2012 lo	03/26/2012 avg
WANLustre	87.3	86.0	86.7	86.3	83.2	85.3	87.6	86.5	86.9	85.7	81.9	84.2	87.4	84.3	85.4	83.6	82.2	83.1	86.5	83.8	85.4

From host: PSC_Client1_VM

To host:	04/01/2012 hi	04/01/2012 lo	04/01/2012 avg	03/31/2012 hi	03/31/2012 lo	03/31/2012 avg	03/30/2012 hi	03/30/2012 lo	03/30/2012 avg	03/29/2012 hi	03/29/2012 lo	03/29/2012 avg	03/28/2012 hi	03/28/2012 lo	03/28/2012 avg	03/27/2012 hi	03/27/2012 lo	03/27/2012 avg	03/26/2012 hi	03/26/2012 lo	03/26/2012 avg
WANLustre	17.8	11.1	15.2	25.5	23.1	23.9	15.7	13.6	14.6	102.5	12.7	70.5	96.8	92.9	94.3	92.1	65.7	77.6	93.3	67.8	76.3

From host: PSC_Client2_VM

To host:	04/01/2012 hi	04/01/2012 lo	04/01/2012 avg	03/31/2012 hi	03/31/2012 lo	03/31/2012 avg	03/30/2012 hi	03/30/2012 lo	03/30/2012 avg	03/29/2012 hi	03/29/2012 lo	03/29/2012 avg	03/28/2012 hi	03/28/2012 lo	03/28/2012 avg	03/27/2012 hi	03/27/2012 lo	03/27/2012 avg	03/26/2012 hi	03/26/2012 lo	03/26/2012 avg
WANLustre	20.0	gen	19.6	27.5	gen	27.5	16.6	gen	16.6	91.3	13.9	64.3	85.9	gen	85.9	gen	gen	N/A	gen	gen	N/A

From host: PSC_Client3A_VM

To host:	04/01/2012 hi	04/01/2012 lo	04/01/2012 avg	03/31/2012 hi	03/31/2012 lo	03/31/2012 avg	03/30/2012 hi	03/30/2012 lo	03/30/2012 avg	03/29/2012 hi	03/29/2012 lo	03/29/2012 avg	03/28/2012 hi	03/28/2012 lo	03/28/2012 avg	03/27/2012 hi	03/27/2012 lo	03/27/2012 avg	03/26/2012 hi	03/26/2012 lo	03/26/2012 avg
WANLustre	26.4	17.2	20.3	24.9	21.7	23.6	21.4	16.7	18.5	97.6	14.3	57.2	94.0	94.0	94.0	NO DATA			NO DATA		

From host: PSC_Client3_VM

To host:	04/01/2012 hi	04/01/2012 lo	04/01/2012 avg	03/31/2012 hi	03/31/2012 lo	03/31/2012 avg	03/30/2012 hi	03/30/2012 lo	03/30/2012 avg	03/29/2012 hi	03/29/2012 lo	03/29/2012 avg	03/28/2012 hi	03/28/2012 lo	03/28/2012 avg	03/27/2012 hi	03/27/2012 lo	03/27/2012 avg	03/26/2012 hi	03/26/2012 lo	03/26/2012 avg
WANLustre	98.9	98.6	98.8	100.0	84.9	94.2	99.8	94.5	97.8	100.1	74.9	91.2	98.5	81.9	90.7	99.7	77.0	92.1	97.6	85.4	93.0

From host: UF_Ultralight

To host:	04/01/2012 hi	04/01/2012 lo	04/01/2012 avg	03/31/2012 hi	03/31/2012 lo	03/31/2012 avg	03/30/2012 hi	03/30/2012 lo	03/30/2012 avg	03/29/2012 hi	03/29/2012 lo	03/29/2012 avg	03/28/2012 hi	03/28/2012 lo	03/28/2012 avg	03/27/2012 hi	03/27/2012 lo	03/27/2012 avg	03/26/2012 hi	03/26/2012 lo	03/26/2012 avg
WANLustre	127.3	126.9	127.1	127.1	127.0	127.0	127.2	127.0	127.1	127.2	126.6	127.0	126.9	126.8	126.8	127.2	126.9	127.1	127.2	126.9	127.1

From host: UF_VMUF

Legend
<25MBps
25-50MBps
>50MBps
Login Failure
Operation Timeout
Checksum Failure
Generic Error

All speeds in MegaBytes per second.

ADDENDUM

- ExTENCI Lustre-WAN PSC Documentation
- CMS: MonteCarlo Data Flow
- CMS: Data Tiers

ExTENCIProjectWithOSG < KerbLustre < TWiki - Mozilla Firefox

https://wiki.psc.edu/twiki/viewauth/KerbLustre/ExTENCIProjectWithOSG

PITTSBURGH SUPERCOMPUTING CENTER

You are here: TWiki > KerbLustre Web > ExTENCIProjectWithOSG

r718 - 30 Mar 2012 - 09:00:07 - JosephinAtPscEdu

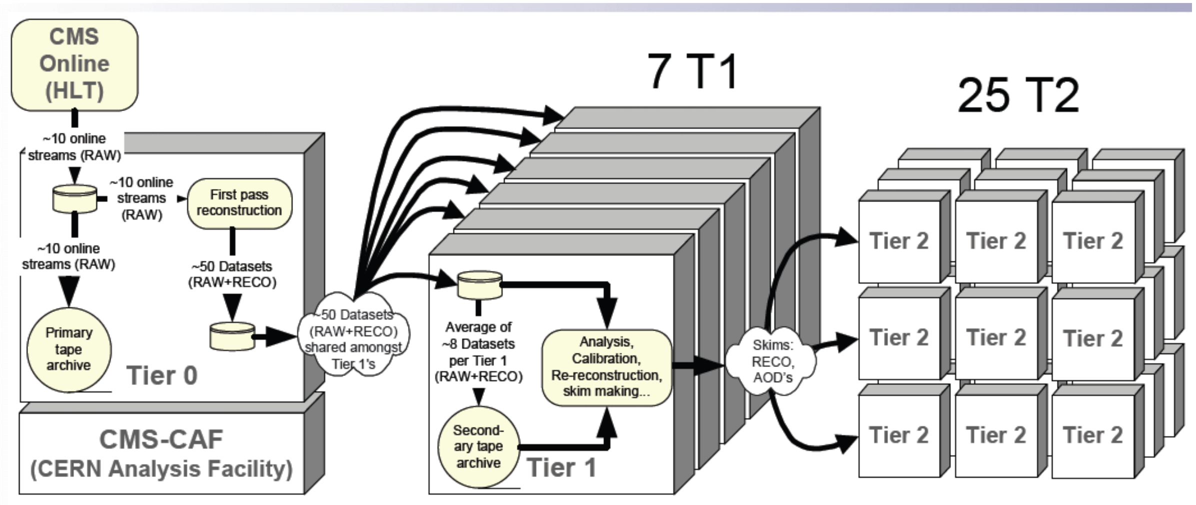
Webs Utilities Options

ExTENCI with the OSG

Introduction

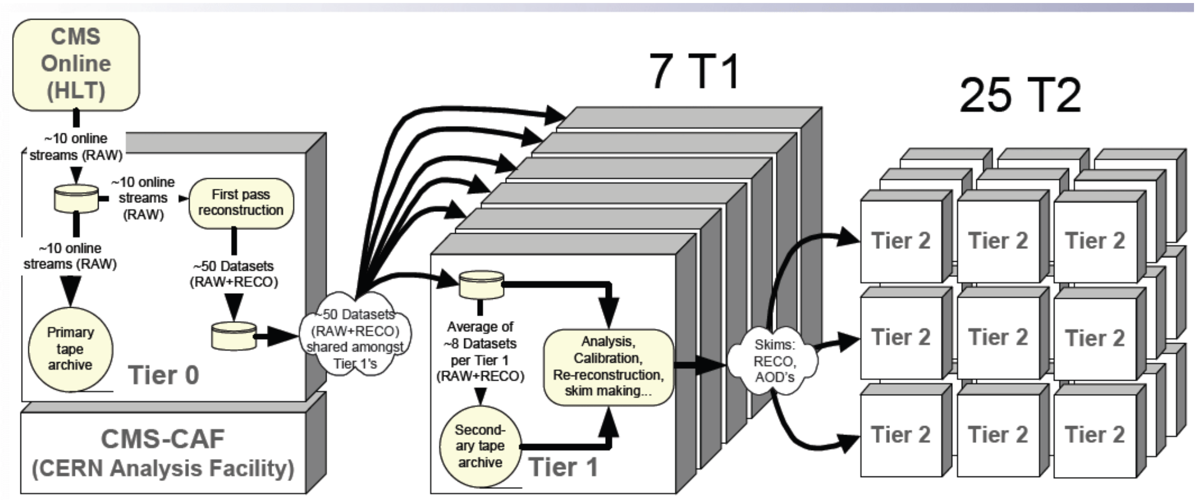
This secure wiki is a resource guide to the kerberos-enabled Lustre 2.* filesystem that we have established over the WAN for the ExTENCI Project as part of the collaborative efforts between the Open Science Grid and the Teragrid.

CMS: Monte Carlo Data Flow



- **T0:** Data received from CMS detector experiment; repacked (unsorted streams are sorted into physics streams of events with similar characteristics). Reconstruction algorithms are ran, AOD produced and RAW, RECO and AOD are exported to Tier1
- **T1:** Redistributing data after reprocessing with improved algorithms
- **T2:** MonteCarlo events generated, detector interactions simulated, events reconstructed, and events moved to tape storage for later use
- **T3:** Users prepare analysis code, run code on data then collect results

CMS Data Tiers



CMS Data is arranged into a hierarchy of data tiers. Each physics event is written into each data tier containing different levels of information about the event.

- **Raw**: full event information from Tier-0 (i.e. from CERN), containing 'raw' detector information (detector element hits, etc); not used directly for analysis
- **RECO** (RECOnstructed data): output from 1^{st} pass processing by Tier-0. This layer contains reconstructed physics objects, but is still very detailed; can be used for analysis but still too big for frequent or heavy use
- **AOD (Analysis Object Data):** "distilled" version of RECO event information expected to be used with good trade-off between size/complexity for speedy data analysis