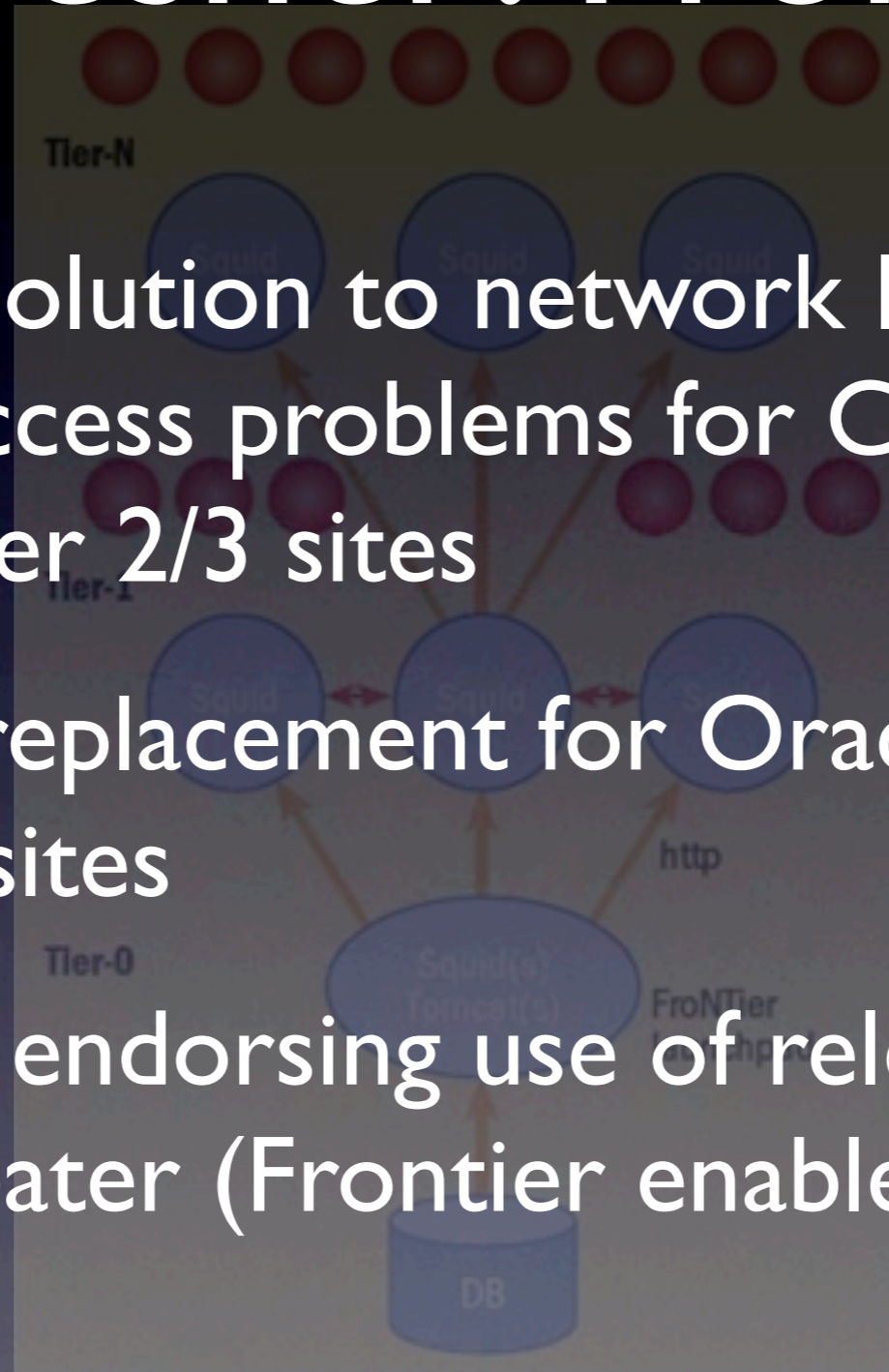


Scalable Access to Conditions Data for Analysis

US ATLAS Tier 2/3 Workshop
University of Texas at Arlington: Nov 10-12, 2009
John DeStefano - BNL

Refresher: Frontier

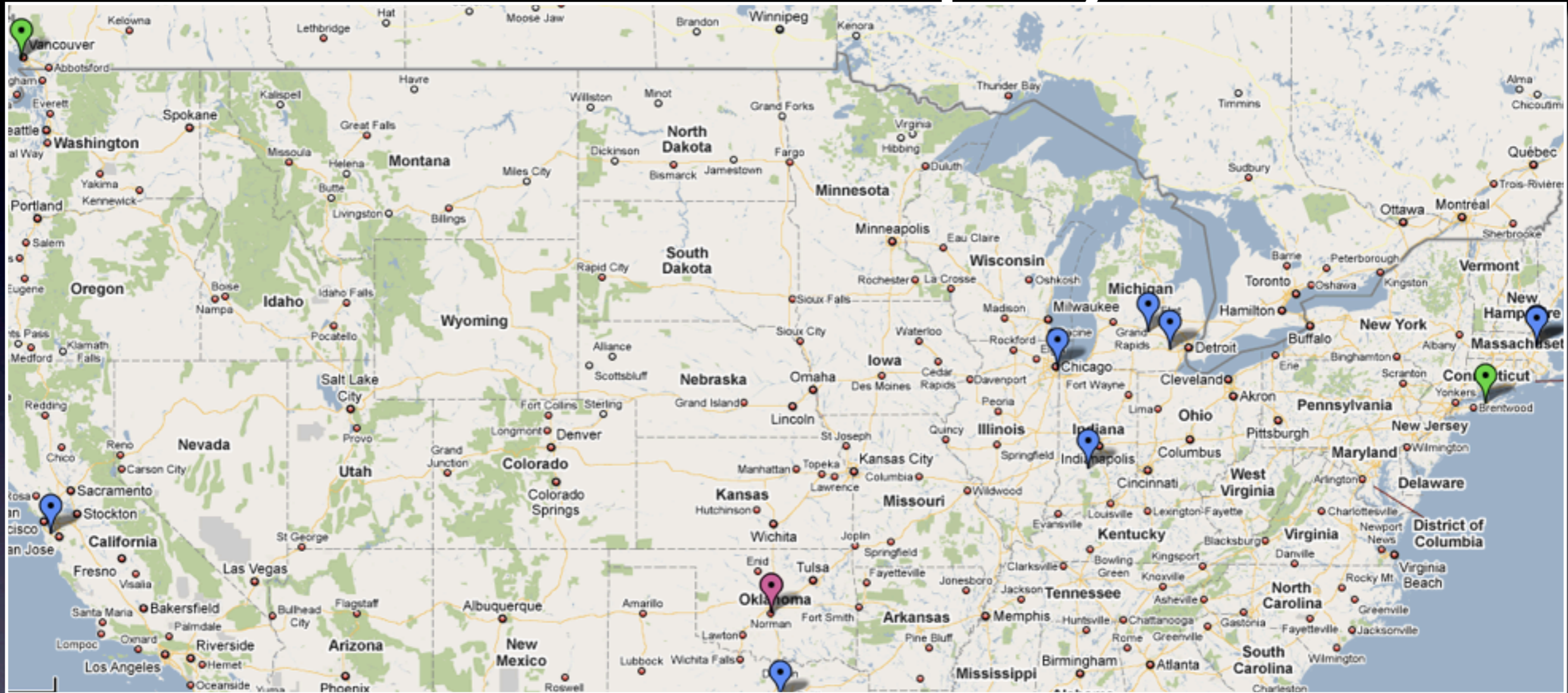
- Scalable solution to network latency and Oracle access problems for Conditions data at Tier 2/3 sites
- Not a replacement for Oracle streams at Tier 1 sites
- ATLAS endorsing use of releases 15.5.1 and greater (Frontier enabled)



Refresher: Squid

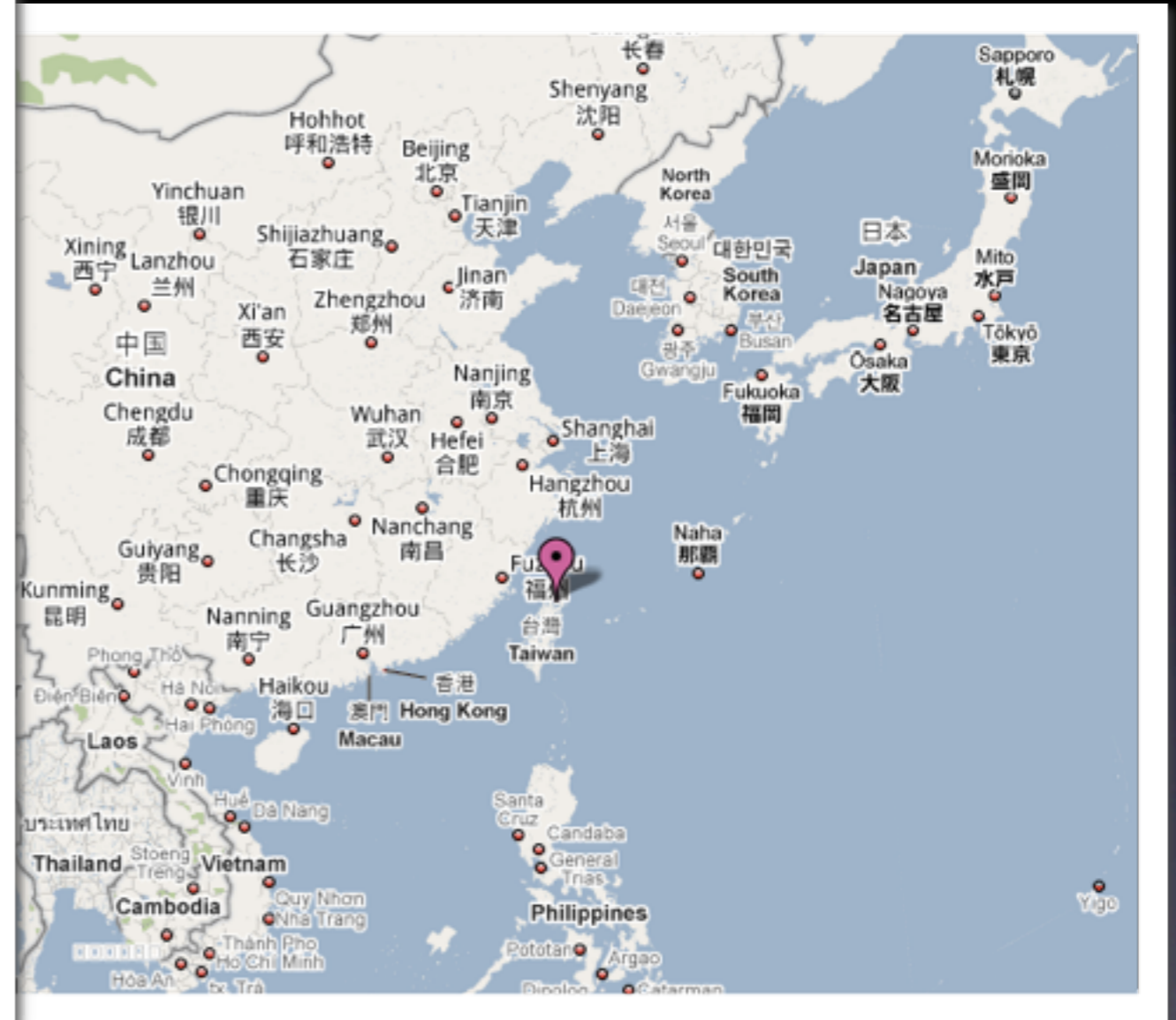
- Local cache for Conditions data at each Tier 2 site (and some Tier 3s)
- ATLAS recommendation: one Squid instance per 500 analysis queue slots at each Tier 2
- Tier 3 deployments at IU, LBNL, Yale
- ATLAS has created and is maintaining Squid RPMs

N.American deployment



- Squids now deployed at seven of eight US Tier 2 sites
- Two of ATLAS's five Frontier sites in North America
- Both BNL and TRIUMF are using cache consistency

ATLAS deployment



- Frontier at FZK, PIC, RAL
- Frontier at CERN (Tier 0) only for development, testing

Frontier site URLs

BNL	squid-frontier.usatlas.bnl.gov:23 28/frontieratbnl/Frontier
FZK	atlassq -fzk.gridka.de:802 /fzk/Frontier
PIC	oratest02.pic.es:3 28/pic-frontier/Frontier
RAL	lcgft-atlas.gridpp.rl.ac.uk:3 28/frontierATLAS/Frontier
TRIUMF	frontier.triumf.ca:3 28/ATLAS_frontier/Frontier

- Not all servers offer open access to all sites

Squid site status

- Initial US site listings almost complete
- Need additional test results
- US on par with most Frontier-enabled clouds
 - <https://twiki.cern.ch/twiki/bin/view/Atlas/T2SquidDeployment>

CA	<input type="checkbox"/>	Show...
DE	<input type="checkbox"/>	Show...
ES	<input type="checkbox"/>	Show...
FR	<input type="checkbox"/>	Show...
IT	<input type="checkbox"/>	Show...
ND	<input type="checkbox"/>	Show...
NL	<input type="checkbox"/>	Show...
TW	<input type="checkbox"/>	Show...
UK	<input type="checkbox"/>	Show...
US	<input type="checkbox"/>	Hide

US	John DeStefano							
Site	Squid	T1 auth	fnget test	backup	setup.sh	schedconfig	1 job	
BNL-T1	http://squid-frontier.usatlas.bnl.gov:23128/frontieratbni/Frontier	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	TRIUMF				
AGLT2-Michigan	http://squid.aglt2.org:3128	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	MWT2-Chicago				
MWT2-Chicago	http://uct2-grid2.uchicago.edu:3128	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	MWT2-Indiana				
MWT2-Indiana	http://iut2-grid2.iu.edu:3128	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	MWT2-Chicago				
NET2-Boston	http://atlas.bu.edu:3128	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	NET2-Harvard				
NET2-Harvard	http://heroatlas.fas.harvard.edu:3128	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	NET2-Boston				
SWT2-Texas	http://gk07.atlas-swt2.org:3128	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	WT2-SLAC(?)				
SWT2-Oklahoma	use SWT2-Texas (?)	<input checked="" type="checkbox"/>		WT2-SLAC(?)				
WT2-SLAC	http://atl-dbcache1.slac.stanford.edu:23128	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	SWT2-Texas(?)				

Testing data access from BNL

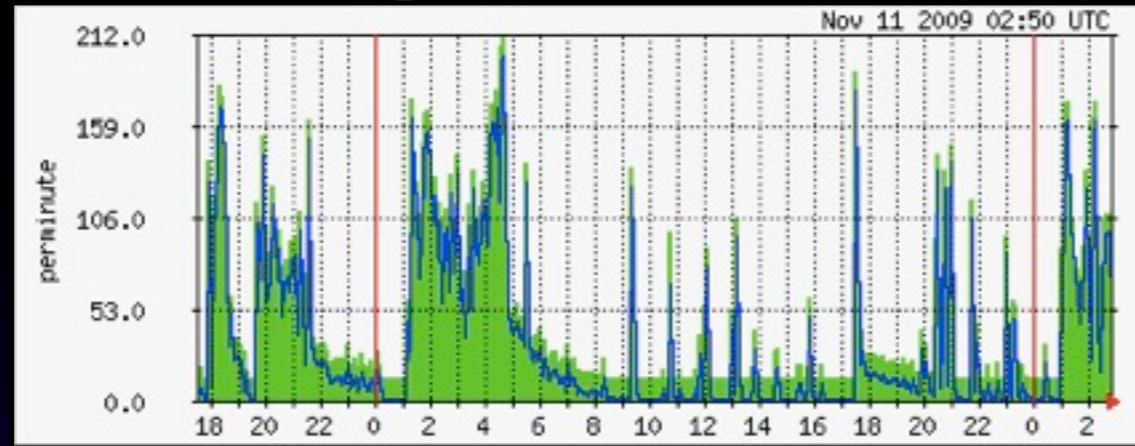
	RTT	Oracle	Frontier	Squid
Boston	7 ms			
Chicago	24 ms			
Harvard	n/a			
Michigan	35 ms			
Indiana	29 ms			
OU	n/a			
SLAC	84 ms			
UTA	58 ms			

Testing: the Luehring method

Fred breaks stuff ... in the name of Physics

- ~650 job Frontier test from MWT2 to BNL using multiple data sets
 - 2 jobs failed due to long-running query
- Another ~650 job run, with all slots empty at MWT2
 - 540 jobs failed, most due to failed Local Site Mover gets via dCache
 - Charles will analyze errors, may change LSM code
 - Some errors due to runAthena priority changes made by Tadashi (since reverted)
- Encouraging testing from Yale T3
 - Uncovered transform problem that prioritizes SQLite over other replicas (Richard, Sarah, Rochelle)

Central Squid monitoring



- Port 3401 (UDP) must be opened in firewalls, Squid ACL rules to remote Squid monitoring host (frontier.cern.ch)
- Default Squid ACL rules for monitoring destination need to be enabled (if disabled)
- Notify David Front or me of Squid IP address
- Detailed instructions for ACL changes (Step 7):
 - <https://www.racf.bnl.gov/docs/services/frontier/squid-installation>
 - <http://frontier.cern.ch/squidstats/indexatlas.html>

Frontier updates

- Long-running ATLAS SQL query revealed via MWT2-to-BNL Frontier tests (Fred, Dave, Gancho, Andrea, Carlos)

```

SELECT /*+
QB_NAME(MAIN) INDEX(@MAIN "COOL_I3"@MAIN ("CHANNEL_ID"
"IOV_SINCE"
"IOV_UNTIL")) LEADING(@MAIN "COOL_C2"@MAIN "COOL_I3"@MAIN)
USE_NL(@MAIN
"COOL_I3"@MAIN) INDEX(@MAX1 COOL_I1@MAX1 ("CHANNEL_ID"
"IOV_SINCE"
"IOV_UNTIL")) */ "COOL_I3"."OBJECT_ID", "COOL_I3"."CHANNEL_ID",
"COOL_I3"."IOV_SINCE", "COOL_I3"."IOV_UNTIL",
"COOL_I3"."USER_TAG_ID",
"COOL_I3"."SYS_INSTIME", "COOL_I3"."LASTMOD_DATE",
"COOL_I3"."ORIGINAL_ID", "COOL_I3"."NEW_HEAD_ID",
"COOL_I3"."HVCHVOLT_RECV", "COOL_I3"."HVCHCURR_RECV" FROM
"ATLAS_COOLOFL_DCS"."COMP200_F0023_CHANNELS" "COOL_C2",
"ATLAS_COOLOFL_DCS"."COMP200_F0023_IOVS" "COOL_I3" WHERE
"COOL_I3"."CHANNEL_ID"="COOL_C2"."CHANNEL_ID" AND
"COOL_I3"."IOV_SINCE">=COALESCE(( SELECT /*+ QB_NAME(MAX1) */
MAX(COOL_I1."IOV_SINCE") FROM
ATLAS_COOLOFL_DCS.COMP200_F0023_IOVS COOL_I1
WHERE COOL_I1."CHANNEL_ID"="COOL_C2"."CHANNEL_ID" AND
COOL_I1."IOV_SINCE"<=12237390000000000000
),12237390000000000000) AND
"COOL_I3"."IOV_SINCE"<=12237396000000000000 AND
"COOL_I3"."IOV_UNTIL">12237390000000000000 ORDER BY
"COOL_I3"."CHANNEL_ID"
ASC, "COOL_I3"."IOV_SINCE" ASC;

```

Frontier updates

- Long-running Oracle query (continued):
 - Takes 10+ seconds to respond on first query
 - Savannah #57703
 - Will not be addressed by ATLAS
 - Frontier updated to send continuous keep-alives while awaiting first response to query
 - Also includes fixes for cached time-outs, matching table names containing SQL keywords

Frontier updates

- Oracle Conditions database modified at BNL, RAL, TRIUMF for updates to ATLAS triggers for Frontier cache consistency checking
- Pilot for detailed T1 launchpad monitoring via Squid logs at BNL:
- <http://frontier.cern.ch/awstatsatlas.html>

Site data requirements

Site elements needed for running jobs with Frontier:

- ➔ Site Squid
- ➔ DDM subscription
- ➔ PFC generation
- ➔ Site listing in Tiers of ATLAS

Site environment

- Site setup for enabling Frontier access via automated installation “job” from CERN:
 - ✓ Generates POOL file catalog (PFC) XML for local POOL files
 - ✓ Applies any needed library fixes, replacements according to local ATLAS site storage elements (dCache, DPM)
 - ✓ Creates environment script (setup.sh) for Frontier, PFC, other site-specific values
- Jobs source site setup script before running
- Environment automation working for ~80% of EGEE sites
- Alessandro, Fred, Rod, Xin collaborating on comprehensive solution

Open issues

- Service access policies
 - ▶ Who should be able to access site Squids? All T2/3s? All clients?
 - ▶ What about Frontier launchpads? All T1/2/3s? Open access?
 - ▶ Balance between ease of use and maintainability
- Service level agreements
 - ▶ What are the expected up-times for Frontier? Squid?

Next steps

- Complete testing at each site
 - Test with simple, 'ReadReal', Ganga jobs
- Enact fail-over strategy at each site
- Load, functional testing on new Frontier servers
- Complete and test site environment automation

Meeting info

- Tuesdays at 10 AM ET
- Now expanded to include all of ATLAS
- Using Alcatel teleconference system
- Meeting minutes and more information:
 - <https://www.racf.bnl.gov/docs/services/frontier/meetings>