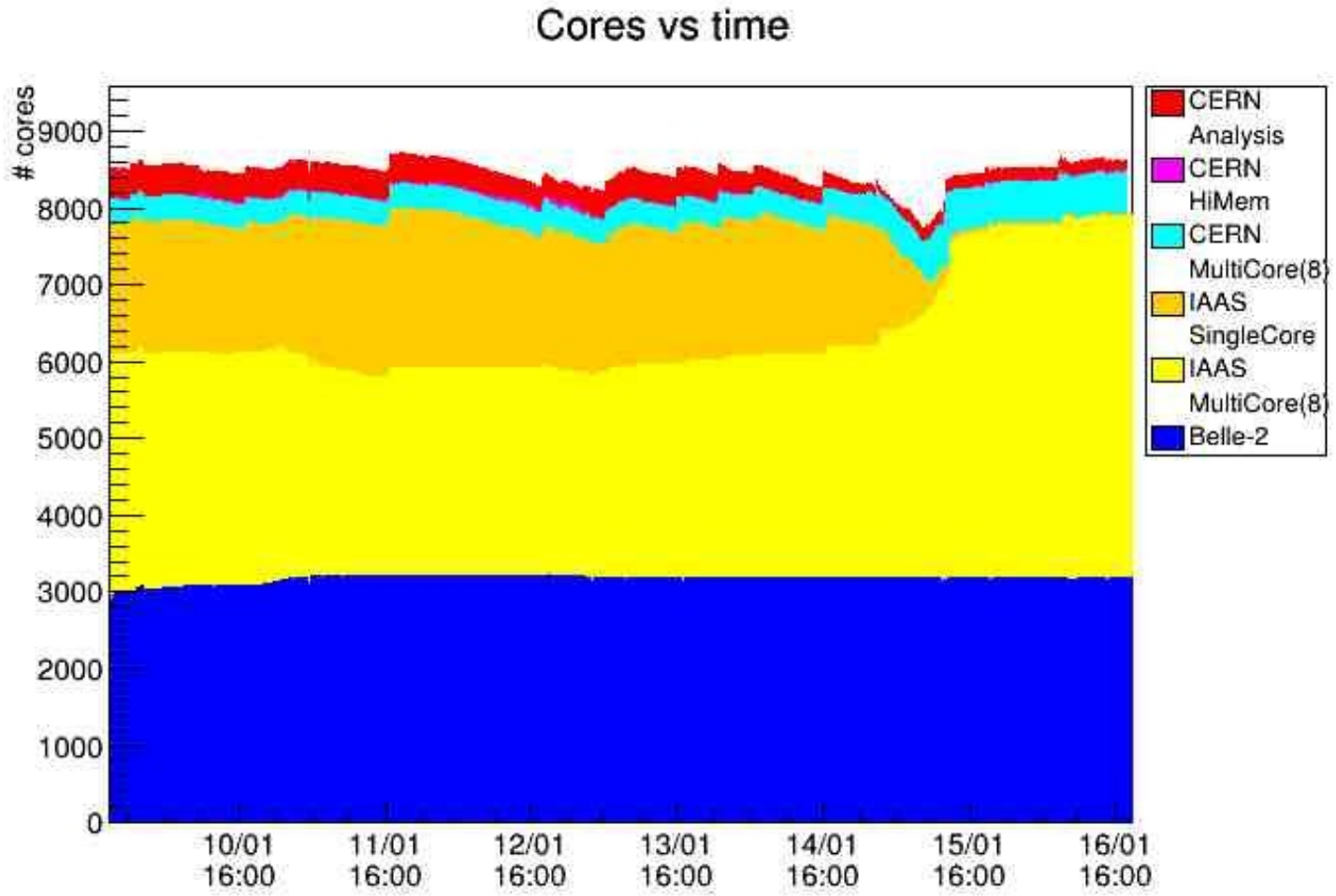
UVic Clouds:
Dynafed and research connectivity

Overall Status
Dynafed
Research Connectivity
Context Awareness
Further news

R. Seuster (Uvic)
Site Jamboree
18. Jan. 2017

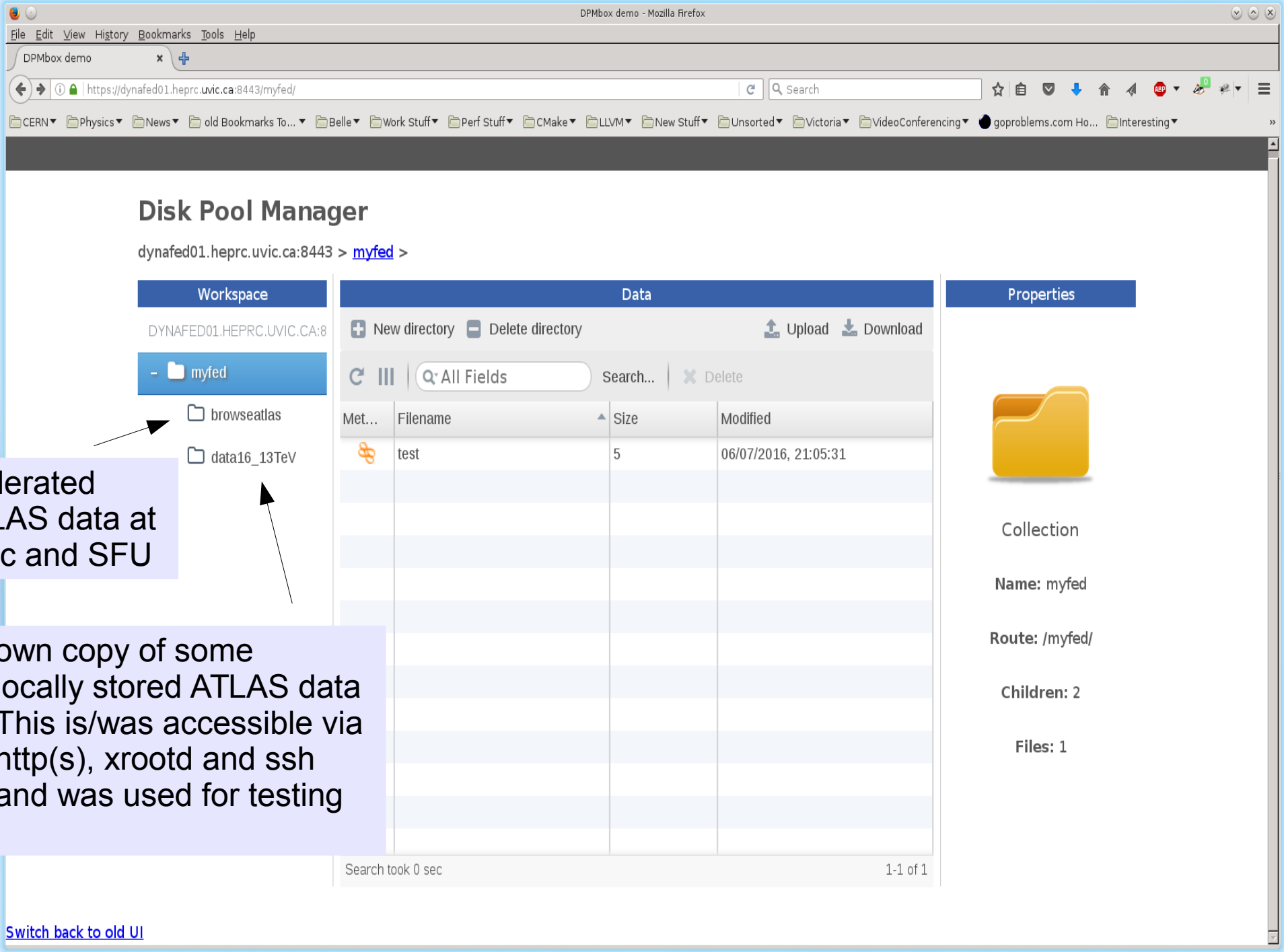
Introduction: overall status

- ~ 8600 cores running on 15 clouds in NorthAmerica + Europa
- No outages or mayor operational issues for months



Introduction: dynafed

- during ATLAS computing week last summer was agreed to test dynafed which utilizes standard code and protocols like http(s), webdav(optional), S3(optional), etc.
  http://lcgdm.web.cern.ch/dynafed-dynamic-federation-project
- unique front-end for all protocols
  e.g. webdav or S3 authentification hidden from user
  - interfaces also to our protocols like xrootd, ...
- Installation of dynafed was easy, and initially federated two Canadian sites on the westcoast. Very easy to add more sites or to declare own storage areas (later more)
- Web location, protected with x509:
  https://dynafed01.heprc.uvic.ca:8443/myfed/
  added local storage recently, but disabled due to RAID issues



Federated ATLAS data at UVic and SFU

own copy of some locally stored ATLAS data
This is/was accessible via http(s), xrootd and ssh and was used for testing

Status

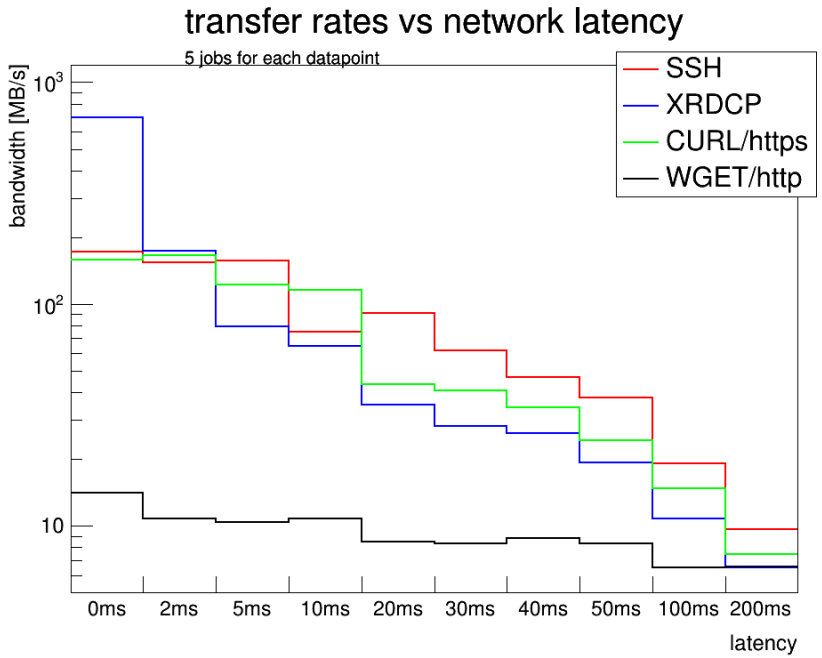
shown already in TCB August 29th

- Initial installed as dynafed and opened access via other protocols for comparisons: own webdav space, xrootd, and http(s) access via apache
  - cleaned up after initial tests to bare minimum
- very first tests of just curl'ing files worked, also with re-director
  - downloaded file from SFU via dynafed redirector
- as a first test, compared access to same file via different protocols: download via xrdcp, ssh, curl (https) and wget (http)
- copied file between two closeby machines, but introduced known latency on file-server via netem package https://wiki.linuxfoundation.org/networking/netem
  - delays of 2,5, 10, 20, 30, 40, 50, 100 and 200 ms

plain transfer times

shown TCB, August 29th

- plain wget (http) very slow
  - using apache (default conf)
- other protocols comparable, xrdcp fastest for very small network latencies
- ssh in general quite good performance
- slowdown expected due to tcp protocol and limited buffersizes, matches expectations fairly well



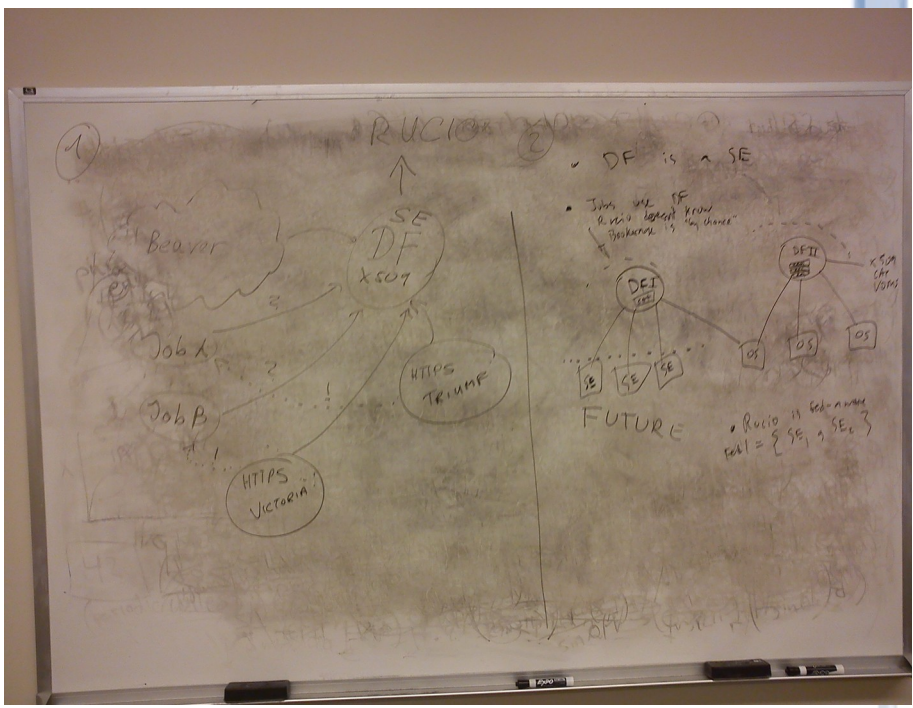
- x-axis shows the artificial network latency introduced with the netem package of the linux kernel

Status
since then

- tested HammerCloud tests, failed immediately
- were waiting for new pilot utilizing new site-movers to be in production, then CHEP slowed down work
- some important follow up after CHEP also distracted from dynafed
- Frank Berghaus joined our team in January, will be stationed (initially) at CERN and started to continue this work, will work in parallel with strong interaction between each other

Our Workplan for 2017+

- setup two(*) dynafed servers: one at Uvic, one at CERN
  - attach own, free (object) storage
    - e.g. CEPH with S3 front end
- declare them as SE in rucio
  - can rucio fill a dynafed SE ?



- create additional new storage and add it to federation
  - jobs initially on few local clouds, aim for small duplication of data to test GeoIP works for production jobs

(*) Why two? CERN and UVic teams can cross check each other and run tests independently, can/will federate/merge when everything works fine past experience: 9h time difference very bad for direct collaboration

How we would like to use dynafed

in future (months to years ?)

- now becoming more important, as we commission the second cloud in Europe for our system
- want to federate 'whole world'
  (at least sites close to clouds we use)
- declare to panda, that we have resources available close to SE e.g. in Europe and receive jobs for it
- current CloudScheduler would submit jobs randomly
  - rewrite planned during next few months
- how does panda send jobs to sites ?
- how would cloudscheduler and harvester interact ?

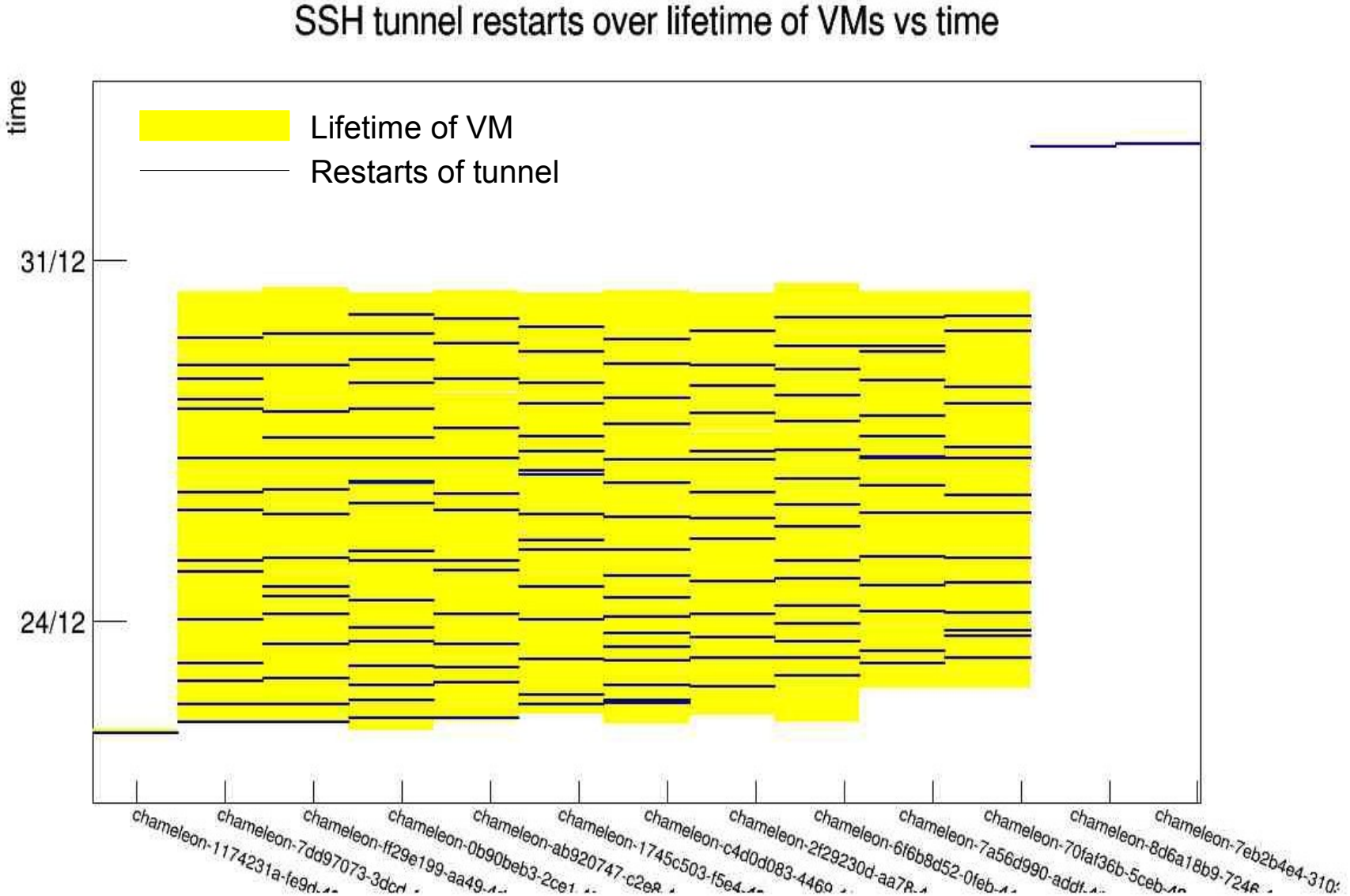
Research Connectivity

- one of our clouds is particularly problematic:
  - restrictive DNS wouldn't resolve some KEK machines
  - often disconnects and lost heartbeats in condor; probably caused by edge router
- investigated 'virtual router' with simplistic approach:
  - create fancy ssh tunnel to route 'all' traffic back to UVic
    used sshuttle (author seemed to be from RIM)
    https://github.com/apenwarr/sshuttle
  - can configure which subnets (not) to tunnel
- maybe these slides give ideas to other people to solve similar problems with their setup → please contact me for further discussions / help / instructions

Research Connectivity first results

- running Belle II jobs, as their system is more resilient for all kinds of failed jobs (e.g. VMs just 'disappear')
- ran O(10) VMs over Christmas w/o any supervision, but seems all jobs finished successful !
- looked at logfiles after Christmas break and made plots shown on next slide

Research Connectivity plots



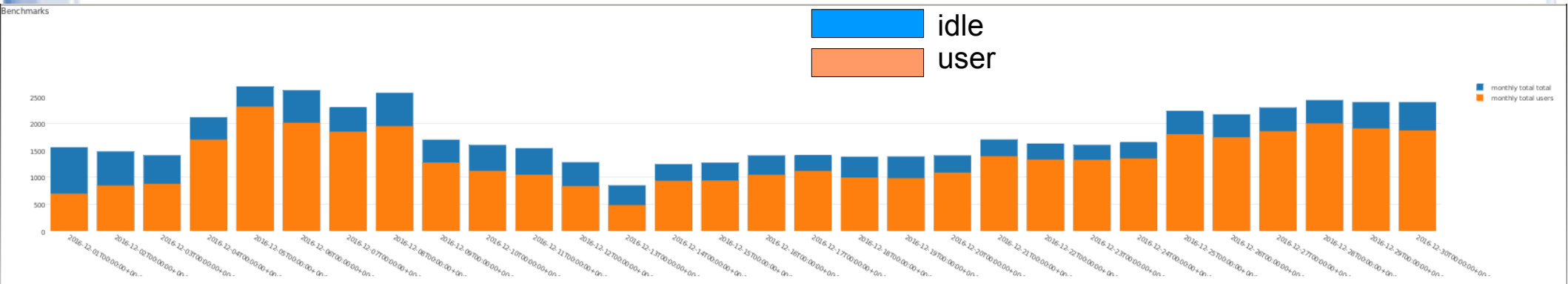
→ Unfortunately no clear pattern for required restarts

Further News

- context aware intervention: stalled java processes require automated retiring and killing of VMs
  → CHEP talks 2016
  - condor runs java benchmarks if it finds working java env
    trying to address this now
- datacentred in UK: 'commissioning' phase
  - required ports for condor were closed earlier last year
- should be ready to start VMs on LRZ cloud this week from within cloudscheduler
  - able to start CERNVM, but then hit quota (other power user was in our group by mistake), resolved now

benchmarking

- code from benchmarking WG at CERN on CVMFS
- used to get VM's performance, monitor uptime and job runtimes (15min time resolutions) to calculate machine efficiencies
- overall good efficiency (small blue bar on to off orange bar) but some days stick out ...



Our Questions

- Further questions for dynafed

  Can we use 'volatile' storage ?
  Is this storage not managed by rucio, but can contain accessible data ?
  Might be needed if we see problems rucio talking to dynafed server

Questions ?