CMS Federation

XCache

CPU Performance comparison based on MINIAOD reading options: local versus remote

J. Balcas

May 11, 2023

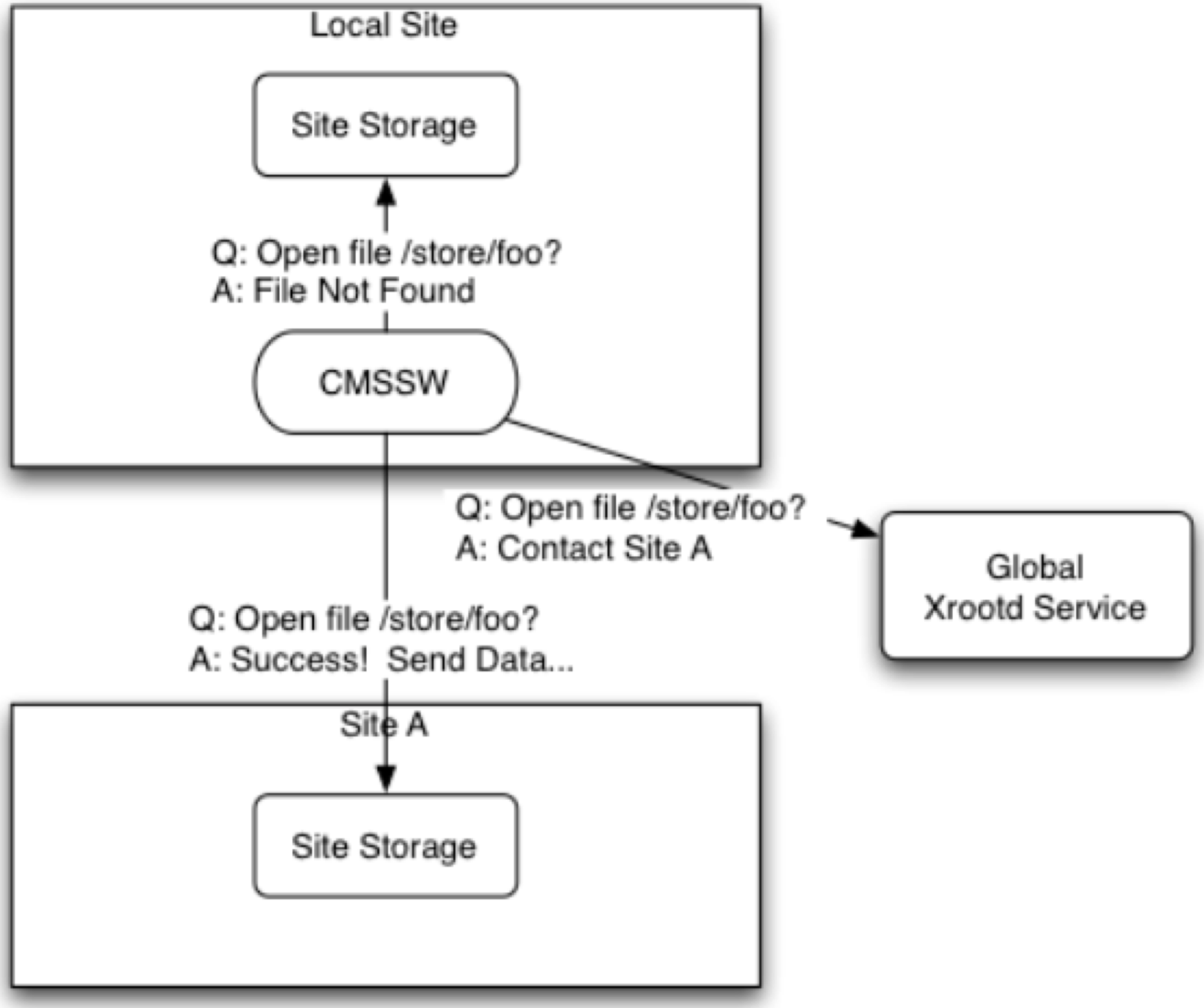
hadoop

ceph

XRootD

How Jobs are accessing data today?

Caltech



J. Balcas "CPU Performance comparison based on MINIAOD reading options: local versus remote", CHEP 2023

Usual CMSSW Job Running at Site do this:

- Try to open local file, if "File Not Found" go to next step;
- Ask Regional XRootD Redirector to find where the file is. If it contains location (usually IP of another site), Open that file.

But what happens when it reads via one type of storage or another and how good performance is?

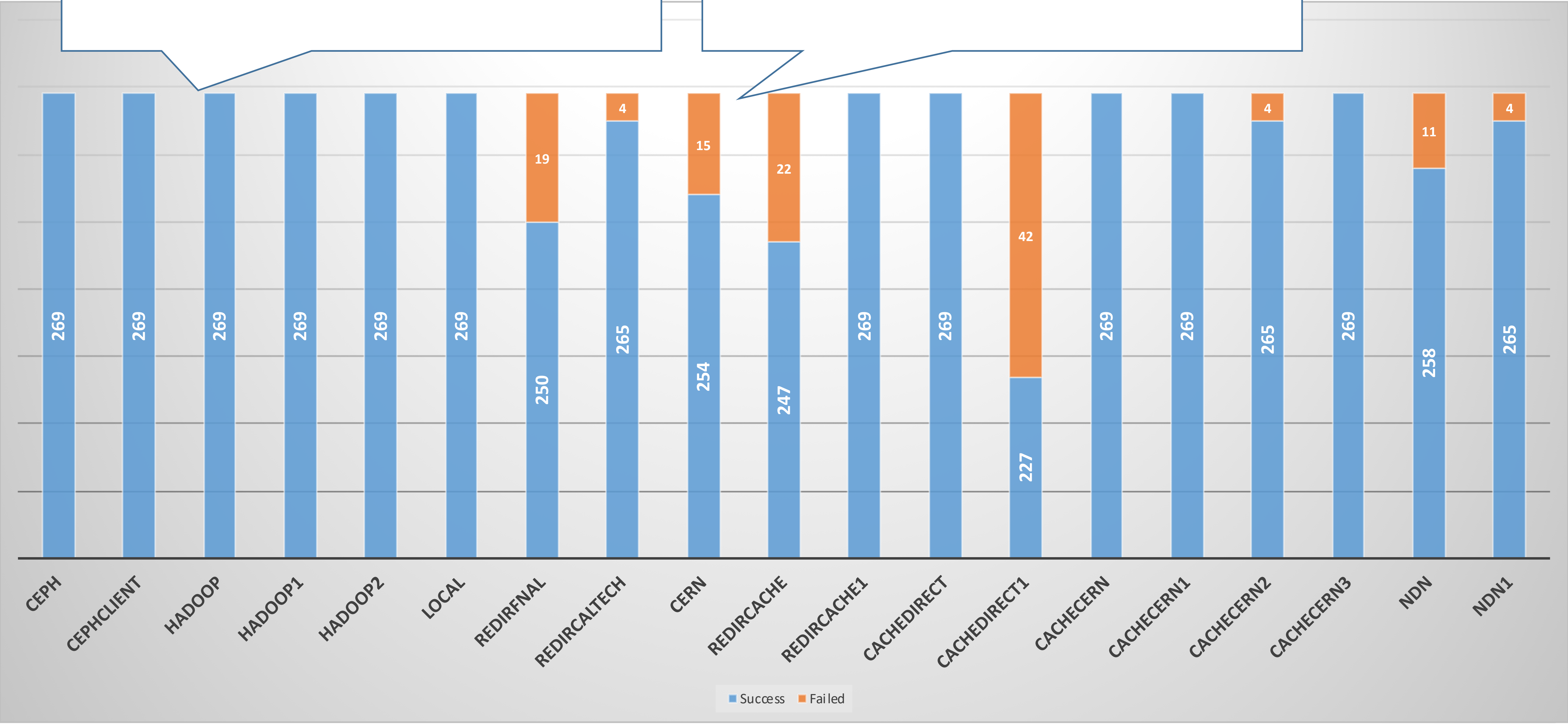
Storages to test

- Pset: Read MINIAOD File and process it to output NANOAOD File
- Input Dataset: /DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3_ext2-v1/MINIAODSIM
- Each Jobs read whole single file and it's runtime varies from ~4hr to ~8hrs
  - Job runtime varies depending on the storage solution we use
- Each test is running 269 jobs to test the specific Storage. (Most of the test were repeated several times)
  - Only 1 test runs at the time.
  - Each job opens 1 file and reads all file.
  - CMSSW Uses Application based caching.
  - No Fallback. Uses only specific rule!
- Jobs always run on same list of machines: blade-1.tier2 to blade-8.tier2. No other CPU Load

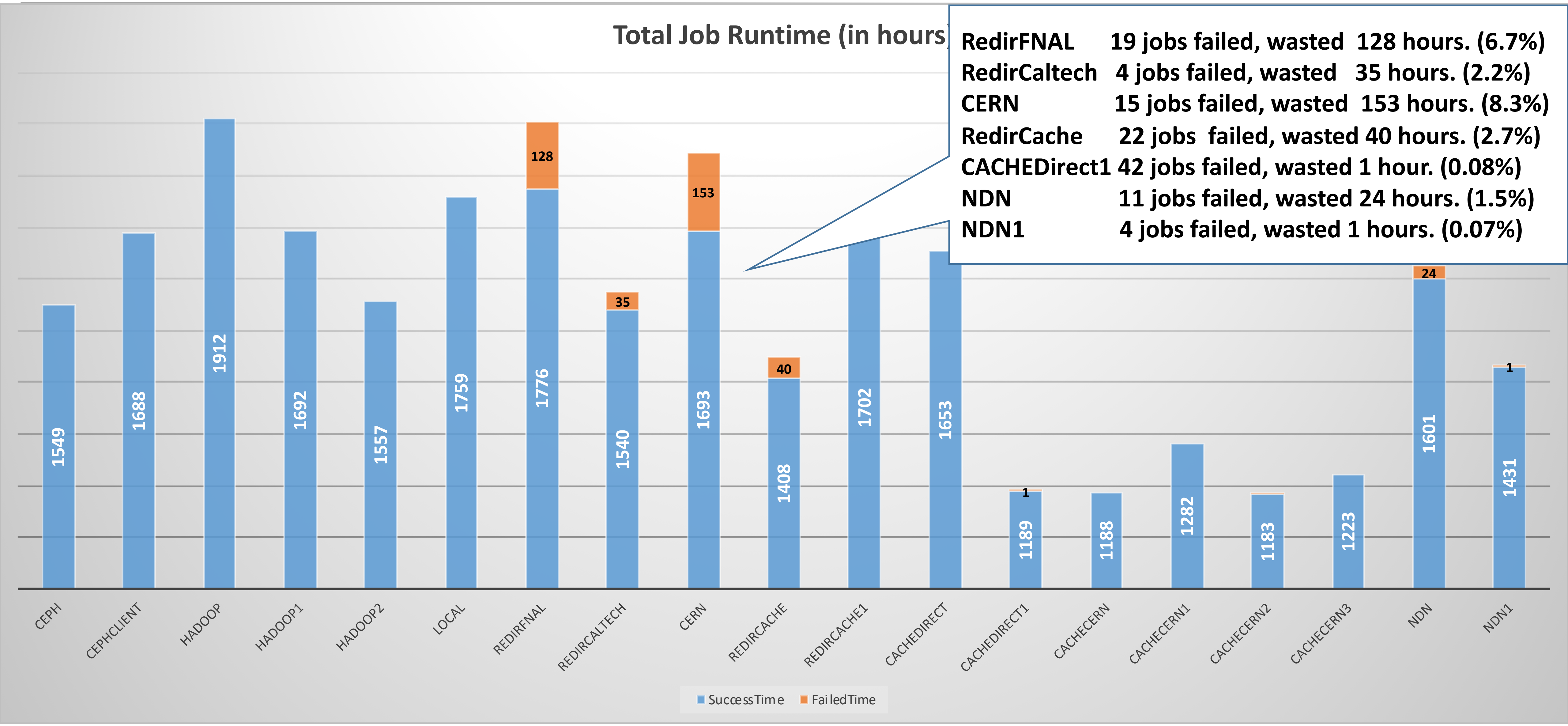
F

0 Failures on Local reads.

Most of the times remote connections show failures.



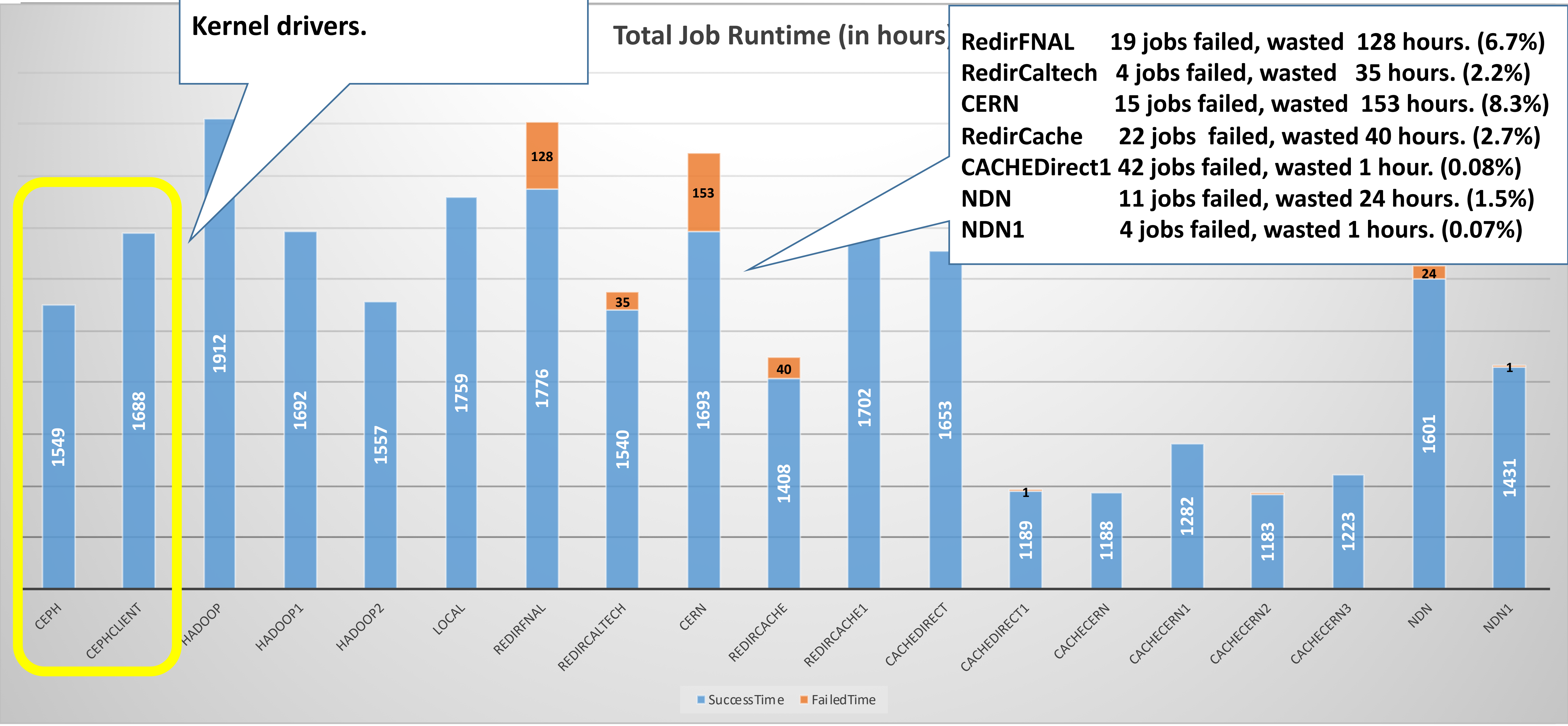
J. Balcas “CPU Performance comparison based on MINIAOD reading options: local versus remote”, CHEP 2023



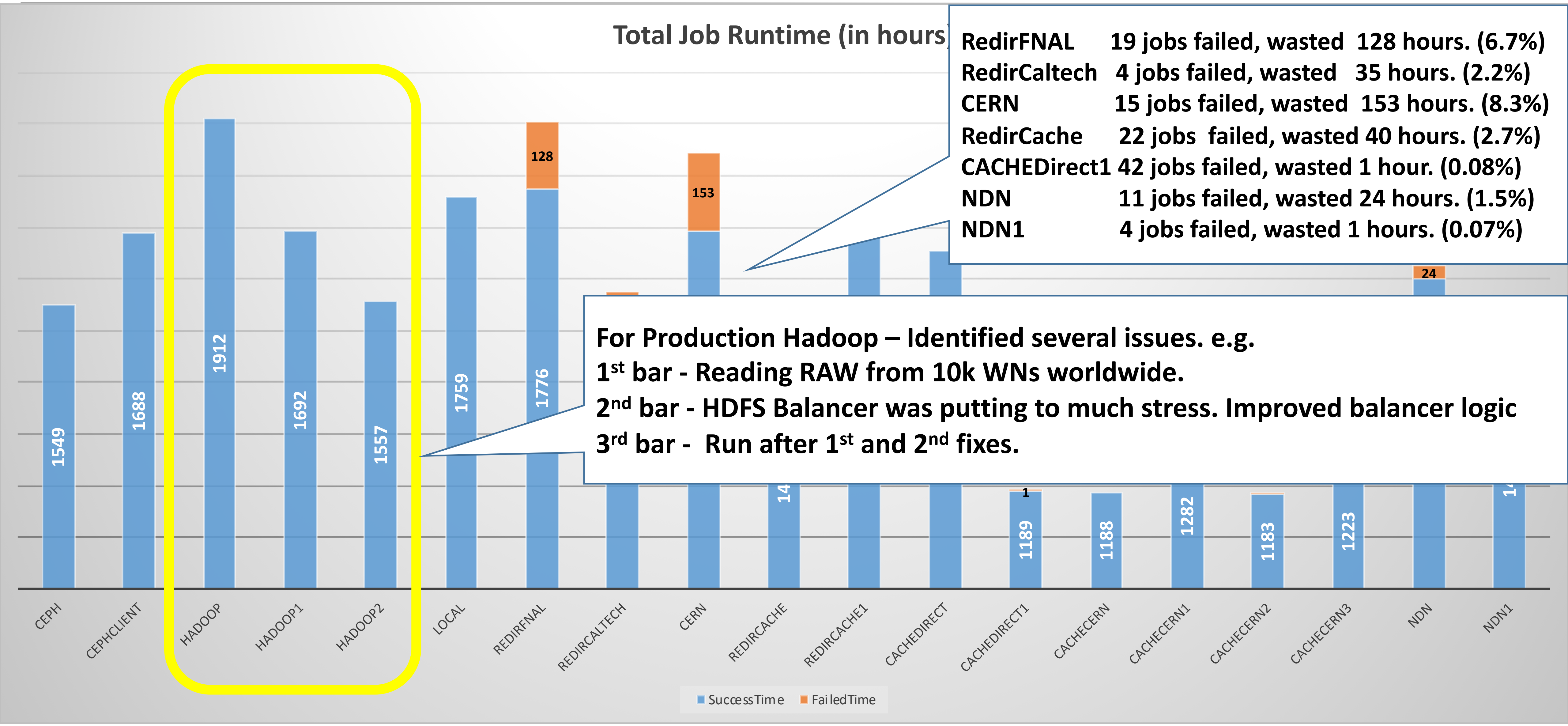
RedirFNAL	**19 jobs failed, wasted 128 hours. (6.7%)**
RedirCaltech	**4 jobs failed, wasted 35 hours. (2.2%)**
CERN	**15 jobs failed, wasted 153 hours. (8.3%)**
RedirCache	**22 jobs failed, wasted 40 hours. (2.7%)**
CACHEDirect1	**42 jobs failed, wasted 1 hour. (0.08%)**
NDN	**11 jobs failed, wasted 24 hours. (1.5%)**
NDN1	**4 jobs failed, wasted 1 hours. (0.07%)**

J. Balcas “CPU Performance comparison based on MINIAOD reading options: local versus remote”, CHEP 2023

Interesting finding that using CEPH Fuse client is slower than Kernel drivers.



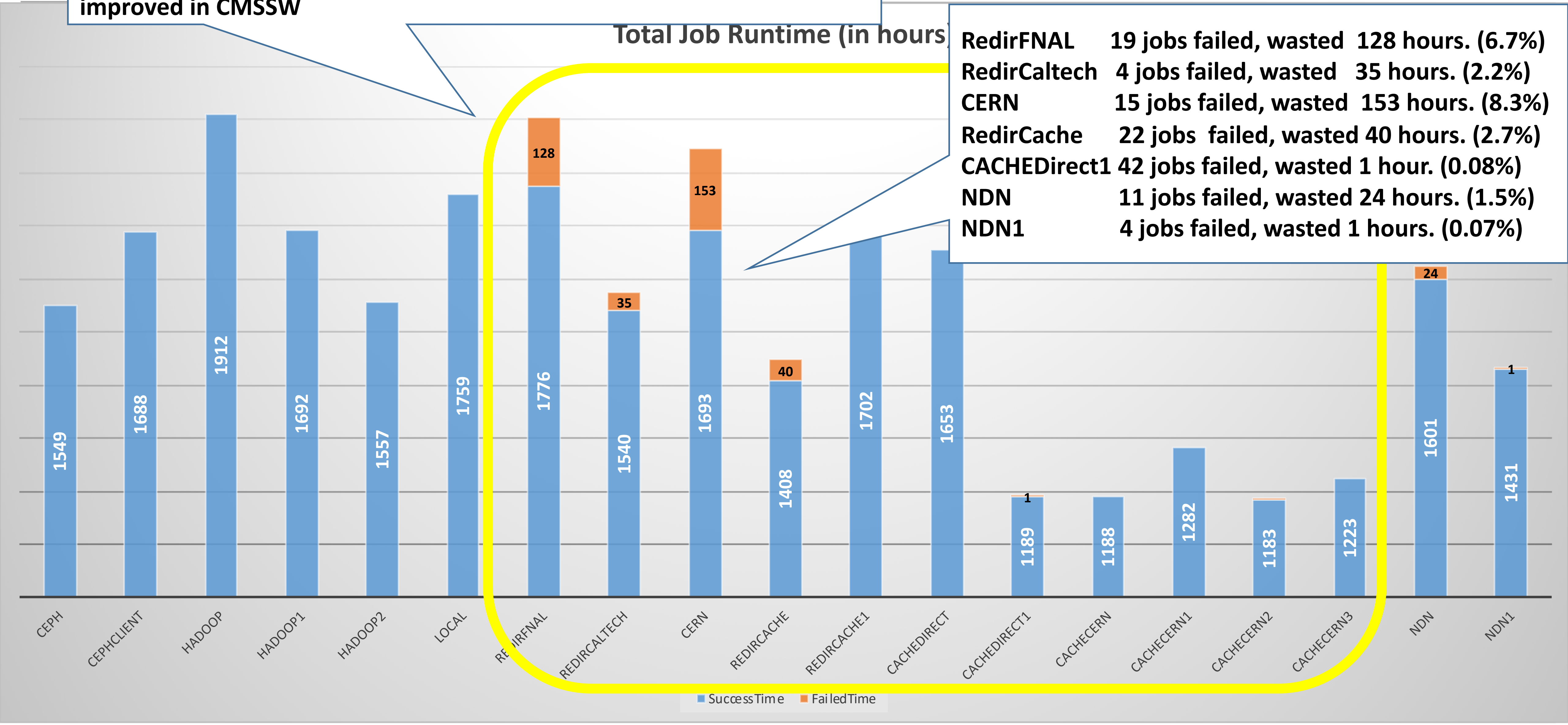
RedirFNAL	19 jobs failed, wasted 128 hours. (6.7%)
RedirCaltech	4 jobs failed, wasted 35 hours. (2.2%)
CERN	15 jobs failed, wasted 153 hours. (8.3%)
RedirCache	22 jobs failed, wasted 40 hours. (2.7%)
CACHEDirect1	42 jobs failed, wasted 1 hour. (0.08%)
NDN	11 jobs failed, wasted 24 hours. (1.5%)
NDN1	4 jobs failed, wasted 1 hours. (0.07%)



RedirFNAL	19 jobs failed, wasted 128 hours. (6.7%)
RedirCaltech	4 jobs failed, wasted 35 hours. (2.2%)
CERN	15 jobs failed, wasted 153 hours. (8.3%)
RedirCache	22 jobs failed, wasted 40 hours. (2.7%)
CACHEDirect1	42 jobs failed, wasted 1 hour. (0.08%)
NDN	11 jobs failed, wasted 24 hours. (1.5%)
NDN1	4 jobs failed, wasted 1 hours. (0.07%)

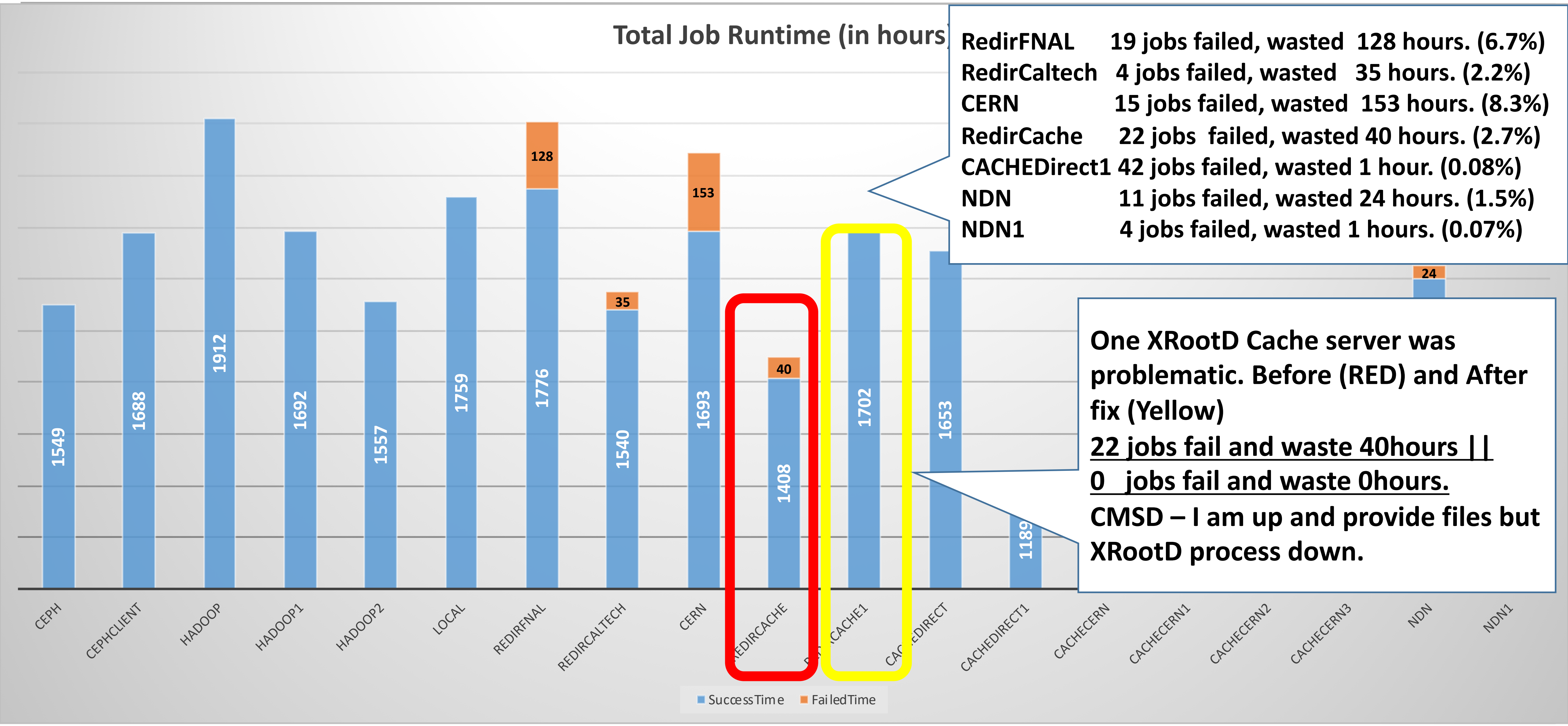
For Production Hadoop – Identified several issues. e.g.
1st bar - Reading RAW from 10k WNs worldwide.
2nd bar - HDFS Balancer was putting to much stress. Improved balancer logic
3rd bar - Run after 1st and 2nd fixes.

CMSSW XRootD source selection is not reconsidering old sources once they are tagged as bad. (The load changes on all of the XRootD Servers over time and they could be reconsidered again after X minutes.) To be improved in CMSSW



RedirFNAL	19 jobs failed, wasted 128 hours. (6.7%)
RedirCaltech	4 jobs failed, wasted 35 hours. (2.2%)
CERN	15 jobs failed, wasted 153 hours. (8.3%)
RedirCache	22 jobs failed, wasted 40 hours. (2.7%)
CACHEDirect1	42 jobs failed, wasted 1 hour. (0.08%)
NDN	11 jobs failed, wasted 24 hours. (1.5%)
NDN1	4 jobs failed, wasted 1 hours. (0.07%)

J. Balcas "CPU Performance comparison based on MINIAOD reading options: local versus remote", CHEP 2023



RedirFNAL	19 jobs failed, wasted 128 hours. (6.7%)
RedirCaltech	4 jobs failed, wasted 35 hours. (2.2%)
CERN	15 jobs failed, wasted 153 hours. (8.3%)
RedirCache	22 jobs failed, wasted 40 hours. (2.7%)
CACHEDirect1	42 jobs failed, wasted 1 hour. (0.08%)
NDN	11 jobs failed, wasted 24 hours. (1.5%)
NDN1	4 jobs failed, wasted 1 hours. (0.07%)

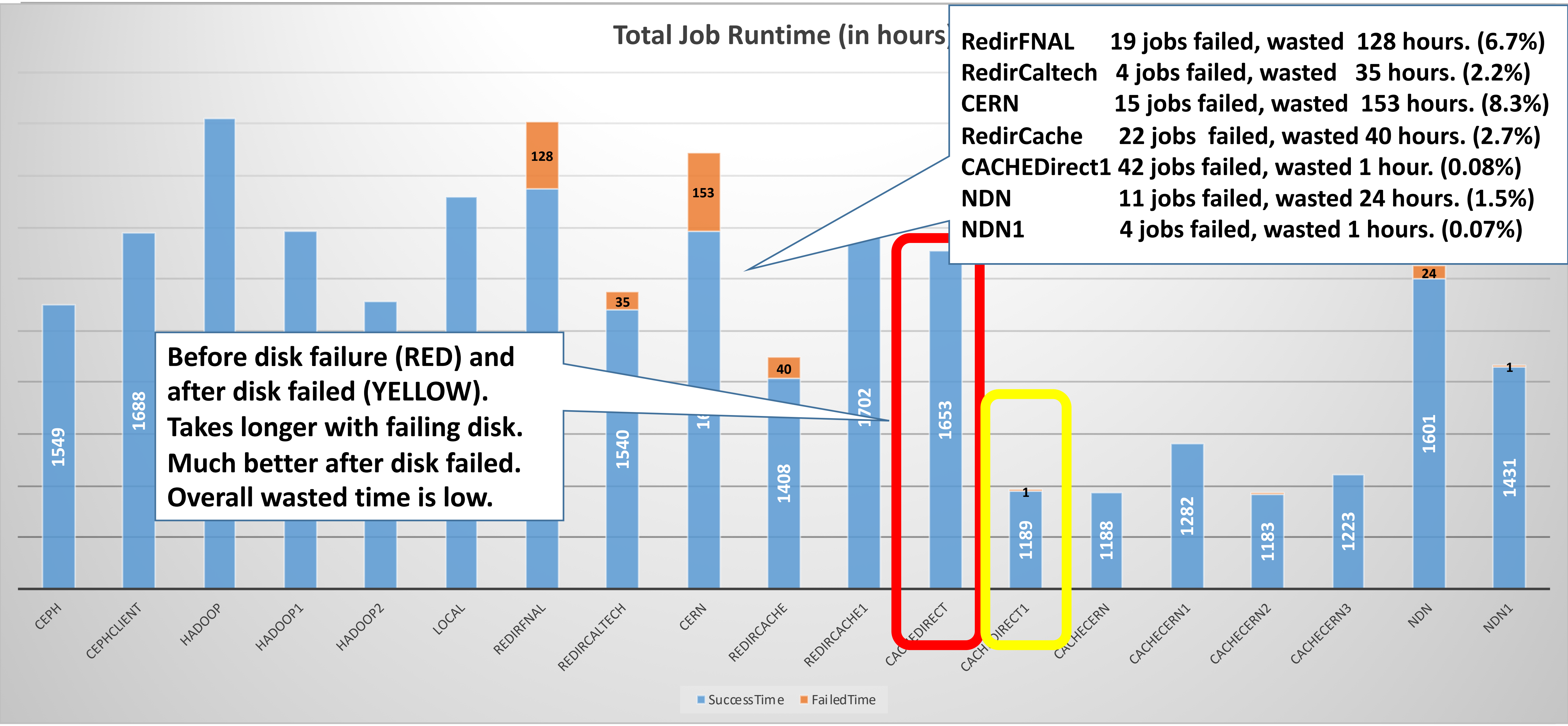
One XRootD Cache server was problematic. Before (RED) and After fix (Yellow)

<u>22 jobs fail and waste 40hours ||</u>

<u>0 jobs fail and waste 0hours.</u>

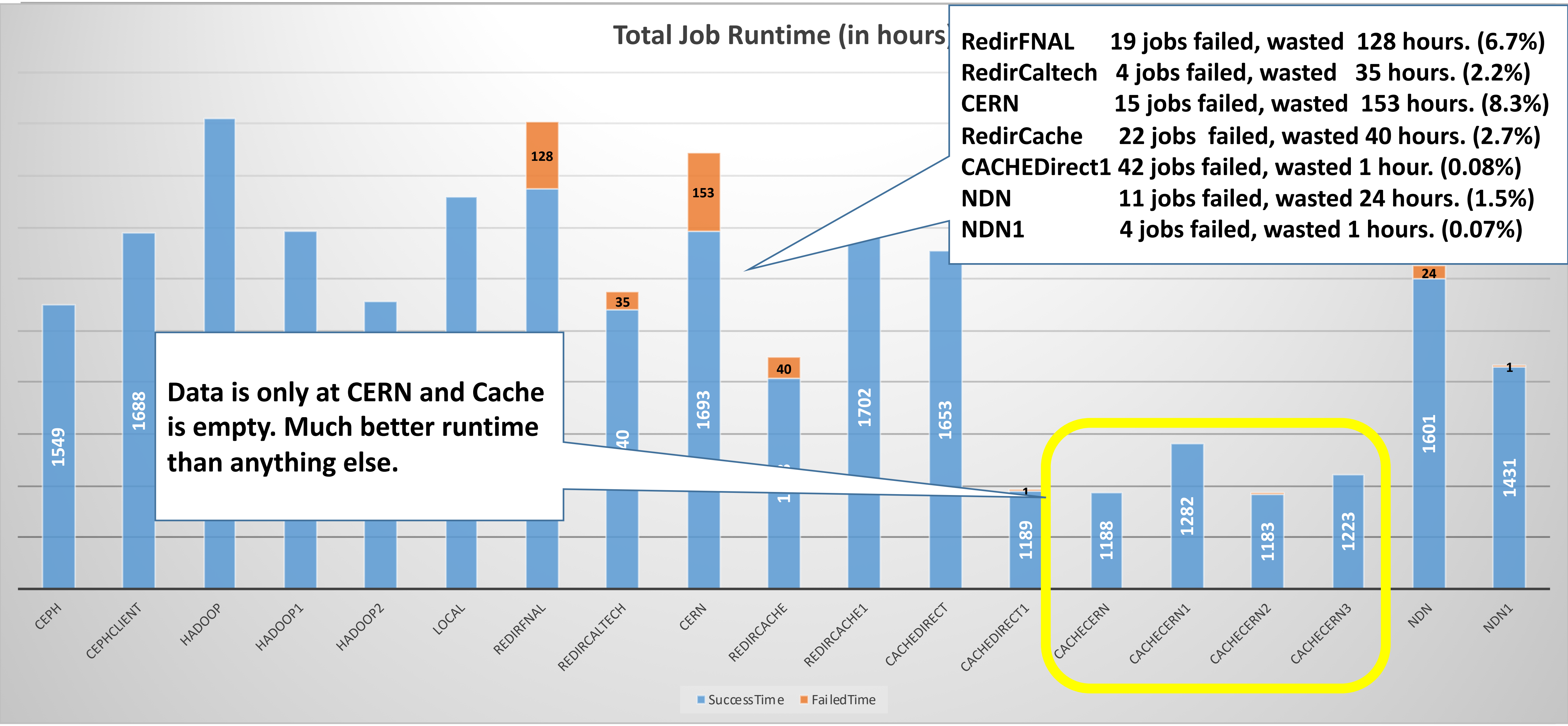
CMSD – I am up and provide files but XRootD process down.

J. Balcas “CPU Performance comparison based on MINIAOD reading options: local versus remote”, CHEP 2023



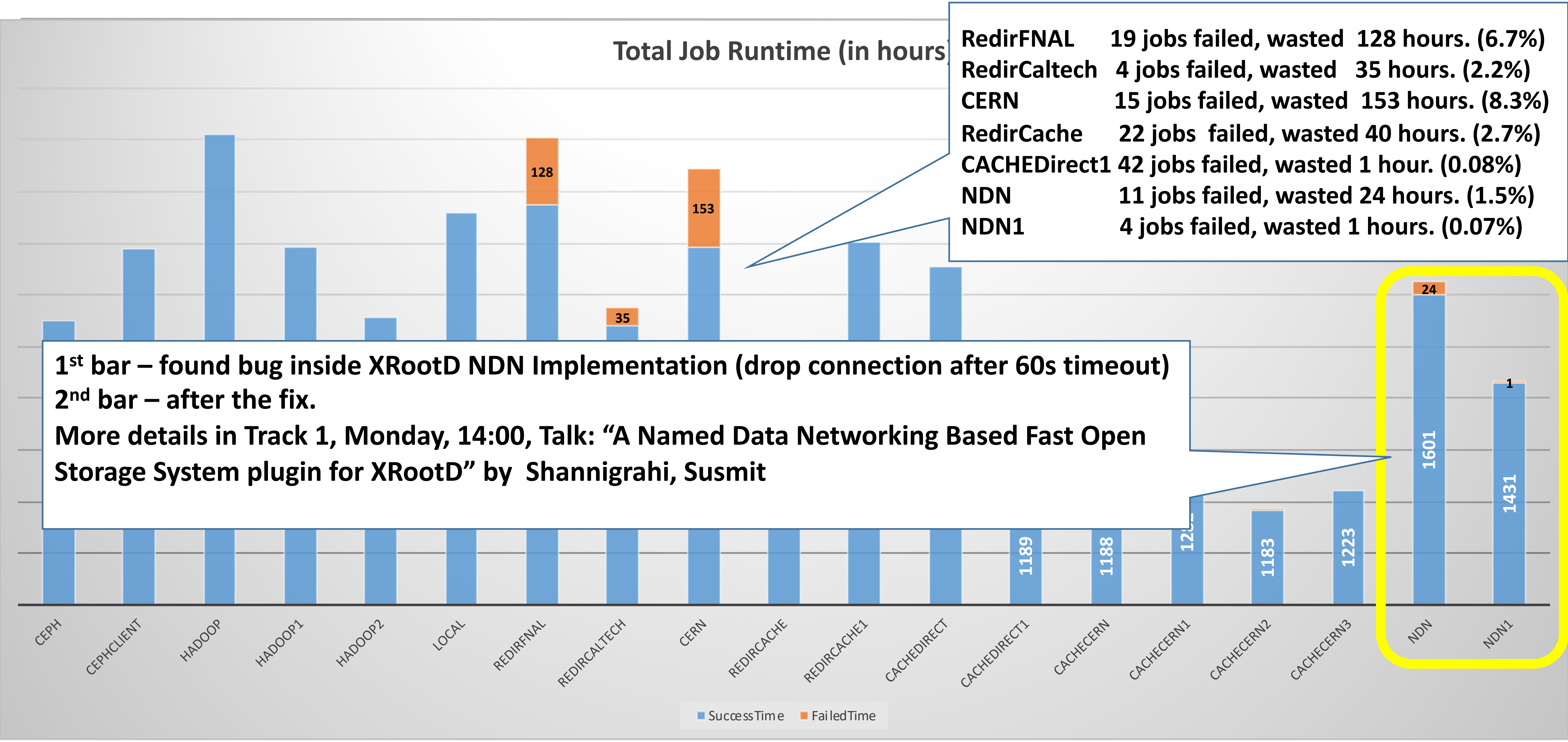
RedirFNAL	19 jobs failed, wasted 128 hours. (6.7%)
RedirCaltech	4 jobs failed, wasted 35 hours. (2.2%)
CERN	15 jobs failed, wasted 153 hours. (8.3%)
RedirCache	22 jobs failed, wasted 40 hours. (2.7%)
CACHEDirect1	42 jobs failed, wasted 1 hour. (0.08%)
NDN	11 jobs failed, wasted 24 hours. (1.5%)
NDN1	4 jobs failed, wasted 1 hours. (0.07%)

Before disk failure (RED) and after disk failed (YELLOW). Takes longer with failing disk. Much better after disk failed. Overall wasted time is low.



RedirFNAL	19 jobs failed, wasted 128 hours. (6.7%)
RedirCaltech	4 jobs failed, wasted 35 hours. (2.2%)
CERN	15 jobs failed, wasted 153 hours. (8.3%)
RedirCache	22 jobs failed, wasted 40 hours. (2.7%)
CACHEDirect1	42 jobs failed, wasted 1 hour. (0.08%)
NDN	11 jobs failed, wasted 24 hours. (1.5%)
NDN1	4 jobs failed, wasted 1 hours. (0.07%)

Data is only at CERN and Cache is empty. Much better runtime than anything else.



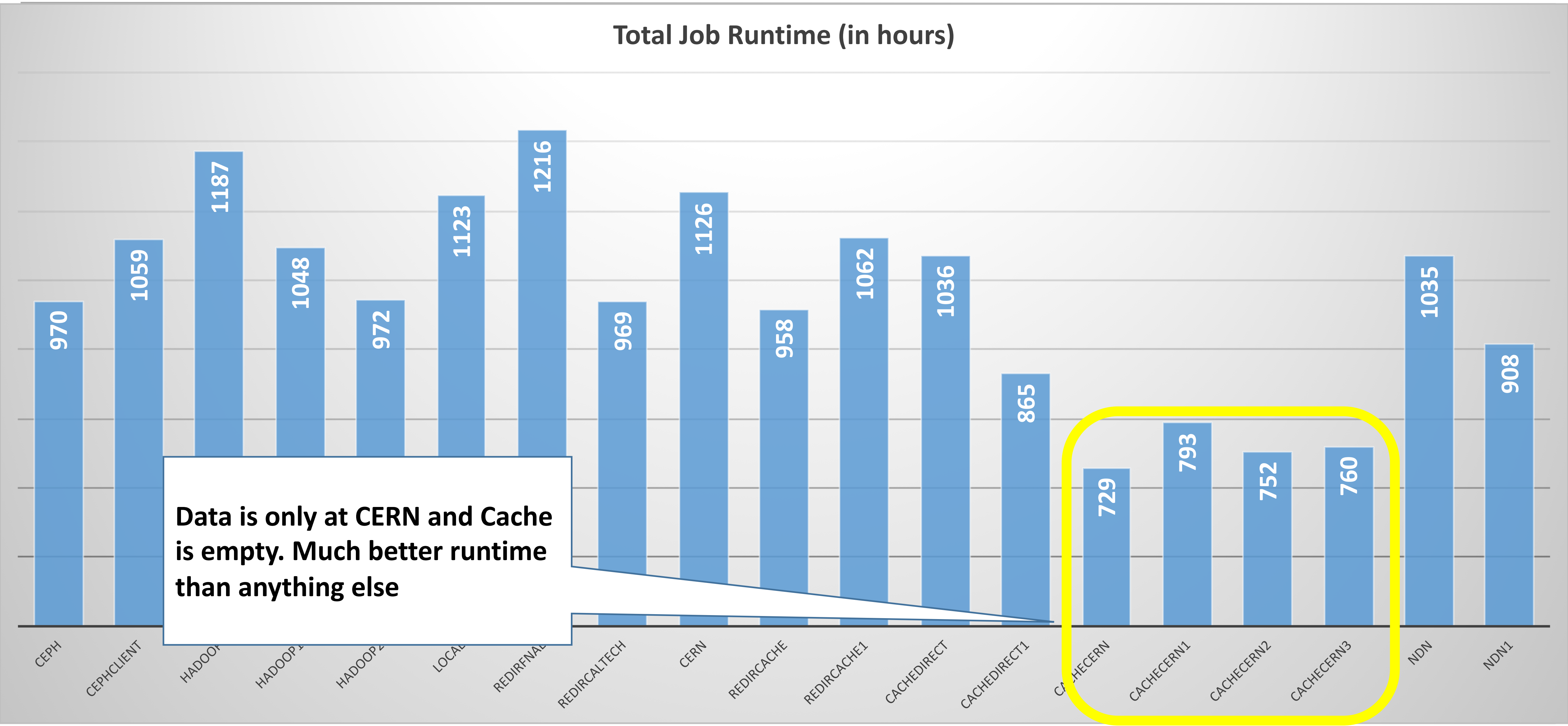
RedirFNAL	19 jobs failed, wasted	128 hours. (6.7%)
RedirCaltech	4 jobs failed, wasted	35 hours. (2.2%)
CERN	15 jobs failed, wasted	153 hours. (8.3%)
RedirCache	22 jobs failed, wasted	40 hours. (2.7%)
CACHEDirect1	42 jobs failed, wasted	1 hour. (0.08%)
NDN	11 jobs failed, wasted	24 hours. (1.5%)
NDN1	4 jobs failed, wasted	1 hours. (0.07%)

1st bar – found bug inside XRootD NDN Implementation (drop connection after 60s timeout)
2nd bar – after the fix.
More details in Track 1, Monday, 14:00, Talk: “A Named Data Networking Based Fast Open Storage System plugin for XRootD” by Shannigrahi, Susmit

Highlights till now

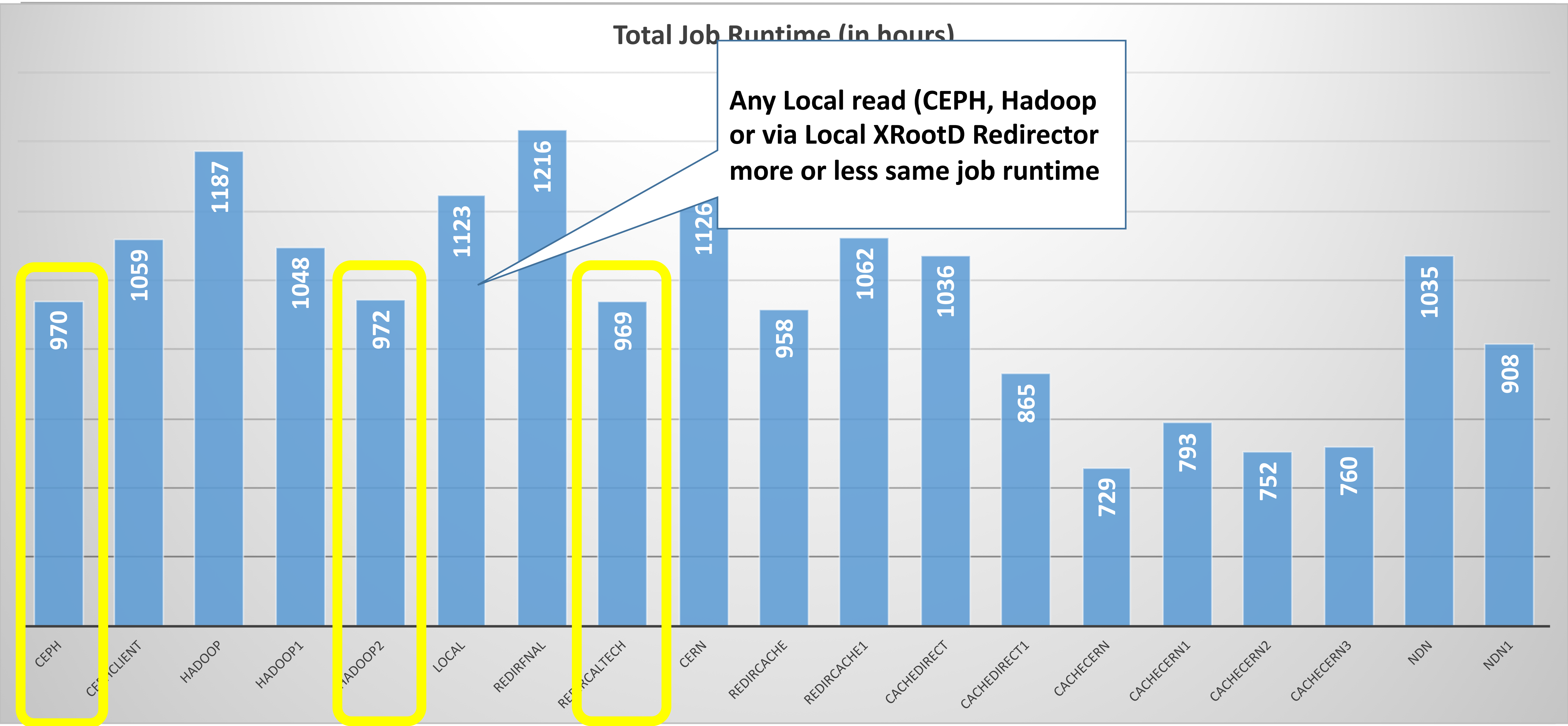
- Kernel mount for CEPH performs better compared to CEPH Client Fuse (~10% faster)
  - Tests have been retried several times.
- Hadoop is 23% slower than CEPH Kernel and 13% slower than CEPH Client Fuse in an overloaded environment.
  - **It is not equal comparison as HDFS is used in production.**
  - We can see that with decreased load on HDFS it performed equally as CEPH Kernel mount.
- CMSSW XRootD algorithm drops good servers from it's list. It could reconsider servers again after X minutes. To be implemented in CMSSW.
- Source reselection on caches is not good. Wasted storage space and increased network usage.
- Caches depend a lot on the data sources (if it is bad, job will fail.) Small wasted time for failed jobs (XRootD client timeouts after 90s).
- Single good source location (like CERN) and reading via cache shows very nice throughput and no big failures.

Same plot as before, but only equally successful jobs (169)



Data is only at CERN and Cache is empty. Much better runtime than anything else

Same plot as before, but only equally successful jobs (169)

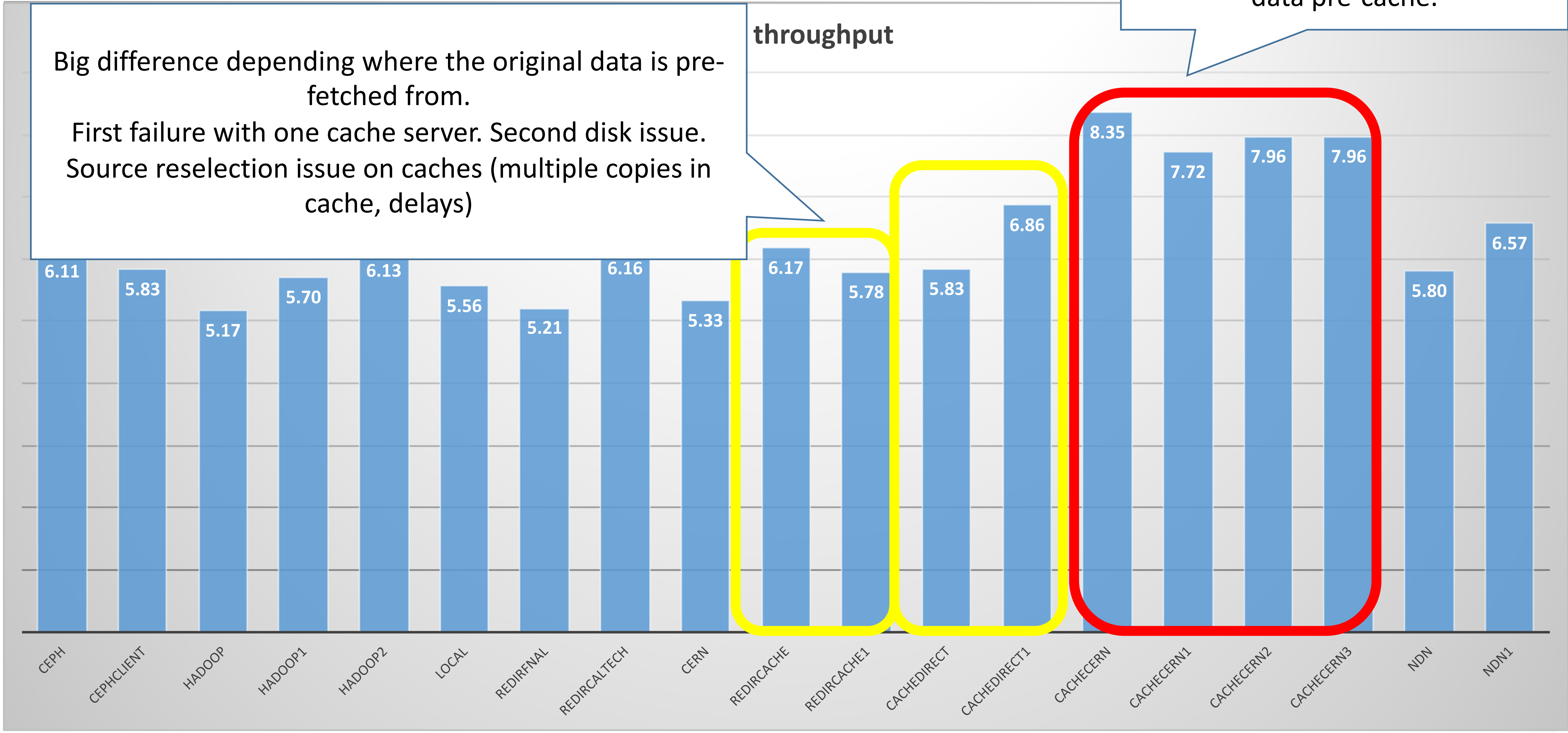


Any Local read (CEPH, Hadoop or via Local XRootD Redirector more or less same job runtime

Event Throughput (only 169 jobs)

More than 25% Event throughput vs others. Good Source and fast data pre-cache.

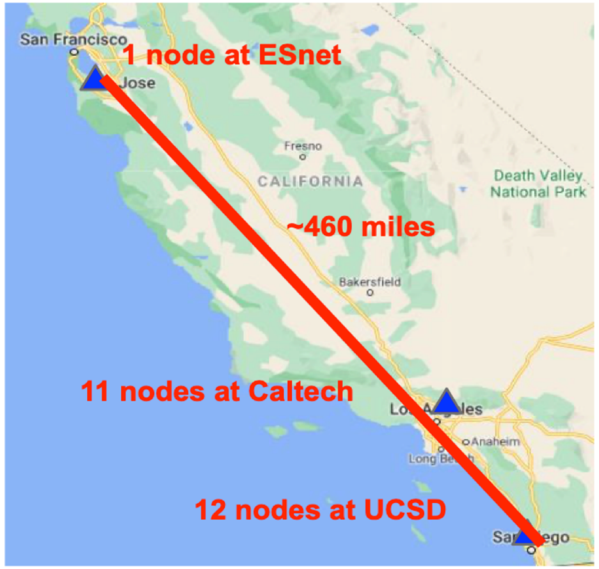
Big difference depending where the original data is pre-fetched from.
First failure with one cache server. Second disk issue.
Source reselection issue on caches (multiple copies in cache, delays)



Southern California Petabyte Scale Cache (SoCal Repo)

SoCal Repo consists of 24 federated storage nodes for US CMS

- 12 nodes at UCSD: each with 24 TB, 10 Gbps network connection
- 11 nodes at Caltech: each with storage sizes ranging from 96TB to 388TB, 40 Gbps network connections
- 1 node at LBNL (by ESnet): 44 TB storage, 40 Gbps network connection
- Approximately 2.5PB of total storage capacity
- ~100 miles between UCSD and Caltech nodes, round trip time (RTT) < 3 ms
- ~460 miles between LBNL and UCSD nodes, RTT ~10 ms



1 node at ESnet

~460 miles

11 nodes at Caltech

12 nodes at UCSD

Sunnyvale–San Diego is the relevant distance scale

Statistics about US CMS data analysis with MINIAOD/NANOAOD

- Analysis Object Data (AOD):
  - 384 PB of RAW
  - 240 PB of AOD

  } Mostly on Tape: accessed a few times per year

  - 30 PB of MINIAOD
  - 2.4 PB of NANOAOD

  } **Mostly on disk: heavily re-used by many researchers**
- More than 90% of analyses work with either MiniAOD or NanoAOD

500 Miles is an interesting distance for regional caches !!!

J. Balcas "CPU Performance comparison based on MINIAOD reading options: local versus remote", CHEP 2023

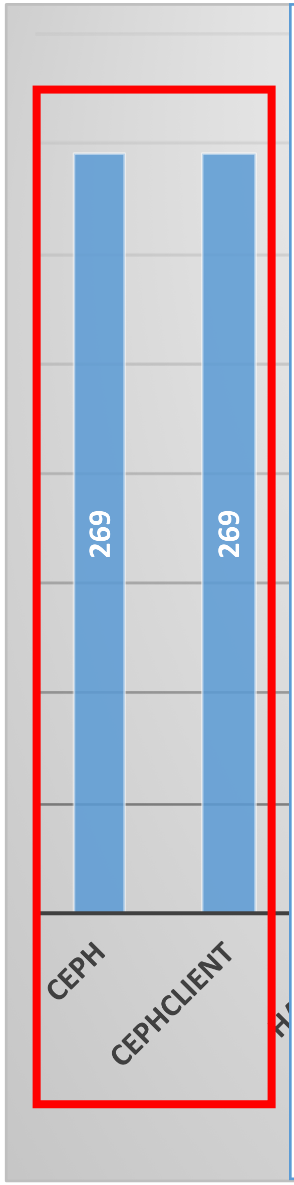
Highlights/Future

- Caches perform better than any other storage solutions:
    - Problem is to make sure all data sources provide files correctly. (Seen very good performance even if source is **170ms away**). Cache will pre-cache whole file in advance.
    - Even the source providing data is bad, wasted time is low. Job fallbacks to read old way via Global redirector.
- CMSSW/CMSSW-XRootD code improvements:
    - Reconsider bad sources after X Minutes.
    - CMSSW no reselection on caches. (Do not duplicate data on caches)
- Publish cache content to all AAA*, but not allow cache recalls. Inform Rucio about cache content.
- Test and allow Rucio to prefetch needed data to cache in advance. (Virtual placement)
- My wish: Do not allow users to **CONTROL Overflow flag. Leave it for Operators/Global Pool Rules to make decision (Controllable environment)**

*AAA – Any Data, Any Time, Anywhere

BACKUP

Job Success/Failures



CEPH Test Instance (no any other load to it. Details on Setup in Backup slides.)

CEPH

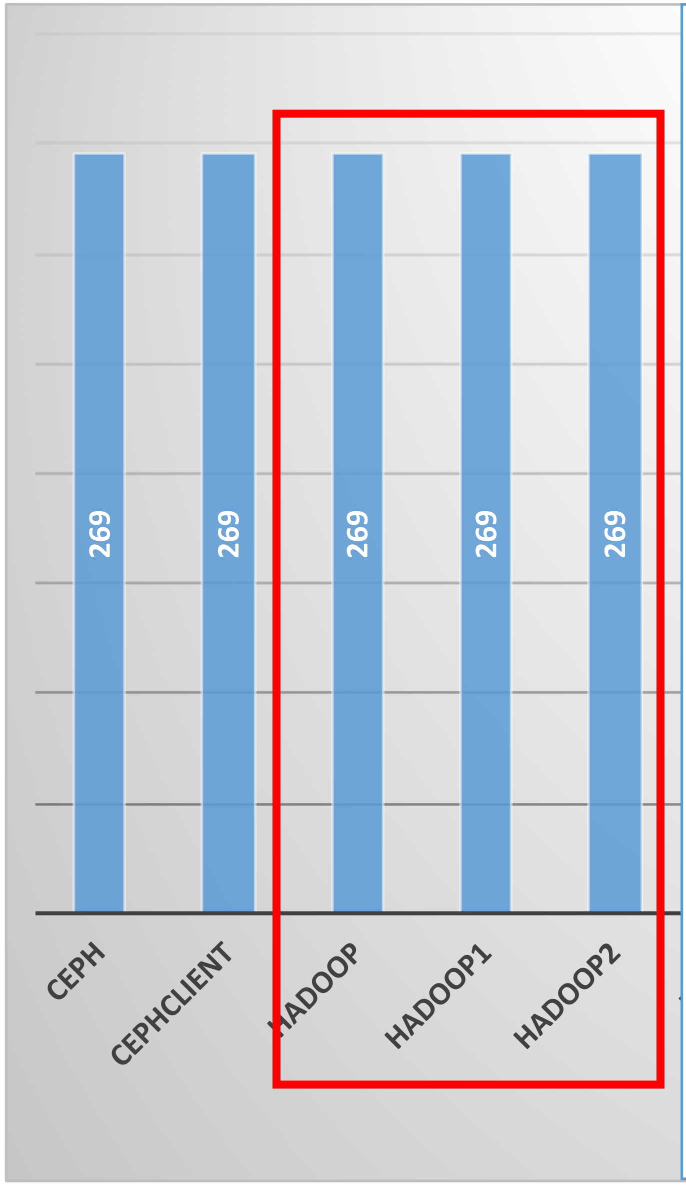
Reading data from CEPH Storage using **Kernel driver.**

CEPHCLIENT

Reading data from CEPH Storage using CEPH **Fuse client.**

For Ceph details, look https://indico.fnal.gov/event/22127/contributions/194938/

Job Success/Failures

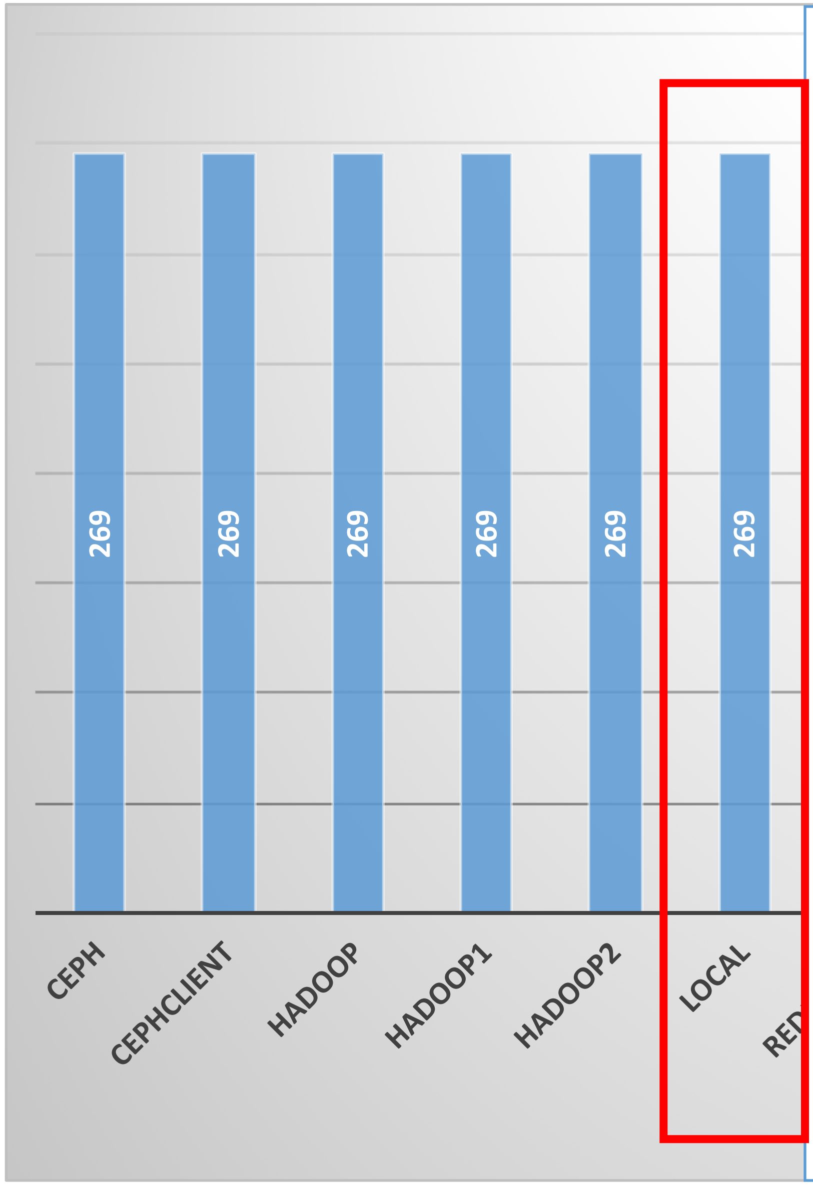


HADOOP (Production, 8PB RAW, Rep 2)

Production Hadoop Cluster. Load from Production, Analysis and Remote reads/writes. Tests were repeated several times and we will cover reasons why tests were repeated later.

Job Success/Failures

Caltech

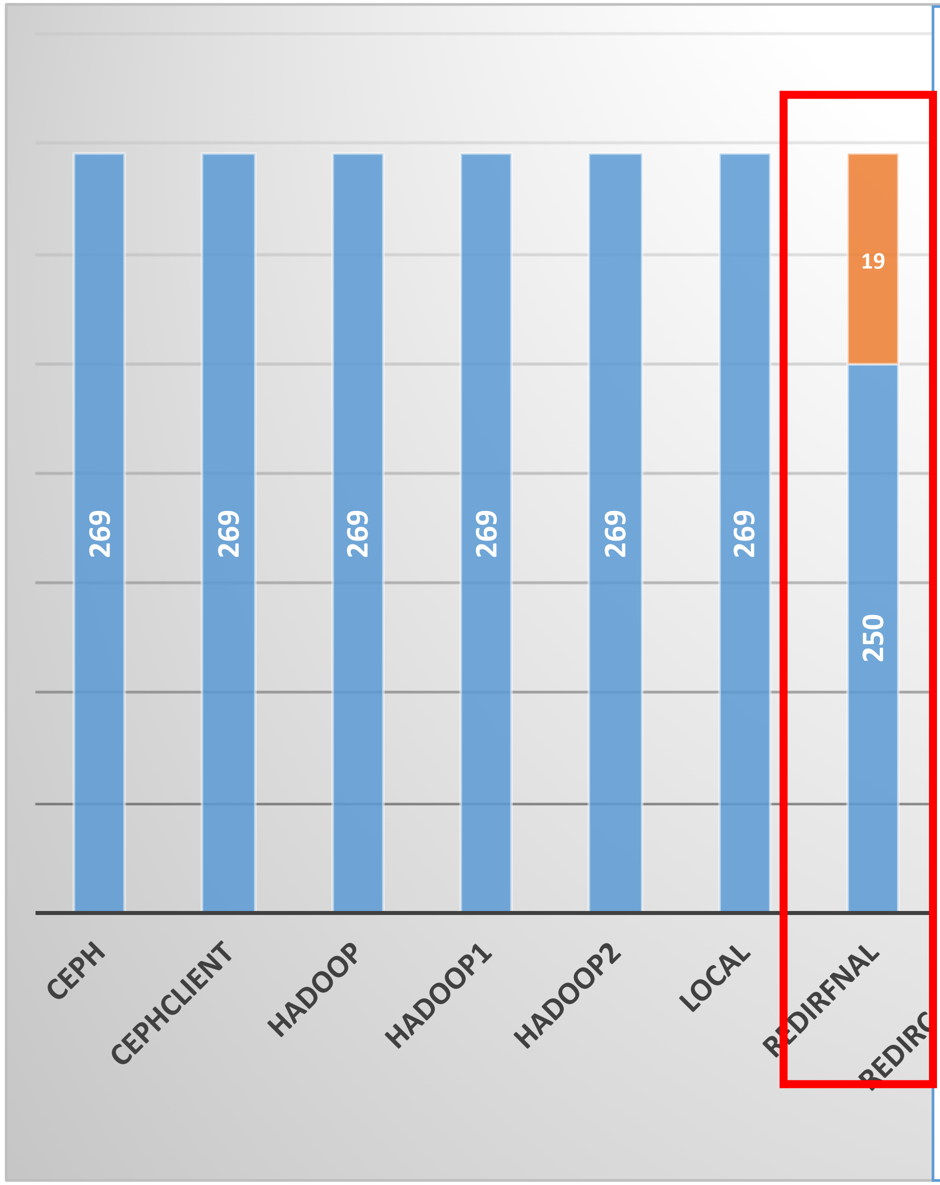


LOCAL
Download file to local scratch space, and run over the local file. (Means no remote open/read done, using same file from scratch directory)

For reference, calculations later will assume only the job runtime and not the time it takes to download file to local disk;

48 jobs simultaneously run on each node. Read/Write happens on same disk. (HDD)

Job Success/Failures

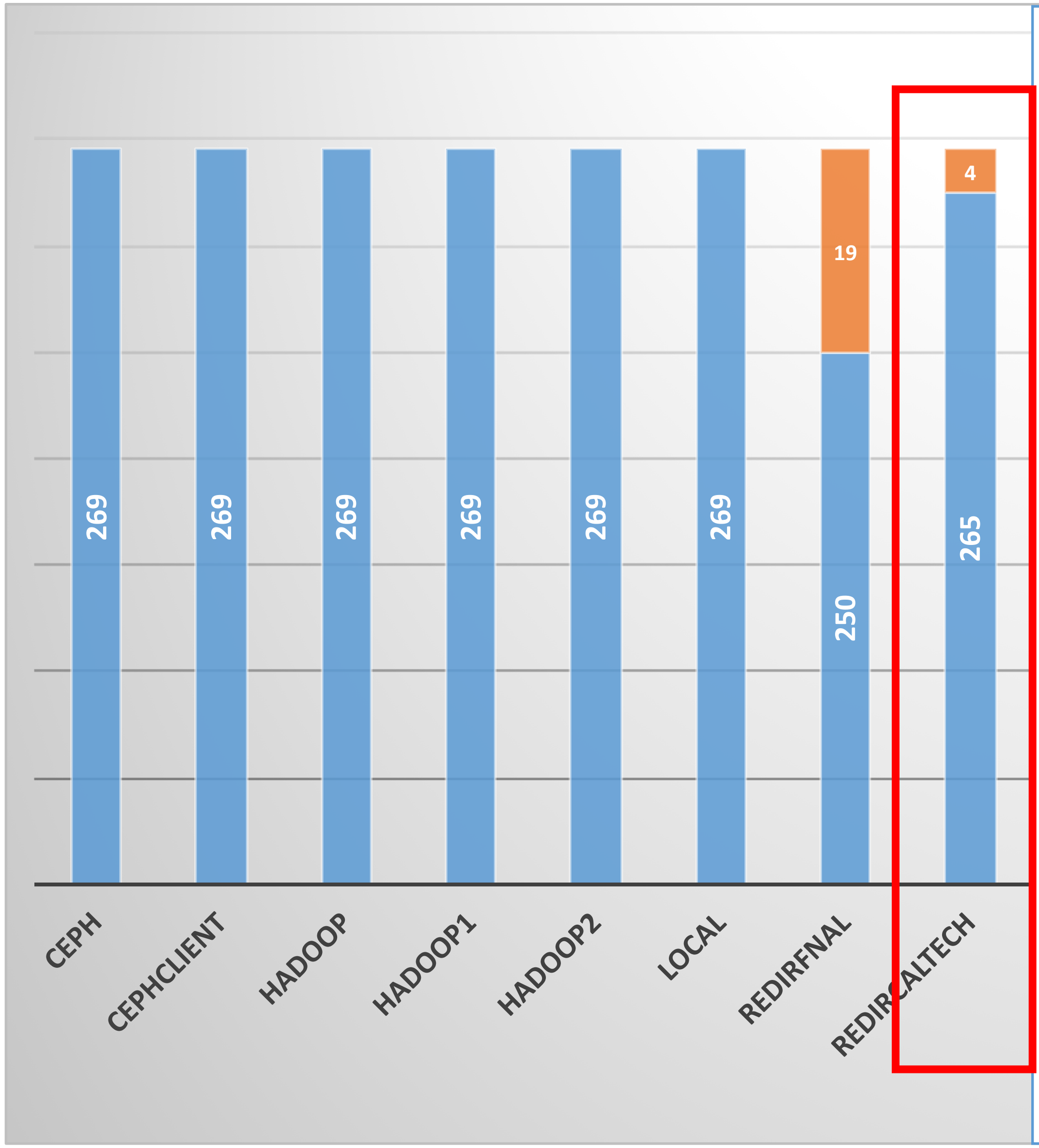


REDIRFNAL

Reading data via AAA Fermilab Redirector. At the time of run data is available at 6 Disk sites: T2_CH_CERN, T2_US_Caltech, T2_US_Nebraska, T1_US_FNAL_Buffer, T2_FR_GRIF_LLR, T1_US_FNAL_MSS, T2_ES_IFCA, T2_DE_DESY

Reading is based on Multisource algorithm. More details here: https://github.com/cms-sw/cmssw/blob/master/Utilities/XrdAdaptor/doc/multisource_algorithm_design.txt

Job Success/Failures

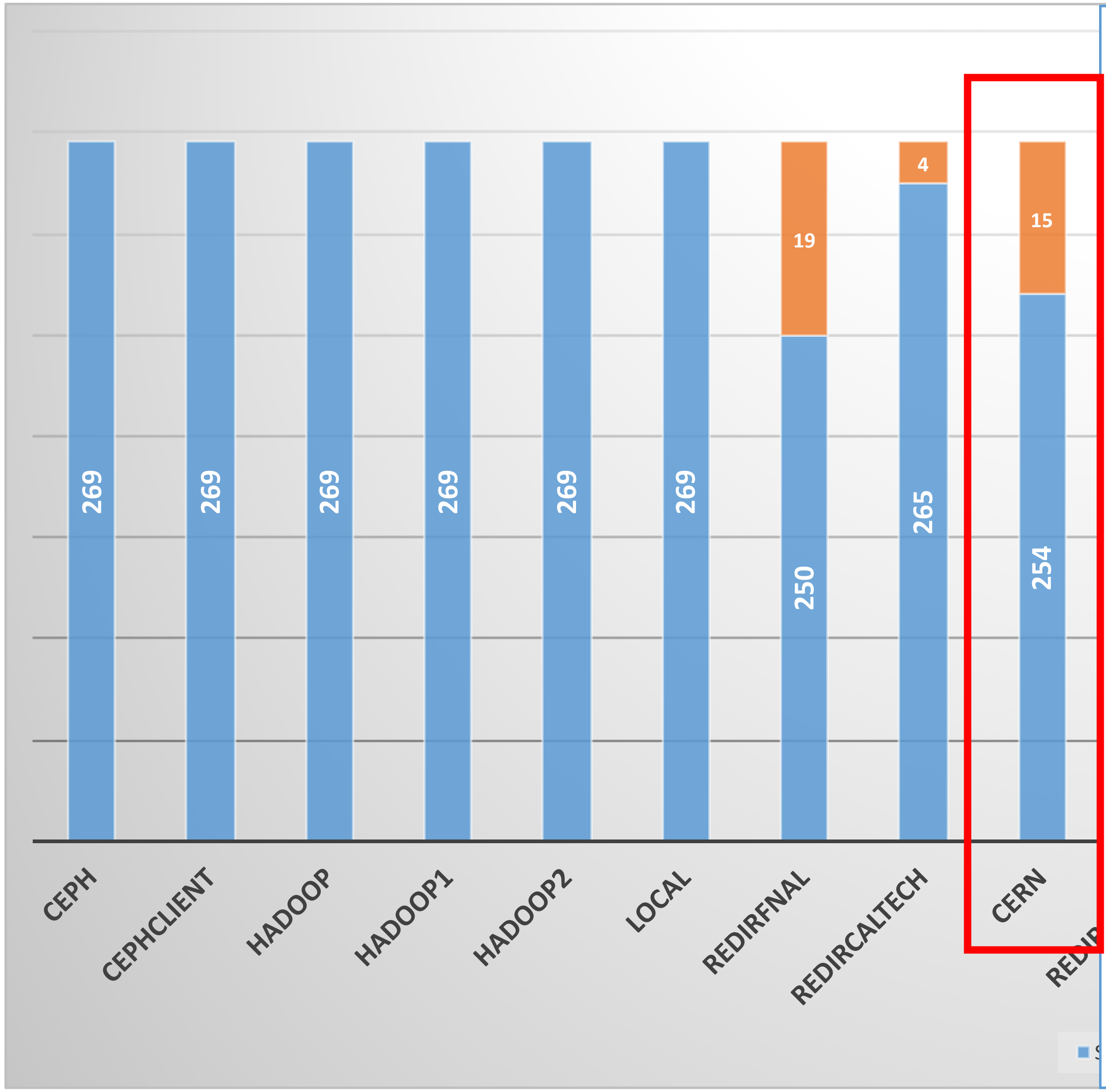


RedirCaltech

Reading data via Caltech Redirector (9 XRootD servers behind redirector). Data is on Hadoop storage;

Reading is based on Multisource algorithm. More details here: https://github.com/cms-sw/cmssw/blob/master/Utilities/XrdAdaptor/doc/multisource_algorithm_design.txt

Job Success/Failures

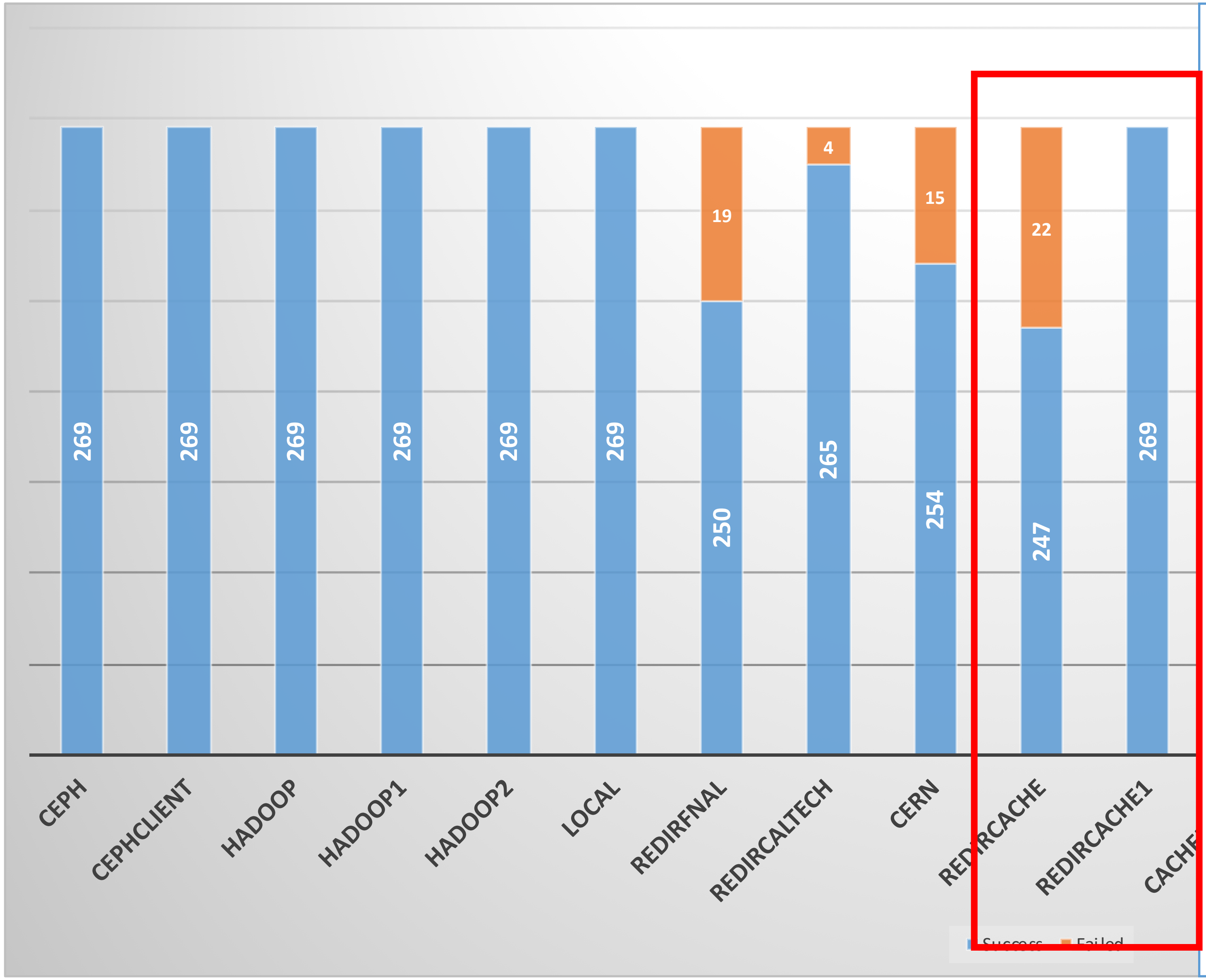


CERN

Reading data from CERNs XRootD Redirector. Data is under /store/user (means no other site has it.) RTT ~170ms

Reading is based on Multisource algorithm. More details here: https://github.com/cms-sw/cmssw/blob/master/Utilities/XrdAdaptor/doc/multisource_algorithm_design.txt

Job Success/Failures



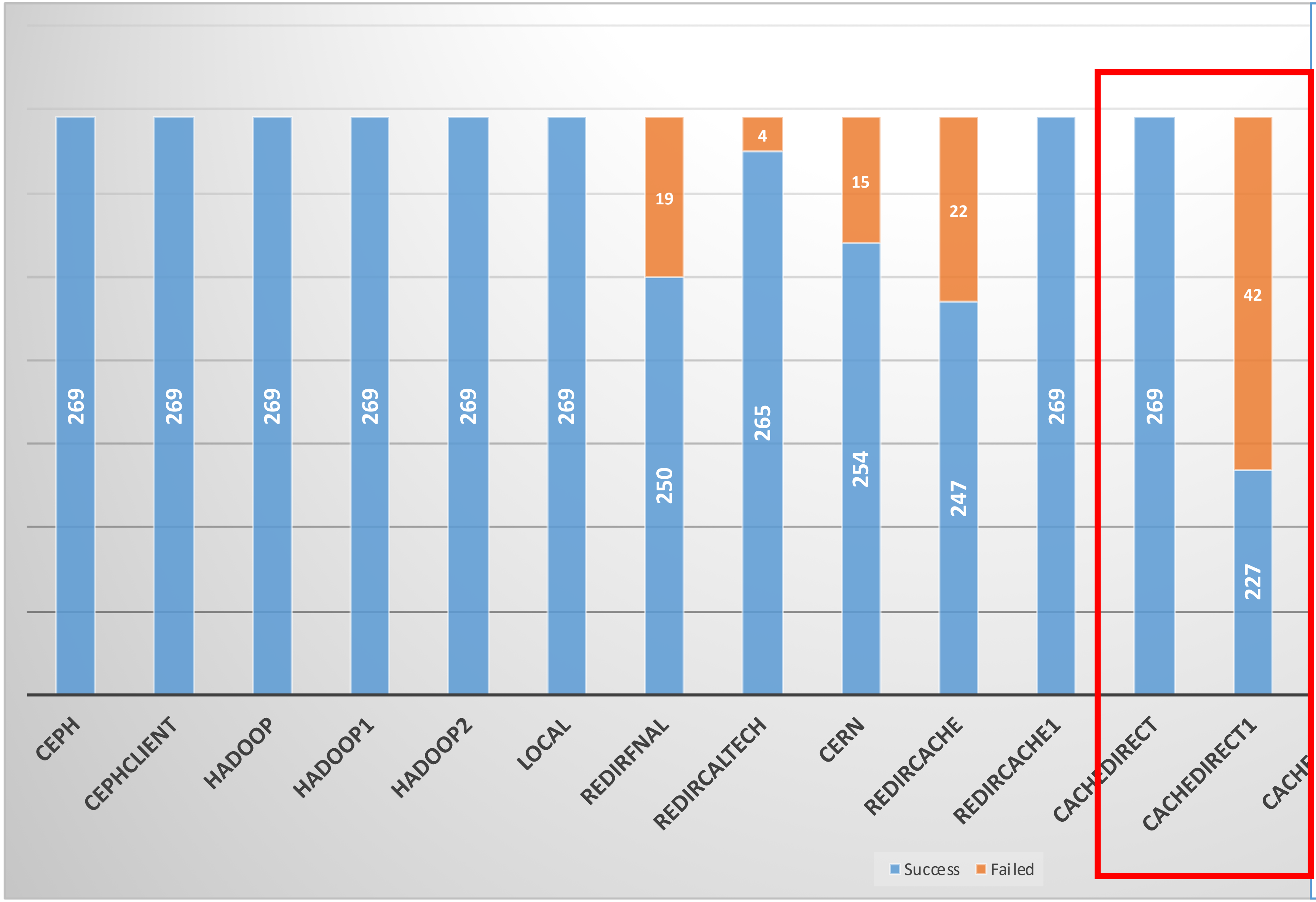
REDIRCACHE (cache empty)

Reading via SoCal cache redirector (14 nodes behind redirector)

Reading is based on Multisource algorithm. More details here: https://github.com/cms-sw/cmssw/blob/master/Utilities/XrdAdaptor/doc/multisource_algorithm_design.txt

What is SoCal cache? More details: https://docs.google.com/presentation/d/1vofUOf9dK7R1j75IF9EY1tcbCB9WTqTonhgnX4k8yg8/edit?usp=sharing

Job Success/Failures



CACHEDirect (cache empty)

Reading only via 1 cache server directly and data avail in 7 Sites.

Reading is based on Multisource algorithm. More details here: https://github.com/cms-sw/cmssw/blob/master/Utilities/XrdAdaptor/doc/multisource_algorithm_design.txt

What is SoCal cache? More details: https://docs.google.com/presentation/d/1vofUOf9dK7R1j75IF9EY1tcbCB9WTqTonhgnX4k8yg8/edit?usp=sharing

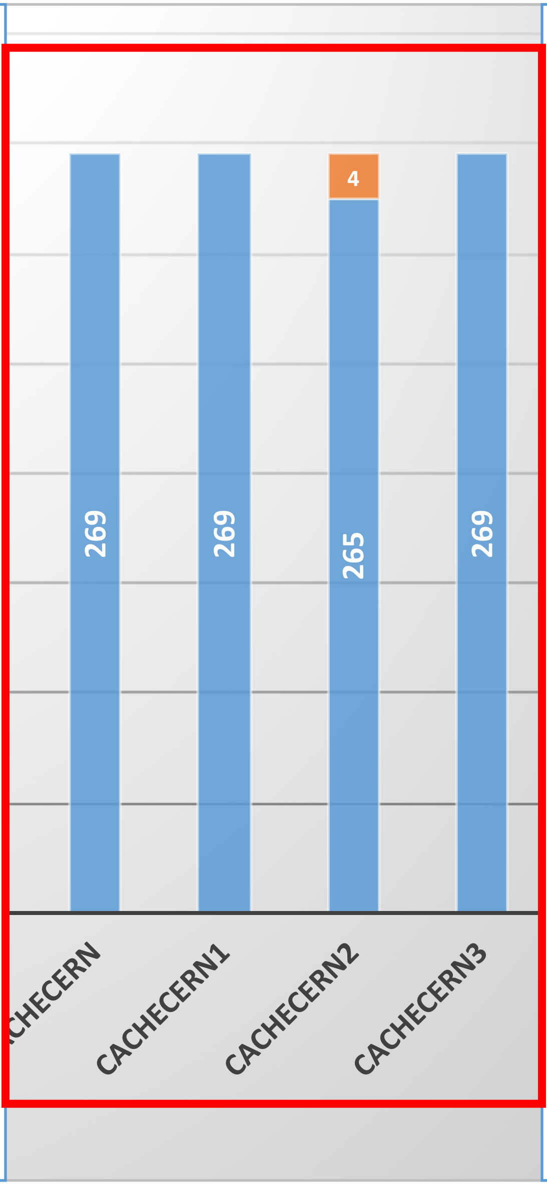
Job Success/Failures

CACHECERN (cache empty)

Reading only via 1 cache server directly and data is available only at CERN (/store/user/jbalcas…)

Reading is based on Multisource algorithm. More details here: https://github.com/cms-sw/cmssw/blob/master/Utilities/XrdAdaptor/doc/multisource_algorithm_design.txt

What is SoCal cache? More details: https://docs.google.com/presentation/d/1vofUOf9dK7R1j75IF9EY1tcbCB9WTqTonhgnX4k8yg8/edit?usp=sharing

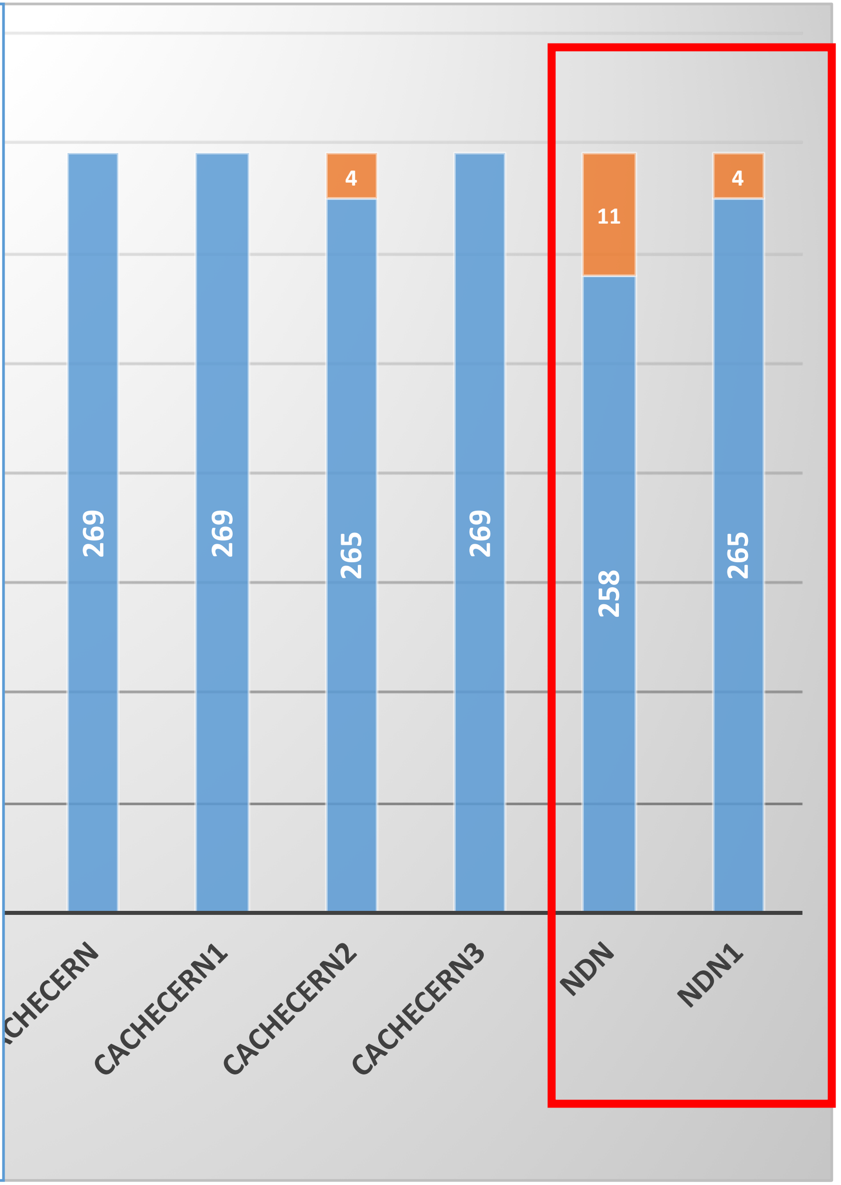


Job Success/Failures

NDN

Research Project building XRootD plugin to access data via NDN Network.
Data is on the same LAN Network (less than 1ms RTT, nodes connected at 100Gbps)

More description what it is:
https://sc19.supercomputing.org/app/uploads/2019/11/SC19-NRE-035.pdf


CHECERN
CACHECERN1
CACHECERN2
CACHECERN3
NDN
NDN1