XRootD Refresher
History, Concepts & Architecture

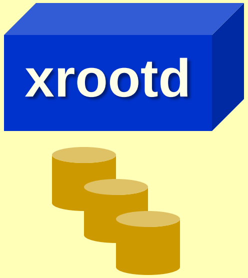
XRootD Workshop
Tokyo
November 9, 2016

Andrew Hanushevsky, SLAC

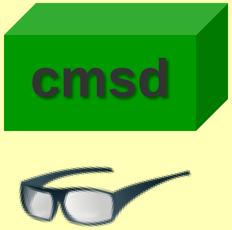
http://xrootd.org

Basic XRootD Concept

- A system for scalable cluster data access



xrootd



cmsd

- Not a file system
- Not *just* for file systems
- If you can write a plug-in you can cluster it

Brief History of XRootD

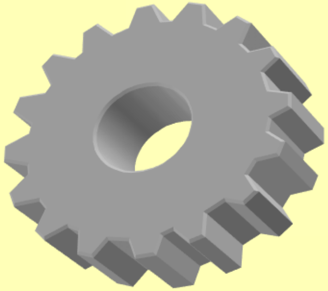
- 1997 – Objectivity, Inc. collaboration
  - Design & Development to scale Objectivity/DB
    - First attempt to use commercial DB for Physics data
    - Successful but problematical
- 2001 – BaBar decides to use root framework vs Objectivity
  - Collaboration with INFN, Padova & SLAC created
    - Design & develop high performance data access system
    - Work based on what we learned with Objectivity
- 2003 – First deployment of XRootD system at SLAC
- 2015 – Wide deployment with several implementations
  - ALICE, ATLAS, CMS, EXO, Fermi, LSST; among others
  - Protocol available in dCache, DPM, and EOS

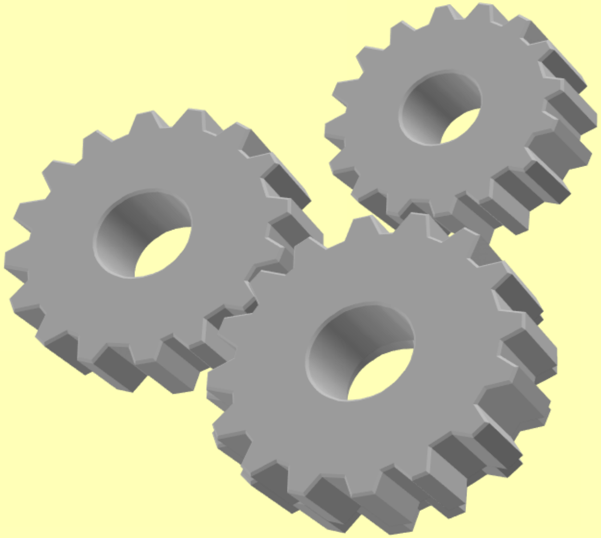
Only One Shot To Succeed!

- A lot of negativity floating around
  - Objectivity deemed a failure why not this?
- So, we focused on the specific problem
  - The High Energy Physics analysis regime
    - Write once read many times access mode
    - Thousands of parallel batch jobs
    - Small block sparse random I/O
    - Not enough money for hardware
  - This explains **XRootD** fundamental features
    - Why the name?

Solution: Over-Engineer It!

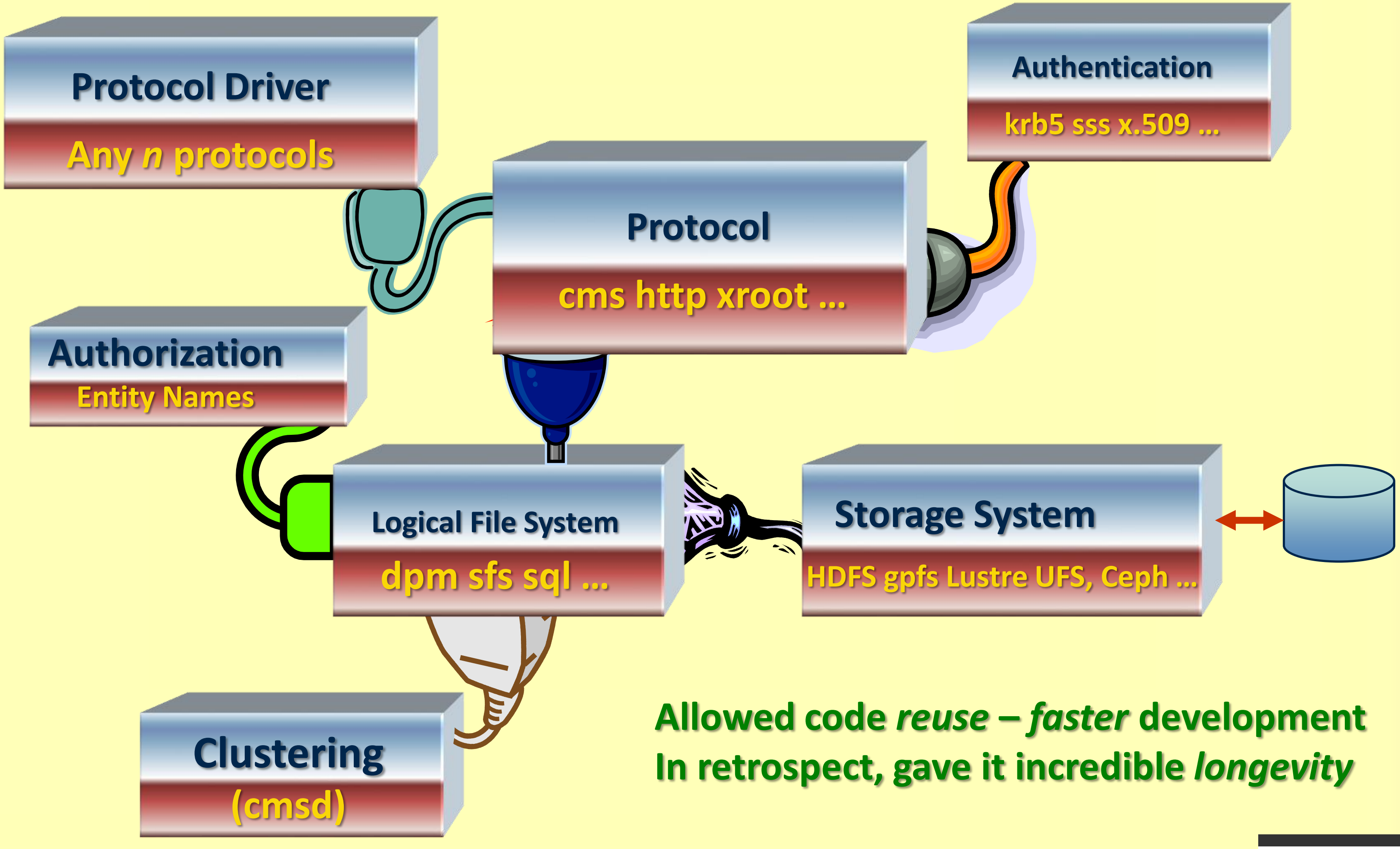
Minimize latency
Minimize hardware requirements
Minimize human cost
Maximize scaling





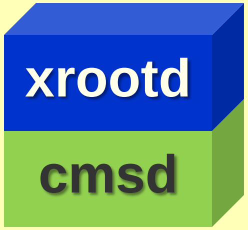
The synergistic approach made XRootD one of the *fastest* cost-effective systems even to this day

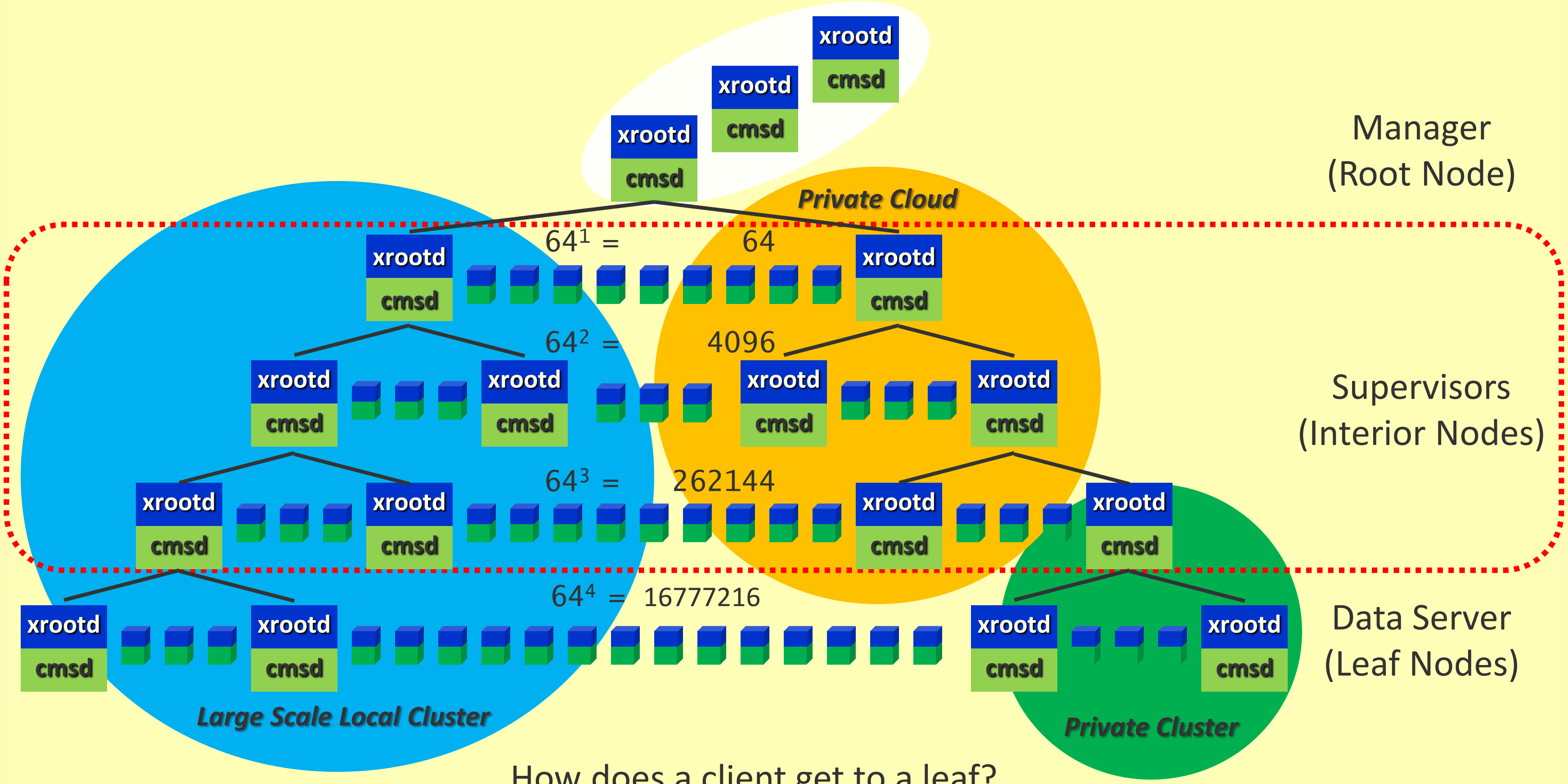
XRootD Plug-in Architecture



Allowed code *reuse* – *faster* development
In retrospect, gave it incredible *longevity*

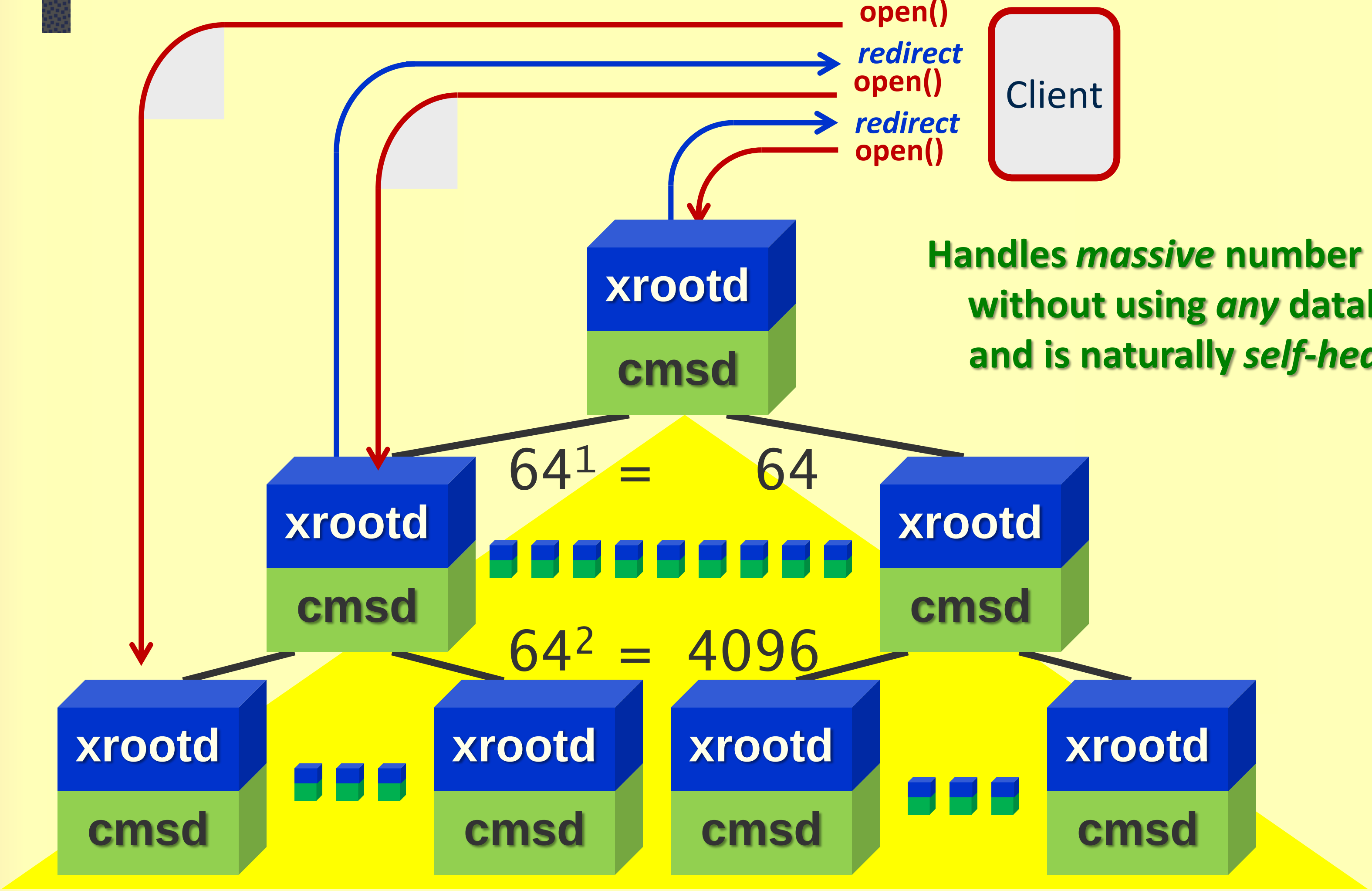
Super Scaling Using B⁶⁴ Trees





How does a client get to a leaf?

Routing Client To The Data



Handles *massive* number of files without using *any* database and is naturally *self-healing*

Conclusion

- XRootD is simple, flexible, and effective
  - Numerous plug-ins developed
    - Latest ones: Ceph, CASTOR, and mySQL
    - Others: File and block caching, Mass Storage, multiple file system (e.g. HDFS, DPM, etc)
- Scaling and latency are ideal for HPC’s
  - Though that path is not obvious
- We’ll explore this with BNL in 2017