Storage federations

Why & How : site perspective

Jeff Templon

Pre-GBD october 2012

1. CDN vision for wLCG
2. Relation to storage federation
3. “what would we need to do to make this ready for sites to deploy?”

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group

What problems need solving?

- "Menial Labor"
- Inefficient use of storage
- Jobs failing due to
  - Disk server overloaded (***NOT AN ERROR***)
  - Actual problem with disk server
  - Experiment catalog out of sync
- Protocol Zoo

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group

CDN division of labor

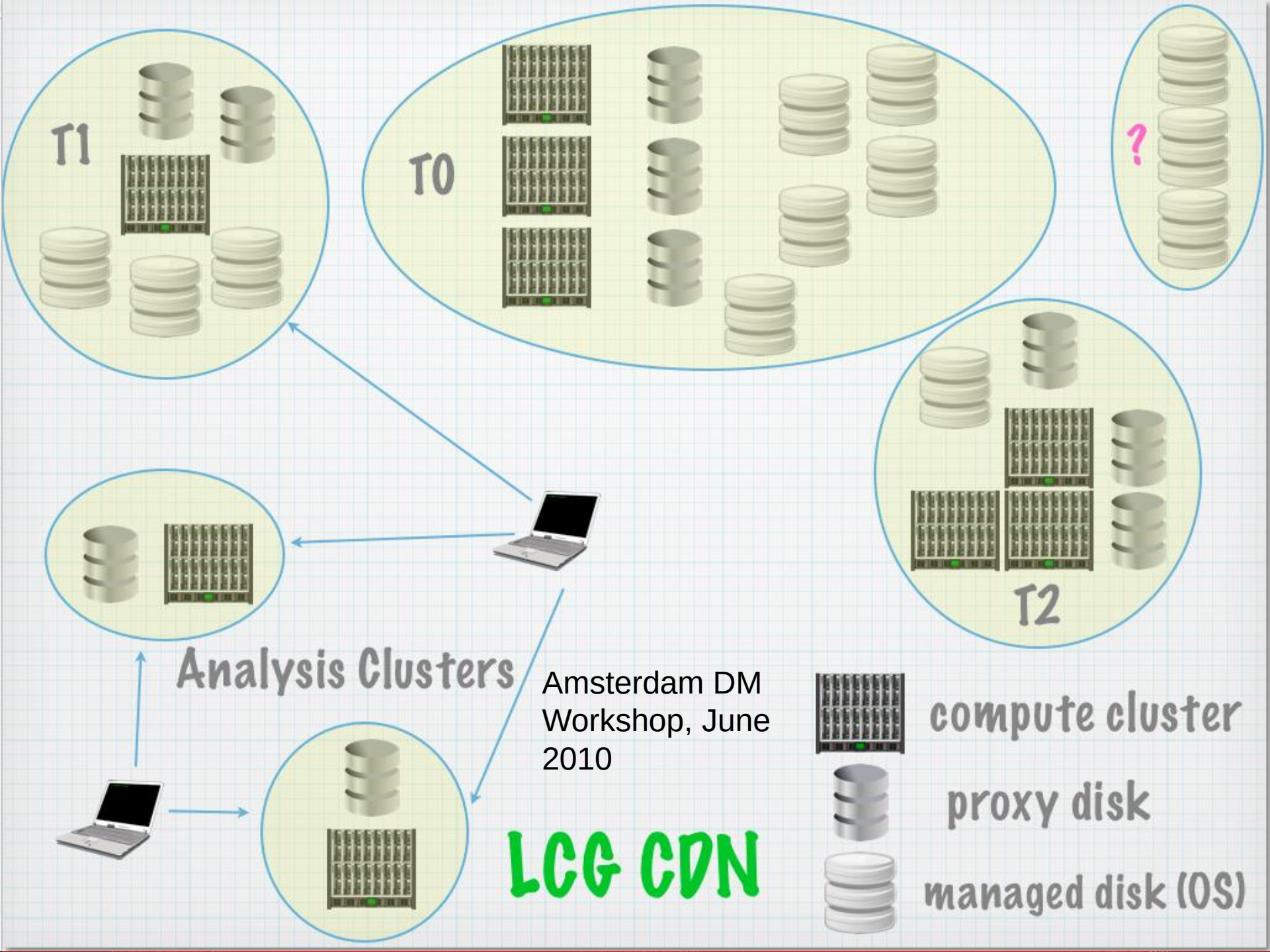
- *Archiving:* Storage of data. "Home" for a specific file. "Master" copy. "origin server"
  Like tape at T1, what else?
- *Data delivery:* focus on performance. Permanence not expected. "proxy/cache"
  Such servers "everywhere"? Esp. at T2s, T3s, local analysis clusters …

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group

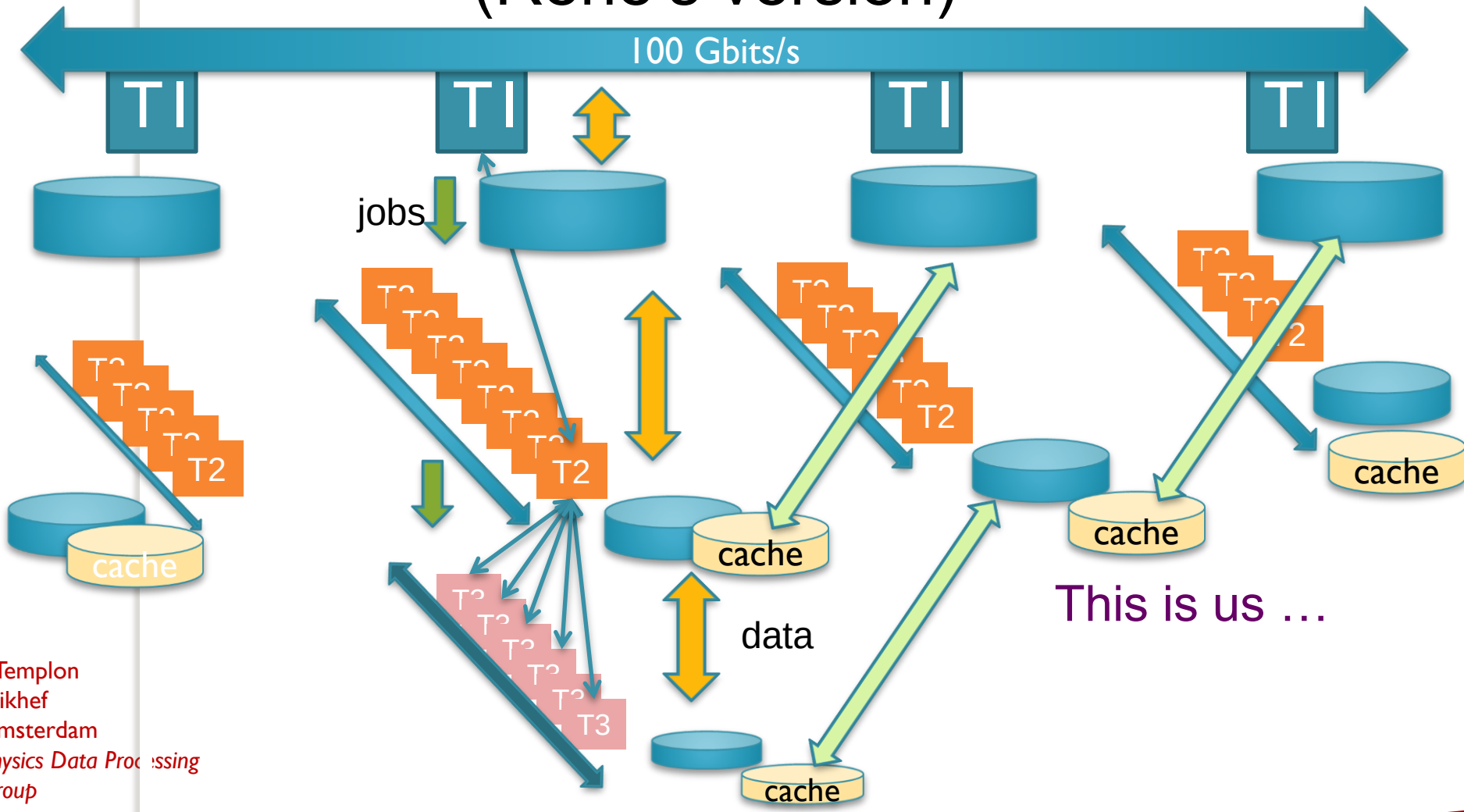
Consequences of model

- Clients never access origin servers; they become less complex (do one thing well)
- Clients may not even notice origin server in maintenance (served directly from delivery layer)
- Popular files rapidly everywhere, boring files only on origin server
- Caches are read-only from user perspective: simpler. Focus on delivery.

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group


T1
T0
?
T2
Analysis Clusters
Amsterdam DM Workshop, June 2010
LCG CDN
compute cluster
proxy disk
managed disk (OS)

The vision (René's version)



J. Templon
Nikhef
Amsterdam
Physics Data Proc ssing Group

CDN + P2P

- Some CDNs overlay on P2P network
- "Catalogue search" is a query operation on a distributed hash table
  - No centralized catalog to fail
  - DHT is resilient: not a "nothing works unless everything works" system.
- Delivery nodes can collaborate – client redirection or inter-node retrieval, avoids traffic to origin servers

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group

SF & CDN

- SF seems to be a milder version of CDN
- Origin servers are explicitly part of the federation.

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group

Time for questions before next part

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group

Site Perspective : how

- Focus on operational aspects: normally ignored by our middleware developers
- How much work is it to install? How much "crap" gets dragged in, how "specific" are packages?
- How much work is it to configure? How easy to automate? How much is redundant (pref. none). How stable is config?

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group

More site how

- How easy is it to tell that the service is correctly working?
- How easy is it to tell that the service is **not** correctly working?
- How easy is it to tell **why** the service is **not** correctly working?
- How easy is it to start from a user-reported error and trace back to a cause?

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group

More more site how

Observation : developers focus on a different goal for logging: that is, to verify that the program follows the expected sequence of events / steps. Unfortunately this is operationally …

the least interesting case.

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group

Example : logging
EMI RT #1202

- Clear time stamps in std format
- Make log file strings easy to parse
- For multithreaded or multi-machine actions, provide a uniqueID that will be present in all messages
- Follow conventions for log levels. Std level should for success produce the minimal output needed for auditability, but all errors and serious warnings should appear

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group

Are all sites configured in same way?

- NO!!!
- Sites *respond* in the same way to queries. How site implements the responding mechanism is unimportant as long as it works and performs well enough.

J. Templon
Nikhef
Amsterdam
Physics Data Processing Group