# Computing model for 2011/12

I.Ueda

ATLAS S&C Workshop 2010.12

# We must also discuss: Changes to the Computing Model

- Our current resources are nearly full with 2010 data
- We don't know how much increase we get for 2011
- More beam and luminosity in 2011
  - 200 days of uninterrupted running
- Likely even more so in 2012
  - To be decided before the April RRB
- New energy (8 (?) TeV) in 2011
  - To be decided in Chamonix in January
- Request for higher trigger rates
  - 400 Hz and 600 Hz need to be considered
- We need revolution, evolution is not enough

*Needs to be discussed at many sessions and at coffee*
*Can we decide which options to pursue on Friday ? Ikuo?*

12

BOS, Kors ; 29-Nov-2010

# Data Volume

data10_7TeV total

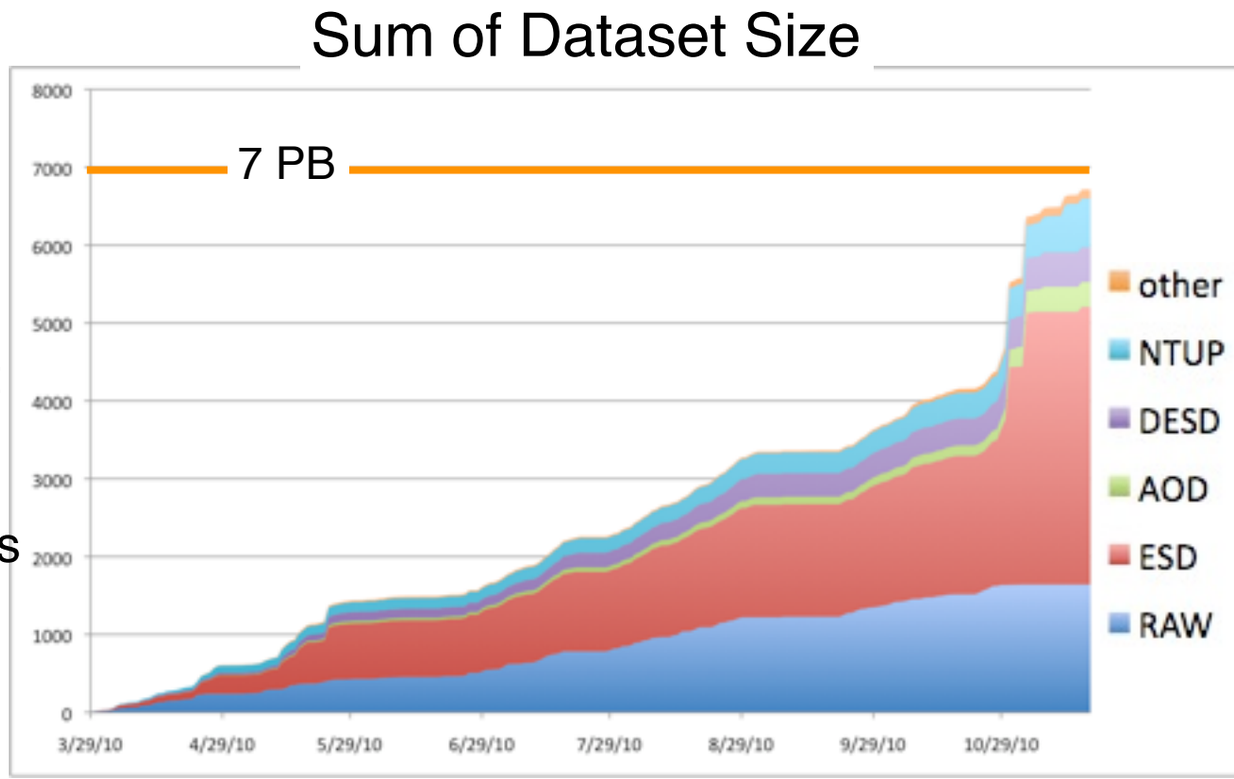
- RAW=1.6 PB, ESD=3.5 PB

data10 Oct average event size

- RAW: 1.40 MB/event, ESD: 1.48 MB/event

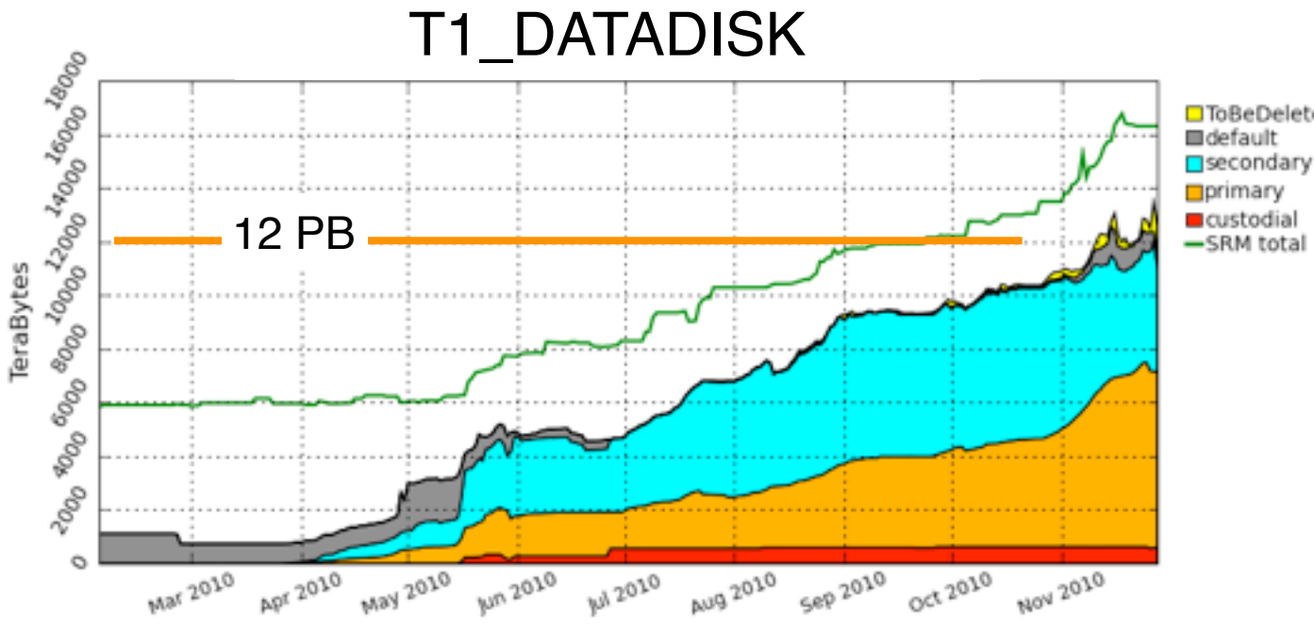
data10_hi (as of Dec 2.)

- RAW: 1.48 MB/event, 300 TB, 202M events
- ESD: 2.01 MB/event, (400 TB)




2011 prospects (a naive calculation)

- 400 Hz x 200 days x 50% = 3.5 Gevts
- RAW (1.4 MB) : 4.8 PB
- ESD (1.48 MB) : 5.1 PB
- **2 repro (merged) = 10 PB ESD**
- **1 repro (recon+merge) = 10 PB ESD**




Note:

- data size depends on the stream
- change of trigger rate could affect average event sizes

# Data Volume

Pledges 2010

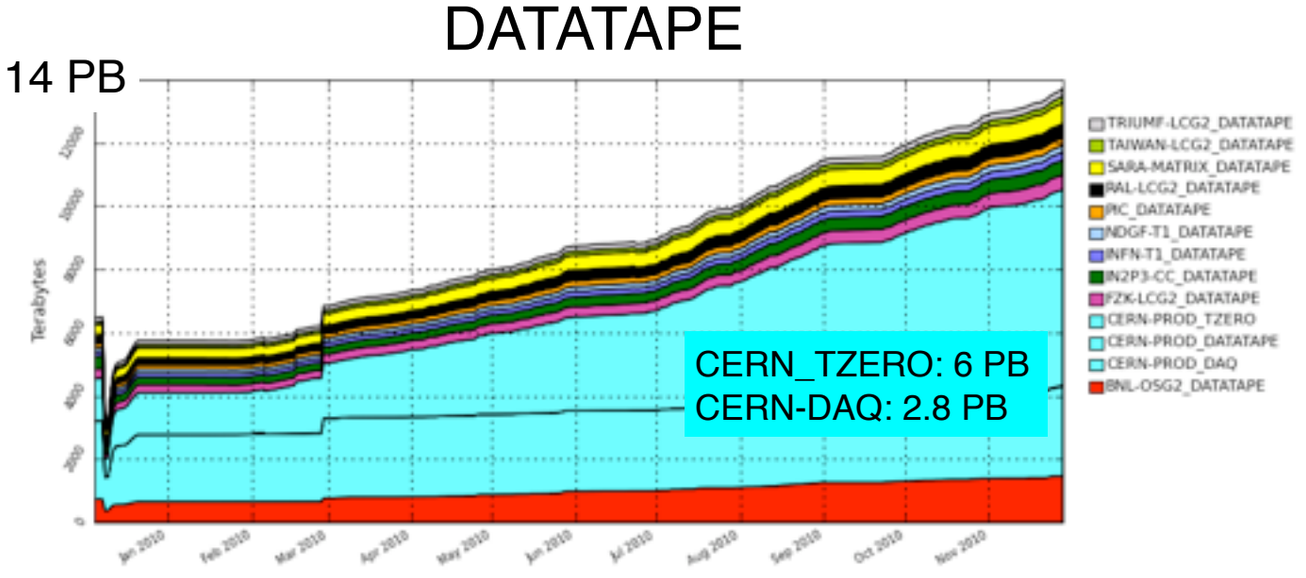
- T1_Disk=22 PB, T1_Tape=15 PB, T2_Disk=21 PB

Pledges 2011 (tentative)

- T1_Disk=26 PB, T1_Tape=32 PB, T2_Disk=34 PB
- eg. datadisk=15 PB, mcdisk=5 PB

**Do not keep ESD on disk (put them all onto tape)**

- we have not used T1_TAPE much




Concequences

- Analysis jobs would not be able to run on ESD because they are on tape.
  - No PD2P of ESD to T2s.
- Group productions run on ESD and provide dESD/D3PD to user analysis

# Tier2 Disks

Having both pre-defined distribution + PD2P is problematic for many Tier-2s.

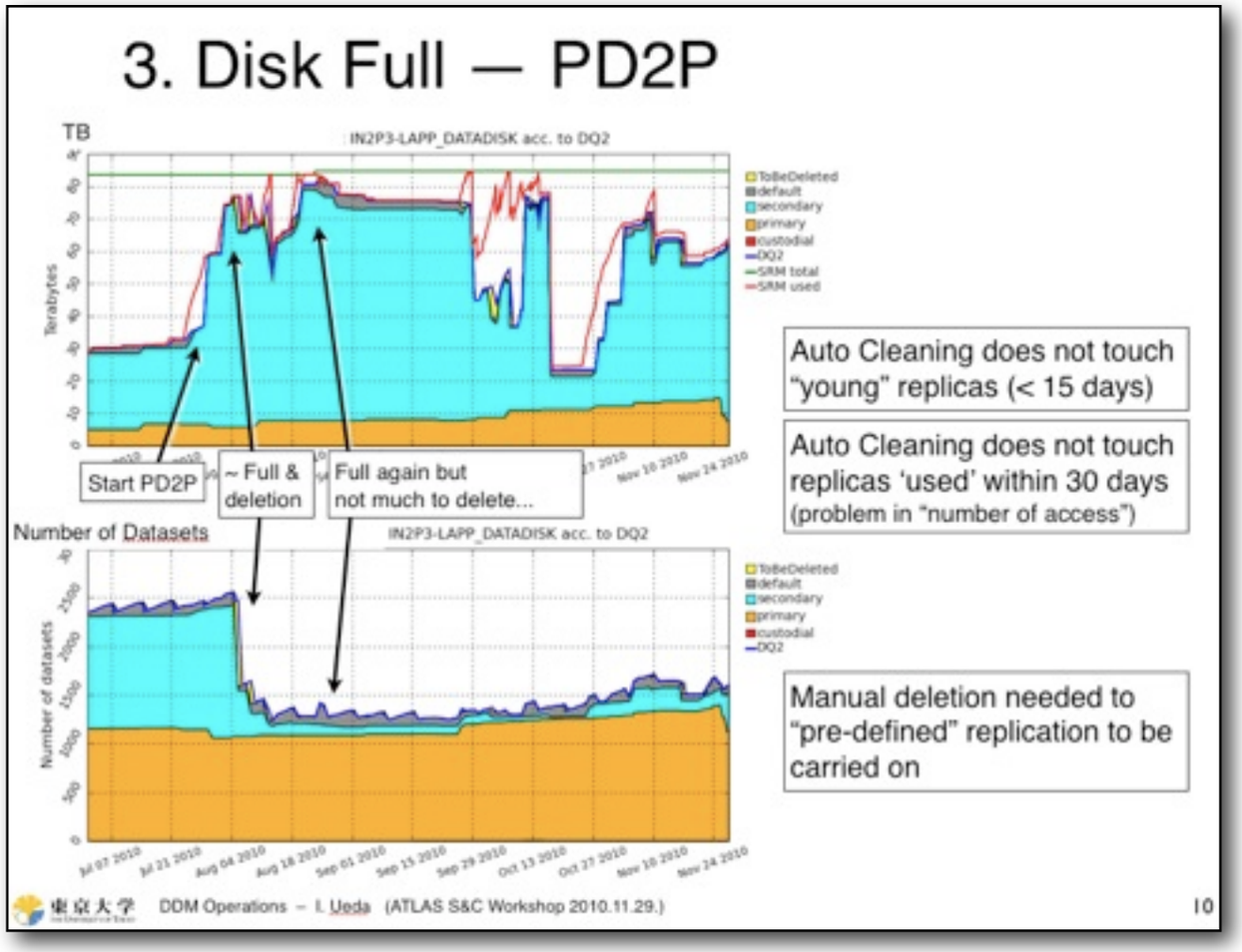
questions posed.

- “Why do we need ‘primary’ at Tier-2s?”
- “What are ‘primaries’ at Tier-2s?”

It was to ensure for the users to have access to the data

- but now we have PD2P and its extension soon

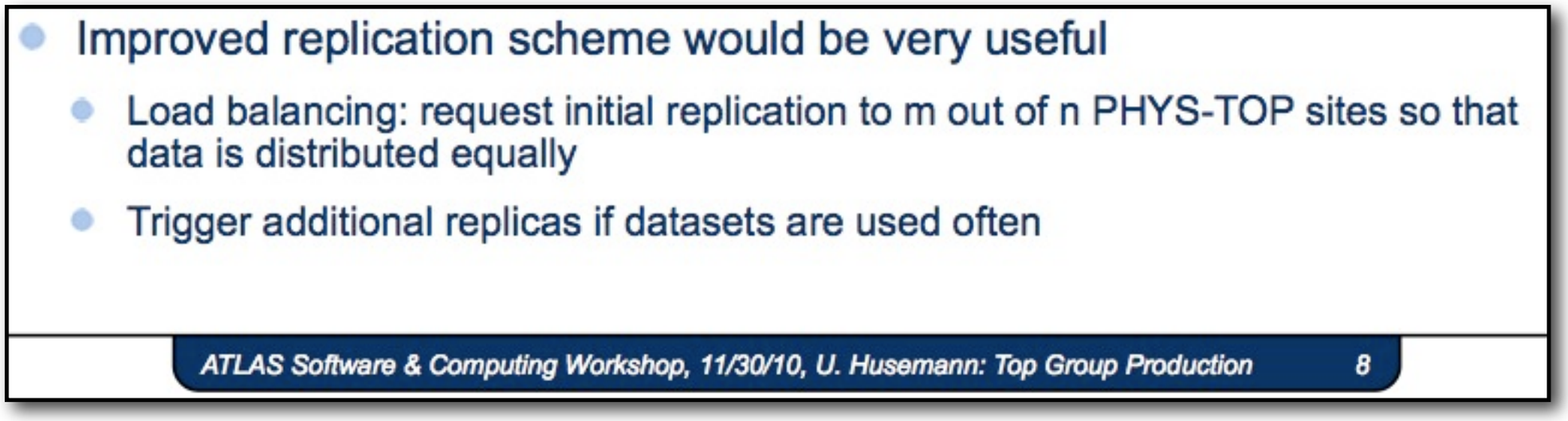
No more pre-defined distribution to Tier-2s (and Tier-3s)




- All the T2_DATADISK and T2_MCDISK spaces can be used for PD2P and on-demand replication (DaTRI)

# Group Disks

Group space/production managers are tired of having small pieces of group spaces at several sites




Global ‘quota’ rather than a collection of small spaces

PD2P for group data

# Group Disks

Why we put groupdisks on Tier-2s?

- data on groupdisks are the main input for user analysis
- sub-groups localized at some sites

But...

- some group people have realized they do analysis on the grid anyway = not much necessity to have data “locally”
- some issues in availability

and now,

- ‘global quota’ + PD2P == choice of sites for the data placement is not a group decision

Tier-1s host group data (persistent store)

- Tier-2s host popular group data (PD2P + DaTRI)

# Space Tokens

Space token (reserved space) has been used as a substitute for “quota”

- now we have some accounting tools, and better accounting system is coming soon (see DDM session)
- No more T2_datadisk,T2_mcdisk and T2_groupdisk but a single PD2P cache
- “Global quota” on group data on T1s would require flexibility in per-site spaces
  - Changing the quota centrally is better than asking sites to adjust the shares among the spaces

Merge datadisk, mcdisk and groupdisk into one

- probably would be called as ‘datadisk’
- even proddisk?

# Summary

## Do not keep ESD on disk (put them all onto tape)

- start immediately once decided, or before start of run next year

## Merge datadisk, mcdisk and groupdisk into one

- starting with the new pledges (2011)

## T1_DISK

- data/mc/group shares controlled centrally
  - with the new accounting system (Jan-Feb 2011)
- Tier-1s host group data (persistent store) with ‘global quota’
  - can start with quota per T1 with auto-distribution. need some development. need the new merged space.

## T2_DISK

- No more pre-defined distribution to Tier-2s (and Tier-3s)
  - can stop even now. wait for extended PD2P (and renamed)
- PD2P and on-demand requests to fill the space (incl. group data)
  - PD2P extension on-going. (Jan-Feb 2011)

# backup

# 5. Transfers Stuck — slow links

## Pre-defined replication

- Some sites request for shares taking only their capacity into account
- not enough network bandwidth

## PD2P

- queueing too much data to 'slow' channels

## Group Production

- group responsibles cannot know the site connectivity...

All the above can interfere 'essential' transfers (eg. DBReleases)

# Tier-2s as ‘repository’

Could help in the sense of disk space

Worries

- Connectivity — not only “to site” but “from site”
  - Replication easier for T1-T1’-T2’ than T2-T1’-T2’
- Reliability — lower service level than Tier-1s
  - Often we did migration/recovery as “start from scratch”

Repository Tier-2s (if we need any) should be large and well connected sites

and tape?

- if we don’t have enough disk, the “second primary” could be on T1_TAPE but we don’t have T2_TAPE

# Revolutionary ideas needed

- Reduce event size from Calo, Indet, Trigg, ..
  - Factor 3 (?) in RAW size, gives also smaller ESD
- Use PD2P everywhere (maybe slower, but ..)
  - And make it more intelligent
- Run bulk processing on the grid
- Get more T1s (stable site with tape) (CERN, ..)
- Custodial copies more distributed:
  - RAW on tape: 1 @CERN and 1 @T1s
  - HITS on tape: 1 @T1s , @T2s
  - ESD on tape: 1 @T1s
- Fewer primary copies on disk:
  - ESD 1 copy @T1s
  - AOD, DESD, etc. on disk: 2 (3?) copies @T1s
  - No a priori data @T2s (rely on PD2P)
- Do away with space tokens (at least in T2s)
  - Just one big cache
  - Will dramatically improve life for Groups

18

BOS, Kors ; 29-Nov-2010