Tier2s connectivity requirements

22 Novembre 2010

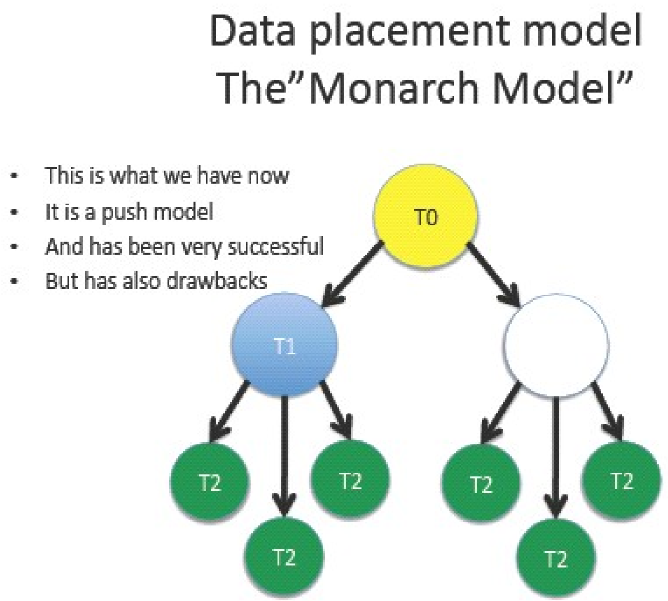
S. Jézéquel
(LAPP-ATLAS)

Outline

- **MONARC model**
- **Current usage of disk and network**
- **Foreseen changes in data organisation and network connection**

Year 2000s : Monarc model

- History
  - network expected to be limited : Monarc model
    - Manage replication centrally from pre-defined algorithm matching bandwidth availability
    - Avoid chaotic transfers as much as possible
    - Jobs go to data
    - Only ATLAS implemented it completly
    - LHCb : No T1 → T2 for analysis



Data placement model
The"Monarch Model"

- This is what we have now
- It is a push model
- And has been very successful
- But has also drawbacks

  - Other implementations
    - ALICE : Distant access by analysis jobs if needed
    - CMS : Data spread over T1s and T2 connects to all T1s

  - Number of replicas : Expected file popularity
  - Position of replicas : Close to users to get back output

Transfer path

- Chosen depending on the monitored transfer rate
  - Not based on some input from network tables
  - only sensitive to network saturation and single file transfer rate
  - No preference within Europe if effective bandwith is

- CMS:
  - T1-T1 + specific T1-T2
  - Channels are constantly commissioned

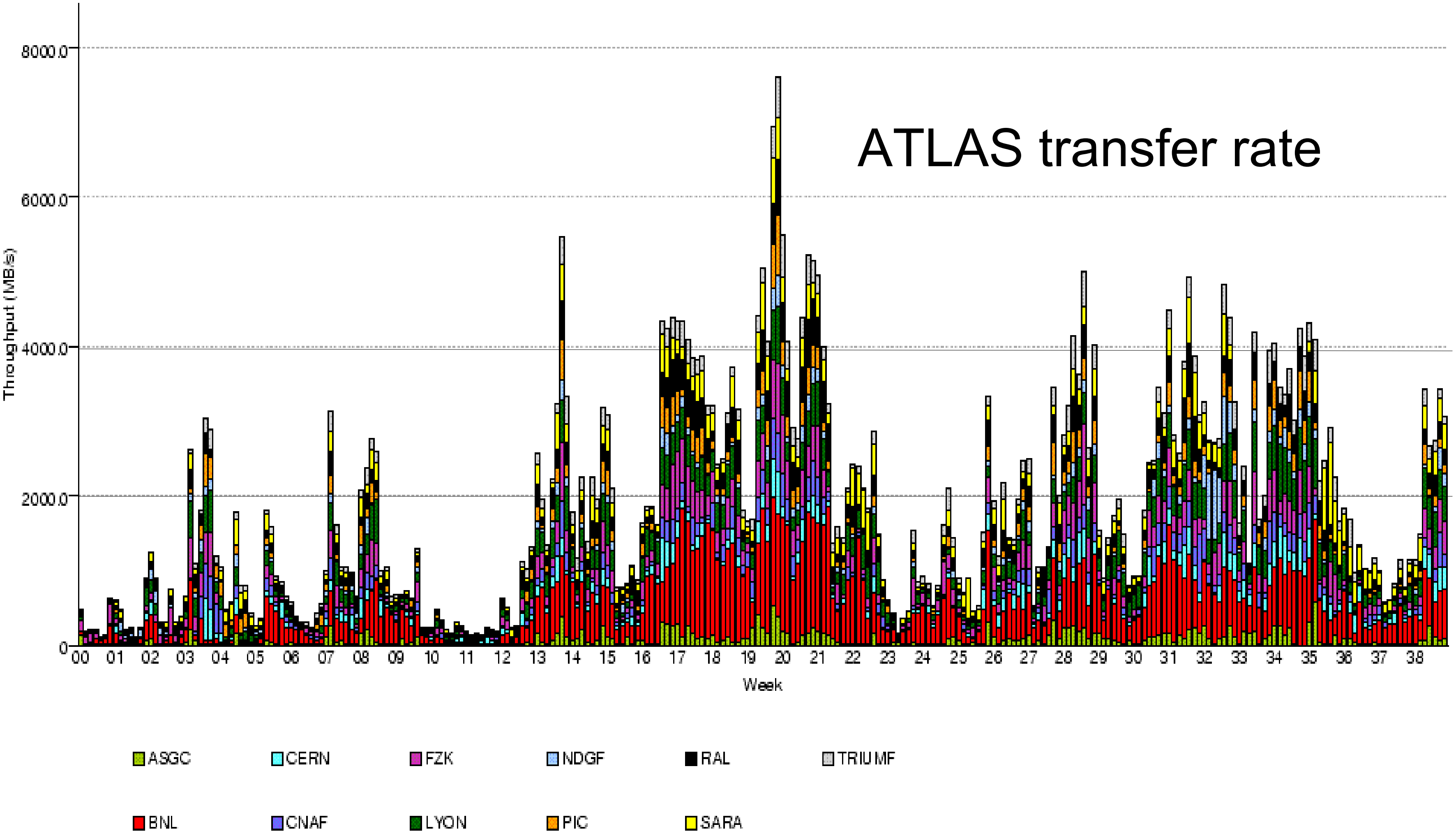
- ATLAS
  - T1-T1 + T1-T2 within same cloud and T2 → T1 outside cloud (user files)
  - T2-T2 path : T2(cloud a) → T1 (cloud b) → T2 (cloud b)
  - Path could be different based on file size
    - < 100 MB : Transfer limited by srm negociation → direct T2-T2
    - > 1 GB : Transfer is limited by transfer speed → Compute fastest path

Network usage

- Network is the most powerfull and reliable part of Grid

  → Increased the number of replicas for redundancy between sites

  → Transfer rate higher than expected



4 GB/s

2 GB/s
(Design)

Data effective usage

- Strong pressure from users to maximise the possibilities
  - Constraint : Jobs go to where data are prepositioned
    - Exception : ALICE with xrootd : Possibility to read distant data
  - Maximise replicas a priori → Maximise the possible CPUs
  - Was applied and successfull in 2009 and 2010
  - But :
    - Effective usage of replicated data is small
    - Lifetime of data is O(month) → Permanent replication activity

→ Go to a push/pull model

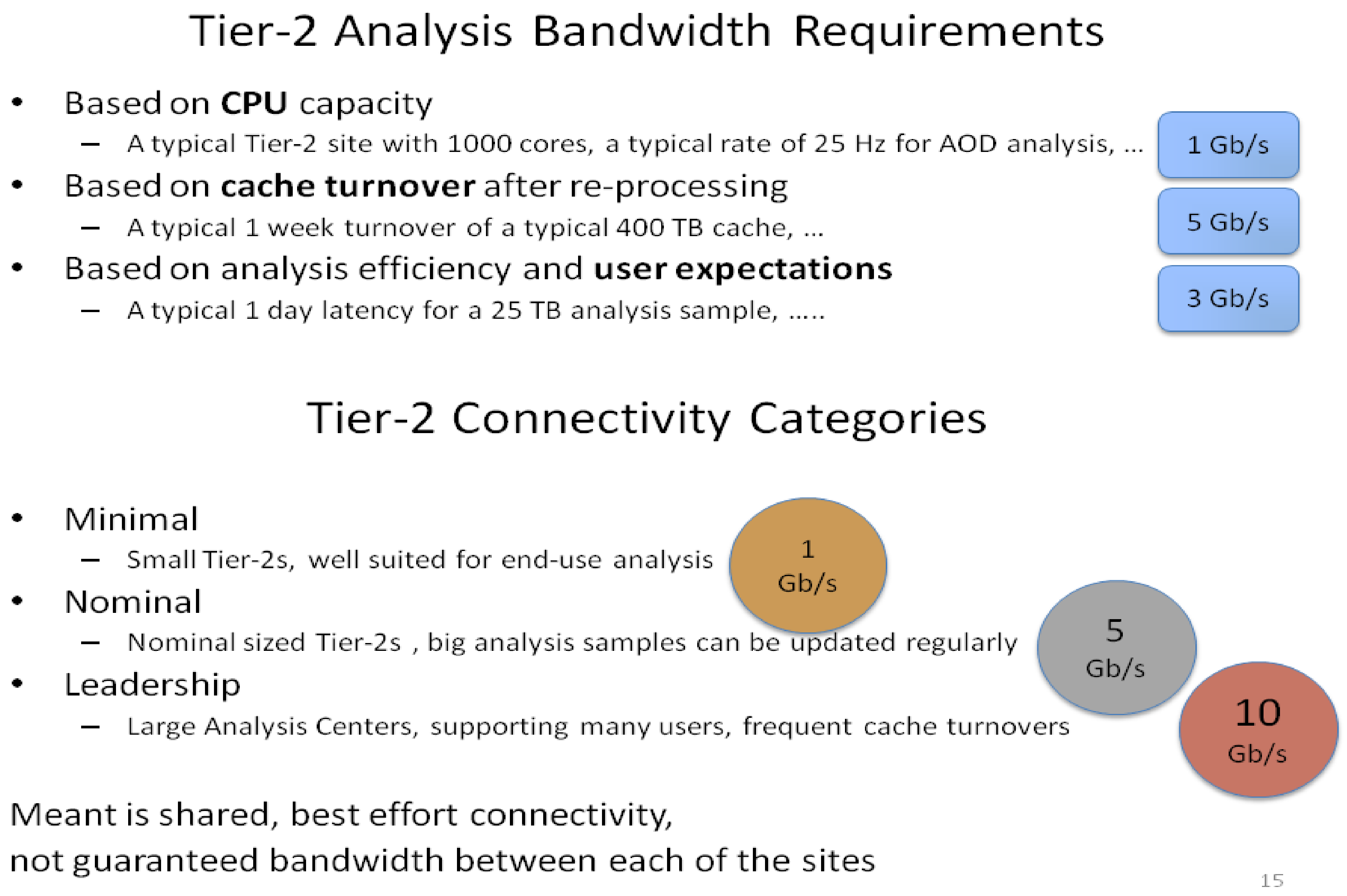
- Data are pushed to places when it is usefull
- Data are pulled for usage on demand

→ Request good reactivity for transfers

T2↔T1/T2 transfers : Current usage

- Distribute data between sites hosting same physics groups
  - Huge amount of data to replicate
- Collect data processed by physics groups → local processing
  - Can be huge depending on T2 size at destination
- Collect data produced by users → local processing
  - Can be huge depending on T2 size at destination

Request for T2 connection



Tier-2 Analysis Bandwidth Requirements

- Based on **CPU** capacity
  - A typical Tier-2 site with 1000 cores, a typical rate of 25 Hz for AOD analysis, … 1 Gb/s
- Based on **cache turnover** after re-processing
  - A typical 1 week turnover of a typical 400 TB cache, … 5 Gb/s
- Based on analysis efficiency and **user expectations**
  - A typical 1 day latency for a 25 TB analysis sample, ….. 3 Gb/s

Tier-2 Connectivity Categories

- Minimal
  - Small Tier-2s, well suited for end-use analysis 1 Gb/s
- Nominal
  - Nominal sized Tier-2s , big analysis samples can be updated regularly 5 Gb/s
- Leadership
  - Large Analysis Centers, supporting many users, frequent cache turnovers 10 Gb/s

Meant is shared, best effort connectivity,
not guaranteed bandwidth between each of the sites

15

Transfer trafic : Future

- T1/T0 ↔ T1 transfers
  - Trafic not expected to increase significantly
  - Pic trafic scales with data reprocessing speed (nb CPU at T1)
- T1/T2 ↔ T2 will increase

→ Generic network will not be sufficient

Conclusion

- Network is reliable → more load on this component
- T2 : Data preplacement → Popularity policy

  → More transfers of temporary files
- Transfer to T2s will increase with time
  - Depend on expected activity (scales with site size)

→ T2 classification for bandwidth connection