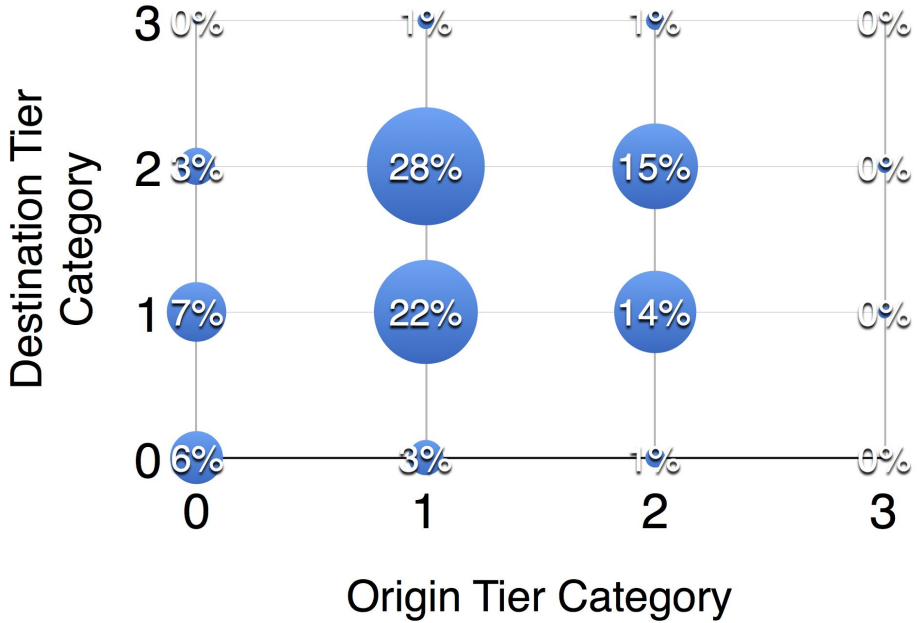
Storage topology for Run3+ in ATLAS

Data placement in ATLAS

- Run 1 start up :
  - ~40 copies of a given dataset (x per cloud)
  - Storage at every site - costly!
- Run 2 :
  - 2 copies
  - Dynamic (popularity) data placement - Network !
  - Remote access should increase
- Run 3 +
  - Caching & concentration of storage

Transfer matrix (data volume)

The hierarchical historical model is gone



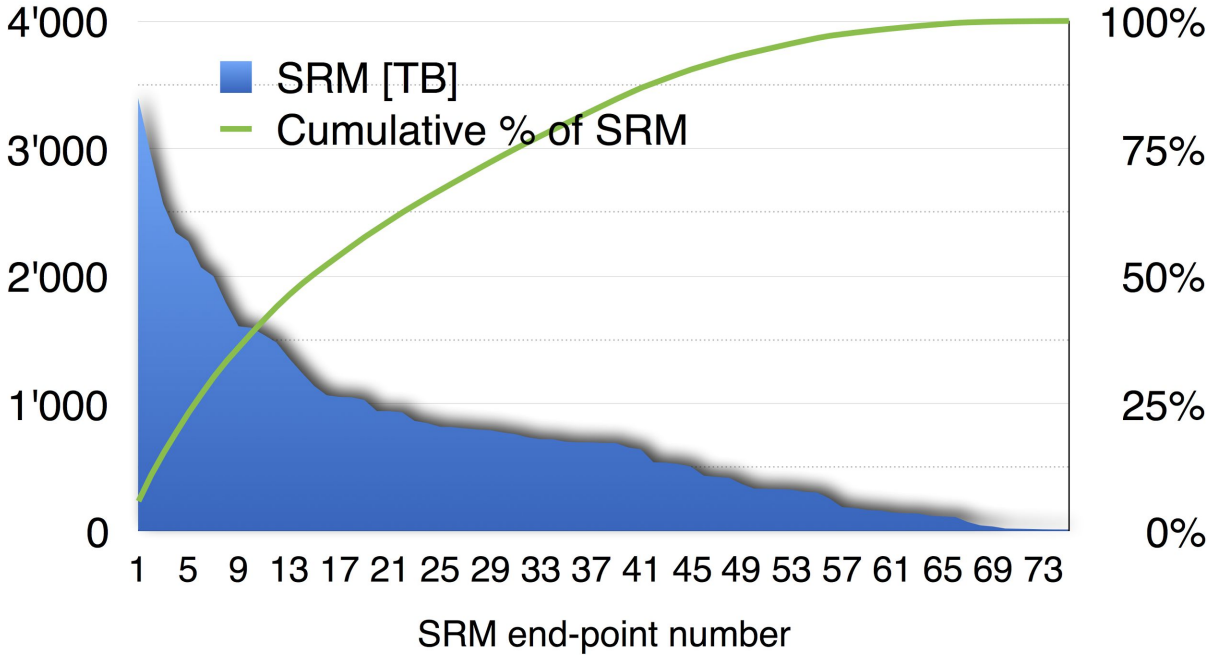
The network is our best friend

WAN access will increase!

Possible evolutions of computing model

(Too) many storage and computing end-points

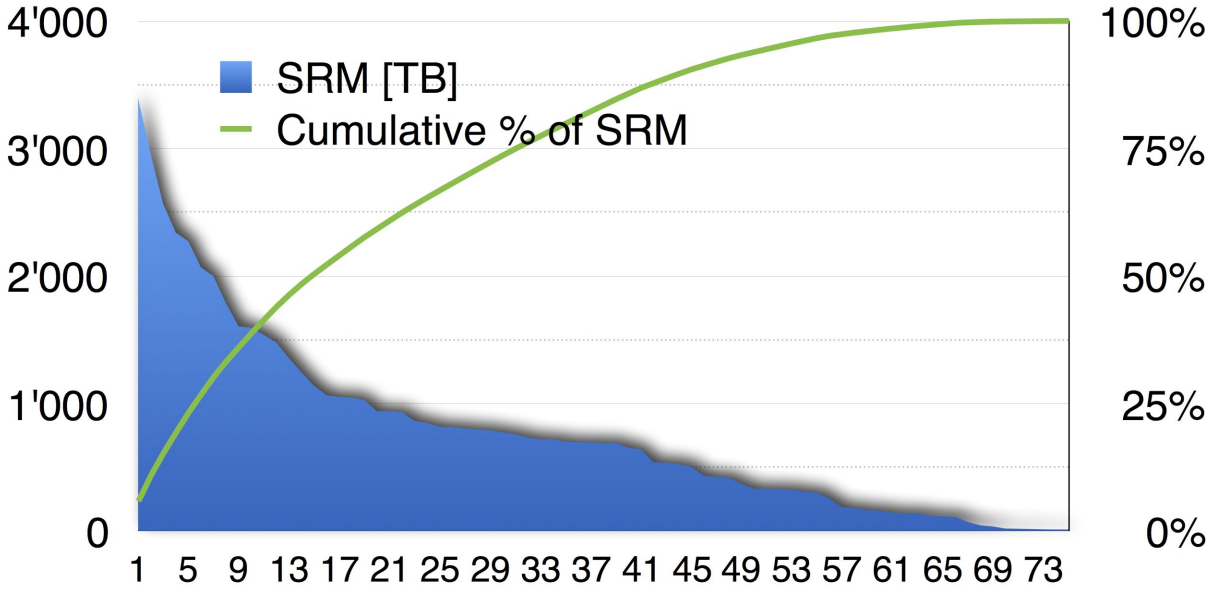
Available storage at Tier 2 sites



75% of Tier 2 available storage in ~30 end points

Large disparity in size of Tier 2s

Available storage at Tier 2 sites



More efficient to have larger and fewer storage end-points
2 possible categories : 'Cache based' & 'large' Tier 2s

Possible storage evolution

- Sizeable (TBD) regional centres
  - True federations
  - One storage entry point by centre
  - National & trans-national regional centres to match the scale

- **Technical solution to be worked out within WLCG**
  - (additionally) How to move PB of data to new system?

- Cache based (TBD) sites for those not part of a regional center

Possible Topologies

- Run3+ time scale
  - Today's investments are there for 5+ years

- **Tier 1 are the backbone**
  - Tapes (usage will grow)
  - High quality of services (storage, availability, support…)
  - Flexibility of workflows

- Data centres topology
  - Several (fewer?) T1s seen as one
  - T1 + nearby T2s
  - Aggregation of several T2s

Possible evolutions of computing model