Documenting BW requirements for use cases from the Computing Model at CMS-Tier1s

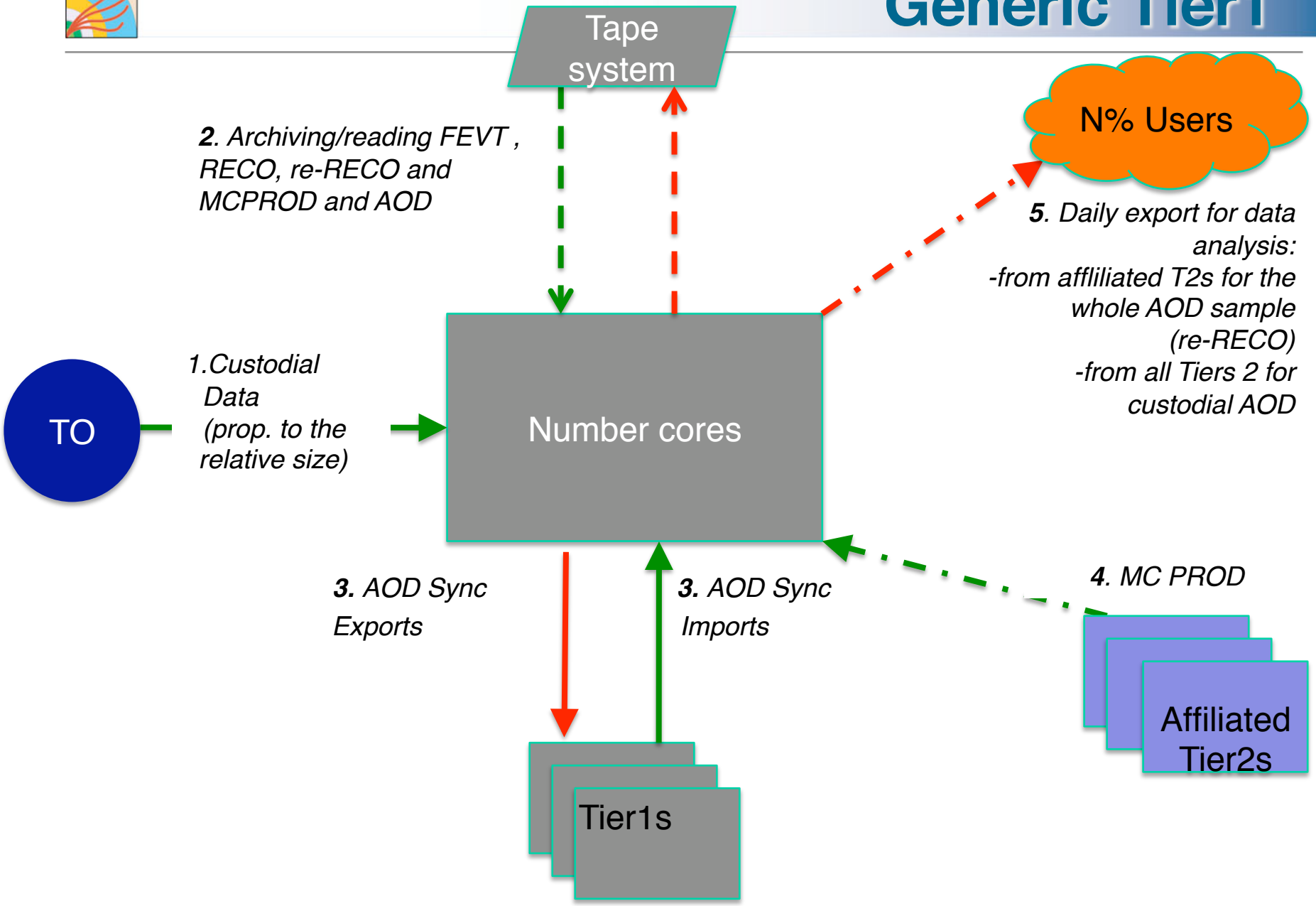
M. C. Sawley

ETH Zurich

12 May 09

Disclaimer: **the numbers hereby given are not to be considered final before approval by the CMS Computing Resource Board.**

Generic Tier1



Parameters 2009-2010

- The proposed model evaluates the requirements for data transfer during the 1st data taking period (2009-2010) of 200 days.
- 100 days of data taking at 20% duty cycle followed by 100 days at 50%
- Resources for each site are those pledged for 2009
- Raw numbers calculated with no security factor
  1. Bandwidth between T0 and T1s during data taking
  2. Tape I/O at T1s when writing FEVT, SimFEVT and RECO, and reading RECO, simRECO and re-RECO (in progress)
  3. Between T1s for AOD synchronization
  4. T2s uploading MCPROD at parent T1
  5. T1s exporting selected data for analysis to users

STARTING POINT

CONSTANTS (Identical to those taken the CMS resources request for 2009/2010, cf. M. Kasemann))	
Trigger rate	300 Hz
RAW Size	1.5MB
SimRAW	2MB
RECO size	0.5MB
AOD size	0.1MB
Total number of events	2590 MEvents
Total number of simulated events	2030MEvents
Overlap between PD	40%
Total size of RAW	3810 TB
Total size RECO	1270 TB
Total Primary AOD	254 TB

1. Exporting custodial data: methodology

- T0—> T1s : exporting FEVT
  - BW=(RAW+RECO) x Trigger frequency x (1+overlap factor). For the chosen parameters, this yields:

  BW= 2MB x 300Hz x 1.4 = 840 MB/sec, or 6.7 Gb/sec.

- Each T1 receives a share according to its relative size in CPUs

- Proportional to the trigger rate, event size and Tier-1 size

2. Tape I/O: methodology

- The maximum rate is computed for concurrent tasks:
  - Writing FEVT
  - Writing MCPROD
  - Writing re-RECO
- Period considered for re-reconstruction: 1 month (stringent!)
- Proportional to data taking period and inversely proportional to period imposed for re-reconstruction; correlated with MCPROD
- Data rate for writing:
  - (FEVT (1 month) + MCPROD (1 month) + re-Reco+AOD)/1 month
- Data rate for reading:
  - (RECO+ re-RECO+simRECO+AOD)/(1 month)

3. AOD synchronization: methodology

- When performing re-reconstruction, each Tier-1 produces the RECO and AOD format corresponding to the RAW data it has in its custody and exports it to all other Tier-1s.

- The full export to all sites should not take more than 2 weeks.

- For each exporting period of 2 weeks, the bandwidth from the exporting site is computed by:

  (6*fraction of AOD/(2 week-period))

- Proportional to full data sample size

4. T1s receiving MCPROD: methodology

- All year around T2s to produce a number of MC events of 2030 Mio Events
- Half of the CPU are devoted for MC production, time for producing simulated event
- Volume aggregated at T1s according to child-parent affiliation
- Proportional to full data sample size and to the cpus size

5. Simulating data analysis: methodology

- Each T1s has a share of RECO and AOD coming from 1st re-reconstruction
- Each T1 will receive requests for data skimming from a share of 1000 CMS Users according to its size
- The analysis pattern is assumed to be:
  - Half of the users will perform twice per month a “standard” analysis (i.e. importing 1% data)
  - 10% of the users will perform once per month a large dataset analysis (importing 10% data)

data set analysis being re-RECO/AOD (from any Tier 1) and full AOD (from parent site)

- Data rate evaluated as daily export set/1day
- Linear with the full event size and the number of active users (1000 in this calculation)

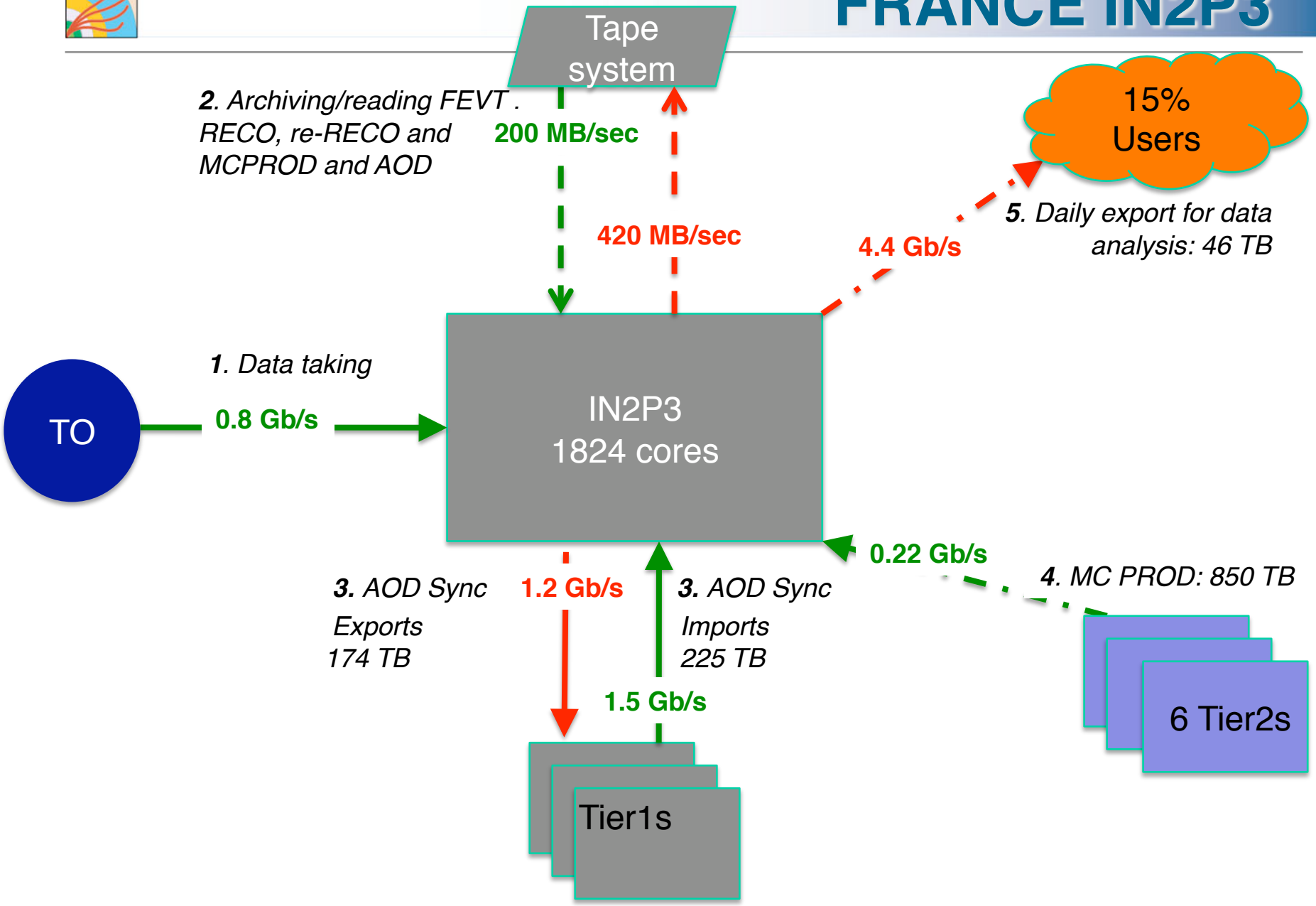
The results

- 1 slide per regional Tier1
- Pledged Cores are given in the old kSI2k units
- Remember: this are really raw values
- Links:
  - Solid line: sustained bandwidth (data taking ONLY)
  - Broken line: peak bandwidth (may happen at any time: numbers shown is the total if it all happens at the same time)
- For each Tier 1, the fraction of served users for analysis is a combination based on
  - Relative size child T2s for analyzing the share of 1srt AOD at considered Tier1
  - Relative size of T1 for analyzing the full AOD

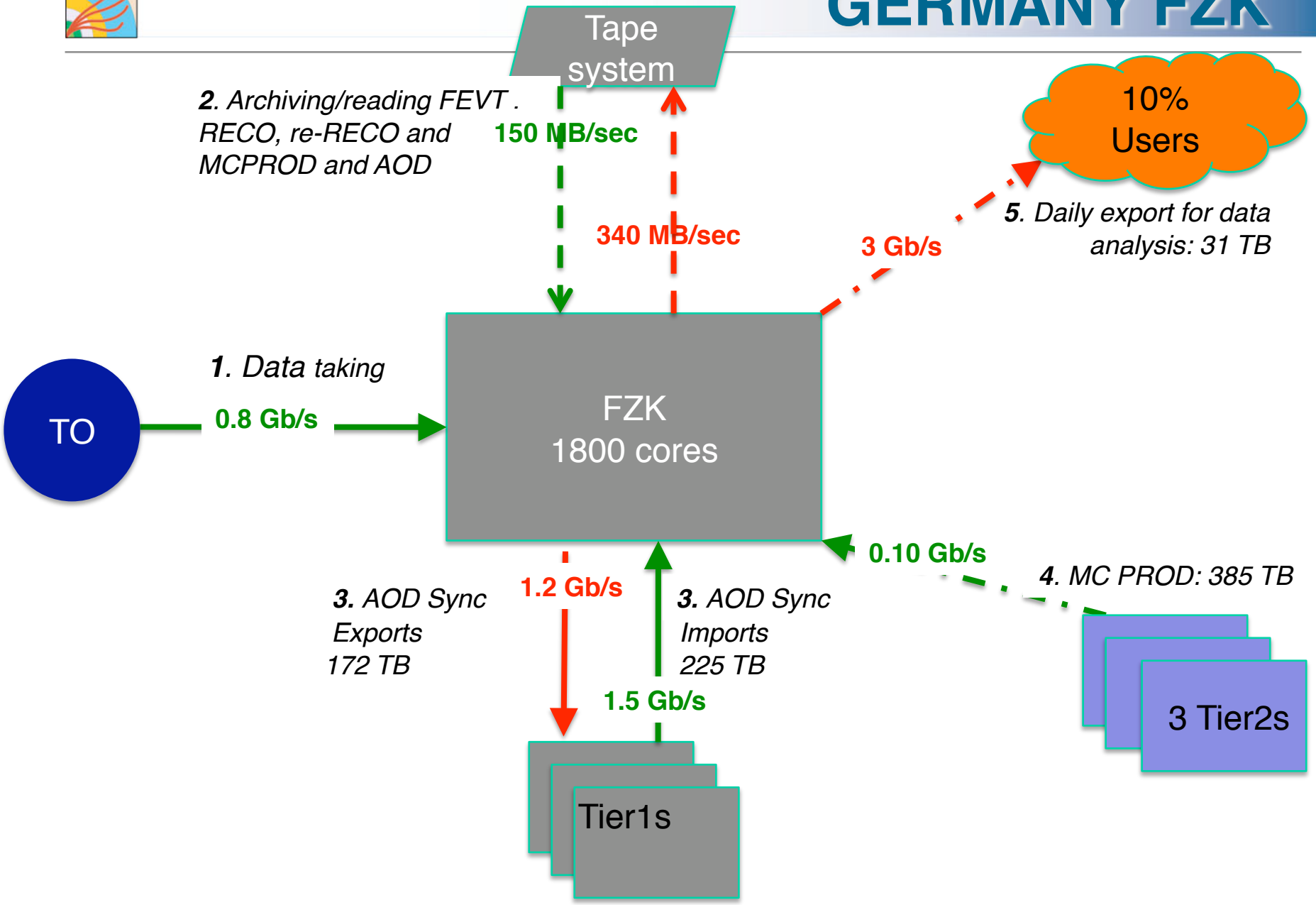
Next steps

- We are discussing with each Tier 1 individually to see how far apart we are
- We need to refine the model
- These two steps need to be don before the CMS Compute Resource Board of June 18th
  - Final green light for publication of the 2009-2010 edition
- For the future: to be reviewed at the end of each data taking period

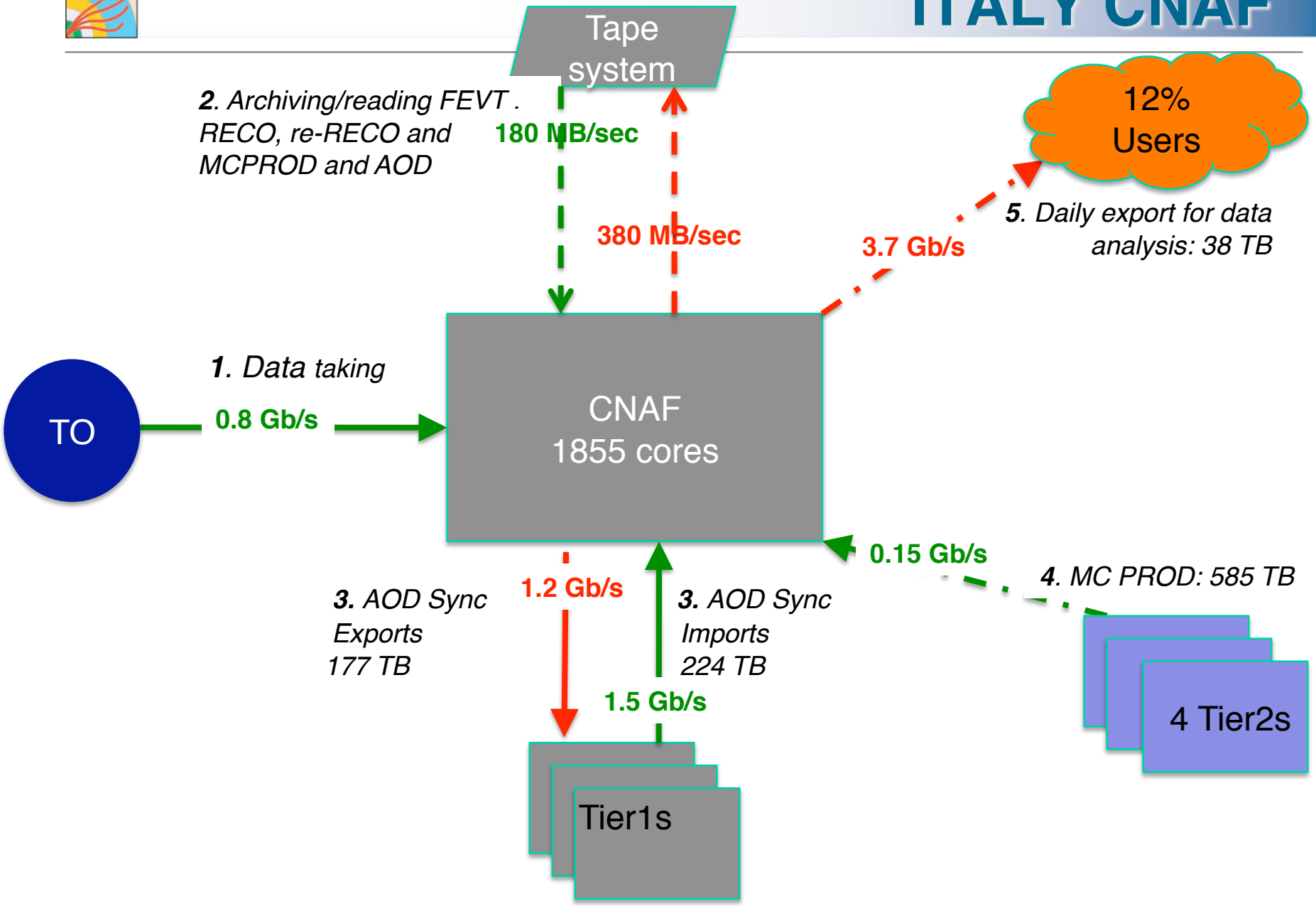
FRANCE IN2P3



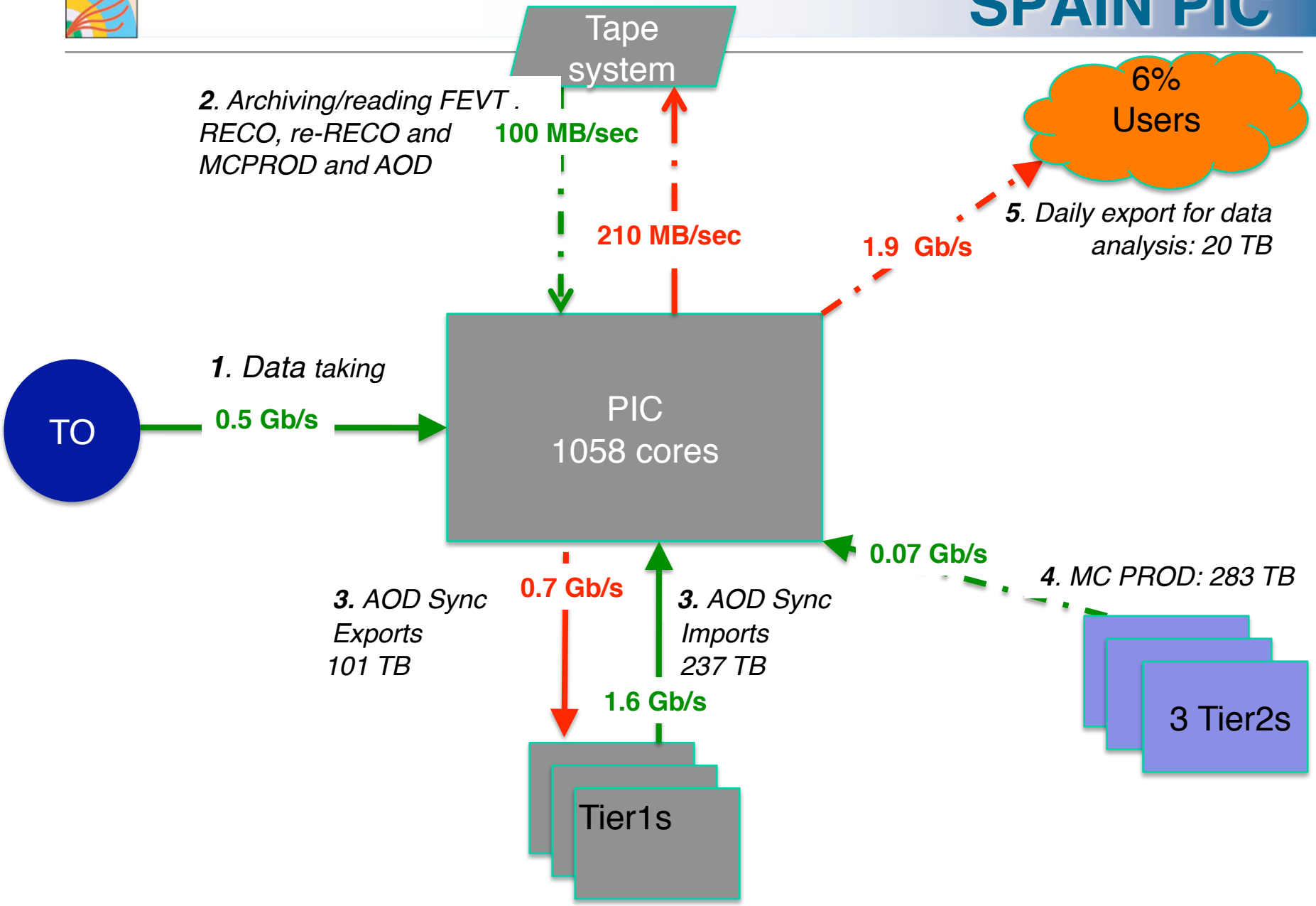
GERMANY FZK


Tape system
2. Archiving/reading FEVT . RECO, re-RECO and MCPROD and AOD
150 MB/sec
340 MB/sec
10% Users
5. Daily export for data analysis: 31 TB
3 Gb/s
1. Data taking
TO
0.8 Gb/s
FZK
1800 cores
0.10 Gb/s
4. MC PROD: 385 TB
3. AOD Sync Exports 172 TB
1.2 Gb/s
3. AOD Sync Imports 225 TB
1.5 Gb/s
3 Tier2s
Tier1s

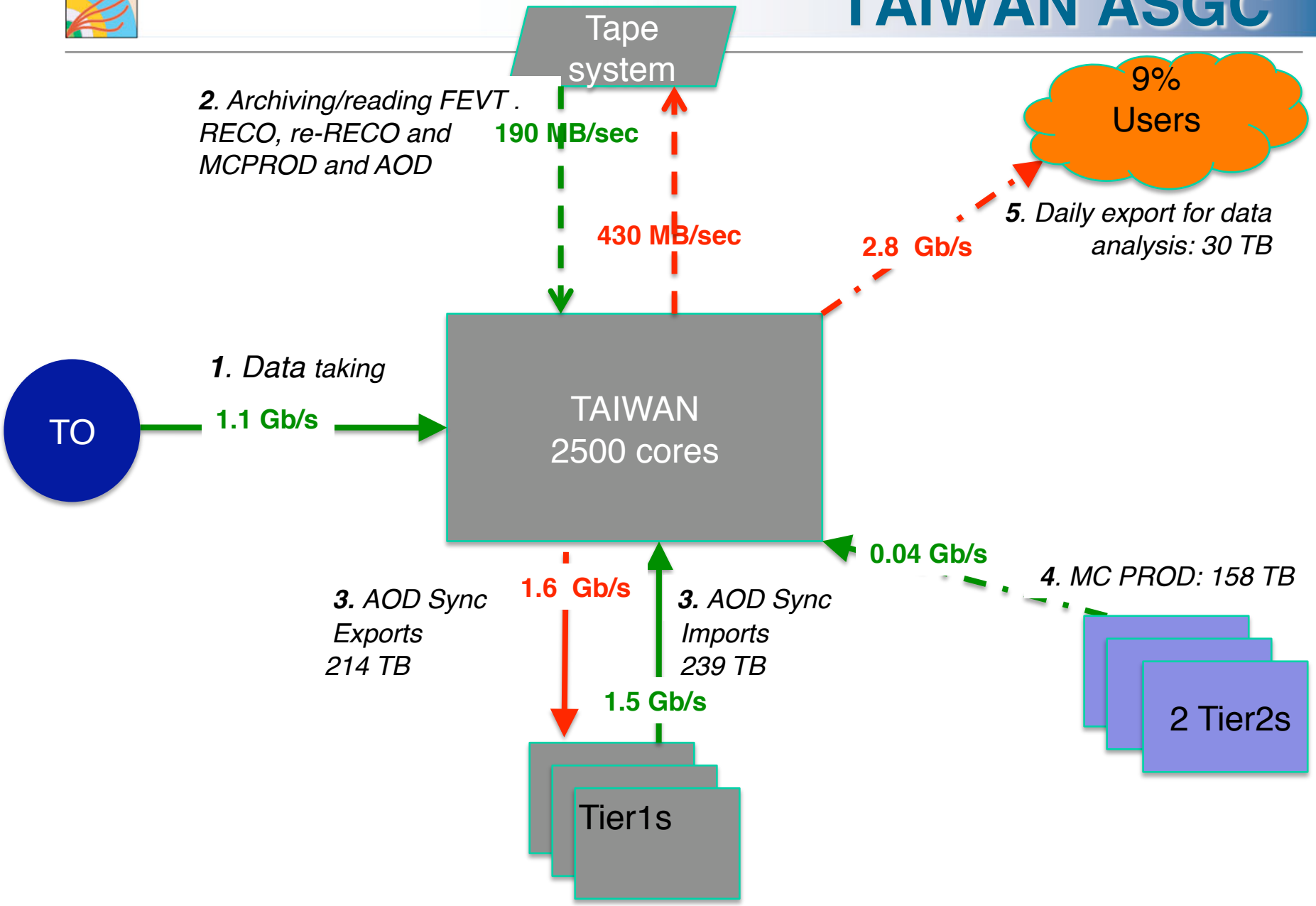
ITALY CNAF


Tape system
2. Archiving/reading FEVT . RECO, re-RECO and MCPROD and AOD
180 MB/sec
380 MB/sec
12% Users
5. Daily export for data analysis: 38 TB
3.7 Gb/s
1. Data taking
TO
0.8 Gb/s
CNAF
1855 cores
0.15 Gb/s
4. MC PROD: 585 TB
3. AOD Sync Exports 177 TB
1.2 Gb/s
3. AOD Sync Imports 224 TB
1.5 Gb/s
4 Tier2s
Tier1s

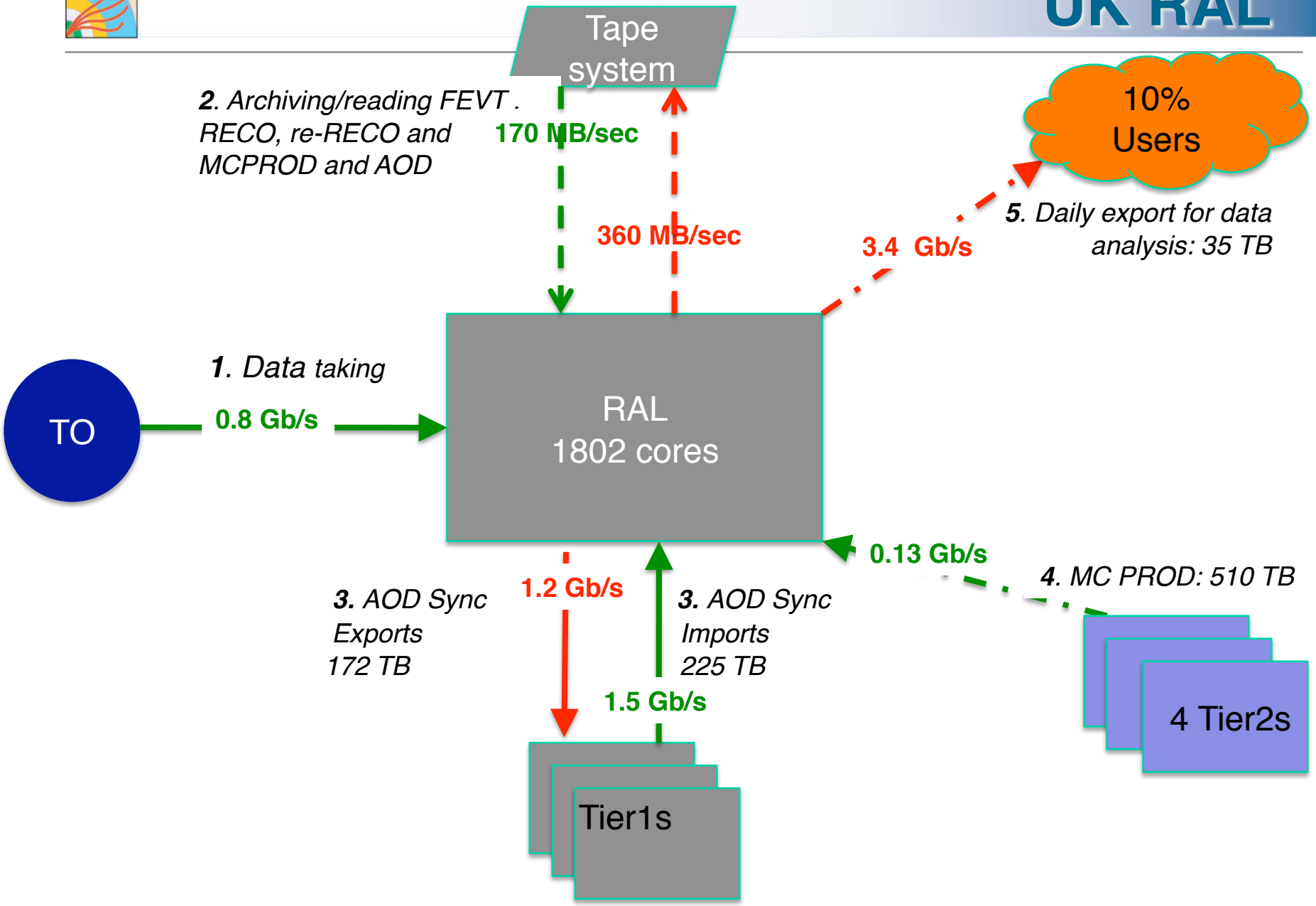
SPAIN PIC


Tape system
2. Archiving/reading FEVT . RECO, re-RECO and MCPROD and AOD
100 MB/sec
210 MB/sec
6% Users
5. Daily export for data analysis: 20 TB
1.9 Gb/s
1. Data taking
TO
0.5 Gb/s
PIC
1058 cores
0.07 Gb/s
4. MC PROD: 283 TB
3. AOD Sync Exports 101 TB
0.7 Gb/s
3. AOD Sync Imports 237 TB
1.6 Gb/s
3 Tier2s
Tier1s

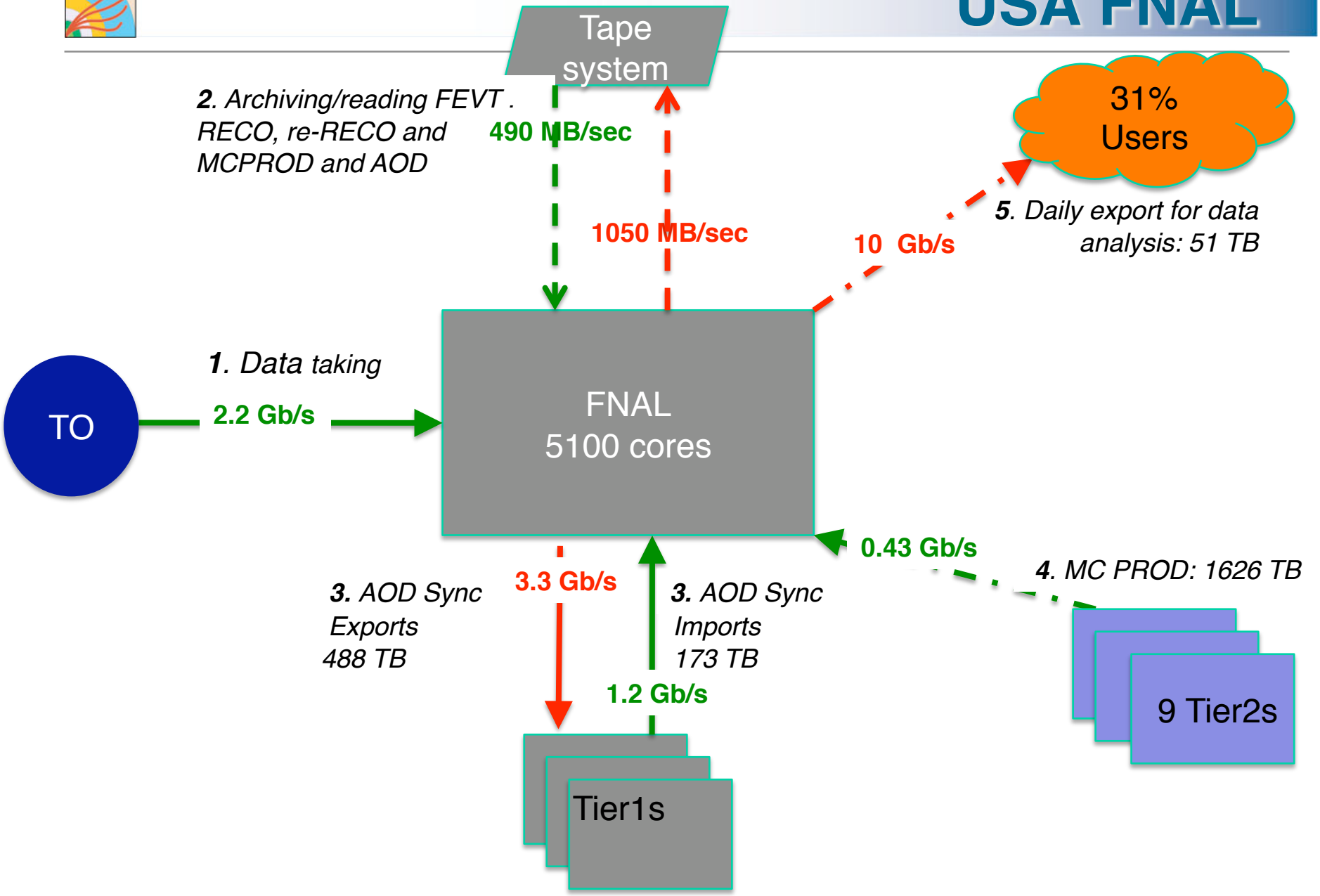
TAIWAN ASGC


Tape system
2. Archiving/reading FEVT . RECO, re-RECO and MCPROD and AOD
190 MB/sec
430 MB/sec
9% Users
5. Daily export for data analysis: 30 TB
2.8 Gb/s
1. Data taking
TO
1.1 Gb/s
TAIWAN 2500 cores
0.04 Gb/s
4. MC PROD: 158 TB
3. AOD Sync Exports 214 TB
1.6 Gb/s
3. AOD Sync Imports 239 TB
1.5 Gb/s
2 Tier2s
Tier1s

UK RAL


Tape system
2. Archiving/reading FEVT . RECO, re-RECO and MCPROD and AOD
170 MB/sec
360 MB/sec
10% Users
5. Daily export for data analysis: 35 TB
3.4 Gb/s
1. Data taking
TO
0.8 Gb/s
RAL
1802 cores
0.13 Gb/s
4. MC PROD: 510 TB
3. AOD Sync Exports 172 TB
1.2 Gb/s
3. AOD Sync Imports 225 TB
1.5 Gb/s
4 Tier2s
Tier1s

USA FNAL


Tape system
2. Archiving/reading FEVT . RECO, re-RECO and MCPROD and AOD
490 MB/sec
1050 MB/sec
31% Users
5. Daily export for data analysis: 51 TB
10 Gb/s
1. Data taking
TO
2.2 Gb/s
FNAL
5100 cores
0.43 Gb/s
4. MC PROD: 1626 TB
3. AOD Sync Exports 488 TB
3.3 Gb/s
3. AOD Sync Imports 173 TB
1.2 Gb/s
9 Tier2s
Tier1s