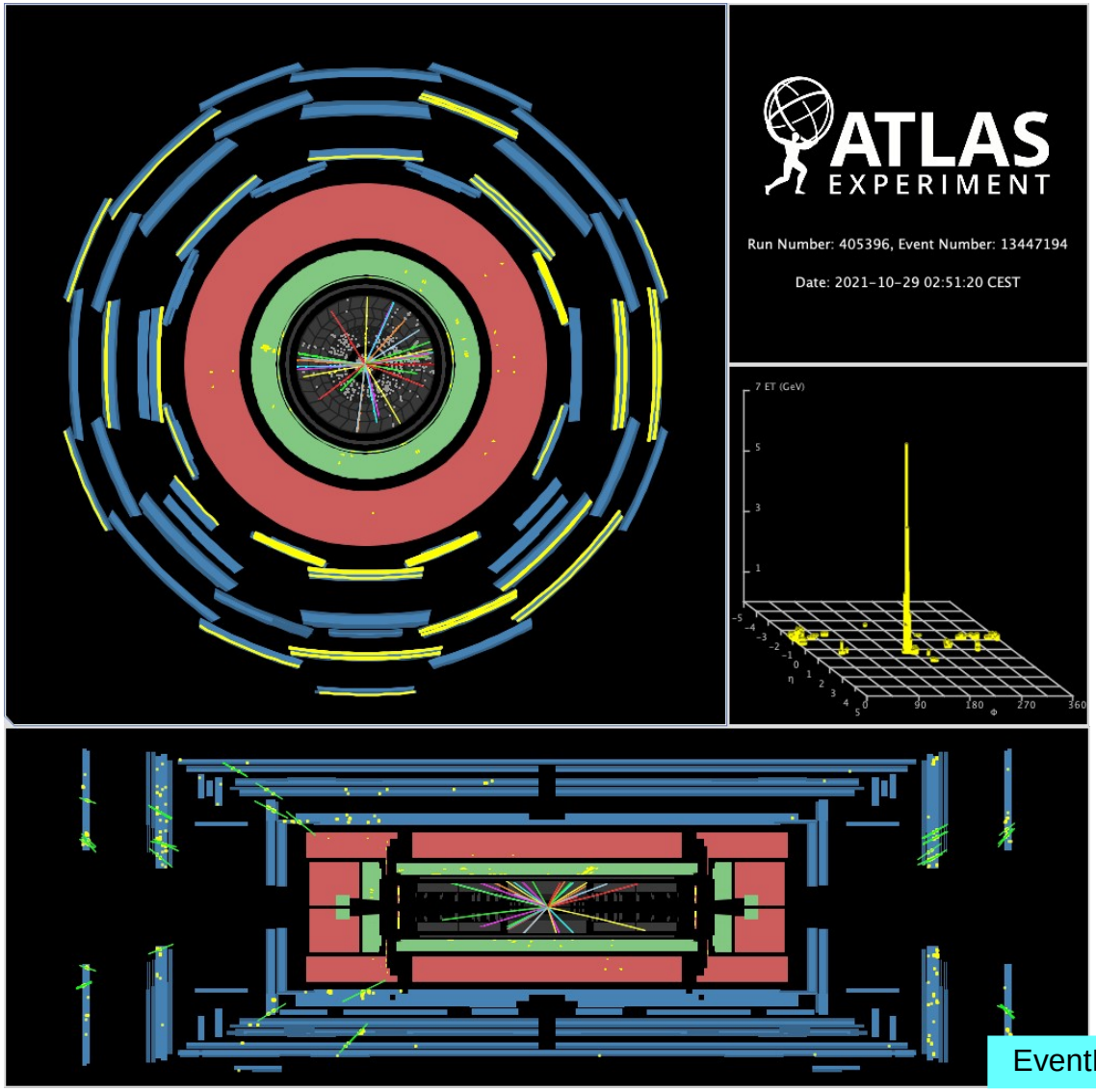
# S&C news from the ATLAS experiment

***F. Derue, LPNHE Paris***

Réunion des sites LCG France

17th November 2021, CC-IN2P3 Lyon



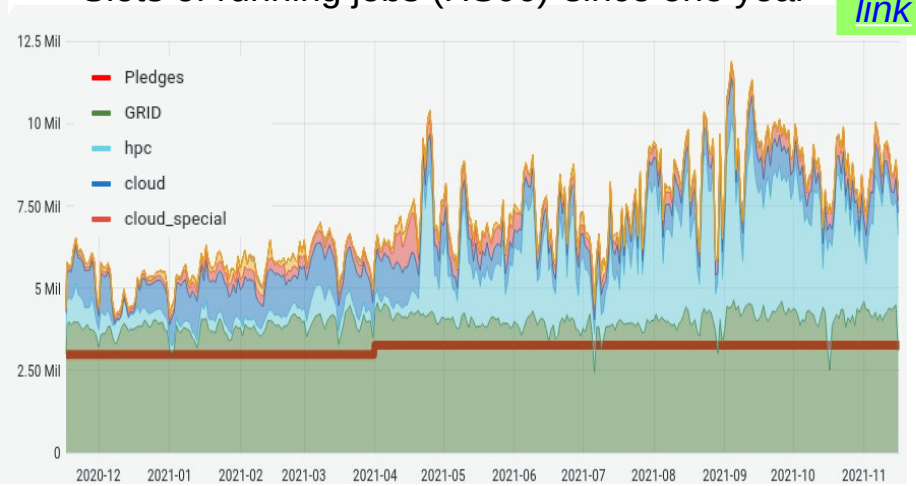

EventDisplayRun3Collisions

# Status and performance of this year

# ATLAS CPU/GPU resource usage

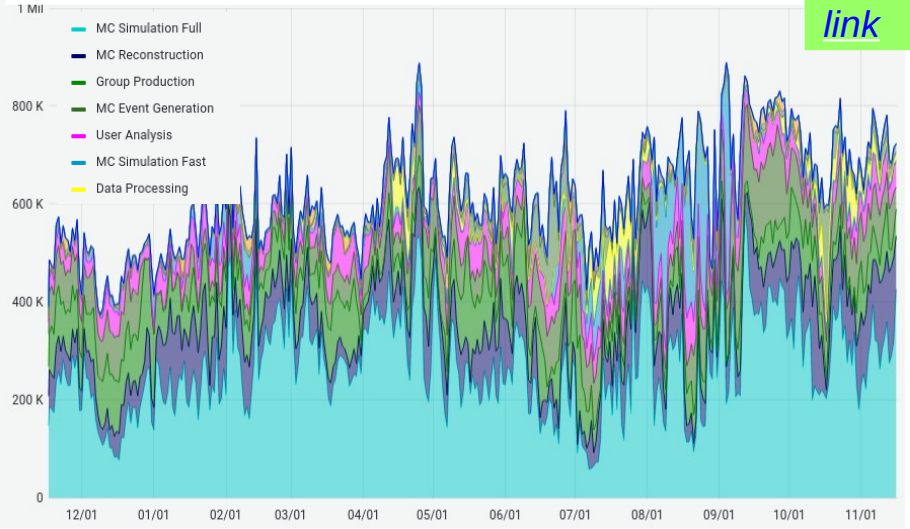
Slots of running jobs (HS06) since one year

*link*




Slots of running jobs (HS06) since one year

*link*

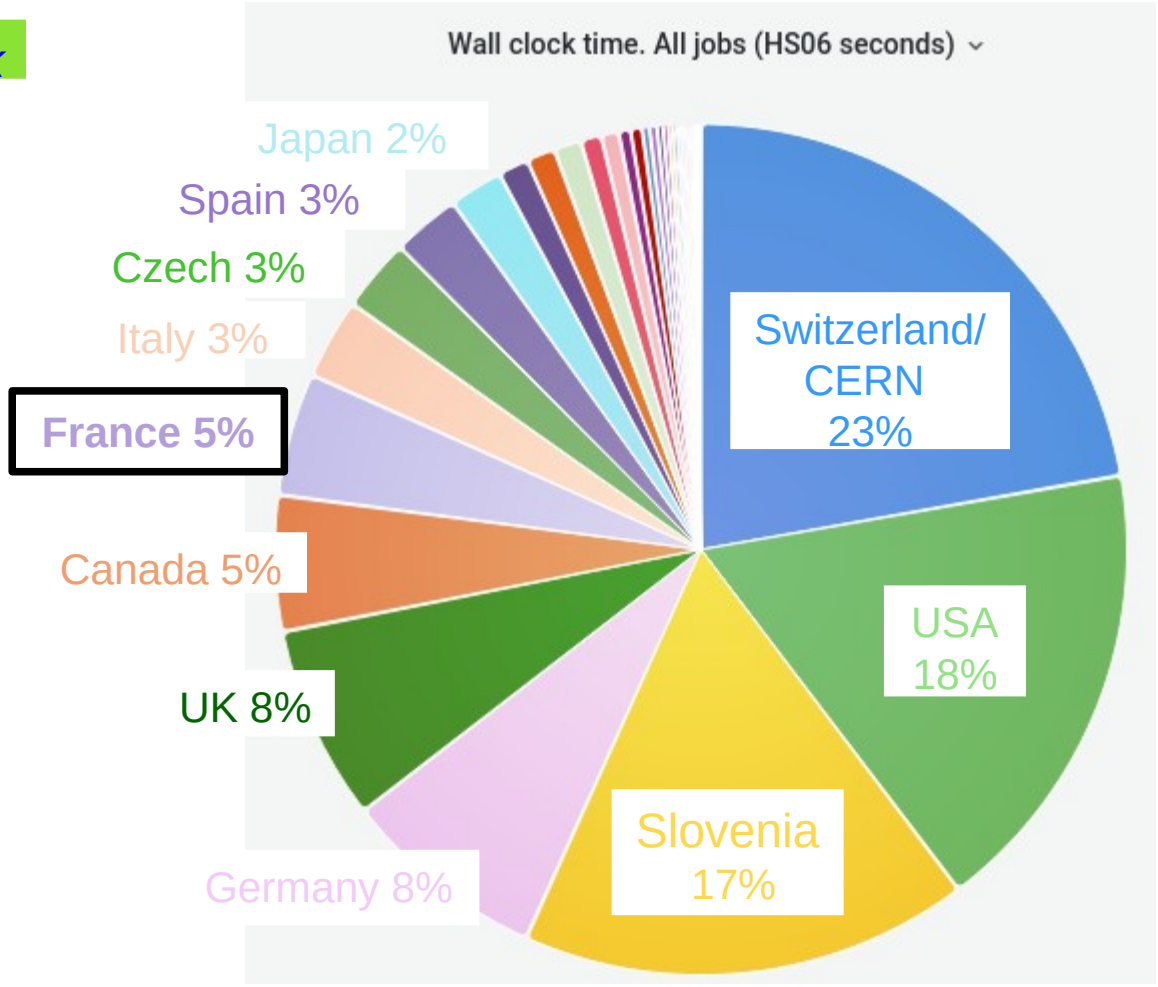



- **Excellent performance of the distributed computing infrastructure**
  - on the grid, HPC, cloud, HLT farm, T0
  - opportunistic resources (HPCs and HLT farm) are doubling our pledge
    - CPU only – no additional disk from these sites
    - opportunistic – may disappear at any time
  - 500-800 k jobs per day
  - many activities in parallel
    - ~3/4 of MC simu, reco, evgen
    - Run 2 data reprocessing just starting, with Run 2 MC repro that will follow “soon”

# CPU/GPU usage per country

- **Computing usage (HS06) realized on Tier 0, grid, HPC and cloud (pledge+opportunistic) by each country since one year**

*link*



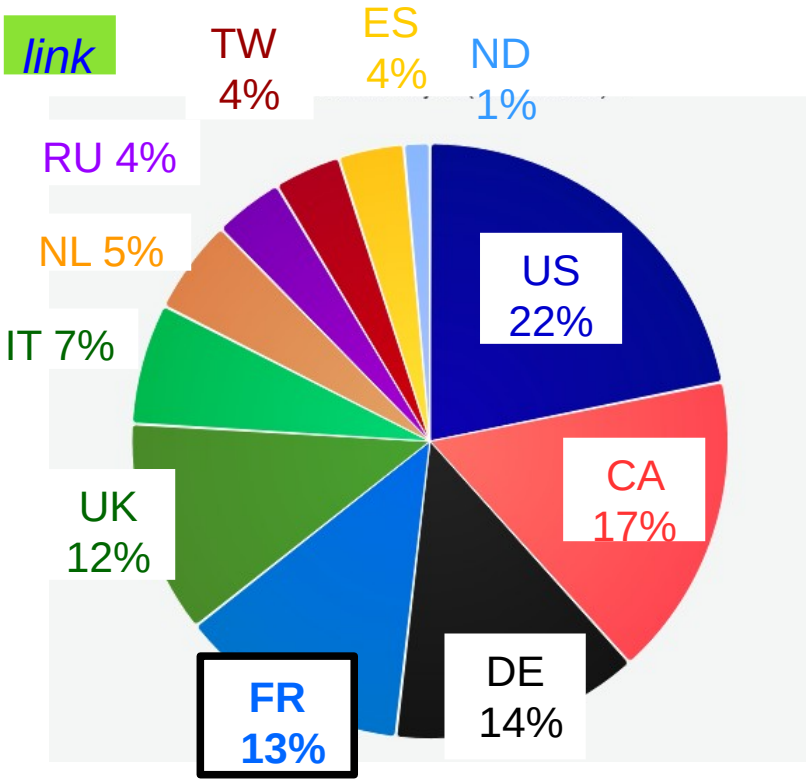

France has realized 5% of all ATLAS usage of computing resources this year

# CPU usage on grid in Tier 1

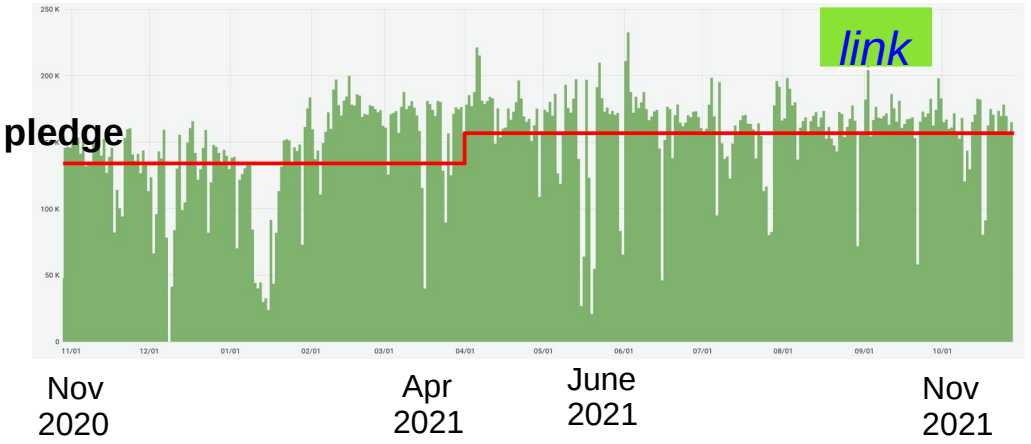
- **Pledge of CC-IN2P3 (see *cric*)**
  - represents 12.6% of all T1s in 2021
  - for 2022 pledge increases by 15% will represent 13.9% of all T1s

| Pledge 2021 (HS06) | Pledge 2022 (HS06) |
|---|---|
| 156780 | 180700 |

- **CPU realized on grid by each T1 since one year**




Slots of running jobs (HS06) at CC-IN2P3



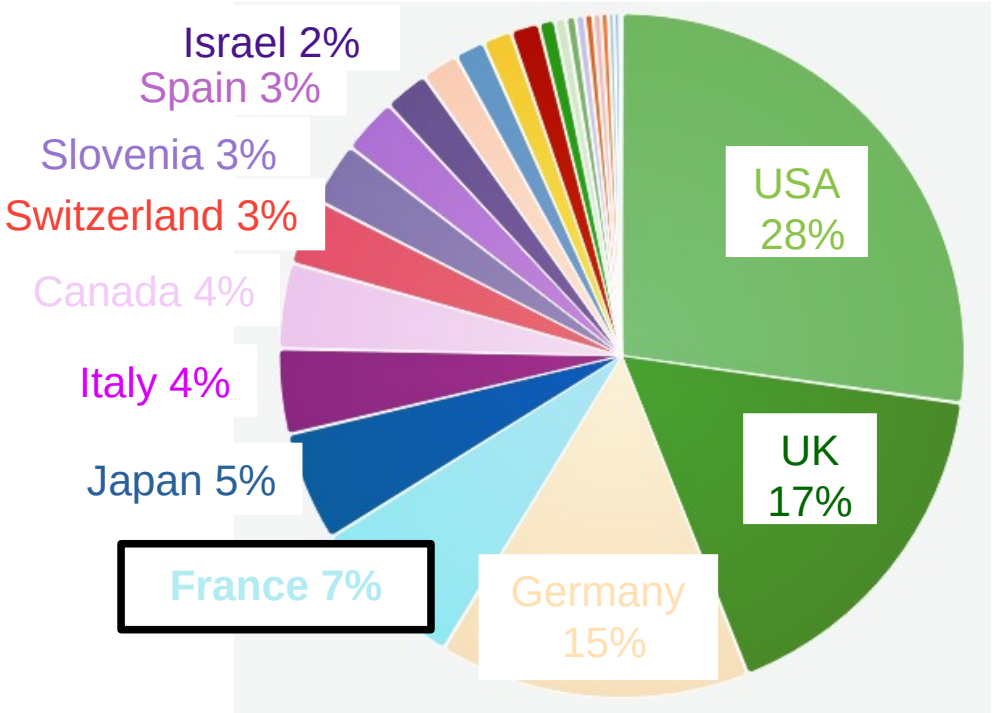

95% of its pledge since a year (drop in January)

France has realized 13% of T1 ATLAS usage of computing resources this year

# CPU usage on grid in FR-cloud Tier2s

- **Pledge (see *cric*)**
  - FR-cloud : France, Japan, Romania, China
    14.9% of T2 in 2021, 14% in 2022
  - France : 8.6% of T2 in 2021, 9% in 2022
    - pledges increase by 7% in 2022
- **realized by each country since one year**

*link*

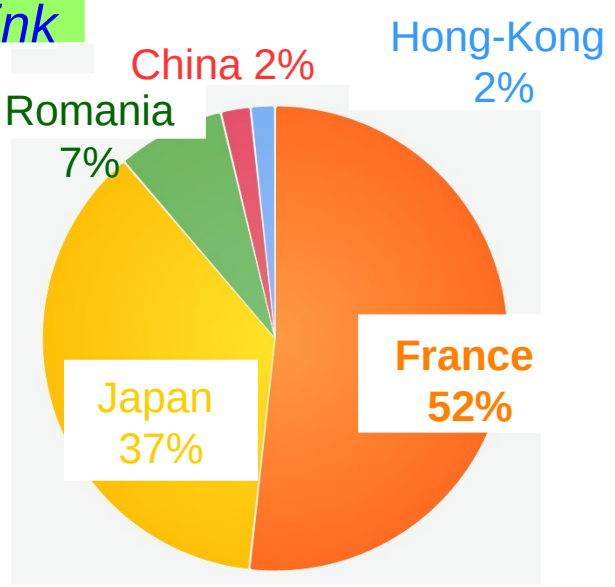



| Site | Pledge 2021 HS06 | Pledge 2022 HS06 |
|---|---|---|
| **GRIF** | 53320 | 59391 |
| **Tokyo** | 48000 | 52000 |
| **LAPP** | 30000 | 36000 |
| **GRIF-IRFU** | 24000 | 25900 |
| **CPPM** | 24000 | 24000 |
| **RO-LCG** | 35000 | 24000 |
| **LPC** | 19000 | 24000 |
| **GRIF-LAL** | 16600 | 18400 |
| **GRIF-LPNHE** | 13200 | 17200 |
| **LPSC** | 13300 | 11900 |
| **Beijing** | 8000 | 8000 |
| **Hong-Kong** | 1000 | 1000 |

France has realized 7% of T2 ATLAS usage of computing resources this year

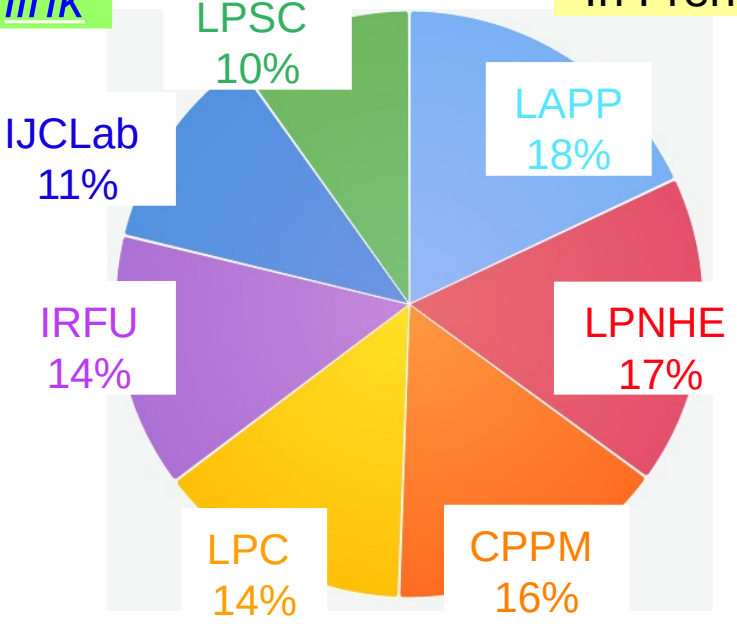
# CPU usage on grid in FR-cloud and French Tier2s

link



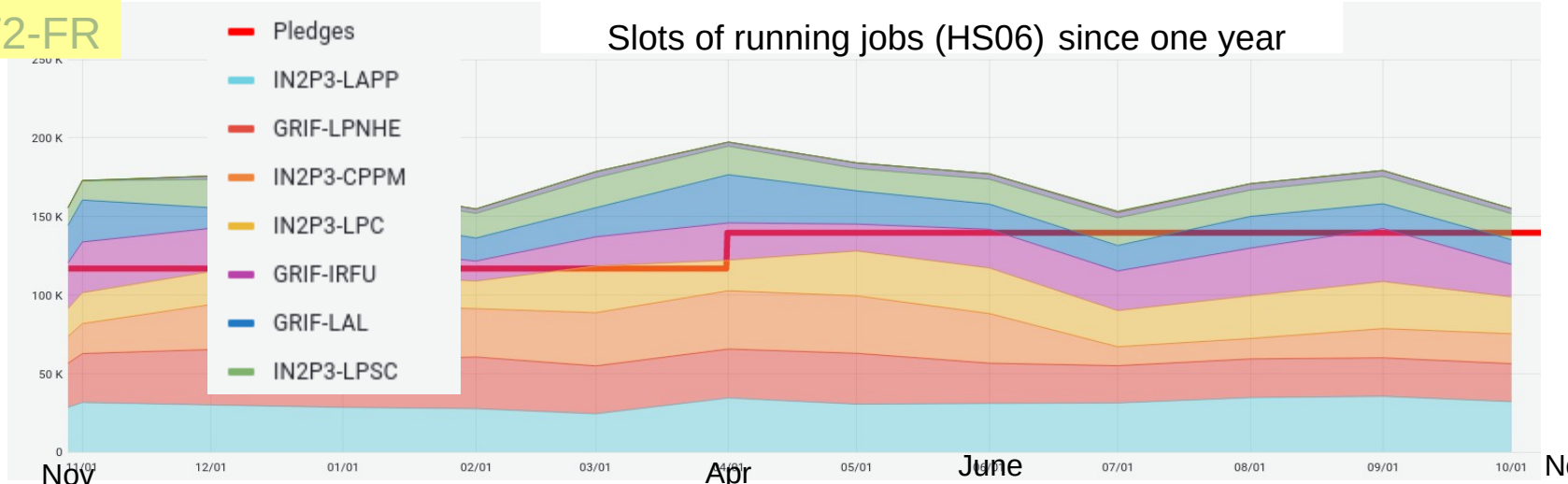

link

In French sites




T2-FR

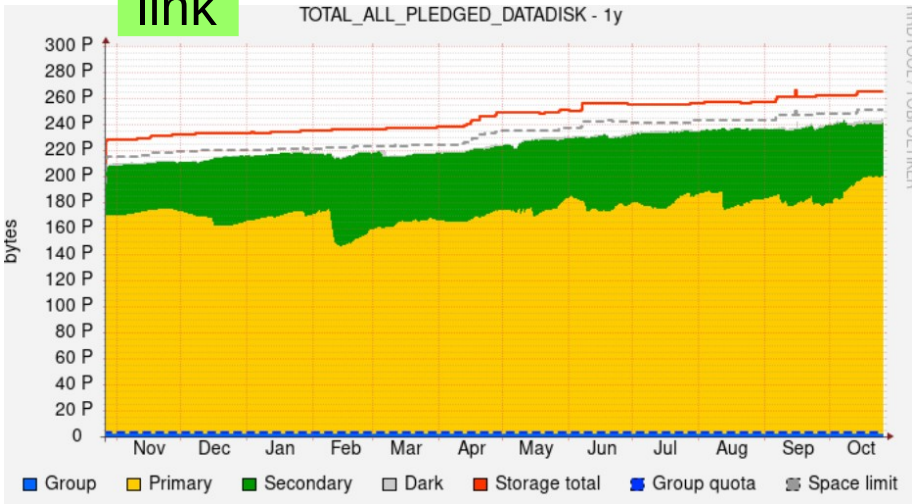
Slots of running jobs (HS06) since one year

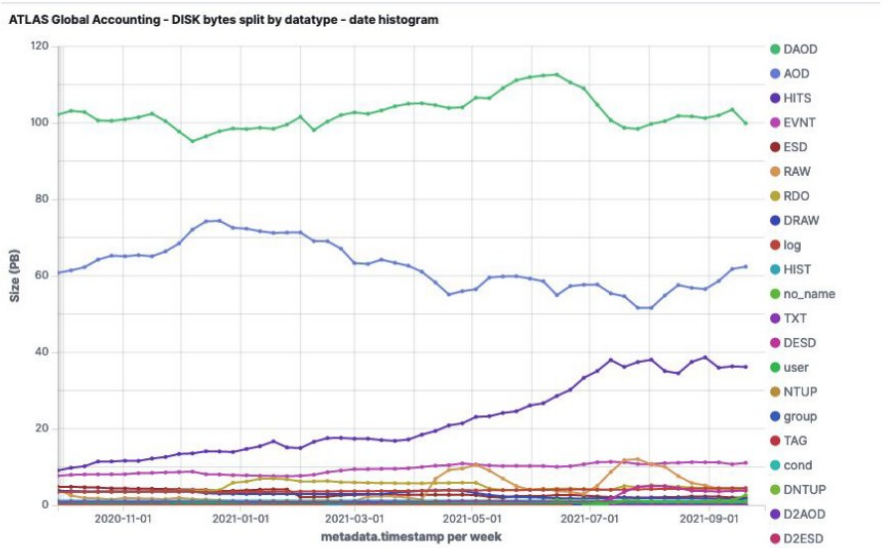



France has realized 134% of its ATLAS T2 pledge of computing resources this year

# ATLAS storage resource usage

link

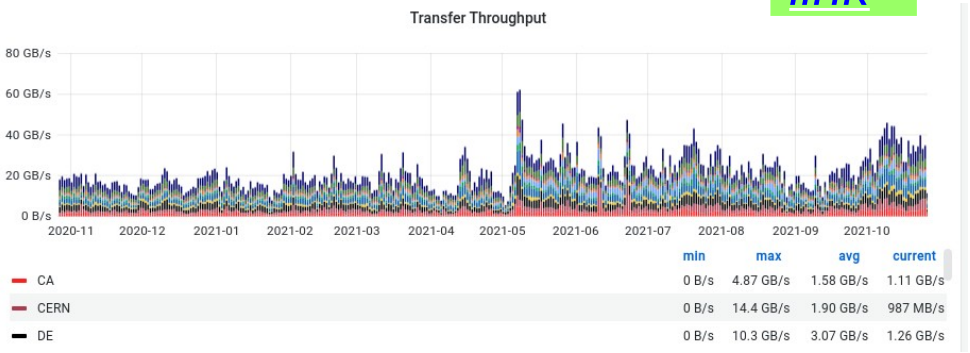



- **Storage evolution**
  - >500 PB disk+tape used
  - disk pledge of 270 PB – 230 TB filled
  - AOD policy change, no longer keep a disk replica: larger disk buffer of secondary data
  - further use of data carousel beyond RAW and HITS, now less popular (mainly MC) AOD inputs

- **Data movement (per day)**
  - moving 2 PB/d (1.5M files) @20 GB/s with peaks at 60 GB/s
  - deleting ~1-2 PB/d

*link*

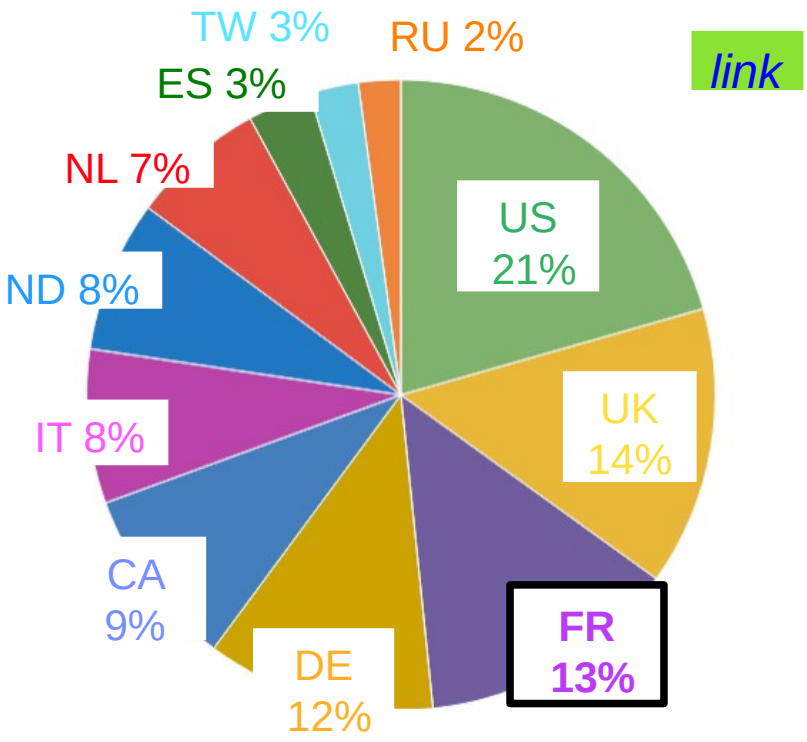


# Storage usage in Tier 1

- **Pledge of CC-IN2P3 as a Tier-1 (see *cric*)**
  - represents 13.3% of all Tier 1 in 2021
    (12.2 % for disk, 13.9% for tape)
    14.3% in 2022 (14.8 % for disk, 14% for tape)
  - in 2022 pledge increases by 17% for tapes and by 28% for disk

| | Pledge 2021 (TB) | Pledge 2022 (TB) |
|---|---|---|
| **Disk** | 14175 | 16240 |
| **Tape** | 33605 | 40256 |

- **Storage for each Tier 1 (last month)**
  DATADISK+SCRATCHDISK+
  DATATAPE+MCTAPE




France has realized 13% of T1 ATLAS usage of storage resources this year

# Storage usage in FR-cloud Tier2s

- **Pledge (see *cric*)**
  - FR-cloud : 18% of T2s in 2021, 19% in 2022
  - France : 9% of T2s in 2021, 10% in 2022
    - pledges increase by 16% in 2022
- **LOCALGROUP** *live-storage*
  - non pledged resources
  - local storage for our analyses

| | Total (TB) | Free (TB) |
|---|---|---|
| **CPPM** | 245 | 26 |
| **GRIF** | 1014 | 219 |
| **GRIF-IRFU** | 396 | 103 |
| **GRIF-LAL** | 26 | 12 |
| **GRIF-LPNHE** | 592 | 104 |
| **LAPP** | 147 | 110 |
| **LPC** | 22 | 11 |
| **LPSC** | 82 | 22 |
| **Beijing** | 18 | 7 |
| **Tokyo** | 1300 | 249 |
| **RO-07** | 11 | 6 |

*live-storage*

| | Pledge 2021 (TB) | Pledge 2022 (TB) | Installed (TB) |
|---|---|---|---|
| **CPPM** | 2200 | 2200 | 2264 |
| **GRIF** | 4962 | 5510 | 5240 |
| **GRIF-IRFU** | 1945 | 2124 | 2140 |
| **GRIF-LAL** | 1507 | 1646 | 1590 |
| **GRIF-LPNHE** | 1510 | 1740 | 1510 |
| **LAPP** | 2940 | 3700 | 2940 |
| **LPC** | 1500 | 2188 | 1735 |
| **LPSC** | 734 | 734 | 1160 |
| **Beijing** | 400 | 400 | 370 |
| **Hong Kong** | 1050 | 1050 | 880 |
| **Tokyo** | 7200 | 8000 | 8000 |
| **RO-07** | 2500 | 3500 | 2640 |

France has realized 11% of T2 ATLAS usage of storage resources this year

# Readiness for Run 3

# ATLAS usage of opportunistic resources

- **HPC providing now significant resource to ATLAS**
  - US, JP, PRACE, EuroHPC … : last ½ year, ~30% of CPU power
  - universal case (like Vega) : transparent for all workflows
    HPC must provide CE, cvmfs, squid, outbound connectivity - basic grid requirements
  - difficult to determine exactly how many HPC resources we will have in 2022
    expect more EuroHPC to come online next year - but also a much reduced allocation on Vega

- **Sim@P1 on the HLT farm**
  - several hard/software improvements done to prepare the HLT farm for data taking
  - can we efficiently utilse the gaps between the fills during Run 3 for simulation jobs?
    - duty cycle expected to be several hours, longer in 2022
    - lends itself to running AF3, with dedicated, low priority queue
      also need “FS mode” for longer machine breaks
    - also need to improve ramp up on both panda/harvester and hardware sides

- **Tier 0 status**
  - T0 will also resume its normal rôle :hardware upgrades + recommisioning are on track
  - software updated to include latest functionality, features; new testing and integration instance available

# Data reconstruction

- **Data Challenges : Storage (June)**
  - Tier 0 « processed » biggest 3 physics streams (14 TeV numbers)
    - RAW inputs: physics_Main (1.7 kHz, 3.5 GB/s) + BLS and HH delayed streams (1.1 kHz, 2.2 GB/s)
    - Another ~2 GB/s of other delayed streams + calibration, TLA ...
  - output AODs ~2GB/s with an additional ~2 GB/s from derived outputs (DRAW, performance DESD, DAOD)

- **Not clear yet if/how the delayed streams can be promptly processed during data taking**
  - began discussions with data prep to get more info on which delayed streams are important

- **If delayed stream processing to be done sooner than normal (e.g EoY), alternatives to Tier 0 :**
  - spillover to grid – onlyever really « tried in anger »
    - trig
    - triggered by conditions green light, when RAW still on disk buffer
    - tracking SW changes from T0 : same tags
    - long standing problem of DQ histograms
    - reserve for « real » physics_Main spillover ?
  - standard « reprocessing mode » via ProdSys
    - open-ended request : add incoming RAW dataset to a container, produce task to process it
    - Data Carousel makes rule to move and keep RAW on disk
    - Tier 0 tag connection is lost, or is manual

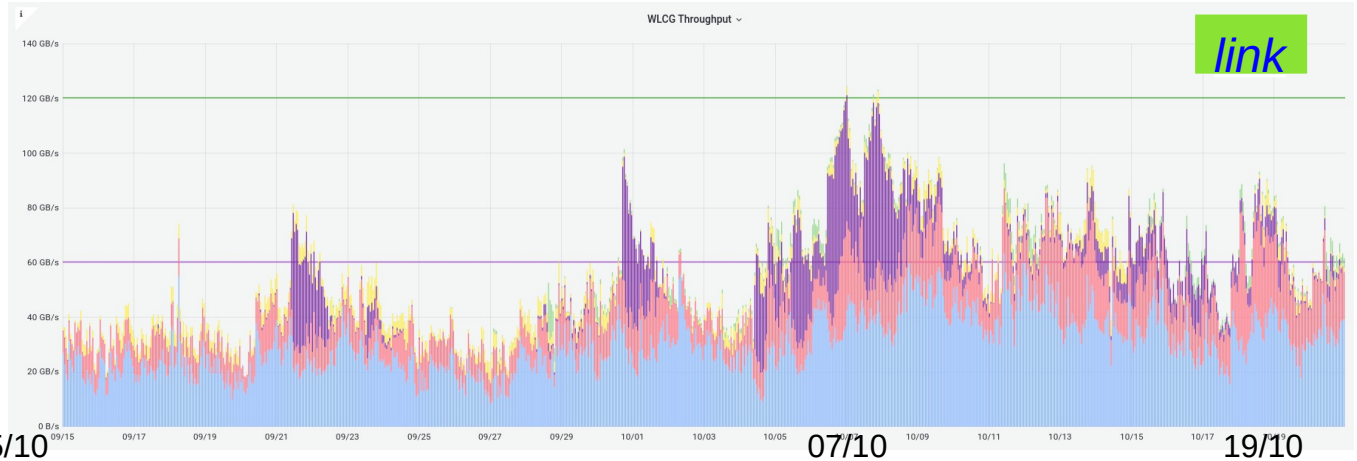
# Analysis Model

- **New analysis model in place**
  - most DAOD formats replaced by a single format called DAOD_PHYS + run 4 prototype format called DAOD_PHYSLITE
  - only combined performance groups and physics groups who are
  - unable to use DAOD_PHYS(LITE) will be allowed to produce dedicated DAODs
  - possibility to centrally skim DAOD_PHYS
- **But ...**
  - full reprocessing of all run 2 data/MC in release 22 will happen later this year
  - full set of CP recommendations/calibrations won't be complete until (probably) late 2022
  - can't expect new analyses to start on a data sample that doesn't have a full set of calibrations
  - so this means unfortunately that new analysis on run 2 data will continue to use the old analysis model and release 21 AOD until probably the end of next year
  - ongoing Run 2 analyses won't move over in any case
  - Run 3 analysis will use the new model (but early on this will mostly be combined performance)
  - so we need to be prepared to support both analysis models (and both r21 and r22 AODs) for at least 2022 and 2023

# Data challenge : network/disk

- **Main goals**
  - commission HTTP-TPC
  - demonstrate we can fill 10% the bandwidth that is requested at HL-LHC scale
  - set of challenges to gradually test bandwidth between start of Run 3 and Run 4 increasing the bandwidth expectation each year
- **Challenge running all together**
  - ATLAS using real data, mainly shipping between the T1s
  - aim : 60 GB/s minimal, 120 GB/s flexible

| T1 | Data Challenge target 2027 (Gbps) | Data Challenge target 2025 (Gbps) | Data Challenge target 2023 (Gbps) | Data Challenge target 2021 (Gbps) |
|---|---|---|---|---|
| CA-TRIUMF | 98 | 59 | 29 | 10 |
| DE-KIT | 312 | 187 | 94 | 31 |
| ES-PIC | 93 | 56 | 28 | 9 |
| FR-CCIN2P3 | 281 | 169 | 84 | 28 |
| IT-INFN-CNAF | 336 | 202 | 101 | 34 |
| KR-KISTI-GSDC | 25 | 15 | 7 | 2 |
| NDGF | 71 | 43 | 21 | 7 |
| NL-T1 | 94 | 56 | 28 | 9 |
| NRC-KI-T1 | 62 | 37 | 19 | 6 |
| UK-T1-RAL | 296 | 177 | 89 | 30 |
| RU-JINR-T1 | 52 | 31 | 15 | 5 |
| US-T1-BNL | 227 | 136 | 68 | 23 |
| US-FNAL-CMS | 454 | 273 | 136 | 45 |
| (atlantic link) | 681 | 408 | 204 | 68 |
| Sum | 2400 | 1440 | 720 | 240 |




- **Successfully achieved flexible rate targer**
  - Majority of transfers coming from ATLAS
- **Post-challenge analysis and further iterations to follow. really targetting Run 4**

# TPC migration

- **Started migration in August 2020**
  - 80% of the sites migrated to https-tpc by May 31st 2021
  - now only one T2 still remains to be moved, as well as nine T3s to handle
- **Protocol zoo being reduced to 2 protocols at most sites**
  - https and root
    - read/write/delete on lan/wan
  - support for globus offline for HPC
- **SRM being kept in the medium term**
  - SRM+gsiftp → SRM+https
- **TAPE REST API**
  - ongoing work to replace SRM completely in the long term

TPC migration : successful move away from gsiftp to https

# Tokens migration

- **Transition to tokens was supposed to be a gradual change during Run 3**
  - but policy changes in an external provider (HTCondor dropping all support for X.509 by Sept. 2022) have forced our hand to some extent

- **Harvester job submission to HTCondor CE with tokens is already supported**
  - this must be rolled out to all HTCondor CE sites by next summer
  - along with submission to ARC REST interface
  - note still with X.509 delegation, so jobs still use X.509 proxy to contact storage, Rucio, PanDA, etc.

- **Timeline for sites (suggestion)**
  - OSG sites support tokens by Feb 2022
  - other HTCondor CE sites support tokens by June 2022
  - ARC-CE sites provide REST interface by June 2022
    - no tokens necessary, job submission can still use condor SSL authentification
  - all done before Sept. 2022, EOL of condor v9.0.x

# Tape challenge

- **Stress test readiness of WLCG tape sites (T0/T1s) and relevant central services (e.g FTE) at the Run 3 scale, for all experiments**
  - happening in mid-October : all experiments and all stape sites joined
    - ATLAS/CMS: both write and read test
    - ALICE/LHCb: write test only

- **Alongside production activities, add test streams on top as needed**
  - driven by production ADC workflow + data management systems
  - Data-Taking (DT) mode:
    - write test: Tier-0 export to T1s at ~10 GB/s
    - read test: keep staging at low level (Run 2 repro)
  - After-Data-Taking (A-DT) mode:
    - read test: staging (DC for Run 2 repro) + test streams if needed
    - write test: keep writing from T0 to T1s, at lower A-DT target rate

Numbers from CRSG report + agressive trigger scenario [*link, ATLAS only*]

ATLAS tape throughput estimates for Run 3

| Site | Reads (DT) GB/s | Writes (DT) GB/s | Reads (A-DT) GB/s | Writes (A-DT) GB/s | participation in the October tape challenge |
|---|---|---|---|---|---|
| CERN | No reprocessing during data taking | 10GB/s (RAW + AOD) | If reprocessing and all activity read from CTA: 12GB/s | - | |
| all T1s (avg) | 2.5 | 9.6 | 8.4 | 5.1 | |
| BNL | 0.6 | 2.2 | 1.9 | 1.2 | Yes |
| CNAF | 0.2 | 0.9 | 0.8 | 0.5 | Yes |
| IN2P3 | 0.4 | 1.4 | 1.2 | 0.7 | Yes |
| KIT | 0.3 | 1.2 | 1.0 | 0.6 | Yes |
| NDGF | 0.1 | 0.5 | 0.5 | 0.3 | Yes |
| NLT1 | 0.2 | 0.7 | 0.6 | 0.4 | Yes |
| RRC-KI | 0 | 0 | 0 | 0 | Yes |
| PIC | 0.1 | 0.4 | 0.3 | 0.2 | Yes |
| RAL | 0.4 | 1.4 | 1.2 | 0.7 | Yes |
| TRIUMF | 0.3 | 1.0 | 0.8 | 0.5 | Yes |

- **Run 3 data taking rate will be ~10 GB/s, if we write a whole run to a single T1, can site resources (tape disk buffer + tape drives) handle it ?**
  - planning in earnest today, also involving reprocessing coordination

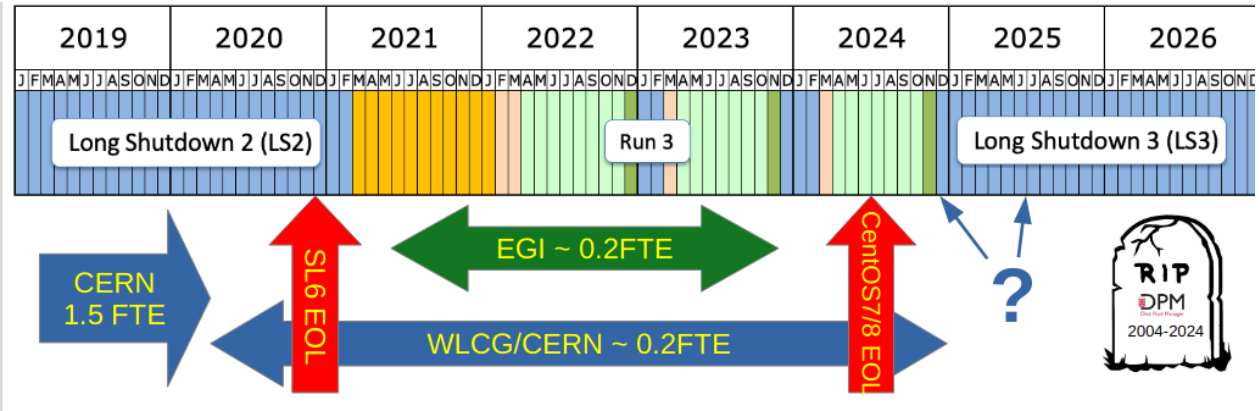
# ATLAS Database and Metadata (ADAM)

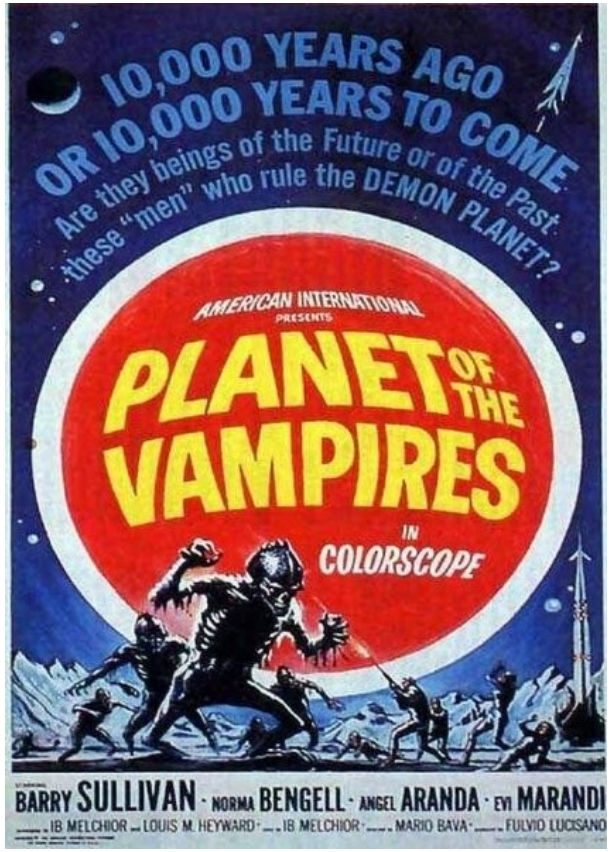
- **Database infrastructure**
  - Oracle19c : long term support extended until April 2027: covers the entire Run 3
  - but in 2023, the current Oracle Campus License contract comes to an end.
    - new licensing model is expected to change to Oracle processor based cost proportional to number of CPU cores deployed in DB servers
    - due to licensing cost, Tier1s will probably phase out
    - we need to guarantee we can run with Oracle (and Frontier launchpads) at CERN only

- **Application area : AMI**
  - Oracle DB migration from CC-Lyon to CERN
    - primary AMI service will move @CERN
    - removal of Golden Gate replication from CC-Lyon to CERN
  - ongoing activities
    - prepare AMI DB snapshots
    - Openstack cloud AMI nodes @ CERN ready to be used on the snapshot
  - migration procedure
    - expected for second half of October
    - simple update of DB connection string
    - transparent for the users in terms of clients (web and pyAmi)

- **Frontier/squid**
  - stable operations, no hardware or software changes
  - stress tests (study for Oracle @ CERN only), second half of October
    - setup : Lyon and TRIUMF forwarding traffic to CERN launchpads

# Beyond DPM !


2019
2020
2021
2022
2023
2024
2025
2026
Long Shutdown 2 (LS2)
Run 3
Long Shutdown 3 (LS3)
CERN
1.5 FTE
SL6 EOL
EGI ~ 0.2FTE
WLCG/CERN ~ 0.2FTE
CentOS7/8 EOL
?
RIP
DPM
2004-2024


10,000 YEARS AGO
OR 10,000 YEARS TO COME
Are they beings of the Future or of the Past
these "men" who rule the DEMON PLANET?
AMERICAN INTERNATIONAL PRESENTS
PLANET OF THE VAMPIRES
IN COLORSCOPE
BARRY SULLIVAN · NORMA BENGELL · ANGEL ARANDA · EVI MARANDI
IB MELCHIOR · LOUIS M. HEYWARD · IB MELCHIOR · MARIO BAVA · FULVIO LUCISANO


AMAZING!
M·G·M PRESENTS
FORBIDDEN PLANET
IN CINEMASCOPE AND COLOR
STARRING
WALTER PIDGEON · ANNE FRANCIS · LESLIE NIELSEN
With WARREN STEVENS
And Introducing ROBBY THE ROBOT
Screen Play by CYRIL HUME · Based on a Story by IRVING BLOCK and ALLEN ADLER · Photographed in EASTMAN COLOR · Directed by FRED McLEOD WILCOX · Produced by NICHOLAS NAYFACK
AN M-G-M PICTURE

# C-RSG - Resource Request 2023

| Stable beam fraction | Average <μ> | Stable beam seconds | Integrated luminosity ($fb^{-1}$) | Main physics trigger rate (kHz) | Main physics stream events | Delayed trigger rate (kHz) | Delayed stream events |
|---|---|---|---|---|---|---|---|
| 50% | 50 | $6 \times 10^6$ | <100 | 1.7 | 10.2 billion | 1.6 | 9.6 billion |

*Table 2: summary of the LHC operating programme in 2023 and the number of events that ATLAS expects to record.*

| | | 2022 agreed @ April 2021 RRB | 2023 Request @ March 2022 RRB | Balance 2023 wrt 2022 |
|---|---|---|---|---|
| CPU | T0 (kHS06) | 550 | 740 | 34.5% |
| CPU | T1 (kHS06) | 1300 | 1536 | 18.1% |
| CPU | T2 (kHS06) | 1588 | 1877 | 18.2% |
| CPU | SUM (kHS06) | 3438 | 4152 | 20.8% |
| Disk | T0 (PB) | 32 | 40 | 25.0% |
| Disk | T1 (PB) | 116 | 136 | 17.5% |
| Disk | T2 (PB) | 142 | 167 | 17.4% |
| Disk | SUM (PB) | 290 | 343 | 18.3% |
| Tape | T0 (PB) | 120 | 174 | 45.0% |
| Tape | T1 (PB) | 272 | 353 | 29.8% |
| Tape | SUM (PB) | 392 | 527 | 34.4% |

*Table 6: Summary of the requests for computing resources in 2022.*

# Lost and founds

- **GRIF-IRFU**
  - after many years of good and loyal services,
    Frédéric Schaer has left GRIF duties at the beginning of the year
    Jean-Pierre Meyer has left / is leaving GRIF / CAF duties this year

**A HUGE thanks to both of them !**

- **LPSC**
  - not yet ! Foreseen for January 2023 ...
  - we started mid-October to remove the LPSC storage from the PandaQueues
  - it will let time to DDMOps to empty the DATA+SCRATCHDISK
    and move files elsewhere

# Conclusion

- **ADC in ATLAS and France**
  - CC-IN2P3 as a Tier-1 represents ~13% of ATLAS cpu and storage on grid T1s
  - French Tier-2s represent ~7% of cpu and 12% of storage of Tier-2s on grid
  - we are totally absent from HPC/cloud

  → France represents ~5 % of computing resources used by ATLAS

- **S&C in ATLAS, preparation of Run 3**
  - new analysis model will be used, in particular fewer/smaller DAODs
  - many recent challenges on data, network, tapes have been performed
  - still work to be done for full readiness for Run 3 but much already done !

- **Support / person power**
  - we have been suffering from a lack of operations support for a while
  - lack of commitment to ops and maintenance of R&D projects once they are established
  - the Run 3 physics program will suffer if the situation does not improve

- **Computing ATLAS France (CAF) WG**
  - next annual CAF-user meeting on 9th December
  - morning session: report from CAF, reports from each lab/group
  - afternoon session: discussions with S&C conveners + ICB chair

    software status, tape challenge