# ATLAS Computing Needs

So we can plan for the spending in the next CG can you answer the following questions for you area, if the question is not relevant for your area you can ignore it:

1. How many staff, PhD students and project students are expected to work on your project from 2026 to 2029? Same as now, more or fewer in each of the three categories.
   - We expect similar number of staff and students to now, but with a larger contribution to upgrade activities

2. What resources will they need? Same as now, more CPU/disk or are they OK with just an IT services provided laptop and an internet connection?
   - ATLAS make significant use of local HEP computing resources, both for the final stages of analysis and for upgrade. The analysis requirements are likely to increase by a factor of 2-3, commensurate with the increased luminosity in run-3 compared to run-2. On the upgrade side, the usage is likely to stay the same.

   a. What local CPU, GPU and disk space (but explicitly not local grid resources) will they require?
      - On the analysis side, we use local resource to run over ntuples and do the final stages of analysis, including making histograms, increasingly training ML, and statistical analysis
         - Disk: we currently have 60 TB on bundle, which is 90% full. Given that the ntuples for a single analysis can be up to 10 TB and the 2-3 factor increase in luminosity, then there will be a significant increased need for space. Also, /bundle is rather slow which makes training MLs slow and hence a large fast scratch disk (1TB) would useful for temporary data storage to make running MLs more efficient
         - CPU: Interactive nodes are fine but we could do with some more ALMA9 batch resources; if we convert the remaining Centos7 notes to ALMA9 this would be sufficient.
         - GPU: GPU resources are roughly sufficient for current ML needs but likely to increase
      - On the upgrade side we will need CPU/Disk to run over analysis ntuples and data for the commissioning of the pixel endcap detector
         - CPU: the current/local interactive machines are sufficient
         - Disk: will require a few (max 10) TB
      - 

   b. What grid resources will they use? What have the Grid allocated to the project?

- ATLAS has large grid resources, which is used for all central MC simulation and MC + Data reconstruction, along with central ATLAS production of derivations and analysis-specific skimming/slimming of these derivations to produce final analysis ntuples.

c. What IT services provided resources will be used? Including the 2TB of disk space allocated to any project on request.
  - The 2TB of disk space per project from the central IT is not useful for analysis since, as far as we are aware, there is no way to access this from the HEP nodes for analysis. However, we could and should investigate if this can be used for the upgrade, which currently uses the HEP-provided WISP (20 GB per account) for backups of shared file storage for several cleanroom systems (pull test, smartscope, Canon images, Keyance images).

3. What production or other tasks that generate large amounts of local data that need to be stored either for a few days or a few years. What if any offsite replication strategy is planned?
   - On the analysis side, ntuples are generally replicated on the grid and/or eos too so the only unique data is final histograms, ML models and statistical results.
   - On the upgrade side the pixel endcap build will produce several (max 10) TB of data as outlined above. This will be the only copy short-to-medium term (months) while it is analysed. Summary information will then be uploaded to the central ATLAS production DB but it will remain the only copy of the more detailed lower-level information that could be needed for analysis of issues in the future.
   - On the upgrade strip side, some space will be needed for module electrical test data during production,, but here the disks on the dedicated local machines we currently have are sufficient and it doesn't require backup as it will be uploaded to the central ATLAS production database and can be relatively easily and quickly reproduced if needed
   - The shared files for clean room data mentioned above are stored on local storage with a backup on WISP.

4. What software other than the "LCG" stack do you need? Specifically any dedicated mechanical design, electrical circuit design, 2D and 3D CAD, TCAD etc.
   - On the analysis side there is an increasing need for up-to-date python ML tools, which tend not to be in LCG views in a timely manner.
     - Since we rely more and more on conda + pip + venv for installation, we end up with large installations of python+packages on some user owned storage. These can be 10s of GB per user in many cases and end up taking a lot of space on /user or /hepstore
   - On the upgrade side we require the following, which are also used by non-ATLAS projects:
     - Mechanical design: Creo Parametric (University License Server)
     - Electrical design: Altium (10 seat licence purchased and shared with nuclear structure (5 each); managed by Ashley)

- - Cleanroom: dedicated software to operate equipment
    - Smartscope
    - Pull Tester
    - Keyence Microscope

a. What resources in terms of computers, disk space, databases and licences servers are needed to support this?
   - The drawing office is in need of a dedicated server for long term archiving engineering models which is backed up.
     - Not sure under whose remit this falls but mentioning it here for completeness.
   - Several activities, such as Ansays (Analsis sw for FEA and mechanical engineering analysis) and Microsoft project require higher-spec individual desktops/laptops e.g. with significant RAM (32 GB) in the former case and with large monitors in the later case.
   - Some cleanroom equipment requires specialised machines to operate e.g. older, unsupported windows OSs and specified motherboards to run YARR DAQ
     - TODO: further details can be found and provided as needed
   - See above for details on licences

b. Are there issues with current provision? I heard that the 3D models of the ATLAS inner tracker have been problematic to open.
   - TODO: Helen investigating

5. Do you need public facing websites or other data repositories that cannot be managed by IT Services? Do you have anyone to design, implement and maintain them? Either production databases or other resources, including outreach.
   - On the analysis side, many people use the www areas for viewing results efficiently sharing them
   - On the upgrade side there is a production database maintained by ATLAS.
     - Strip: Require local db and tools that are maintained locally; some systems require grafana/influxDB access for machines within cleanroom/university network

6. Do you expect to invite external users to use our local resources? For example, people visiting to help with detector production that need access to Liverpool resources.
   - Yes, for the upgrade we need to give access to several people who visit regularly for ITk production.
     - In this regard, Ling Meng (Lancaster) and Graham Miller (Manchester) already have HEP accounts.

7. What amount of local data is the only copy, what fraction could be rebuilt from external repositories? What problems would occur if the data was lost due to a hardware failure?
   - As mentioned above, on that analysis side ntuples are replicated on the grid and could be re-downloaed and used to remake final results such as

histograms.  However, this could lead to 1-2 week delay which could mean missing conference deadlines etc

- For the pixel endcap upgrade, as mentioned above, there will be data with the only copy at Liverpool short-to-medium term and remaining the only copy of lower level information long term.
  - If the data is lost in the short-to-medium term before upload to the central production DB we would need to redo all the associated QC analysis, which could take weeks.
  - If the longer-term low-level data is lost it would not be possible to redo the analysis and we would lose the ability to do detailed investigations into any issues/problems that arise.
- For the engineering drawings mentioned above, they can in principle be pieced back together from the specifications on CERN's EDMS but this would take a long time and we would lose the revision management.
  - Hence the requirement of a dedicated backed up server for the drawing office listed above.

8. If a service such as the scratch server, local indico licence server or gitlab (to give some examples) were to be offline for an extended period (weeks) would that prevent significant amounts of work being completed? What, if any, disaster plans are in place or will be in place?
   - Indico/gitlab wouldn't be a big issue as mostly use the CERN one
   - For analysis, a disk server would be a big issue, particularly at certain times e.g. build up to conferences as times to reproduce results / recopy inputs would be too long
   - For the upgrade, HEP user authentication server downtime would prevent usage of all linux based test systems.  Similarly loss of the pixel endcap data would would require the data to be retaken and the QC analysis redone.  In the worse case, either of these could case direct delays to the international project.
   - If machines that host the licences above (e.g. in drawing office and clean room)  go down, this would be a major issue and prevent all work in related areas.
     - It would be helpful to investigate the best way to have redundancy in such systems.
   - Our FTAG institutional commitment relies on scratch to store intermediate ntuples (produced on the grid) for further local skimming (since they are too large to go on bundle) until that further reduced files are checked.  Running on this is rather quick but often needs several iterations, hence being run locally.  Without scratch or other large temporary storage this would not be possible in the current workflow.  If the intermediate ntuples are lost they can be redownloaded from the grid but this would cause a delay in producing central ATLAS FTAG recommendations, especially if scratch is offline for a significant period.