

IRIS Use CASE: LZ Experiment

D. Bauer, S. Fayer, A. Vacheret, D. Colling on behalf of the LZ collaboration

November 2, 2018

1 Administrative Details

Organisation: LuxZeplin (LZ)

Website: <http://lz.ac.uk/>

Contact for this proposal:

Daniela Bauer, Blackett Lab, Imperial College, daniela.bauer@imperial.ac.uk.

Please *cc* all correspondence to David Colling d.colling@imperial.ac.uk and Antonin Vacheret antonin.vacheret@imperial.ac.uk.

2 Summary of Approved Science Programme

LZ (LUX-Zeplin) is a Dark Matter experiment based at the Sanford Underground Research Facility in South Dakota/USA. It is currently under construction and aims to start data taking in 2020 and continue until 2024. At present nine UK institutes are members of the LZ collaboration [1]. Once commissioned the LZ detector will be the foremost detector to search for WIMP dark matter scattering via spin-independent interactions, offering an an order of magnitude improvement with respect to the expected sensitivities of currently running LXe dark matter experiments. In addition the ultra-low background and unprecedented scale of LZ mean it will be capable of measuring a variety of other rare phenomena such as: other WIMP interactions, including spin dependent interactions, effective field theory operators, and inelastic interactions; keV-scale axions in the electron-recoil channel through the axio-electric effect; measurement of the solar pp neutrino flux; a first observation of coherent scattering of solar $8B$ neutrinos with Xe nuclei; exotic nuclear decays of Xe isotopes, including a search for neutrinoless double beta decay of ^{136}Xe and double electron capture of ^{124}Xe ; and real-time detection of nearby core-collapse supernovae within our galaxy.

3 Report on the use made of IRIS resources in the previous year

In 2018 all UK based computing resources for LZ were provided by GridPP. No IRIS resources were used in 2018.

4 The LZ Computing Model

An overview of the LZ computing model is shown in Fig. 1. Its defining feature is the two data centres, one in the USA (USDC) and one in the UK (UKDC), both holding a complete copy of the experiment's data. During stable periods of running both data centres plan to concentrate on different aspects of reconstruction and analysis. However, both data centres are expected to be able to handle all aspects of data processing, including user analysis, to provide a fail-over should one of the data centres be temporarily unavailable. Unlike the USDC where both storage and processing facilities are located at NERSC, the UKDC uses a distributed approach based on GridPP-provided computing resources.

For the UKDC, the data themselves will be hosted at the Tier-2 at Imperial College London, while the processing and analysis will take place at various UK Tier-2s and possibly other collaborating European institutes. Data access is provided via SRM (GridFTP) and XRootD.

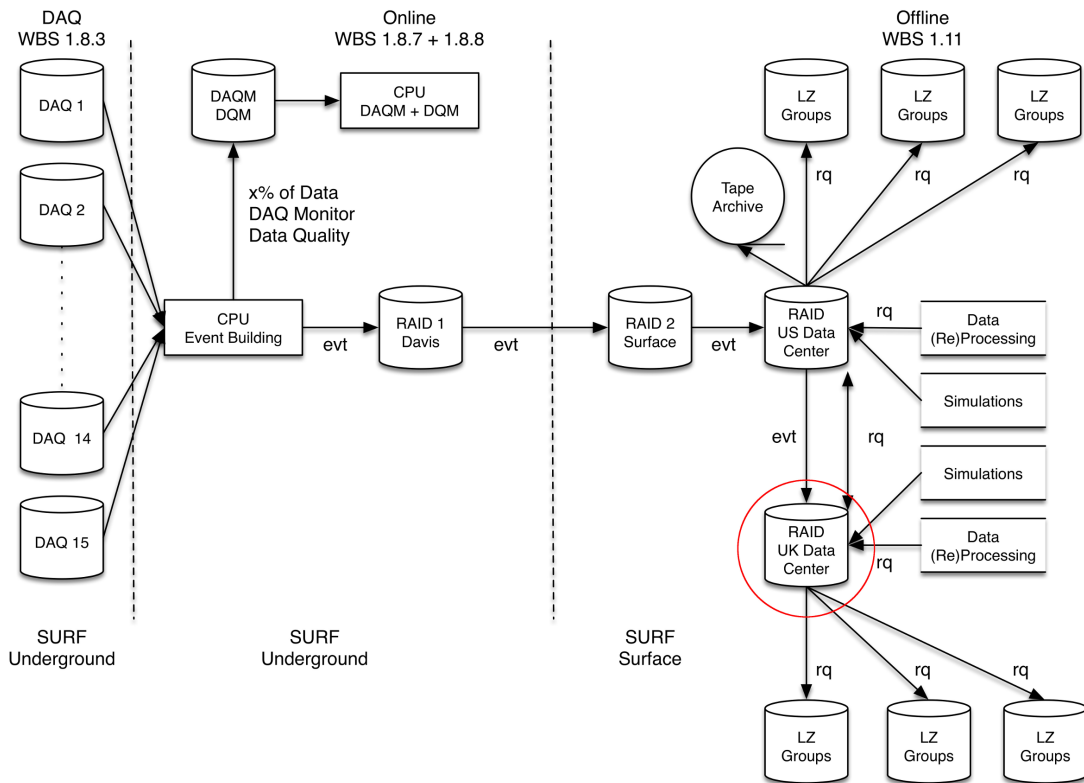


Figure 1: The LZ computing model [2]. The UKDC is circled in red.

4.1 Current technical implementation

4.1.1 Compute

LZ's main computing activity is currently Monte Carlo Generation and analysis. All LZ compute jobs are single core, though they will occasionally use multi-core slots to get around limitations in memory provision on GridPP resources. LZ has been assessing using machine learning/GPUs for background suppression.

4.1.2 Access to computing and storage

LZ uses the GridPP DIRAC instance [3] to access GridPP resources. DIRAC is build on standard grid middleware (ARC-CE, CREAM-CE, HTCondorCE) and uses X.509/voms for authentication and authorisation. LZ Monte Carlo data is currently moved between the data centres via GridFTP and access for reconstruction and analysis is via XRootD.

Making use of DIRAC's Python API, LZ has built a custom Job Submission Interface (JSI) website, which is currently being used for their Monte Carlo production and reconstruction workflows. The JSI provides additional features beyond those provided by DIRAC such as referencing the LZ software and code repositories (CVMFS, gitlab) to ensure that production requests conform to LZ software constraints. It also provides progress monitoring for the various LZ production workflows. A detailed description of the JSI within the context of the UKDC can be found in [5].

4.2 Data Management

LZ relies on its two data centres holding a complete copy of their data. The data is transferred between the two data centres using SPADE [6] which is currently under development to include globus tools for the actual file transfers. The central data catalogue will be hosted at NERSC with a local copy located at Imperial College. All data is accessible to any LZ member via SRM (GridFTP) and XRootD.

4.3 Software

LZ distributes its software via CVMFS and its configuration files via its own git repository (<https://lz-git.ua.edu/>). It also uses standard software packages, especially ROOT and XRootD from non-LZ CVMS repositories. LZ software currently uses CentOS6 as its base release and is in the process of transitioning to CentOS7. This transition is expected to be complete by the time IRIS resources become available.

LZ has been using CernVM [7] based containers (using Singularity) at their Imperial College site for the past year. This should enable LZ to make a straight forward transition to cloud resources.

5 Estimate of the Resource Request

5.1 Provide a firm estimate of the Resources you already have access to in 2019

An overview of the predicted CPU and storage usage from the LZ TDR [2] can be found in Table 1. While considerable effort has gone into providing realistic requirements we would assume a revision of all forecasts once data taking has started. It should be noted that these numbers indicate CPU usage averaged over a year; peak demand can substantially vary from this. Please see Figs. 2a and 2b for numbers gleaned during the LZ Mock Data Challenges in 2017 and 2018. For the UK, the compute and storage requirements are currently being met by GridPP. It should be noted that LZ UK has no influence on allocations in the United States.

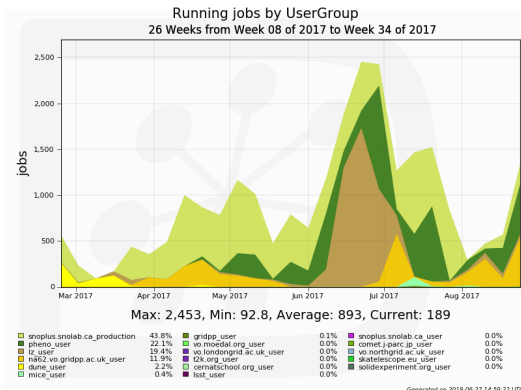
Financial year	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Raw data	–	–	–	–	–	560	1680	2800	3920	5040	6160
Calibration data	–	–	–	–	–	160	480	800	1120	1440	1760
Simulation data	40	80	80	100	100	200	200	200	200	200	200
Processed data	20	40	40	50	50	172	316	460	604	748	892
User data	20	40	40	50	50	55	134	213	292	371	451
Total data	80	160	160	200	200	1147	2810	4473	6136	7799	9463
USDC: Disk space	40	220	220	220	220	1360	3360	5360	7360	9360	11360
USDC: CPU cores	–	–	175	350	350	390	830	1270	1710	2150	2590
UKDC: Disk space	150	220	220	270	650	1597	3260	4923	6586	8249	9913
UKDC: CPU cores	150	175	350	350	350	390	830	1270	1710	2150	2590

Table 1: Planned storage (in TB) and processing power by US fiscal year at the US and UK data centres [2].

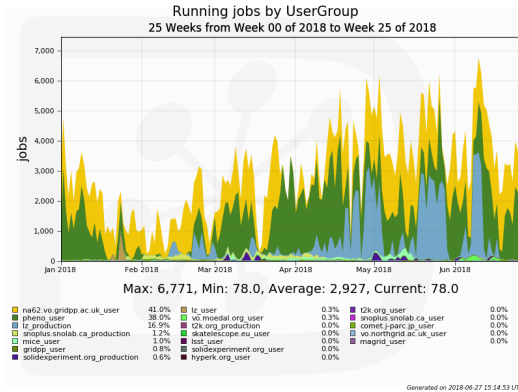
5.2 Resources requested from IRIS

LZ requests the following resources from IRIS for the years 2019 and 2020:

- 300 CPUs with 12 GiB/core. No further predictions for the years after this can be made at this time, though it is to be expected that LZ will need access to high memory resources in the years up to 2025 as well.
- LZ requests 8 GPUs for 2019. As there are no provisions for GPU resources within the GridPP supported computing for LZ it is likely that LZ will need a similar amount of GPUs up until 2025.
- LZ request 300 TB of storage, accessible via XRootD for 2019 and 2020. While the actual size of the raw data for obvious reasons is not yet known, it can be assumed that GridPP will meet the storage requirements for the experiment’s data. It is likely that that LZ will request further storage from IRIS from 2021 onwards for specific projects. It is unlikely that these requests will exceed 0.5 Pb/year.



(a) MDC 2017: LZ jobs are marked in light brown. Running jobs from other VOs currently falling under GridPP’s ‘non-LHC VO’ remit are also shown. All the VOs at least nominally shared the same quota, leading to contention of resources if two productions runs coincide.



(b) MDC 2018: LZ jobs are marked in light blue. At this point LZ had a finished setting up their production infrastructure, hence the jobs submitted through the JSI are marked ‘LZ production’. Note the different scale of the y-axis when compared to Fig.2a.

Figure 2: Number of running jobs during the LZ Mock Data Challenges (MDCs) in 2017 and 2018.

- LZ does not request any tape storage.

6 Requested Enabling Infrastructure

LZ would prefer to continue to use the GridPP DIRAC instance for submission to IRIS resources. LZ and GridPP have personnel in place to implement an GridPP DIRAC - IRIS cloud interface as this can be potentially used by other experiments that GridPP supports. In addition LZ workflows need access to storage via XrootD and the appropriate bandwidth if the analysis facilities are not co-located with the storage. (Actual numbers to follow, but so far we had to make no special provisions within GridPP, so this is unlikely to be an issue.)

In case of cloud resources LZ is able to use a generic CernVM based image or will be able to supply their own image(s) if needed.

7 What will the Requested IRIS Resources be used for?

LZ requests IRIS resources for three distinct cases: Large MC events, background suppression studies and detector commissioning.

- CPUs with 12GiB/core: Monte Carlo production for events with a large number of photons. Due to the high memory requirements of these events GridPP resources are not well suited.
- GPUs to further develop background suppression software
- CPUs with 12GiB/core: Commissioning data reconstruction and analysis. It is expected that the initial data will contain large amounts of noise, requiring more memory per core

than offered by the standard GridPP resources.

- Storage (300 TB): To be able to reprocess commissioning data on short time scales.

References

- [1] *LZ UK Collaboration* <http://lz.ac.uk> [accessed: 2018-08-15]
- [2] LZ Collaboration, *LZ Technical Design Report*, arXiv:1703.09144 (physics.ins-det)
- [3] D. Bauer, S. Fayer, *GridPP DIRAC: Supporting non-LHC VOs on LHC centric resources*, <http://iopscience.iop.org/article/10.1088/1742-6596/898/5/052003>, Journal of Physics: Conference Series, Volume 898, 2017
- [4] Rob Taylor, http://teacher.pas.rochester.edu:8080/wiki/pub/Lux/AWGworkshop180422_Agenda/Multiple_Scatter_Discrimination_in_Z_1.pdf, (unpublished, available on request, apologies)
- [5] D. Bauer, D. Colling, S. Fayer, E. Korolkova, A. Richards, A. Vacheret *The LZ UK Data Centre*
- [6] SPADE [software] <http://nest.lbl.gov/projects/spade/html/index.html> [accessed 2018-09-06] <https://indico.cern.ch/event/587955/contributions/2937236/>, CHEP 2018
- [7] CernVM [software] <https://cernvm.cern.ch/> [accessed 2019-11-02]