PROCEEDINGS OF SCIENCE

The new “Gauge Connection” at NERSC

Massimo Di Pierro[*]
School of Computing - DePaul University - Chicago
E-mail: mdipierro@cs.depaul.edu

James Hetrick
University of the Pacific
E-mail: jhetrick@pacific.edu

Shreyas Cholia
Lawrence Berkely Laboratory
E-mail: scholia@lbl.gov

James Simone
Fermi National Accelerator Laboratory
E-mail: simone@fnal.gov

Carleton DeTar
University of Utah
E-mail: detar@physics.utah.edu

In this paper we present a new web interface (NERSC3) for the Gauge Connection, one of the largest repositories of lattice gauge configurations in the US. The Gauge Connection is hosted at the National Energy Research Scientific Computing Center (NERSC), and stores more than 16TB of data using the High Performance Storage System (HPSS). The new interface allows users to search, download, and view statistics on lattice QCD ensembles hosted at NERSC. It also aggregates metadata from other major lattice archives (ILDG), and allows one to search the metadata in a single place. Additionally, users can tag, annotate, and bookmark the ensembles (http://qcd.nersc.gov).

31st International Symposium on Lattice Field Theory - LATTICE 2013
July 29 - August 3, 2013
Mainz, Germany

PoS(LATTICE 2013)427

[*] Speaker.

1. Introduction

The Gauge Connection is one of the most popular repositories of lattice QCD ensembles. It has been operated by the National Energy Research Scientific Computing Center [1] (NERSC) since 1998, and currently stores over 16 Terabytes of data. The popularity of the Gauge Connection can be attributed to the amount of data it stores, and its easy-to-use web interface which lowers the barrier of entry for lattice QCD researchers.

At the core of the Gauge Connection architecture is the High Performance Storage System [2] (HPSS) data archive. The HPSS system provides a tape archival storage library for the QCD data. The content of the archive is exposed using an FTP interface in conjunction with a CGI-based HTTP download service.

The original web interface consisted of static HTML files linked against this CGI service that would pull data from the HPSS archive. In 2011 the web interface underwent a major revision and the static pages were replaced by dynamically generated pages. The content of the pages comprised of free text mined from the original HTML files, which was made editable in a wiki-like interface. The pages were augmented with metadata obtained from the file catalog itself and stored in a database. The new interface allowed users to search the data more easily, comment on the data, and download ensembles in batch.

In this paper we present a new version of this service which we will refer to as NERSC3. Apart from user interface changes, the main functional change is an expanded search capability which allows users to search both local and remote repositories of ensembles.

Lattice QCD physicists formed the International Lattice Data Grid (ILDG) [3] collaboration in 2000, and developed protocols and infrastructures to annotate and share lattice QCD ensembles. They created a file format to store gauge ensembles along with standardized metadata, and a network of ILDG metadata catalogs and file catalogs. This enabled searching against the ensembles and their metadata, and the ability to download the files.

The ILDG infrastructure is powerful but compared to the NERSC archive, it presents two major barriers of entry: 1) By design, the metadata catalogs do not inter-operate but provide a common search API; therefore search is decentralized. 2) The file catalogs are built on top of GridFTP [5] and can only be accessed using ILDG-tools, and only after obtaining a valid X509 grid certificate.

In our latest revision of the NERSC archive we have been addressing the first issue. The new system periodically connects to all of the ILDG metadata catalogs and copies the ensemble metadata to a local database. This allows a visitor of the site to search for both local ensembles and remote ILDG ensembles in one place. While the local data can be retrieved directly from the web site or via the provided batch download script, remote data is only referenced, and the user must use the ILDG tools to query the remote file catalog and use GridFTP to download the data.

Moreover, conversion to the ILDG data format has been slow: the bulk of the data is stored in other formats, e.g. NERSC or MILC while only a fraction is stored in ILDG format. Additionally, one of the key improvements to NERSC3 includes the ability to convert file formats post-download.

We have designed the new website to provide additional capabilities beyond data access, including wiki capabilities to annotate the data, link external work derived from the data (derived

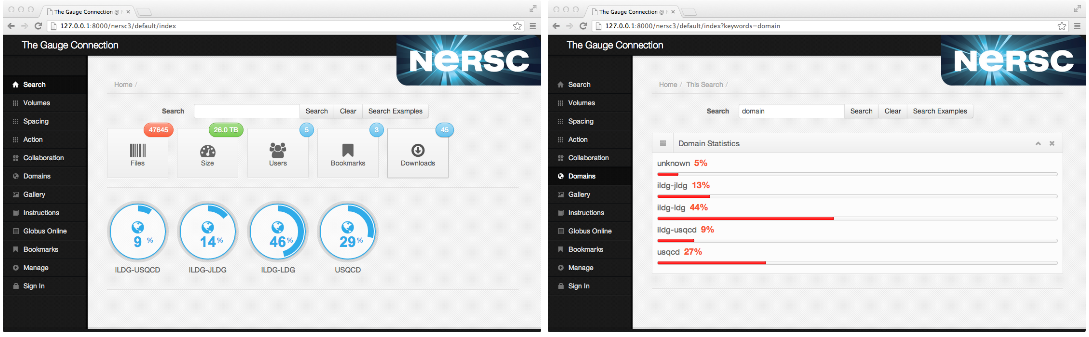


Figure 1: The main web page for `qcd.nersc.gov` allows browsing and searching for gauge ensembles (left) and visualizing statistical data about the ensembles (right).

data, software, tutorials, and publications), and bookmark interesting data. The new system keeps usage statistics for analytics purposes.

Figure 1 and figure 2 show some screenshots from NERSC3.

2. Architecture Overview

The Gauge Connection is hosted at NERSC on the High Performance Storage System (HPSS). HPSS is a hierarchical tape storage system designed to store multiple petabytes which be accessed using a custom command line client (HSI), a C API, an FTP interface, a parallel FTP interface, as well as a GridFTP interface. The latter can also be accessed using Globus Online, a service that enables scheduled data transfers between GridFTP endpoints.

NERSC3 periodically performs directory dumps of the HPSS tapes and reproduces the folder structure in the database. NERSC3 has no direct access to the files because staging them from the tape archive is expensive. Yet it can infer some information about the content of the files from the names of the folders, the names of the files, and their sizes. Most of the files stored at NERSC are produced by the MILC collaboration which is one of the major sources of public gauge ensembles in the world. The MILC collaboration adopts a practical naming convention for the file names which include information about the lattice volume, the kappa value, and the quark masses (for dynamical ensembles) and a progressive configuration counter. NERSC3 can link this information to additional metadata obtained from other sources which is inserted in the system manually e.g. links to papers describing the algorithm used to generate each ensemble.

The NERSC3 database stores information about two types of ensembles: the ensembles stored at NERSC described above and ILDG ensembles stored at other locations.

The ILDG infrastructure consists of the following components:

- A description of metadata which accompanies each ensemble. This includes name of the group/collaboration, description of the action and algorithms used to generate it, simulation parameters (lattice volume, lattice spacing, dynamical quark masses) and links to relevant papers. The information is encoded in a custom XML-based schema.

- A file format for storing data along with metadata. ILDG uses a custom encoding protocol similar to TAR, called LIME.
- A metadata catalog service which exposes SOAP services to query the catalog for local ensembles, their metadata, and their file content.
- A file catalog service which exposes additional SOAP services, protected by WS-Security, to convert logical filenames into actual URLs for downloading the files.

ILDG is comprised of 5 realms: CSSM (Australia), LDG (Europe), JLQCD (Japan), UKQCD (UK), USQCD (US). Each realm runs at least one metadata catalog service and one file catalog service.

In ILDG each ensemble is identified by a Unique Resource Identifier (URI) starting with the `mc:` prefix. Each logical file name is identified by a URI starting with the `lfn:` prefix. The actual file names are not URIs but they are URLs. They usually have one of the following prefixes: `http:`, `ftp:`, `https:`, `ftps:`, and `srm:`. `srm:` is a GridFTP based web service protocol which operates over `HTTP` but is designed to act as a referrer for multiple endpoints and effectively load-balances transfers over multiple nodes, allowing transfers to scale. It is very similar in scope to the BitTorrent protocol, i.e. it splits data transfers over multiple connections. The SRM protocol is supported by the *dCache* [7] tape storage system used at Fermilab.

The metadata catalog services are public. NERSC3 periodically connects to these services and downloads a complete list of available ensembles, their metadata, and their logical filenames. This information is stored in a local database. In this way, when using the NERSC3 web interface one can search both the local ensembles and all the ensembles made available via ILDG.

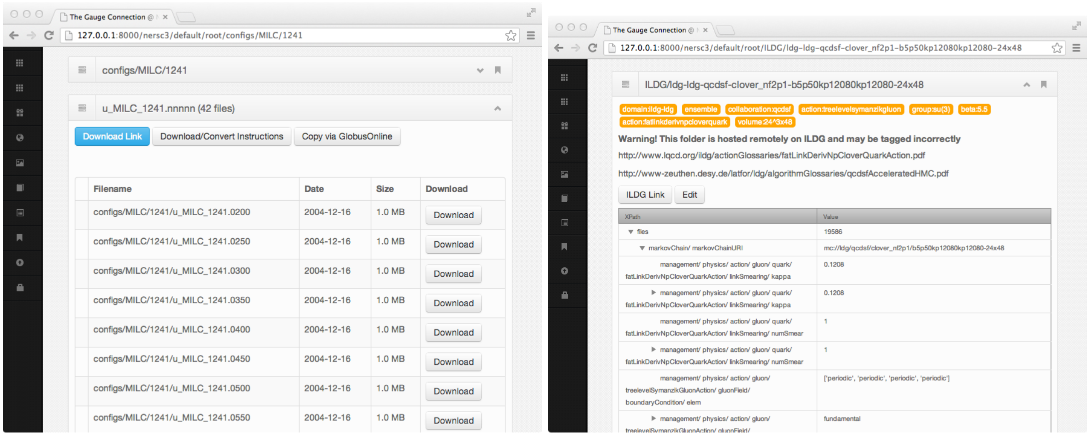
The ILDG file catalogs are not public. This means that the conversion between logical filename to actual URLs cannot be performed without a valid GridFTP certificate, and each realm has its own policies and procedures for assigning certificates. Accessing the actual URLs also requires a certificate. Thus, while NERSC3 allows for searching of remote data and can point the user to a specific remote ensemble, it does not help the user with accessing the data itself.

Only the data stored at NERSC is truly public in the sense that it can be accessed by anybody without explicit permission from an ILDG certificate authority.

Figure 3 shows an overview of the architecture described here.

In NERSC3 tags are associated with ensembles, rather than individual files. Each tag is divided into two parts separated by a colon (`:`). e.g. `volume:24^3x48`. The first part of the tag (in this example `volume`) is a key, while the second part (in this example `24^3x48`) is a value. Tags are all optional and key names are not dictated by the system but chosen by the users. Some keys may not have a corresponding value. More examples of tags include `beta:2.1`, `collaboration:milc`, and `flavor:2+1`. When clicking on a tag, the user is presented with a list of ensembles sharing the same tag.

Users can search by key (any key) and they are presented with a list of all possible values for that key along with statistical information. They can search by tag (key:value) and they are presented with a list of ensembles with the corresponding tag. They can search for multiple tags and they are presented with a list of ensembles that have all the requested tags. Here are some examples of search strings:



127.0.0.1:8000/nersc3/default/root/configs/MILC/1241

configs/MILC/1241

u_MILC_1241.nnnnn (42 files)

Filename	Date	Size	Download
configs/MILC/1241/u_MILC_1241.0200	2004-12-16	1.0 MB	Download
configs/MILC/1241/u_MILC_1241.0250	2004-12-16	1.0 MB	Download
configs/MILC/1241/u_MILC_1241.0300	2004-12-16	1.0 MB	Download
configs/MILC/1241/u_MILC_1241.0350	2004-12-16	1.0 MB	Download
configs/MILC/1241/u_MILC_1241.0400	2004-12-16	1.0 MB	Download
configs/MILC/1241/u_MILC_1241.0450	2004-12-16	1.0 MB	Download
configs/MILC/1241/u_MILC_1241.0500	2004-12-16	1.0 MB	Download
configs/MILC/1241/u_MILC_1241.0550	2004-12-16	1.0 MB	Download

127.0.0.1:8000/nersc3/default/root/ILDG/ldg-ldg-qcdsf-clover_nf2p1-b5p50kp12080kp12080-24x48

ILDG/ldg-ldg-qcdsf-clover_nf2p1-b5p50kp12080kp12080-24x48

domain:ildg-ldg ensemble collaboration:qcdsf action:treelevelsymanzikgluon group:su(3) beta:5.5 action:fatlinkderivnpcloverquark volume:24^3x48

Warning! This folder is hosted remotely on ILDG and may be tagged incorrectly

http://www.lqcd.org/ildg/actionGlossaries/fatLinkDerivNpCloverQuarkAction.pdf

http://www-zeuthen.desy.de/latfor/ldg/algorithmGlossaries/qcdsfAcceleratedHMC.pdf

XPath	Value
▼ files	19586
▼ markovChain/ markovChainURI	mc://ldg/qcdsf/clover_nf2p1/b5p50kp12080kp12080-24x48
management/ physics/ action/ gluon/ quark/ fatLinkDerivNpCloverQuarkAction/ linkSmearing/ kappa	0.1208
▶ management/ physics/ action/ gluon/ quark/ fatLinkDerivNpCloverQuarkAction/ linkSmearing/ kappa	0.1208
management/ physics/ action/ gluon/ quark/ fatLinkDerivNpCloverQuarkAction/ linkSmearing/ numSmear	1
▶ management/ physics/ action/ gluon/ quark/ fatLinkDerivNpCloverQuarkAction/ linkSmearing/ numSmear	1
management/ physics/ action/ gluon/ treelevelSymanzikGluonAction/ gluonField/ boundaryCondition/ elem	['periodic', 'periodic', 'periodic', 'periodic']
▶ management/ physics/ action/ gluon/ treelevelSymanzikGluonAction/ gluonField/	fundamental

Figure 2: For each local ensemble, the NERSC3 interface can list the ensemble files (left). For remote ensembles, the NERSC3 interface can list the ILDG metadata and the FLN URI.

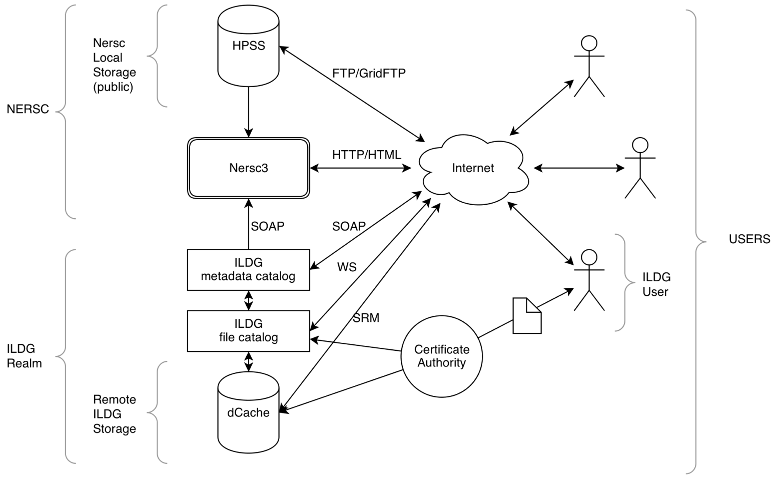


Figure 3: Overview of the NERSC3 and ILDG architecture.

```
volume
volume:12^3x12
volume is 12^3x12
volume is 12^3x12 and domain is usqcd and collaboration is milc
```

In this case the domain "usqcd" refers to the set of ensembles that are hosted at NERSC as opposed to the ILDG ensembles.

The search string understands the "is" and "and" english words. A search string can also be the name of an ensemble.

Figure 4 shows the database architecture of the NERSC3 application. It includes tables to store users, groups, memberships and permissions. We implement role based access control for administrators. It includes a table `catalog_folder` (which represents an ensemble) and a table `catalog_file` which stores links to individual files (the files themselves are in NERSC HPSS or ILDG). `ildg` metedata, `accesslogs`, `bookmarks`, and `tags` reference the `catalog_folder`

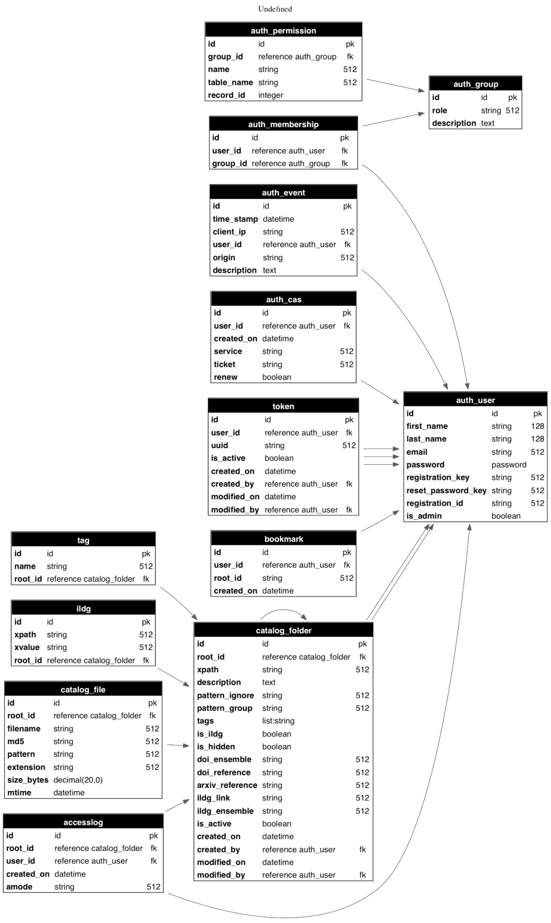


Figure 4: Overview of the table structure of NERSC3. The table `catalog_folder` represents ensembles and the table `catalog_file` represents files in the ensembles.

table. Logged in users are also issued `tokens` which are used to authenticate users for batch downloads. NERSC3 is based on the web2py [8] framework.

3. The batch download script

NERSC3 allows users to download local ensemble files individually or in batch. Individual downloads can be performed directly from the web pages. For batch downloads, the web page of each local ensemble provides a link to a JSON file containing a list of files in the ensemble. A user does not need to open this file, but only needs to copy its URL, and can then use the `qcdutils_get` utility from `qcdutils` [9] to download all files referenced by the link. For example if the link is `http://qcd.nersc.gov/nersc/api/files/demo` then the user would download the files with the following command:

```
> qcdutils_get.py http://qcd.nersc.gov/nersc/api/files/demo
http://qcd.nersc.org/nersc/api/files/demo
target folder: demo
total files to download: 1
```

```
downloading demo.nersc
demo.nersc 100% |################| Time: 00:00:00 654.52 K/s
completed download: 1/1
```

The `qcdutils_get` script will download the `demo` file, read its contents, and sequentially download all the files listed in there. It shows download progress and keeps a record of completed downloads. If restarted, it resumes from the last completed download.

It also has the ability to convert ensembles to different formats. For example, here we ask to convert the most recent downloads to single precision IDLG format.

```
> qcdutils_get.py --convert ildg --float demo/demo.nersc
converting: demo/demo.nersc -> demo/demo.nersc.ildg
  (precision: f, size: 4x8x8x8)
100% |#####################################|
```

It keeps a log of all completed operations to avoid duplication of work. One can query `qcdutils` for a log of the completed tasks. For example:

```
> qcdutils_get.py demo/qcdutils.catalog.db
demo.nersc created on 2011-06-17T13:42:30.876812
   [14e7cf9106bfcc16388aeac285ccdad9]
```

We refer to the `qcdutils` [9] code and manual for details.

4. Conclusions and Outlook

In this paper we have presented a new interface to the Lattice QCD Gauge Connection at NERSC. The purpose of the new interface is to lower the barrier of entry for lattice QCD researchers, and to allow users of the archive to search both local and remote ensembles of lattice gauge configurations. It also allows scientists to access metadata about the ensembles, and to contribute to the metadata by editing wiki entries associated to them.

Acknowledgments

This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

[1] `http://qcd.nersc.gov`

[2] `http://www.hpss-collaboration.com`

[3] Maynard, Chris M., arXiv preprint arXiv:1001.5207 (2010).

[4] Beckett, Mark G. *et al.*, Comp.Phys.Comm. 182 (2011) pp.1208-1214

[5] Yildirim, Esma *et al.*,High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:. IEEE, 2012.

[6] Di Pierro, Massimo, Nucl. Phys. B-Proceedings Supplements 106 (2002): 1034-1036.

[7] Agarwal, A., et al. Journal of Physics: Conference Series. Vol. 219. No. 7. IOP Publishing, 2010.

[8] Di Pierro, Massimo, Computing in Science and Engineering 13.2 (2011): 64-69.

[9] Di Pierro, Massimo, arXiv preprint arXiv:1202.4813 (2012).