Contents lists available at ScienceDirect

ELSEVIER

Computer Physics Communications



www.elsevier.com/locate/cpc

Building the International Lattice Data Grid

Mark G. Beckett^a, Paul Coddington^b, Bálint Joó^c, Chris M. Maynard^{a,*}, Dirk Pleiter^d, Osamu Tatebe^e, Tomoteru Yoshie^e

^a The University of Edinburgh, Edinburgh, United Kingdom

^b e-Research SA and School of Computer Science, University of Adelaide, Adelaide SA 5005, Australia

^c Scientific Computing Group, Jefferson Lab, 12000 Jefferson Avenue, Newport News, VA 23606, USA

^d Deutsches Elektronen-Synchrotron DESY, 15738 Zeuthen, Germany

^e Center for Computational Sciences, University of Tsukuba, Tsukuba 305-8577, Japan

ARTICLE INFO

Article history: Received 4 June 2010 Received in revised form 17 January 2011 Accepted 31 January 2011 Available online 15 February 2011

Keywords: ILDG Data grids Lattice QCD

ABSTRACT

We present the International Lattice Data Grid (ILDG), a loosely federated grid-of-grids for sharing data from Lattice Quantum Chromodynamics (LQCD) simulations. The ILDG comprises of metadata, file-format and web-service standards, which can be used to wrap regional data-grid interfaces, allowing seamless access to catalogues and data in a diverse set of collaborating regional grids. We discuss the technological underpinnings of the ILDG, primarily the metadata and the middleware, and offer a critique of its various aspects with the hindsight of the design work and the two years of production.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

In this paper, we present the International Lattice Data Grid (ILDG) project, a mostly volunteer effort within the Lattice Quantum Chromodynamics (LQCD) community to share data worldwide and to thus amortise the very high computational cost of producing the data. In terms of organisation it is a data-grid, but it is also a loosely federated grid-of-grids.

Large data sets require significant scientific endeavour to amass them. In the case of LQCD, the resources are both intellectual such as the scientific ideas and algorithmic development — as well as other resources, such as the manpower required to write the computer code and the resources to procure/develop and operate a large supercomputer. Why then do scientists wish to share this valuable data? It is precisely because this data is so valuable that scientists make it available for others to use. A mechanism is required whereby those who generate shared data can receive credit for so doing.

For the LQCD community there are two compelling reasons to share data. First, fully exploiting the data requires computing and manpower resource. A particular group may generate a data set to compute a target physical quantity with sufficient precision to

* Corresponding author.

E-mail addresses: george.beckett@ed.ac.uk (M.G. Beckett), paul.coddington@adelaide.edu.au (P. Coddington), bjoo@jlab.org (B. Joó), c.maynard@ed.ac.uk (C.M. Maynard), dirk.pleiter@desy.de (D. Pleiter), tatebe@cs.tsukuba.ac.jp (O. Tatebe), yoshie@ccs.tsukuba.ac.jp (T. Yoshie). have an impact on experimental results, and yet not have sufficient resources or even the expertise to calculate many other possible quantities on that data set. At this stage, rather than waste some of the scientific potential of the data, a group may give the data away freely, provided some basic use conditions are met such as citing a certain paper in any resulting publication. Second, the resources required to generate more scientifically valuable data sets require ever greater resources, outstripping the computational gains of Moore's Law. This forces different groups to collaborate, to jointly bear the cost of data generation.

Quantum Chromodynamics (QCD) is a theory of sub-atomic particles (specifically, quarks and gluons) and their interactions. Lattice QCD (LQCD) is a version of QCD where space-time is discretized, making the theory amenable to calculation by computer. LQCD computations are of utility in a variety of theoretical particle physics contexts including Nuclear Physics and High Energy Particle Physics, and have historically consumed a large fraction of available computing cycles worldwide. The interested reader can find several excellent books and review articles on LQCD in the literature, for example [1,2] and [3].

LQCD computations are based on Markov Chain Monte Carlo methods (see [4] for a review) and typically the primary data from such calculations are samples of the QCD vacuum known as *gauge configurations*. The Monte Carlo process will generate an *ensemble* of such configurations for each set of physical and algorithmic input parameters. At the time of writing, the typical cost of generating an ensemble is O(1)-O(10) Teraflop years depending on the precise formulation employed, and this cost is expected to grow to

^{0010-4655/\$ -} see front matter © 2011 Elsevier B.V. All rights reserved. doi:10.1016/j.cpc.2011.01.027

the Exaflop-year scale as one simulates lattices with finer lattice spacings, larger physical volumes and physically light quarks.

A set of ensembles is amenable to many different types of secondary analysis. One can, for example, perform calculations of nuclear structure on the same configurations one also uses to perform calculations of fundamental parameters of the Standard Model of Particle Interactions. Alternatively, an ensemble generated to measure Nuclear Energy spectra may also be useful in the study of the nuclear strong force binding together nucleons into atomic nuclei.

Since the generation of ensembles is very demanding in terms of effort, and since the ensembles can facilitate multiple uses, it makes sense to share them amongst the LQCD community to get maximum value from each generation project. The ILDG infrastructure discussed in this paper is designed to promote and facilitate such data sharing.

The LQCD community has a history of sharing data that predates the formation of the ILDG. The MILC Collaboration¹ has pioneered the approach of freely giving away the data, after publishing results for their target quantities. This conservative approach is necessary for scientific prudence. The data has been very widely used, and the MILC Collaboration policy of data release is seen as successful and beneficial to the collaboration. There are many examples of different groups collaborating together to share the burden of generating the data. In some sense, the ILDG is similar to other kinds of data archives and Science Gateways, of which there are now many throughout the world. However, it does present some particularly unusual aspects, which stem from it being a *gridof-grids*.

In 2002, different groups were starting to make use of grid technologies to store and retrieve data, primarily within their own collaboration — see, for example, [5]. A proposal by Richard Kenway [6], at the annual lattice QCD conference, to use grid technologies to store and share data was well received and supported. Subsequently, the ILDG was formed from five interested groups that were willing to share data — specifically, the Center for the Subatomic Structure of Matter (CSSM) in Australia; the Japan Lattice Data Grid (JLDG); the Latfor Data Grid (LDG) for continental Europe (primarily Germany, Italy, and France); the regional grid of the UKQCD Collaboration in the UK; and the regional grid of the USQCD Collaboration in the United States.

There is no central authority forcing policy on the member collaborations: rather the ILDG is a collaboration of groups that are prepared to commit some resource to a central service. This idea of an aggregation or grid-of-grids is a powerful one that allows each group to retain control of its own resources, and to maintain existing infrastructure and grid middleware. Other activities – for example, consider the LHC – have elected (in some instances) to *move the computation to the data*, to avoid the overhead of transmitting large amounts of data over the Internet. However, since the compute resources used in LQCD research are typically much less uniform and the costs for data transfer is relatively small compared to the computing requirements, it has been decided that moving the data to the compute resource is appropriate.

Ensembles of gauge configurations that predate ILDG are typically identified using locally agreed naming conventions. For example, a particular configuration might be identified by a combination of the Unix path to the file and the hostname of the server on which it resides. While this approach may be suitable for a small group of researchers working in a particular collaboration, it is inadequate for a community like ILDG that is loosely coupled and distributed across multiple research groups.

¹ http://www.physics.indiana.edu/-sg/milc.html - accessed 29/APR/10.

What is required is a method for assigning a *unique and persistent identifier* to each file (that is, each gauge configuration) that is to be held within the infrastructure. In addition, there needs to be an equivalent method for identifying each ensemble.

1.1. Data curation

For a configuration (or an ensemble) to be useful to a researcher, it must be apparent what it represents in scientific terms. This information is provided by *metadata* — literally, data about data. Metadata may be captured in a number of different ways. For example, a widely used approach is based on descriptive filenames that follow an agreed naming format. For Lattice QCD, the detail required to describe a data set is too great to be realistically encoded in its filename, especially considering the various different formulations of QCD available, all with different parameters. For this reason, the process of scientific annotation has warranted a more sophisticated approach.

ILDG researchers require a scientific annotation that thoroughly and unambiguously describes a configuration (or ensemble of) for other members of the community. The annotation should be extensible: that is, it should support the introduction of new descriptive elements. This may be required — for example — to accommodate new science.

A user should easily be able to search the catalogue of scientific annotations and, complementing this, the generation of metadata should be a lightweight and straightforward process. Where possible, elements of the description should be populated automatically.

As well as having an agreed mechanism for describing data, one must also be able to read the binary files that hold the data. This motivates convergence to a common file format (for gauge configurations, at least). At the inception of ILDG, a number of different file formats existed, based on the conventions used in the most popular LQCD codes. Alongside the formalisation of the scientific metadata, it has been decided that a community-wide, flexible, extensible binary file format is required. By choosing a common data format, the requirement for an on-the-fly data format translation service between the many different formats is avoided.

1.2. Infrastructure

Pre-dating the formation of ILDG, the five collaborations that make up the core of the consortium have procured or developed storage facilities to host the ensembles of data that they each generate. These systems are all accessible — in principle — over the Internet, but via different and incompatible access protocols and access-control systems targeted at local (that is, institution-based) users.

The ILDG project does not have the resources to build and deploy a new infrastructure from scatch. Because of this, it has been decided to build on what is already available from the different regional grids. Then, to work around compatibility issues, two specific requirements need to be fulfilled. First, a layer of software is required on top of the local infrastructures, to provide a uniform interface to an end-user. Second, an access-control mechanism needs to be established that permits ILDG members from different collaborations to access designated data at partner institutions/storage facilities.

1.3. Operation and monitoring

To be useful, the ILDG infrastructure must achieve high levels of availability. High availability must be attained in spite of the decentralised and heterogeneous nature of the component elements, and should efficiently exploit the support effort available at the regional grids. It has therefore been decided that an automated monitoring service should be set up within the infrastructure, fulfilling the following specific attributes. The monitoring service needs to:

- be reliable since it is the primary means by which problems and failures are identified;
- be flexible in order that the diversity of ILDG components can be represented and monitored;
- produce accurate and informative alarms, which will allow regional-grid support teams to quickly and effectively diagnose and resolve issues;
- post alarms using email as this is the primary medium over which regional-grid teams communicate;
- maintain a record of system performance, to inform coordinators as to overall reliability and to highlight any specific weaknesses.

For simple control of access to all ILDG resources, a centrally coordinated user management system is required: making all globally registered users known to all local-resource providers. To this end, we have adopted the concept of a Virtual Organisation (VO), with membership being managed by the VO itself. With ILDG consisting of several regional grids, a setup is however needed that allows the decision - as to whether an application for VO membership is to be approved or declined - to be delegated to the regional grids. For users to access ILDG resources only a single sign-on should be required: that is, a single trust domain has to be defined. This domain should include a sufficiently large set of trusted Certificate Authorities that every potential users can be provided with a certificate that is acceptable to any of the resource providers. Moreover, users join both their regional grid, and the ILDG as part of the same procedure, thus avoiding any duplication of procedure.

While it is not envisaged that the regional-grid make-up of ILDG will change in a particularly dynamic manner, it is expected that new collaborations will wish to join the infrastructure, either independently or as part of an existing group. With this in mind, it is important that the infrastructure evolves in a way that does not prevent expansion. Specifically:

- ILDG specifications (for example, service definitions) are thoroughly documented in a manner intended to facilitate the creation of new implementations;
- the technology layer is supported by a test suite, which allows new implementations of ILDG services to be validated against the specification;
- where possible, ILDG uses open (or at least widely adopted) technologies and standards, aiming to increase coverage of user groups and to reduce the risk of systems becoming obsoleted;
- the technological aspect of the infrastructure is specified as a thin layer (that is, focused on a baseline set of functionalities), which can easily be incorporated into existing infrastructures with low levels of development effort.

2. Metadata and file formatting

To motivate the need for metadata, consider an example where there is no metadata. Configurations from different ensembles are all stored in a single directory with potentially random strings for names. Clearly this data is now not accessible. A set of rules for the metadata, known as a scheme, is required to describe the data. As noted above, many groups have in the past constructed ad hoc schemes for describing the data based on filenames and directory structures. Whilst this approach is not without merit, it does not scale when many groups are sharing data. In constructing this scheme there are likely to be several assumptions which are specific to the group that uses the scheme. Another group may well find these assumptions are not valid for their data, and hence their data will not fit into the scheme. Modifying the scheme is only possible where the assumptions used in its construction are still valid. To accommodate several potential different formulations of LQCD, and the needs of different groups, a different approach is required.

Extensibility is a critical requirement of any annotation scheme. Any new data will need new metadata to describe it and the scheme will have to be modified. In an extensible scheme this can be done without breaking the original scheme. That is, the new scheme is an extension of the old one. Furthermore, any document that was valid in the old scheme is valid in the new one, so that the old documents don't have to be updated to be valid in the new scheme.

Data provenance is likewise an important requirement. Can the data be recreated from the metadata? Taken to the limit this question is extremely challenging. In principle the code used to generate the data and its inputs should allow the data to be regenerated. However, this doesn't include any machine, compiler or library information. Moreover, in the context of sharing data, the application belonging to one group may not be able to parse and process the input parameters of the application belonging to a different group. Hence while a full archival of a statically linked code and its inputs should allow recreation of the data, if the original producing machine were to be available, archiving to this level of detail is not practical. Correspondingly some of the data provenance requirements may need to be relaxed in practice.

Lattice QCD metadata is hierarchical in nature and the annotation scheme should reflect this. Markup languages combine text and information about the text, and thus are perhaps a natural choice for a language in which to construct the scheme. Semantic or descriptive languages do not mandate presentational or any other interpretation of the markup. XML was chosen as it is the most widely used and best supported markup language. Similarly XML schema was chosen as the schema language to define the scheme.

In order to make sharing lattice QCD data useful and effective, lattice QCD metadata should be recorded uniformly throughout the grid. The metadata working group designed an XML schema called QCDml for the metadata [7]. The primary use case being data discovery via this metadata.

As described above, a key concept for Lattice QCD data is the organisation of the data as configurations and ensembles to which the configurations belong. With this in mind, the metadata is divided into two linked XML schemata, one for the configurations and one for the ensemble. The two schemata are linked together by a unique Uniform Resource Identifier (URI), called the markovChainURI, which lives in the name-space of the ILDG and which appears in the XML instance documents (IDs) of the configuration and the ensemble to which it belongs. This URI can be considered a Uniform Resource Name (URN) as it is not formally resolvable. There is no formal mechanism for ensuring uniqueness, but a simple convention has been adopted whereby the name of the group who generated the data appears in the URI, and responsibility for uniqueness is thereafter delegated to that group.

The separation of the metadata into two pieces, besides reflecting the nature of lattice QCD data, has two advantages. First, metadata capture is potentially simplified, as only the configurationspecific information has to be recorded for each configuration, and the information specific to an ensemble is recorded only once. Second, the performance of searches on the data may be improved since the split represents a factoring of the original more complicated schemata. The metadata scheme is encoded as a set of XML schemata [8] and whilst this does not mandate how the metadata is stored and accessed, for simplicity it is often stored in native XML databases such as *eXist* [9]. It is well known that the speed of access of hierarchical databases, such as native XML database, is vastly inferior to that of relational databases. Scientists are almost always interested in finding an ensemble rather than finding an individual configuration. Therefore, for most cases, the separation of ensemble and configuration XML reduces the number of documents to be searched by $\mathcal{O}(100-1000)$.

In each configuration ID the *logical file name* (LFN) of the data file is recorded. The LFN is a unique and persistent identifier of the file in the ILDG name-space. The ILDG and local grid services then map the LFN to actual file instances.

The data itself is stored in a file format known as LIME [10]. LIME is short for Limited Internet Message Encapsulation, and is a simplified and generalised version of the DIME (Direct Internet Message Encapsulation) [11], which was proposed as an Internet Standard and which is now part of the Microsoft .NET framework. LIME is a record-oriented message format that simplifies and extends the original DIME framework by introducing 64-bit length records instead of the original 32-bit ones, and correspondingly it eliminates the need for continuation records. LIME thus allows the packing of descriptive text records and binary data records into the same file. This format itself is very flexible and extensible since the types and sequence of records are not mandated in the file format itself. For ILDG, it was natural to adopt the LIME format since this format had already been accepted by a large fraction of the community and a considerable amount of software had been available to generate and process LIME files.

The ILDG specifies and requires a set of LIME records, including: a record containing some XML file-format metadata describing the size of the space-time lattice and data precision; a record containing the data itself in a specified data ordering; and a record of the LFN for the data, to allow the linking of ILDG data files to their metadata-catalogue entries. LIME was developed by the USQCD Collaboration through the SciDAC software initiative, and a C-code to read and write LIME files on a serial machine (C-LIME) can be downloaded from the USQCD web-site [10]. The QIO package² also developed by the USQCD Collaboration has facilities for reading ILDG formatted data files on both serial and parallel machines.

The scientific core of the metadata is contained within the ensemble schema. The most important section, from the datadiscovery viewpoint, is the *action*; it contains the details of the physics. Here, the object-oriented ideas of inheritance are used to build an inheritance tree of actions based on the XML schema concepts of extension and restriction as appropriate. This enables users to make both very specific searches and more general searches on the names, types and/or parameters of the actions. The exact details of the physics can only be encoded in mathematics and describing the mathematics in XML was not a requirement of the metadata scheme. A reference to a paper, and the URL of an external glossary document that contain the mathematical descriptions of the physics are included. Clearly an application cannot parse this information, but it is included to avoid ambiguity in the names used in the inheritance tree.

Lattice QCD algorithms are very complex with many different algorithmic components. They are also an active area of research, and changes and improvements are common. This makes designing a scheme, especially an extensible one, rather difficult. The defined names and inheritance-tree ideas used for the action would be too cumbersome for describing the algorithms. QCDml has only small



Fig. 1. Illustration of ILDG infrastructure – stacked boxes indicate services provided on a per-regional-grid basis.

scope for the algorithms limited to name, value pairs for the parameters. Algorithmic details can be expressed mathematically in an external, non-parseble glossary document. This approach further limits the data provenance of QCDml. However, individual groups can import their own namespace with as much detail and structure as they see fit, which can help ameliorate the data provenance issue even if the metadata is no longer universal. More physics-oriented information about the metadata can be found in [12].

3. Middleware

As described above, a grid-of-grids concept has been adopted for ILDG. Each of the regional grids has to provide the following services: a metadata catalogue (MDC) for metadata-based file discovery, a file catalogue (FC) for data-file location, and one or more storage elements (SE) that can serve the data to the user. The design of the ILDG infrastructure is illustrated in Fig. 1.

The user can discover available data sets by sending a query to the MDC of each of the regional grids. On input this search requires an XPath expression based on the QCDml schema. On output the services will return the list of Logical Files Names (LFNs) of those documents for which the XPath expression identifies a non-zero set of nodes. At the time the ILDG standard was adopted, XPath was the best-supported XML query language and it was to be expected that the planned (and — in the meanwhile — completed) definition of XQuery would include XPath.

To identify all copies of a particular file, a query to the file catalogue has to be performed, which takes an LFN on input and returns a source URL (SURL) for each replica of the file. The scheme part of the SURL tells the client whether it can either directly download the file using the transfer protocols HTTP or GridFTP, or whether it has to connect to a Storage Resource Manager (SRM) interface.

The SRM protocol [13] is evolving to an open standard for grid middleware to communicate with site-specific storage fabrics. A particularly appealing feature of the SRM is that this services comprises a standard Web Service interface, allowing the SRM component to fit in easily with the comparatively less complex, ILDG-defined MDC and FC interfaces.

SRM is a system to manage local storage fabrics comprised of several data servers and possibly long-term backup storage. From the point of view of an ILDG transaction, the SRM may be required to: stage a file to some transfer location on a server, negotiate a transfer protocol between the server and the client, and to then arrange for the transfer to occur. The location, where the file is or

² http://usqcd.jlab.org/usqcd-software/ - accessed 29/APR/2010.

may become available, and the transfer protocol are returned to the client by the SRM service in a transfer URL (TURL). The client and the data server can then carry out the transfer independently of the rest of the SRM system. Typically, GridFTP is used as a transport protocol.

The selection of an authentication mechanism has been straightforward. It is fortunate that all of the relevant grid-middleware stacks and regional-grid infrastructures have adopted a mechanism based on X.509 certificates [25]. Further, the International Grid Trust Federation³ provides a collection of certification authorities covering all likely member groups for the project, ensuring a trusted virtual organisation can be established across the ILDG community.

By adopting a grid-of-grids concept with different middleware stacks being used by the different regional grids, interoperability becomes a challenge. Interoperability is required to provide standardised interfaces towards the application layer. While there are clear similarities between the different grid architectures, there are crucial conceptional differences and incompatibilities of the interfaces.

For those services that have not been specifically designed for ILDG, one of two strategies has been applied to overcome this interoperability issue. First, wherever possible, common grid standards supported by all the used middleware stacks have been adopted. One example is the transfer protocol GridFTP. Second, interface services have been defined and implemented. Instead of accessing a service directly, the user connects to the interface service that then processes the request on behalf of the user.

To standardise a web service within ILDG a WSDL description is implemented and additionally a behavioural specification is provided. The WSDL description specifies the structure of the service's input and output data structure, while a functional description of the service is provided by the behavioural specification. Additionally, test suites have been defined and implemented which can be used to verify whether a service conforms to the ILDG standard.

If a service requires authentication, the corresponding interface service has to support credential delegation. Within ILDG we use an implementation of such a service that has been developed within the GridSite project [14] and is now part of the gLite middleware stack. On request of the client the server returns a proxy certificate request, which is signed on the client side and returned back to the server. Since the proxy certificate has only a limited lifetime, the risk due to a compromised server hosting the interface service is considered to be acceptable small.

Access to most services is restricted to members of the Virtual Organisation (VO) ILDG. For the management of this VO we use a VOMRS service [15]. A person that wishes to join the VO has to submit an application and nominate one of their regional grid's representatives. For each regional grid at least two representatives have been assigned, who can accept or reject the request. For each regional grid an individual group has been created. Information on group membership may be used by the regional grids as input to authorisation decisions.

The only other global service that is used within ILDG is the monitoring service. This monitoring service has been implemented using INCA [16], which is a popular tool in the research community for testing and monitoring grid infrastructure and services. In the framework of INCA a set of so-called reporter managers regularly execute test scripts accessing grid resources. The information returned by the reporter managers is collected in a repository. In case of failures a notification email is generated and sent to the regional grid that is responsible for a particular service. Data in the repository can later be used to check the service's availability.

4. Review

4.1. General status

At the time of writing, the ILDG has been in production use for a little over a year. It is comprised of five main-partner regional grids, as detailed in Section 1. The ILDG VO has 113 registered users and the combined ILDG hosts some 207 gauge-configuration ensembles, corresponding to various lattice volumes, gauge and fermion actions. Each single ensemble represents significant portions — potentially years — of human and supercomputing resources. Thus these archives are immensely valuable.

On the management side, the Middleware working group hosts monthly teleconferences to discuss operational exceptions, experiences and future development efforts while at the higher level, the ILDG holds bi-annual video conferences, so that regional partners can discuss more general progress.

4.2. Benefits of sharing

Hosting such a wealth of data has had great benefit on computational lattice QCD worldwide. In the case of some regional grids, the regional grid itself has become the primary means of data distribution, for multi-site projects, prime examples of which are the LDG and UKQCD Collaborations [17].

A number of research activities have been enabled, thanks to the ILDG infrastructure. Scientists in Japan have been using data produced by the MILC Collaboration (in United States) as part of their research [18] and, complementing this, a team at χ -QCD (University of Kentucky) has accessed data from CP-PACS (Japan), in the ILDG community [19,20]. Other examples of ILDG use can be found in [21] and [22] where two groups made use of lattices generated by the German QCDSF Collaboration.

Both inter-collaboration and intra-collaboration activities are enabled by ILDG. In [23], a number of ILDG-enabled activities are noted relating to data sharing across LDG sites. The fact that the ILDG is making a serious impact in international collaboration can also be seen in the fact that physics workshops are being held within the community that focus, not only on the generation of QCD data, but also on accomplishing calculations by sharing the data via the ILDG.⁴

4.3. Critique of the ILDG

While ILDG appears to be operating successfully, there are some aspects of it that could be improved. Using the ILDG to locate and share data is relatively straightforward especially with the easy-toinstall client tools, as described in [24]. However, contributing to the ILDG potentially involves a lot of effort. Depending on the level of involvement, one may need to maintain storage and database resources as well as having to mark up configuration and ensemble metadata.

Ensemble-metadata markup is not straightforward to automate and may need to be done by hand. If a new collaboration wishes to extend the XML schemata to mark up data for which no QCDml exists, the process of standardisation of the markup may take a substantial amount of time.

Marking up configurations may be more straightforward, and may be automated. However, it too involves some amount of postprocessing. The checksum needed in the configuration metadata document is not easy to compute in a parallel program and likewise a unique key; the configuration LFN; needs to be known in order to create both the configuration metadata and in order to

¹²¹²

³ http://www.igtf.net/ – accessed 5/JAN/11.

⁴ http://www.ccs.tsukuba.ac.jp/workshop/EP09 – accessed 29/APR/10.

write a fully ILDG-compliant configuration file as described previously. However, the LFN may not be known at the time of production. Thus typically configuration metadata is generated postproduction, and the configurations typically do not contain the LFN on creation. This has to be added on insertion to the ILDG.

While much of this activity can be automated, the initial goal of the computation producing the configuration metadata and the ILDG compliant configuration at the same time has been sacrificed in order to agree on other aspects. There is thus scope in the data production workflow, for data to lay idle for quite some time before being added to the ILDG with the consequent loss of history and provenance information. Hopefully future software tools can alleviate this problem.

Although it was thought that these difficulties would be a major stumbling block to ILDG participation, in practice metadata creation proved to be less laborious than initially expected. The ensemble metadata typically needs to be created only once, making it worth the effort and, as mentioned previously, the workflow for configuration metadata markup and publication can be substantially automated. Hence while the in-principle issues discussed above remain, at a practical level the bar for participation in ILDG came not from the metadata, but rather from maintaining the middleware stack of the participating organisations such as managing the grid-security certificate infrastructure.

One aspect of the ILDG to remark upon is that it is most definitely a volunteer, and altruistic activity. It receives very little in the way of funding for itself and is usually piggybacked discretely onto other grid-related projects or to regional-grid activities. Correspondingly, it can become difficult to maintain effort focused on the ILDG, which limits large-scale development and essentially forces simple solutions.

4.4. Related work

We can contrast the ILDG with some other related work. Other non-ILDG lattice archives include the Gauge Connection⁵ (at NERSC) and the QCDOC Configuration download site⁶ (at the Brookhaven National Laboratory) which is very similar in structure to the Gauge Connection and we shall treat the two identically below. The Gauge Connection was created before the era of Web Services and Grid services. It hosts files on a single filesystem and one can download all the configurations over HTTP. The file format used is an ASCII header followed by a binary data segment. The header contains rudimentary metadata (for example, information about the creators, a checksum, and some derived measurement). Hence there is no separation between the configuration files and their metadata like there is in our case. Ensembles are not marked up in terms of XML at all, but there is some human-readable description for each one. Authentication and authorisation is done at the web-server level and one needs to register with the site to gain access. This setup, though very simple has worked very robustly and well. On the other hand, it becomes harder to search this archive, since there is no actual metadata catalogue as such. A user must read through a list of available ensembles until they find the one they want. The Gauge Connection served as a guide to the ILDG effort. In particular the layout of the data in the binary part of the Gauge Connection format has been kept in the ILDG data record.

We should also mention in this section the LQCD Archive⁷ (LQA) which is maintained at the Center for Computational Sciences at the University of Tsukuba in Japan. The LQA began development prior to the ILDG to distribute the data of the CP-PACS

Collaboration as a configuration-download service similar to the Gauge Connection. However, upon inception of the ILDG, the LQA was re-developed to be the front-end portal to the data available on the JLDG. It provides metadata search facilities as well as HTTP based download that may be useful to users who do not wish to set up a full grid client infrastructure on their machine. The JLDG data is of course also available through the usual ILDG client tools independent of this portal. To use this service, one is required to register. The portal posts a list of publications to which citations should be made on publication of results that come from the downloaded data sets.

Download services have proved useful to the community however they have several shortcomings. They allow downloading primarily through HTTP which may encounter performance limitations when one considers downloading entire ensembles, especially given the ever growing size of configurations. There has been no attempt to provide a common file format. The individual architectures do not lend themselves to data replication and lack a common security infrastructure (each requiring separate registrations). That having been said, historically the Gauge Connection share their file format while the LQA as noted above has been extensively redeveloped to complement rather than contrast with the ILDG.

One can also compare the ILDG to the concept of a Science Gateway. Quoting from the definition of Science Gateways on the TeraGrid,⁸ "A Science Gateway is a community-developed set of tools, applications and data that is integrated via a portal, or suite of applications, usually in a graphic interface that is customised to meet the needs of a target community." In this sense the gateways have a broader scope than the ILDG, they can offer codes and grid services, as well as access to data collections. As an example, we consider the "Massive Pulsar Surveys Using the Arecibo L-band Feed Array (ALFA)" TeraGrid Science Gateway, which allows one to browse data on pulsars and is similar in scope to the ILDG. One can browse pulsar information, and can download associated data-products. On the other hand, the SCEC Earthworks Gateway actually allows a user to run earthquake simulations on TeraGrid resources.

One unique feature of the ILDG, in contrast to a Science Gateway, is that the ILDG is the result of a collaboration of collaborations. A single Science Gateway would typically consist of a single portal maintained by a group on behalf of a larger community. This group then has some freedom (within community limits) in defining internal formats, markup and can settle on a single set of software tools. The ILDG instead is a loose federation of existing grids, some of which at the inception of the ILDG had no grid infrastructure and some of whom were already heavily invested in their own systems. The worldwide community had to therefore come together in order to define metadata standards, middleware operation and lightweight, easy-to-implement interfaces that could then wrap any potentially existing, underlying infrastructure.

5. Summary and future work

In summary, the ILDG is a loosely federated grid-of-grids to facilitate the sharing of LQCD data worldwide. The technology allows it to operate across regional-grid boundaries, relies on a simple and lightweight layer of middleware standard definitions, and a standardised metadata markup.

In six years of design and implementation, and a little over one year of operation, the ILDG effort has brought together the lattice QCD community and has fostered QCD research and collaboration.

⁵ http://qcd.nersc.gov - accessed 29/APR/10.

⁶ http://lattices.qcdoc.bnl.gov – accessed 29/APR/10.

⁷ http://www.jldg.org/ildg-data/ - accessed 29/APR/10.

⁸ http://www.teragrid.org/gateways - accessed 29/APR/10.

Potential future work focuses on several areas including but not limited to data replication, and the storage and mark-up of secondary large data such as quark propagators.

Acknowledgements

Notice: Authored in part by Jefferson Science Associates, LLC under U.S. DOE Contract No. DE-AC05-06OR23177. The U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce this manuscript for U.S. Government purposes.

References

- M. Creutz, Quarks, Gluons and Lattices, Cambridge Monogr. Math. Phys., Cambridge Univ. Press, Cambridge, 1983, 169 p.
- [2] I. Montvay, G. Munster, Quantum Fields on a Lattice, Cambridge Monogr. Math. Phys., Cambridge Univ. Press, Cambridge, 1994, 491 p.
- [3] R. Gupta, Introduction to lattice QCD, arXiv:hep-lat/9807028.
- [4] M. Peardon, Monte Carlo simulations of lattice QCD, prepared for 3rd International Workshop on Numerical Analysis and Lattice QCD, Edinburgh, Scotland, 30 June-4 July, 2003.
- [5] C.R. Allton, et al., QCDgrid: A grid resource for quantum chromodynamics, J. Grid Comp. 3 (1-2) (2005).
- [6] C.T.H. Davies, A.C. Irving, R.D. Kenway, C.M. Maynard, UKQCD Collaboration, International lattice data grid, Nuclear Phys. B Proc. Suppl. 119 (2003) 225, arXiv:hep-lat/0209121.
- [7] C.M. Maynard, D. Pleiter, QCDml: First milestones for building an International Lattice Data Grid, Nuclear Phys. B Proc. Suppl. 140 (2005).
- [8] E. van der Vlist, XML schema: The W3C's object-oriented descriptions for XML, O'Reilly Media; 1st ed., 25 June, 2002.
- [9] W. Meier, eXist: An Open Source Native XML Database, Lecture Notes in Comput. Sci., vol. 2593, 2002.
- [10] SciDAC Software Coordinating Committee, LIME (Version 1.1) accessed 29/APR/10, http://www.physics.utah.edu/~detar/scidac.

- [11] J.H. Gailey, Sending files, attachments, and SOAP messages via direct Internet message encapsulation MSDN magazine, Microsoft Corporation, December 2000.
- [12] P. Coddington, B. Joo, C.M. Maynard, D. Pleiter, T. Yoshie, Marking up lattice QCD configurations and ensembles, PoS LAT2007 (2007) 048, arXiv:0710.0230 [hep-lat].
- [13] A. Sim, A. Shoshani (Eds.), The storage resource manager interface specification version 2.2, Lawrence Berkeley National Laboratories, Technical Paper from September 2009.
- [14] A. McNab, S. Kaushal, The GridSite proxy delegation service, presentation to the UK e-Science All Hands Conference, Nottingham, September 2006.
- [15] L. Bauerdick, et al., The virtual organisation management registration service, in: CHEP 2006 Proceedings, Proc. International Conference on Computing in High Energy Physics (CHEP06), Mumbai (India), February 2006.
- [16] S. Smallen, et al., User-level grid monitoring with INCA 2, in: Proceedings of the 2007 Workshop on Grid Monitoring, 2007.
- [17] M.G. Beckett, et al., Building a scientific data grid with DiGS, Philos. Trans. R. Soc. A 367 (2009).
- S. Furui, Propagator of the lattice domain wall fermion and the staggered fermion, Few Body Syst. 45 (2009) 51, arXiv:0801.0325 [hep-lat];
 S. Furui, Few Body Syst. 46 (2009) 73 (Erratum).
- [19] T. Draper, T. Doi, K.F. Liu, D. Mankame, N. Mathur, X.f. Meng, Light scalar mesons in 2 + 1 flavor full OCD, arXiv:0810.5512 [hep-lat].
- [20] T. Doi, et al., Nucleon strangeness form factors from N_f = 2 + 1 clover fermion lattice QCD, arXiv:0903.3232 [hep-ph].
- [21] C. Ehmann, G. Bali, Charmonium spectrum including higher spin and exotic states, PoS LAT2007 (2007) 094, arXiv:0710.0256 [hep-lat].
- [22] E.M. Ilgenfritz, M. Muller-Preussker, A. Sternbeck, A. Schiller, Gauge-variant propagators and the running coupling from lattice QCD, arXiv:hep-lat/ 0601027.
- [23] H. Stuben, S. Wollny, Using the mass storage system at ZIB within I3HP, Nuclear Phys. B Proc. Suppl. 153 (2006) 300, arXiv:hep-lat/0512008.
- [24] T. Yoshie, Making use of the International Lattice Data Grid, PoS LATTICE2008 (2008) 019, arXiv:0812.0849 [hep-lat].
- [25] V. Welch, I. Foster, C. Kesselmann, O. Mulmo, L. Pearlman, J. Gawor, S. Meder, F. Siebenlist, X.509 proxy certificates for dynamic delegation, in: Proc. of the 3rd Annual PKI R&1D Workshop, 2004.