Journal of Research in Education and Society
Volume 13, Number 1, April 2022
ISSN: 2141-6753
Published By
International Centre for Integrated Development Research, Nigeria
In collaboration with
Copperstone University, Luanshya, Zambia

Towards Implementing an In-Memory Data Grids for Digital Library Resources

Sulaiman, Abiola Moyosore
Fatiu Akesode Library
Lagos State University, Ojo, Lagos State, Nigeria.
E-mail: biola1000x@yahoo.com

Sulaiman, A. Adesola
Ogun State College Of Health Technology,
Ilese Ijebu Ogun State, Nigeria.
E-mail: soladipo2003@yahoo.com

Akalumhe, O. Kareem
Fatiu Akesode Library, Lagos State University, Ojo, Lagos State, Nigeria

ABSTRACT

*This study focuses on implementing in-memory data grids for digital library resources. Implementing an In-Memory Data Grids for digital library resources is a technological approach to dealing with problems of data in the library. A digital Library is an informal collection of information stored in digital formats and accessible over a network with associated services. Implementing an In-Memory Data Grids (**IMDG**) provides highly available data by keeping it in memory and highly distributed (i.e. parallelized). Thus, offering the very high available data guaranteed speeds by loading data in-memory, and size using scalability structures provided by a cluster. This paper gives an overview of a digital library, In-memory data grids, advantages, challenges and implementation. It also highlights technical possibilities and patterns that make an **IMDG** beneficial.*

***Keywords**: In-Memory Data Grids, digital library resources, scalability*

INTRODUCTION

The library is an organized body that holds collections, digital objects that have been grouped into categories, presumably for access purposes. So, a digital Library is an informal collection of information, stored in digital formats and accessible over a network, together with associated services (Jie Sun and Bao-Zhong Yuan, 2012). The Internet has changed the concept of libraries, processing methods,

Journal of Research in Education and Society
Volume 13, Number 1, April 2022
ISSN: 2141-6753
Published By
International Centre for Integrated Development Research, Nigeria
In collaboration with
Copperstone University, Luanshya, Zambia

storage, transmission and distribution of information. As a result, large numbers of electronic resources are published in all subject areas. Libraries in 21st-centuries are referred to as Digital Library, E-Library or Virtual Library (Kotur and Mulimani, 2019).

A Digital Library is a library in which collections are stored in digital formats and accessible by computers. Libraries are now available with electronic library resources used by computer and other equipment to collect and use information. Digital library resources are a computer-readable file that takes up less space than traditional library resources (Kotur and Mulimani 2019). A digital version of print journal like electronic publication with no print counterpart can be made available via the web, e-mail or other means of internet access. Some web- based e-journals are graphically modelled on the print version.

Electronic resource libraries are key elements and depositories of resources. Internet, Google or www is used extensively for meeting quick retrieval and user requirement of information. Electronic resources consist of data (information representing numbers, texts, graphs, maps, moving images, music, sound, etc.) programs or combination of data and programs (Sunitha and Vanaja, 2012). Some examples of Digital Library resources are databases, E-Journals/E-Books/E-Papers/E- Magazines, Streaming Videos, Online Databases, CD-ROM/DVD ROM Database, Digitized reports/Standards, E-Reference Tools and Web OPAC containing hypertext links though not limited to this. It makes all academic to do research work effectively. Libraries' digital/e-resources or are easily found to be less expensive and more accessible for easy access. Electronic resources are the indispensable tools of higher education. They provide unlimited information round the clock and in large volume as desired (Sudha, 2012).

Features of Digital Resources

The following are the features of digital resources

1. The digital resources are potentially large in volume.
2. 24/7 access to library collections in the database and full content search is feasible.
3. The information is well organized and comprehensive of everything as they are from variety of sources.
4. The speed of its generation is tremendous and its search options are of wide range.
5. The constraints which are common in conventional libraries are barred in e- resources (Sudha, 2012).

Journal of Research in Education and Society
Volume 13, Number 1, April 2022
ISSN: 2141-6753
Published By
International Centre for Integrated Development Research, Nigeria
In collaboration with
Copperstone University, Luanshya, Zambia

Advantages of digital library

According to Chore and Salwe (2010), the benefits of digital library are

(1) Preserve the valuable documents, rare and special collections of libraries, archives and museums.
(2) Protect information source.
(3) Facility for the downloading and printing.
(4) Provide faster access to the holding of libraries worldwide through automated better catalogues.
(5) Help to locate both physical and digitized versions of scholarly articles and books through single interface.
(6) Search optimization, simultaneous searches of the Internet make possible, preparing commercial databases and library collections.
(7) The user can peruse them instant.
(8) Cross references to other documents.
(9) Make the chain from author to users shorter.
(10) Save preparation and conservation cost, space and money.
(11) Afford multiple, simultaneous user from a single original which are not possible for materials stored in any other forms.
(12) Full text search.

In- Memory Data Grid (IMDG)

The global network internet has brought forth new dimension to libraries of modern digital world. In order to keep pace with the cyberspace, librarians are to furnish libraries with latest version of sophisticated technology. In this new library, digital networking and communication infrastructure provides a global platform over which the people and organization devise strategies, interact, communicate, collaborate and search for information (Kavita 2011). This platform includes, a vast array of digitalizable products that is, databases, computers, In- memory data grid and software which are delivered over the digital infrastructure anytime, anywhere in the world. The information available on- line could be looked at as big data. The numbers of users of these resources are also enormous, all making concurrent request on the database. According to Guroob and Manjaiah (2017), the "three V's" of Data are Velocity, Volume and Variety. Being able to easily handle these three considerations are core to a fast data strategy. For example, Fast Data often encompasses the ingestion and analysis of streaming and/or transactional data (velocity). While there is certainly an ever-growing amount of data (volume) and the sources and types of data in the enterprise are expanding (variety), the key notion to understand about Fast Data are the type of the latencies

Journal of Research in Education and Society
Volume 13, Number 1, April 2022
ISSN: 2141-6753
Published By
International Centre for Integrated Development Research, Nigeria
In collaboration with
Copperstone University, Luanshya, Zambia

at which data can be processed to produce actionable insights or machine driven decision-making. Succinctly, the key notion that makes data fast is the actionable analysis of information in near real-time.

In other to maximize the digital resources in a library, we need to deploy In- Memory Data Grid. It offers the following attractions:

- Even when the library is not used frequently by the user, the library can still operate as portal, offering guidance for web-surfing by list of approved links, or tips on surfing techniques and proving access to a broad range of e-resources.
- Access to information from anywhere at any time.
- Ensuring speed, accuracy and effective in retrieval of information.
- Providing multiple accesses to user and sharing the e-resources simultaneously.
- A Boosting for digital libraries
- Improving Speed and Scalability with In-Memory Data Grids.

The goal of implementing In-Memory Data Grids (**IMDG**) is to provide extremely high availability of data by keeping it in memory and in highly distributed (i.e. parallelized) fashion (Gridgain.com, 2013). Thus, offers the very high availability of data and guaranteeing speeds that the user desired. When using data In-memory grids we achieve greater scalability structures provided by a cluster. By loading Terabytes of data into memory IMDGs are able to work with most of the Big Data processing requirements in today's digital library (Guroob and Manjaiah, 2017). An online source at Gridgain.com (2013) states that In-Memory Data Grids (IMDG) help users reduce the cost of running huge volume of data, transactional or analytical applications by enhancing their scalability and performance while maintaining the integrity of data in memory. It is not an in-memory relational database, an NOSQL database or a relational database. IMDG has a completely different architecture. The structures of IMDG can be brief as follows:

- The data are scattered on several servers (Data Fabric).
- The data are stored in the main memory (RAM) of the servers. All servers in the IMDG environment operate their data in the active mode. For increasing the amount of memory, servers can be added or removed. A data model is usually object based and non-relational. This enables the highest application performance by using RAM along with the processing power of multiple computers that run tasks in parallel (Ivanov, 2020). In order to use main memory as a storage capacity, two weaknesses must be overcome:

Journal of Research in Education and Society
Volume 13, Number 1, April 2022
ISSN: 2141-6753
Published By
International Centre for Integrated Development Research, Nigeria
In collaboration with
Copperstone University, Luanshya, Zambia

(i) Limited space of storage: Includes data that exceeds the maximum capacity of the main memory of the server.

(ii) Reliability: includes data that damage in case of a system failure.

IMDGs are especially valuable for applications that do extensive parallel processing on large data sets such as library databases.

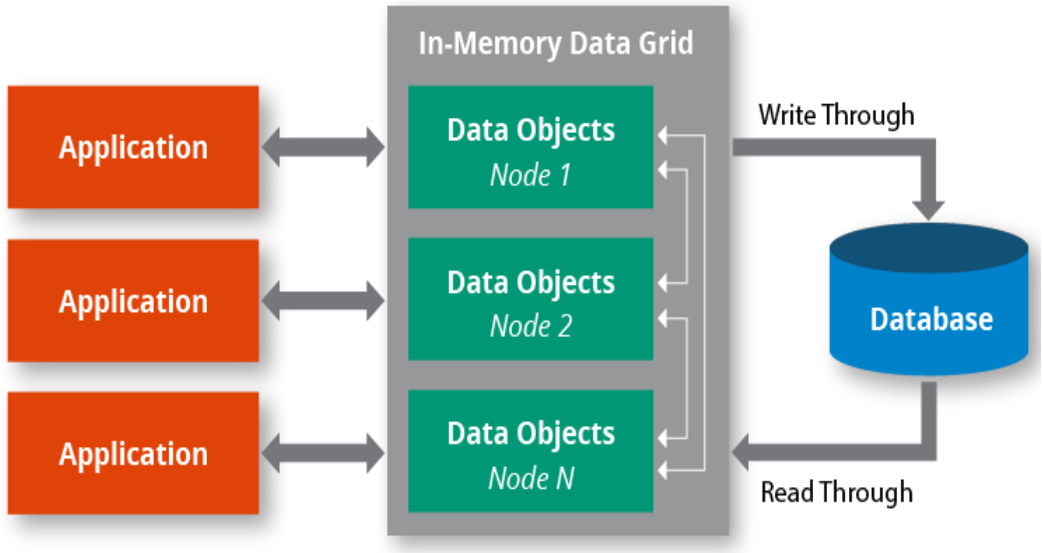


Figure1: In-Memory Data Grid Diagram (Source: Gridgain.com, 2013).

Importance of IMDGs to electronic resources

IMDGs fuel exponential growth in data volumes. Systems process millions of events and generate a massive amount of data. Libraries are also exposing their digital resources as APIs with stringent response time and larger throughput Service Level Agreements (SLAs). These modern applications/services require high-speed access to data to ensure engaging user experience and meet throughput and response time SLAs. To put this in context, consider these applications scenarios:

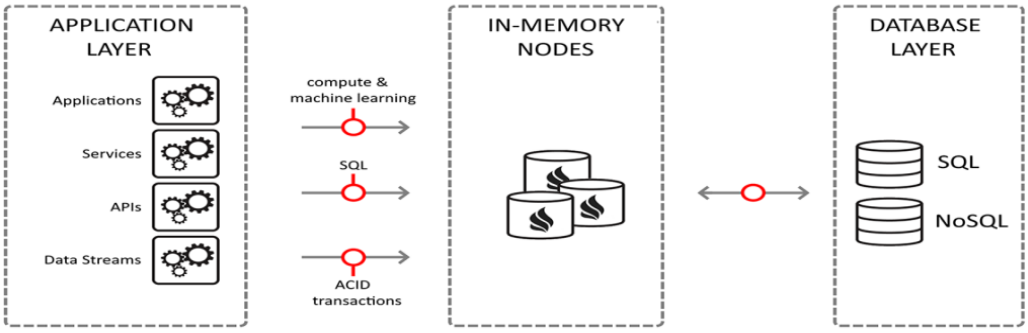
Digital libraries applications - these applications require fast access to book catalogues, managing readers request and query, making recommendations based on items in requested and other criteria, tuning the search results based on user's past browsing /reading pattern , behaviour, etc. The libraries application exposes an API to provide reader portal and other details .It then leverage this API to build

Journal of Research in Education and Society
Volume 13, Number 1, April 2022
ISSN: 2141-6753
Published By
International Centre for Integrated Development Research, Nigeria
In collaboration with
Copperstone University, Luanshya, Zambia

rich mobile applications for readers to read on the go. This results in an exponential rise in the number of request and queries.

Building these applications with traditional databases would involve substantial engineering complexities, especially those related to meeting the non-functional requirements of response time and throughput. By using IMDGs, a fast data layer can be created that will help in data lookup, storing states and also storing and maintaining library operational metrics.

They can also be used as a primary data store and then using its write-behind capability to write the data into backend databases as shown bellow



A usage pattern that we will be observed is depicted from the user below:

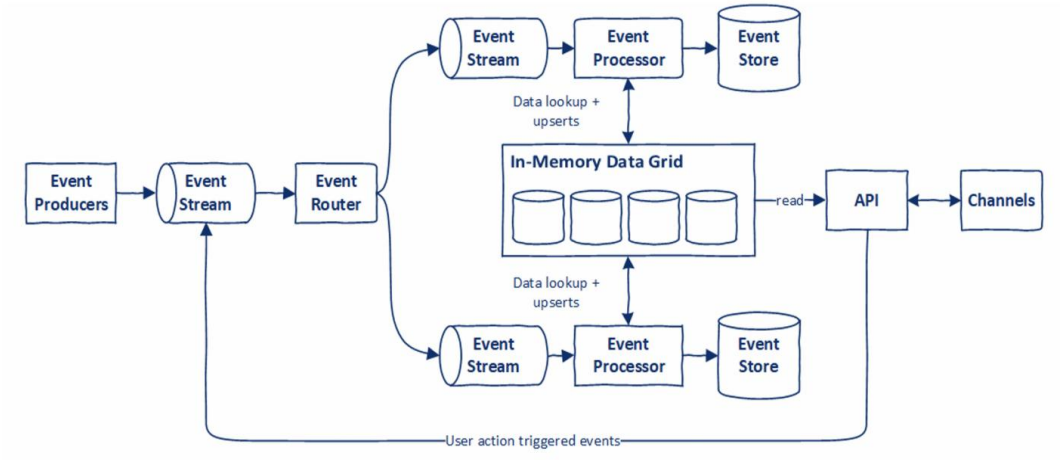


Figure2 — IMDGs and Event Driven Application (EDA) common usage pattern

There are several characteristics of IMDGs that are of attractions:

Journal of Research in Education and Society
Volume 13, Number 1, April 2022
ISSN: 2141-6753
Published By
International Centre for Integrated Development Research, Nigeria
In collaboration with
Copperstone University, Luanshya, Zambia

- Fast and highly concurrent access to data
- Help reduce the load on Systems of Record
- Distributed and fault-tolerant in nature
- Easy to scale as data volumes increase
- Support for cloud-based architectures
- Ease of setup and deploy

Technical Possibilities And Patterns That Make An IMDG Is Beneficial.

i. Reads are more for the "recent" data rather than "old" data.
The user interface (UI) requires data in a representation that is different from the underlying data model. Typically, an aggregation across multiple entities such as Real-times needs to build for display on dashboards.

Response time requirements for reads are aggressive.
It is easy to populate these grids in event-driven and stream processing types of applications. These data streams provide a mechanism to transform/enrich the data, which is more aligned to the consumption requirements (Kafka Streams, 2021).

ii. Faster performance and high concurrency (through partitioning and appropriate colocation strategy) compared to traditional databases. Ease of scaling is an important characteristic. IMDGs are inherently distributed in nature and can be easily scaled horizontally.

iii. IMDGs also support replication, thus making them more resilient to failure.
If using the SQL interface (JDBC, ODBC drivers): (a) queries need to be optimized for performance, (b) proper indexes have been created, (c) try to store data in de-normalized form to avoid joins, (d) if joins are unavoidable, try to ensure data being joined is "collocated".

iv. Real-time Analytics
There are many metrics and insights that need to be generated in near real-time. This is especially important for systems analyzing risk, detecting deviant behaviour, i.e. fraud detection, anomaly detection, and IoT based systems.

v. Risk assessment applications
Pattern and anomaly detection applications such as fraud detection, high-value payment processing, predicting failure, etc.

Journal of Research in Education and Society
Volume 13, Number 1, April 2022
ISSN: 2141-6753
Published By
International Centre for Integrated Development Research, Nigeria
In collaboration with
Copperstone University, Luanshya, Zambia

vi. Machine Learning inference, especially for models that require a large volume of near real-time data. This entails:

- Correlating different types of events
- Aggregating events and data
- Running algorithms to identify patterns
- Fast access to near real-time data to feed into machine learning models

Typical challenges in IMDG implementation

IMDGs are extremely beneficial for providing fast data access and reduce load on the existing traditional systems of record. However, there are certain challenges pertaining to its use that need to be considered:

- Without persistence of in-memory data, there is a risk of data loss in case of IMDG cluster-wide failure. This can be addressed to a certain extent by having a cross data center cluster setup and ensuring replicas are maintained across data centers. A more reliable solution is to have persistence enabled for the IMDG. However, persistence has performance and storage costs associated with it. Persistence can either be on a disk (for example, native persistence in the case of Apache Ignite (2021) or in a 3rd party database such as Cassandra or an RDBMS. IMDGs provide "write-behind" pattern for persistence to improve performance. However, performance issues manifest when there are big bursts of inserts/updates happening on the IMDG.

- Recovering from a cluster-wide crash. When the IMDG cluster crashes, data are lost. If this data are backed up on persistence storage, then it needs to be reloaded before allowing applications to access it. Having optimal recovery time is extremely important to ensure minimum disruption. Recovery time is proportional to the data volume and number of threads involved in recovery. If you are dealing with a large amount of data, it is highly advisable to test recovery scenarios as part of the development and tune the IMDG and the infrastructure to optimize recovery time.

- Memory is still not very cheap and setting up an IMDG has cost implications. Increasing the cluster size as data volumes grow comes at a cost. It is prudent to set up limits on how many days' worth of data needs to be available in the IMDG.

Journal of Research in Education and Society
Volume 13, Number 1, April 2022
ISSN: 2141-6753
Published By
International Centre for Integrated Development Research, Nigeria
In collaboration with
Copperstone University, Luanshya, Zambia

Implementation

The steps of how to use an In-Memory Data Grid as a follow:

1. Install servers in a single site or across multiple sites. Each group of servers within a site is denoted to as a cluster.
2. Install the IMDG software on all the servers and indicate the appropriate topology for the implementation. For multi-site operations there two options a partitioned or replicated cache
3. Setup APIs or GUI interfaces to let replicated between the various servers.
4. Develop data model and the business logic around the model.

Related Products

In recent times, there are many IMDG products, some of them are commercial and others are open source. The most commonly used products according to Guroob and Manjaiah (2017) are:

- GridGainDataGrid (Gridgain.com, 2013)
- Hazelcast (Hazelcast, 2021)
- JBoss Infinispan
- Terracotta Enterprise Suite
- Gigaspaces XAP
- VMware Gemfire
- Oracle Coherence.

CONCLUSION

Implementing In-Memory Data Grids for digital library resources will give the users the best of experiences ranging from fast and highly concurrent access to data of diverse sources and ease in scaling up as data volumes increase.it is very ease to setup and deploy. The architecture support cloud based storage or computing. Though it has challenges of high cost of Memory but overall it is very cheap. Recovering from a cluster-wide crash and performance issues manifest when there are big bursts of inserts/updates happening on the IMDG. Management will need to invest more and educate patrons on how to use the improved digital resources.

Journal of Research in Education and Society
Volume 13, Number 1, April 2022
ISSN: 2141-6753
Published By
International Centre for Integrated Development Research, Nigeria
In collaboration with
Copperstone University, Luanshya, Zambia

REFERENCES

Chore, N. V. and Salwe, S. M. (2010). Library Sources and Service in Digital Environment. Proceeding of state level seminar on role of information technology in library, Karad (April 8-9).

Dean J. and Ghemawat S. (2004). MapReduce: Simplified Data Processing on Large Clusters. Available:https://static.googleusercontent.com/media/research.google.com/en/ /archive/mapreduce-osdi04.pdf. [Accessed: 21-Jan-2021]

Fowler M., "bliki: CQRS", *martinfowler.com*, 2020. [Online]. Available: https://martinfowler.com/bliki/CQRS.html. [Accessed: 23-Dec-2020]

Gridgain.com, (2013). In-Memory Data Grid — White Paper [Online]. Available: https://www.gridgain.com/media/in-memory-datagrid.pdf. [Accessed: 20-Jan-2021]

Guroob A. H. and Manjaiah D. H. (2017). Big Data-based In-Memory Data Grid (IMDG) Technologies: Challenges of Implementation by Analytics Tools. International Journal of Emerging Research in Management &Technology (Volume-6, Issue-5)

Hazelcast (2021). "What is Machine Learning Inference?", [Online]. Available: https://hazelcast.com/glossary/machine-learning-inference/. [Accessed: 22-Jan-2021]

Ignite.apache.org (2021). Distributed Database — Apache Ignite®. [Online]. Available: https://ignite.apache.org. [Accessed: 20-Jan-2021]

Ivanov, N. (2020). In-Memory Data Grid: Explained…, *GridGain Systems* [Online]. Available: https://www.gridgain.com/resources/blog/in-memory-data-grid-explained. [Accessed: 23-Dec-2020]

Jie Suna and Bao-Zhong Yuan (2012). Development and Characteristic of Digital Library as a Library Branch 2012 International Conference on Future Computer Supported Education. IERI Procedia 2 (2012) 12 – 17.

Kafka Streams (2021). Apache Kafka Documentation,. [Online]. Available: https://kafka.apache.org/documentation/streams/. [Accessed: 22-Jan-2021]

Kavita A. J. (2011). Digital library: today's need- a review. International Multidisciplinary Research Journal,; 1(11):17-19.

Kotur M. B. and Mulimani M. N. (2019). Digital Library Resources for the Users: An Over View. Journal of Advancements in Library Sciences.; 6(Special Issue 1): 111s114s

Kotur M.B. and Mulimani M.N. (2019). Digital Library Resources for the Users: An Over View. Journal of Advancements in Library Sciences, 6(Special Issue 1): 111s–114s.

Sunitha K. H. and Vanaja E L. (2012). Electronic Resources and Services in Digital Ambience: A Book or Bane? National Conference on Emerging Trends in User Expectations for Next Generation Libraries, Kuppam; Feb 24-26; India: 2012. 218-222pp.