Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Overview of Data Warehouse architecture, Big Data and Green computing

Milankumar Rana
Department of Information Technology, University of the Cumberlands, Williamsburg, KY 40769, USA
Corresponding Author: Milankumar Rana, **E-mail**: mrana02098@ucumberlands.edu

| ABSTRACT
Enterprise today heavily invests in big data, data warehouses, and green computing to measure performance and make intelligent decisions. By enabling data warehouse architecture, enterprises can store structural and nonstructural data in defined backend systems and transform amounts of data to perform various analyses and make business decisions that put companies on top of their competitors. Data is increasing daily, and businesses want to use all their data to perform advanced business analytics and machine learning and distribute data to their backend algorithms to evaluate and make decisions faster. However, the carbon footprints and global warming have alarmed these organizations to move towards green computing for a better future. Green computing focuses on designing energy-efficient systems, optimizing resource utilization, and reducing the CO2 emissions of data centers and IT infrastructure. This paper reviews how data warehouse architecture, big data, and green computing relate and addresses the challenges and opportunities in achieving sustainable and scalable data management solutions. By integrating energy-efficient practices into data warehouses and big data systems, organizations can make a huge difference globally and set an example for other industries to follow the green computing path. This paper helps to understand the new paradigm of big data and green computing, which helps achieve the best performance, reduce environmental impacts, and achieve the best standards.

| KEYWORDS

reen computing, Big data, Datawarehouse, Business analytics, Green Energy, Data management, IT infrastructure, energy efficient, Advanced analytics

| ARTICLE INFORMATION

ACCEPTED: 02 November 2023 **PUBLISHED:** 28 November 2023 **DOI:** 10.32996/jcsts.2023.5.4.22

1. Introduction

The term data warehouse itself defines what it defines. In the computer world, it is a data unit that connects to multiple data sources of any type and provides information to some other meaningful tool where it can be analyzed and used for a different urpose, e.g., reporting or Artificial machine learning feed for a business model (What Is a Data Warehouse? | Definition, Components, Architecture | SAP Insights, n.d.).

2. Architecture

The Data warehouse has five major components:

Database, ETL process, metadata, query tool, and data warehouse bus architecture (Smith, 2022).

2.1 Datawarehouse Database

The database is a key component in a data warehouse system. As it contains a lot of data, it needs scalability, so it requires parallel computing. This will help in fast querying data from the data warehouse unit. Certain data needs to insert as it is in raw format. So, a non-structural database system is also introduced in the data warehouse when any data can be inserted, and then used that raw data and parsed according to form. E.g., JSON, XML. (Smith, 2022).

2.2 ETL (extract, transform, load)

"The data sourcing, transformation, and migration tools are used for performing all the conversions, summarizations, and all the changes needed to transform data into a unified format in the Datawarehouse. They are also called Extract, Transform and Load (ETL) Tools" (Smith, 2022). In this component, there are different methods or programming involved in parsing and transforming it to meaningful data and feeding it to the proper channel (Smith, 2022).

2.3 Metadata

"Metadata is nothing but data about data" (Smith, 2022). It is a key component as it has more granular information about the data warehouse database and system. "E.g., For example, a line in the sales database may contain:4030 KJ732 299.90 that tells us it was Model number: 4030, Sales Agent ID: KJ732, and Total sales amount of $299.90" (Smith, 2022). Metadata can be helpful in getting information about business in depth.

2.4 Query tools

Query tool helps in getting valuable information from the database and making appropriate decision or report. Query tools are useful in reporting, application development, data mining, and online application processing. Example of reporting tools is brio and oracle. Application development tools can be eclipse or visual studio. Datamining tools are IBM Watson (Smith, 2022).

2.5 Data warehouse bus architecture

"Data warehouse Bus determines the flow of data in your warehouse. The data flow in a data warehouse can be categorized as Inflow, Upflow, Downflow, Outflow, and Meta flow" (Smith, 2022).

3. Data warehouse architecture

Data warehouse architecture is categorized by how data is being presented form the source in the data warehouse system. It can be a one-tier, two-tier, or three-tier architecture (Simic, 2021).

3.1 One tier architecture

"The single-tier architecture is not a frequently practiced approach. The main goal of having such an architecture is to remove redundancy by minimizing the amount of data stored. Its primary disadvantage is that it does not have a component that separates analytical and transactional processing" (Simic, 2021).

3.2 Two tier architectures

Two-tier architectures have an additional layer on top of one tier, which is called staging. in this environment, data is processed using some tool and then it is stored in a data warehouse environment before it is represented (Simic, 2021). "This approach has certain network limitations. Additionally, you cannot expand it to support a larger number of users." (Simic, 2021).

3.3 Three tier architectures

Three-tier architectures have three layers top, bottom, and middle layer. This architecture is preferred by all IT organizations. The bottom layers consist of ETL-parsed data (Simic, 2021). "The middle tier is the application layer giving an abstract view of the database. It arranges the data to make it more suitable for analysis. This is done with an OLAP server, implemented using the ROLAP or MOLAP model" (Simic, 2021). The highest layer is where all data is in a user-readable format where it can be consumed by PowerBI application or Machine learning or AI (Simic, 2021).

4. Current Key Trend in data warehouse

Since data is a prime asset in any business, more technologies are coming along to help businesses use data and transform it to create business models that can profit organizations. All public cloud service provider has their own data analysis and ETL tool that can be used to perform ETL processing. The sudden increase in different ETL tools and more enhancements are coming along with it. Databricks, snowflake, azure datalake, oracle, power, azure Synapse, dedicated serverless tool. These will help to finetune the data and its operation for heavy loads. This will lead to more jobs in the data market (Molander, 2022). "If you do not have a real-time streaming system, you must deal with things like, okay, so data arrives every day. I am going to take it in here. I am going to add it over there. Well, how do I reconcile? What if some of that data is late? I need to join two tables, but that table is not here. So, maybe I will wait a little bit and rerun it" (Ghodsi & Casado, 2021). Current trend also includes data quality and optimization (Molander, 2022).

5. Big data

Big data, in my opinion, is data with heavy size and has all types of data, including XML, JSON, image, video, file, SQL, etc. As it contains various data types, different tools and techniques are used to transform it into a meaningful format. In big data, the type of structure can be non-structural data or structural data. Non-structural data is big in size, and it contains Raw format, e.g., XML, image data. Big data can be used for high-speed services like live events or shows. It needs excellent speed to broadcast to a large audience. Structural data may cost more for storage, while non-structural data is less costly. In my current organization, I have been dealing with a variety of data types, and they are processed using multiple tools. I am currently using oracle golden gate for structural data and Azure data lake for non-structural data. I am also using Azure synapse for ETL and feeding it to ML/AI modeling. For heavy-volume data, I use the Azure cosmos database. In my opinion, there is business value involved in existing data. A data scientist can make a machine learning module and transform it into a business model or enhancement for learning data value. Artificial intelligence can help in provide more hidden value behind data, and it can be used to transform Business and direct it to more optimized operations and cost-cutting. (Woods, 2017). Big data is useful in predictive analytics. Based on historical data, companies are determining their future value. This will also help in identifying the previous mistake, previous costs, and where to improve performance (Woods, 2017). "Some of the most common applications of predictive analytics include fraud detection, risk, operations, and marketing" (Woods, 2017). The current big data trend involves client-specific models and solutions. Those models help in delivering high-end solutions to customers and also improvising business models for the company. High-level visualization is useful in good decision-making and increases profits for the company (Woods, 2017). Earlier data technologies had a huge gap, and it is getting filled up by Big data. The other trend in big data is how to use those data and where to apply them properly. What will be the outcome? (Woods, 2017). "Data functions and quality trends make product and software most useful in big data. There are various data analysis tools and software available on the market right now that can really make a difference in your workplace. For that to happen, however, a data-literate workforce is vital, so it might be worth investing in tutorials to help your colleagues better process the data in front of them" (Woods, 2017). Big data also help the organization secure the data in the most effective way. This can protect data from hacking, and big data can help in identifying the threat and unknown user accessibility. Big data is more capable of governing the data and mitigating risk. (Big Data Examples: 6 Ways It's Transforming Business, 2019). "Big data can be used to analyze every individual action a customer takes when they land on a web page. By examining a user's keystrokes, mouse movements, and clicks, it's possible to predict which moves they'll make next" (Big Data Examples: 6 Ways It's Transforming Business, 2019). Using customer feedback, big companies analyze what customers are expecting and, based on the recommendation, try to improve the business model. (Big Data Examples: 6 Ways It's Transforming Business, 2019). Facebook, Twitter, and Google are all big companies that are targeting customers based on their activities on webpages and feeding them useful content for customers as well as redirecting it to companies' portals to fulfill their needs (Big Data Examples: 6 Ways It's Transforming Business, 2019).

6. Green Computing

A Green data center is a way of optimizing computing that can be routers, internet, electricity, servers, etc. it helps in reducing carbon and contributing to a green environment (Kirvan, 2022). Organizations can go green by following specific guidelines and standards. Organization can upgrade their equipment to meet excellent performance along with less power. Organizations can try to identify the old server, and if it is not in use, they can turn them down or sell them for recycling. As they turn off old servers, they are saving some money and electricity usage. "Reduce carbon footprint: To reduce carbon footprint, green data center service providers can opt for renewable energy sources, recycled materials, and use reclaimed cooling water" (Furtado, 2021). Use public cloud instead of own servers; that way, the organization has capabilities of automation which can be used for turning it on time and turning it off after office hours. It can also help in achieving high scalability if required for an organization to achieve more computers so that they do not need to buy any physical devices and will not be any maintenance over time. Public cloud service provider already has green computing certificate, which is more like cooling the servers in an optimized way and reducing power cost. It also helps monitor server usage and alert for unnecessary usage." The US Green Building Council (USGBC) oversees LEED for data centers, in which the rating system is designed and equipped to meet the needs of server racks, storage systems, and other IT infrastructure" (Furtado, 2021).

6.1 Company that has achieved green computing

Microsoft, a large cloud service provider, has already achieved green computing to help world IT businesses become more environmentally friendly. "Microsoft is now offering to include Emissions Impact Dashboard (opens in new tab), which will allow customers to monitor emissions produced from their use of Microsoft cloud hosting services" (McCurdy, 2022). Microsoft is working with many partners to use its green computing service. Microsoft is targeting to cut carbon emissions by 75 percent by 2030 (Tung, 2019). Microsoft has achieved carbon neutrality by setting up goals and using fundamental principles (Mandel, 2021). Microsoft has worked with green energy companies to set up a cloud facility that is more renewable in nature and generates energy from natural resources. "Since 2012, Microsoft has also implemented an internal carbon fee, currently set at $15 a metric ton, making business units pay for emissions related to their operations and electricity, as well as from business air

travel" (Mandel, 2021). Microsoft has provided a report from its supply chain about carbon emissions and fees (Mandel, 2021). Microsoft worked on different projects, including planting a tree, funding environmental projects, and working with companies that can help remove carbon emissions (Mandel, 2021). Microsoft is also working on its historical carbon emission and optimizing it in a way that can reduce not only carbon but also influence industries to go green. Microsoft has opened a research facility to achieve green computing, and they are actively working on different parameters to achieve the goal (Mandel, 2021). Microsoft is becoming a more significant player in buying renewable energy, which will help them neutralize its carbon emission. Microsoft is now utilizing the underwater system to cool down the server (Gartenberg, 2020). They are since going underwater will naturally help reduce servers' heat and minimize carbon emissions (Gartenberg, 2020). They are also experimenting with how servers will survive underwater and for a long time (Gartenberg, 2020). "Microsoft says the underwater data center had just one-eighth the failure rate of a land-based data center" (Gartenberg, 2020). They are also claiming that this will lead to energy saver design. The team is working on artificial intelligence and machine learning model to deploy it long-term without affecting the environment. This will be a colossal accomplishment if Microsoft successfully achieves full data center capabilities underwater (Sreemany, 2021).

7. Conclusion

This research paper talked about green computing, which is more critical in the existing world and saves the environment. Each organization should consult and achieve this as soon as possible to reduce carbon emissions. In this research, we have mentioned how a company can achieve this by following specific guidelines, e.g., using renewable energy, reducing machine usage, and monitoring smartly. The research paper also talked about data warehouse architecture and its importance. Data is considered digital money, and how one can use different data technologies to use data and turn it into meaningful output, which not only helps the company's performance but also improves customer experience. At the same time, big data is considered a large chunk of data that can be in any form, and how to use those data feeds it to machine learning modules to implement better artificial intelligence. By governing and providing better security, big data can also help reduce data loss. In today's world, data warehouse architecture is an essential part of big data, and if one can achieve it with green computing, the world will be a better place to live.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.

References

[1] *What is a data warehouse? | Definition, components, architecture | SAP Insights*. (n.d.). SAP. Retrieved September 30, 2022, from https://www.sap.com/insights/what-is-a-data-warehouse.html#:%7E:text=What%20are%20the%20key%20components,analyze%20data%20on%20the%20fly.
[2] Smith, J. (2022, July 30). *Data Warehouse Architecture, Components & Diagram Concepts*. Guru99. Retrieved September 30, 2022, from https://www.guru99.com/data-warehouse-architecture.html
[3] Simic, S. (2021, June 10). *Data Warehouse Architecture Explained*. Knowledge Base by phoenixNAP. Retrieved October 7, 2022, from https://phoenixnap.com/kb/data-warehouse-architecture-explained
[4] Molander, O. (2022, February 22). *5 Data Trends in 2022 - VALIDIO*. Medium. Retrieved October 7, 2022, from https://medium.com/validio/5-data-trends-in-2022-4035c099aac2
[5] Ghodsi, A., & Casado, M. (2021, July 28). *Data Alone Is Not Enough: The Evolution of Data Architectures*. Future. Retrieved October 7, 2022, from https://future.com/podcasts/evolution-of-data-architectures/
[6] Woods, P. (2017, July 13). *Why Big Data demands new technology*. ITProPortal. Retrieved October 7, 2022, from https://www.itproportal.com/features/why-big-data-demands-new-technology/
[7] *Big data examples: 6 Ways it's transforming business*. (2019, July 10). www.umassglobal.edu. Retrieved October 8, 2022, from https://www.umassglobal.edu/news-and-events/blog/big-data-examples-ways-its-transforming-business
[8] Kirvan, P. (2022, May 10). *green data center*. SearchDataCenter. Retrieved October 8, 2022, from https://www.techtarget.com/searchdatacenter/definition/green-data-center
[9] Furtado, P. (2021, June 14). *Developing green data centers for a sustainable future*. Retrieved October 8, 2022, from https://www.g2.com/articles/green-data-centers-for-a-sustainable-future
[10] McCurdy, W. (2022, May 18). *Microsoft wants to help your business go green with new sustainability cloud*. TechRadar. Retrieved October 8, 2022, from https://www.techradar.com/news/microsoft-wants-to-help-your-business-go-green-with-new-sustainability-cloud
[11] Tung, L. (2019, April 16). *Microsoft's green plan: Our data centers will run on 60% renewable energy by 2020*. ZDNET. Retrieved October 8, 2022, from https://www.zdnet.com/article/microsofts-green-plan-our-data-centers-will-run-on-60-renewable-energy-by-2020/
[12] Mandel, K. (2021, November 27). *'It's critical': can Microsoft make good on its climate ambitions?* The Guardian. Retrieved October 8, 2022, from https://www.theguardian.com/environment/2021/nov/27/its-critical-can-microsoft-make-good-on-its-climate-ambitions
[13] Gartenberg, C. (2020, September 14). *Microsoft's underwater server experiment resurfaces after two years*. The Verge. Retrieved October 8, 2022, from https://www.theverge.com/2020/9/14/21436746/microsoft-project-natick-data-center-server-underwater-cooling-reliability

[14] Sreemany, A. (2021, December 31). *Microsoft's Project Natick - The Treatise*. Medium. Retrieved October 8, 2022, from https://medium.com/the-treatise/microsofts-project-natick-4965f35da478
[15] Newman, A. (2008, Jun 25). Green computing can save energy -- and save your business money: The increasing global focus on environmental issues has now expanded to include the world of computing. Is "green computing" a marketing tool or is it something real that can help your business and the environment? Asheville Citizen - Times https://www.proquest.com/newspapers/green-computing-can-save-energy-your-business/docview/438759867/se-2
[16] Gibbs, M. (2013). Big Data, Big Business, Big Government, Bigger Brother: Big Data will change business and government and if CISPA is made law, Bigger Brother's will be right around the corner.. Network World (Online), https://www.proquest.com/trade-journals/big-data-business-government-bigger-brother/docview/1312292248/se-2
[17] Saddad, E., El-Bastawissy, A., Mokhtar, H. M. O., & Hazman, M. (2020). Lake Data Warehouse Architecture for Big Data Solutions. International Journal of Advanced Computer Science and Applications, 11(8)https://doi.org/10.14569/IJACSA.2020.0110854
[18] FOCUS 73: Leveraging Entrepreneurship To Harness New Technology | French Chamber of Commerce in Singapore. https://www.fccsingapore.com/news/n/news/focus-73-leveraging-entrepreneurship-to-harness-new-technology.html
[19] Understanding Edge Computing and its advantages - Aplyca. https://www.aplyca.com/en/blog/edge-computing-and-its-advantages
[20] Chambers, I. (2022). The development of visions and strategies for Australia's sustainable future. https://doi.org/10.25911/CASJ-VX13
[21] Chambers, I. (2022). The development of visions and strategies for Australia's sustainable future. https://doi.org/10.25911/CASJ-VX13
[22] Advancing Parliamentary Operations through Artificial Intelligence: A Case Study of the Grand National Assembly of Turkey. https://library.bussola-tech.co/p/advancing-parliamentary-operations