Proceedings of the 24th International Conference on Automation & Computing, Newcastle University, Newcastle upon Tyne, UK, 6-7 September 2018

Integrated Architecture of Data Warehouse with Business Intelligence Technologies

Ch Anwar ul Hassan, Rizwana Irfan, Munam Ali Shah
Department of Computer Science.
COMSATS Institute of Information technology Islamabad, Pakistan
anwarchaudary@gmail.com, rizwana.irfran@comsats.edu.pk, mshah@comsats.edu.pk

***Abstract*—Business Intelligence (BI) is the process to extract information from data then get knowledge from that information to take the decisions. This paper shows the effectiveness of BI technologies with data warehouse, for decision making. In this paper, we deploy the Integrated Proposed Architecture (IPA) on W category hospital in order to manage and monitor the data effectively for analysis and decision making. The accuracy of IPA is 93% in term of information analysis that is 6% better than Traditional Data warehouse Architecture (TDA). The IPA is also able to support dashboard management, multidimensional data model, perform online analytical processing, perform user authentication and generate dynamic reports via BI technologies.**

Keywords-Data warehousing, Business Intelligence, Data Analysis, Architecture

I. INTRODUCTION

Process design and automation are progressively increased to enhance the quality and productivity of the organization. Every organization performed different processes to maintain their operation data in organizational databases. In the organizational databases, the data is growing day by day. In every moment updated data is available for processing. Integrated the available data in a structured form and store in to a single repository. On that repository we perform analytical processing for the decision making so, the concept of Data Warehouse (DW) has been evolved. Where data is processed and then stored for future use.

Data warehouse address two main requirements of business enterprises:

- Data integration
- Decision support system.

Multiple approaches are proposed to build Data Warehouse (DW). Most common approaches are; top-down, bottom-up, hybrid, and federated [1]. Before selecting any approach to implement or build DW, it is essential to determine the organizational standard. On the base of determined standard, pick-up the suitable architecture that satisfies the organization needs [2].

Business Intelligence (BI) technologies in data warehousing environment provide flexibility to manage and monitor the data also provide flexibility to perform effectively analytical processing. Several BI architecture are already proposed. In [3], Author perform Online Analytical Processing (OLAP) and dashboard management was done in [5]. These architectures are varies according to their structures such as their processes, components and relationships Data warehouses are built to collect and integrate the data from heterogeneous sources and then restore the data for analysis and decision making.

Nowadays, organizations required a well-organized data storage and retrieval systems to perform analytical processing and make successful business decisions. For that organizational purpose, we are proposing an integrated DWA to make effective business decision using BI technologies. BI provides current, historical, predictive and perspective view of business processes. We also perform analytical processing, query reporting, data mining, process mining, performance management dashboards, and prescriptive and predictive analysis using BI technologies. In this paper we are proposed integrated Data warehouse Architecture with Business Intelligence technologies. Using the Integrated Proposed Architecture (IPA) various comprehensive analyses could be conducted such as OLAP analysis, reporting, dashboard, slice and dice.

Rest of the paper is organized as follow: section II reflects the state of the art, section III problem formulation. In section IV proposed framework is described, section V analysis and results. Finally section VI conclusion and future work.

II. RELATED WORK

Many researchers and practitioner proposed Data Warehouse Architecture (DWA). In [1], [2] authors, presented multiple DWA such as; single layer architecture, multiple layers architecture and compared them in the terms of efficiency, consistency, integrity and accuracy.

Researchers also integrate BI and DW to provide organizational operational platform for decision making. In [3], Author proposed integrated approach to deploy DW in BI environment to perform analytical processing. Ghosh discussed OLAP, an integrated part of BI and implemented proposed architecture in Fast Moving Consumer Goods (FMCG) Company. Generate efficient reports and performed query processing which incorporates Data Marts (DM), DW and Virtual Data Warehouse (VDW) [3].

From Educational data is growing rapidly, to manage the academic records and data we need a well-organized architecture. Architecture is proposed in Polytechnic Institute of Leiria for undergraduate Informatics

Engineering degree program to manage educational data that demonstrate the proficiency of data warehouse [4]. An EduBI framework is also introduced in [5] to collect educational data from several sources and restore into single Educational Data Warehouse (EDW) to generate reports. However, when we get the data into single respiratory. We can apply multiple data mining techniques to retrieve the desire results. These results would be used to find the deficiencies in educational systems and improved the educational structure. In [6], Author proposed performance dashboard for Romanian universities to improve the education quality and integrate advancement in the scientific and management process of universities. Dashboards display effective information that would help for better decision making. However when developing the performance management dashboard for universities, DW infrastructure in sense of security and query system must be perspective and well structured.

Multiple research areas still need to explore in data warehousing, in this regards researcher also conduct multiple survey and identify these categories; DW architecture, DW security, DW design, DW evolution for future research [7].

Security of the data is also one of the important issues in DW. In this regards authors proposed different security frameworks, security models and performed vulnerability checks to handle the security issues in data warehousing[8]–[10]. These security measures are performed in terms of hardware, internet security, specifically for DW environment [11], [12]. We are adding authentication layer in our proposed architecture to handle the security issues.

From last decades, researchers proposed various solutions to maintain and analyse data effectively in DW. Authors evolve the DW architecture, their components, and analysis tool. Proposed architecture oriented, model-driven approaches [13], [14].

In DW components, ETL is one of the important component which are intensely influenced by the changing and evolution of the business requirements and their complexity. Many researchers and practitioner work on ETL to make it more effective such that authors introduced BPML, spatial ETL using Geokettle, Domain-specific language (DSL) for ETL, Generic transformation from star schemas to big data or for generating ETL conceptual schemas, translating physical schema, star scheme into logical NoSQL schema, , NoSQL Graph-oriented model [15]–[18]. These are the approaches that are used to analyse the data more effectively for future business decisions.

DW analysis tool like OPLA are also optimized to efficiently extract the valuable knowledge and analyse the information. In this regards authors introduces Velocity OLAP, Incremental OLAP, OLPA graphs, OLAP cubes for multi-dimensional data analysis and complex query optimization [19]–[25]. In our proposed architecture, we are using BI technologies to reduce the efforts of integration the external tool. However, DW is still exploring field to maintain the data more precisely, in order to accurately analyse the information. Summary of literature review is describe in Table 1.

TABLE I. CONCISE SUMMARY OF LITERATURE IN THE FIELD OF DATA WAREHOUSING

References	Issue discussed/work done
[1]	Proposed different DWA, However never integrate BI technologies
[2]	Proposed Layered DWA not integrated BI technologies to optimize performance,
[3]	Integrated DWA, not perform real-time analysis, data security issues
[4]	Focus on dimension modeling, using SQL and Graphically language, specific for educational purpose
[5]	Cognos BI Tool, EDWA, not made real-time analysis
[6]	Dashboard Management, Specific for university portal, performance management
[8]	Data warehouse security issues
[9]	Vulnerability check and security models are proposed
[10]	DW security framework
[11]	DW hardware, system, internet security measures
[12]	Security approaches for DW environment
[13]	Architecture approach for DW testing
[14]	Model driven approach for DW requirement
[15]	Spatial DW using Geokettle
[16]	BPML, DSL for ETL
[17]	Transition of conceptual schema into NoSQL logical schema
[18]	Introduces transformation to big data from star schemas
[19]	Introduced VelocityOLAP
[20]	Automatically creation of DW structure and OLAP cube
[21]	Introduced OLAP on graph data
[22]	HaCube: Map reducer for efficient OLAP cube materialization and view
[23]	Effectiveness of DW in e-Governance
[24]	Query optimization to analyze multi-dimensional data
[25]	Introduced IncrementalOLAP

III. PROPOSED ARCHITECTURE

In recent days decision makers demand effective knowledge representative and decision support systems, to take effective business decisions. DW with BI is the ultimate way to design the effective structure for decision makers. In DWA ETL, reporting and analytical tools provides an incorporated atmosphere for making business decisions and can effectively measure, monitor and manage business data.

By the proposed architecture we perform;

- Real-time data analysis.
- GUI Based, Less query typing effort (Easy to analyze for non-technical person).

- More quality data achieved in short time (Chances of better decision increased).
- Simple architecture (Flexible Model).
- Quick response, efficient view and quality results.

In this section, we describe the proposed integrated architecture and concentrate on data quality and flow of information in the system. Architecture consist up on these components; data sources, data integration, data warehouse, data abstraction model, performance management and end users Figure 1.

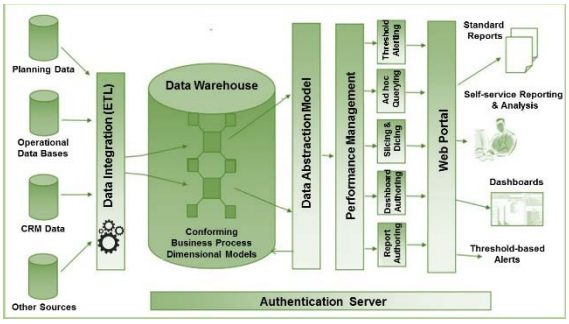

Figure 1. Integrated Proposed Data Warehouse Architecture

A. Data Source

Nowadays, data is growing at every moment. To transform the data into required information and obtain the desire knowledge from that information, we need well organized and structured form of data. Data sources contains structured, unstructured and semi-structured form of data the reason is that the data is collected from heterogeneous sources. We need to convert that data into a structured data for better analysis to take effective and timely decision.

These different formats of data can be attain from two types of sources;

- Internal data source
- External data source

Internal data sources are the data that is taken and sustained by organizational operational systems, data inside the organization, data that linked to the business operations. Data that is initiated outside the organization refer as external data sources. This type of data can be extracted from external sources such as market organizations, Internet, business partners, governments. These data are related to the market, technology, competitors, and environment.

B. Data Integration

In Data Integration, ETL is performed. ETL contain three main processes;

- Extraction
- Transformation
- Loading

Extraction is the process to collect or extract the relevant data from data sources as described above. The data collected from data sources like internal and external data sources are redundant, inconsistent, incomplete and not integrated. Then the extracted data is sent to the data staging area. It is the temporary data storage area, where the data is stored to avoid the data extraction need again in case of any problem. After extraction, data will moved to transformation and cleansing process. Transformation is the process of transforming or converting data into consistent format, according to the business rules that is used for analytical processing. Standardizing data definition and describing business logic for data mapping also includes in data transformation. When data transformed and cleansed, then stored in staging area (temporary data storage) to avoid the data transformation necessity again in case when data loading event fail. Loading is the final phase of ETL process. In loading, the data are loaded into target repository from staging area.

C. Data Warehouse:

Data collected from heterogeneous sources are integrated and converted into structured format then loaded that data into a single respiratory. This respiratory is called data warehouse. Inmon defines data warehouse as a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process".

D. Data Abstraction Model

Data abstraction model introduced the best practice in business intelligence (BI) to take better business decision. Through data abstraction, we achieve more accurate and secure data. It can be classified in three layer;

These different formats of data can be attain from two types of sources;

- Application Layer
- Business Layer
- Physical Layer

The Application Layer provide the platform to the data consumers to consume the data obtained from business layer. The Business Layer is established on the standard to describing significant business entities such as products and customers. In business layer, we define the data model and their relations. Typically data modeler work with experts and data providers define a set of logical view of the data that represent the business entities. These views are reusable components for multiple users or consumer in application layer. Integrated the data sources in physical layer, into abstraction level. Value-added tasks such as value formatting, name aliasing, derived columns, and data type casting, and data quality checks are also described in physical layer [26]

E. Performance Management

Performance Management facilitate executives to measure, manage and monitor organization performance more efficiently and effectively. In Performance Management, we identify and monitor the performance metrics and focus on indicators to perform further analysis at the appropriate detail level. Performance indicator are related to the organizational objective and strategies.

Monitor the individual and organizational performance according to these strategies to understand the current status of business. We also performed ad-hoc query, reporting, online analytical processing, data visualization and dashboard management in performance management area. In few scenarios, web portal is eliminated we directly communicate to performance management instead of web portal. The end user consists of tools that display information for different users in different formats.

F. Authentication Server

Authentication layer is used to secure the information form unauthorized access. Information security is the practice to prevent the information from disruption, disclosure, recording, modification, and destruction.

IV. TRADITIONAL DATA WAREHOUSE ARCHITECTURE

Traditional Data warehouse Architecture (TDA) Figure. 2 is taken from [27] in order to compare the proposed architecture to the traditional architecture. TDA is described in [1], [2], and [27]. In TDA, OLAP server is used for analytics. Data abstraction model, Performance management and Authentication server are not in TDA.

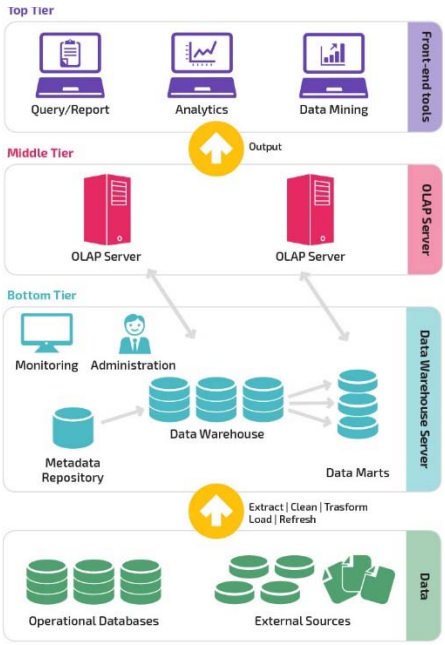

Figure 2. Traditional Data warehouse Architecture [27]

V. IMPLEMENATATION/DEPLOYMENT

We are implementing the Integrated Proposed Architecture (IPA) and Traditional data warehouse (TDA) architecture on medical data of W Category hospital to show the capability of architecture components and effectiveness of the IPA. IPA plays a significant role by optimizing the time to perform current and historical data analysis on medical record.

The medical data of the hospital, which we used for analysis are in heterogeneous formats. So first we performed data integration; extract the data from multiple resources, then transform and clean the data to convert it in uniform format at the end load the data into data warehouse.

Now through data abstraction, we communicate to the data warehouse and design the business models, multi-dimensional model as shown in IPA to support analysis. According to these business models as shown in Figure 3.

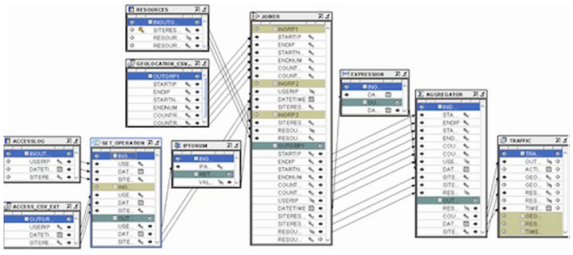

Figure 3. Multidimenstional Design

We utilized the data in data abstraction model (as data abstraction is a three-layer process). When the final data is loaded on data abstraction model continue to the performance management step (accordance to business requirements). In performance management, we perform such as dashboard management, analysis, reporting, OLAP, slice and dice using BI technologies. Web portal utilize to display the desired results according to the demand of user as shown in Figure. 4.

Figure 4. Dashboard Management

Pneumonia, Acute Myocardial infarction, stroke, pacemaker, cholecystectomy, carotid Endarterectomy, PCI and PTCA, Cardiac Surgery. These are the different patient quality measurements. To measure and analyze the patient data, using IPA we analyze the patient data on a single click. As in Figure 4 the data of "Pneumonia" disease is analyzed and described. From the last year data we analyze that mostly patient in W category hospital suffer Pneumonia in the month of March.

Pneumonia is a common illness in all part of the world. It's a major cause of death among all age groups, insufficient treatment of Pneumonia leads to an 11 times higher death rate.

Authentication layer is used for authentication purpose and also to assign the specific privileges to different users to maintain the data security.

VI. RESULT ANALYSIS

In result and analysis, we are comparing our architecture with Traditional Data warehouse Architecture (TDA) and shows the effectiveness of our proposed architecture. TDA takes more time for analysis and also not efficiently and accurately analyze the information that affects the business decision. We are comparing architecture in three different aspects. How much complete and accurate information in aspects of query performance and analysis, how much system is scalable, flexible, manage daily load and backup recovery and also their impact on organization for decision making.

We evaluate the TDA and IPA on three parameters; Effort, Efficiency/ Response time and Accuracy.

In case of effort, we involved 30 employees to perform comparatively analysis on the efforts involved to implement or operate the IPA vs TDA. Each employee like the proposed architecture except 2, 3 employees. These 2, 3 employees are those who have already familiar with ETL. Our Proposed architecture reduce 90% efforts comparing to TDA. The reason is that proposed architecture TDA is on development mode (integrate external tool for analysis e.g. generating reports) and IPA is on clicked based, easy to understand and don't need any external tool for analysis.as shown in Figure 5.

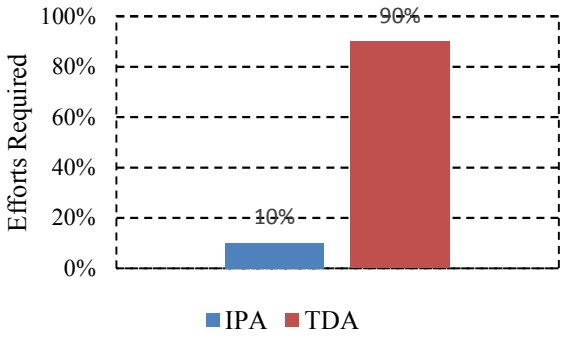

Figure 5. Efforts Rquired for DW

Our proposed integrated data warehouse architecture is more efficient than traditional architectures. IPA performed ETL and analysis, using the BI technologies. So, for ETL and analysis IPA always take less time compared to one on which first you need to perform ETL process then use some external tool for analysis. In order to check the efficiency or response time, we analyze 1GB data on both architectures. Our IPA respond fast enough, just in 122 sec and TDA in 431sec (7min and 11 sec) as described in Figure 6.

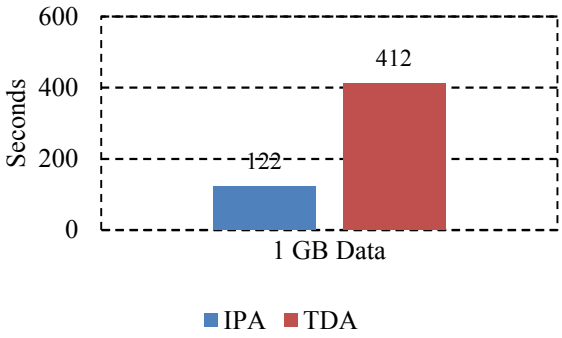

Figure 6. Data analysis w.r.t time

Our results shows that the proposed architecture performed 6 % better than traditional architecture as results as shown in Figure 7. Proposed architecture performed 93% accurately analysis as compared to traditional it gives 87% accuracy. We also performed ad-hoc query, reporting, online analytical processing, data visualization and dashboard management through proposed architecture.

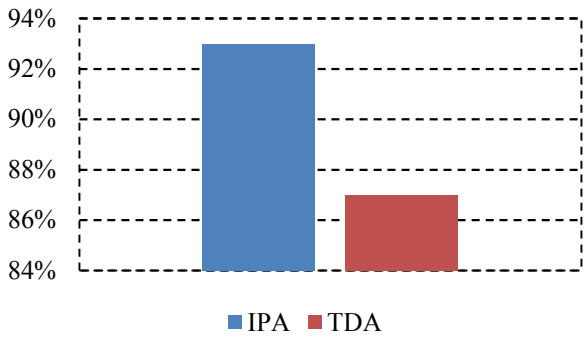

Figure 7. Accuracy in Informaction Analysis

VII. CONCLUSION

The proposed architecture integrate the DW and BI technologies in order to support analysis, reporting and decision making. This architecture deployed in W-category hospital to manage and monitor the health records to perform analysis.

In future, we deploy IPA in cloud environment and integrate it with various hospitals to maintain health records and confirming patient privacy. These health records will be useful in future to diagnose the diseases by some common symptoms also identify the specific number of patients in certain region which was affected by certain disease.

REFERENCES

[1] T. Risch et al., "Data Warehousing Systems: Foundations and Architectures," in Encyclopedia of Database Systems, Boston, MA: Springer US, 2009, pp. 684–692.

[2] G. Blazic, P. Poscic, and D. Jaksic, "Data warehouse architecture classification," in 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2017, pp. 1491–1495.

[3] R. Ghosh, S. Haider, and S. Sen, “An integrated approach to deploy data warehouse in business intelligence environment,” in Proceedings of the 2015 Third International Conference on Computer, Communication, Control and Information Technology (C3IT), 2015, pp. 1–4.

[4] J. V. Ramos and R. Oliveira, “Data warehousing in the context of a Bologna undergraduate degree,” in 2013 IEEE Global Engineering Education Conference (EDUCON), 2013, pp. 540–544.

[5] Aziz, A. A., Jusoh, J. A., et al. “A Framework For Educational Data Warehouse (Edw) Architecture Using Business Intelligence (Bi) Technologies.,” J. Theor. Appl. Inf. Technol., vol. 69, no. 1, 2014, pp. 50–56.

[6] Muntean, M. I. H. A. E. L. A., et al. "Performance dashboards for universities." Proceedings of the 2nd International Conference on Manufacturing Engineering, Quality and Production Systems. 2010, pp. 206-211.

[7] P. Chandra and M. K. Gupta, “Comprehensive survey on data warehousing research,” Int. J. Inf. Technol., vol. 10, no. 2, pp. 217–224, Jun. 2018.

[8] S. Aleem, L. F. Capretz, and F. Ahmed, “Security Issues in Data Warehouse,” pp. 15–20, 2014.

[9] S. L. Gupta, S. Mathur, and P. Modi, “Data Warehouse Vulnerability and Security,” Int. J. Sci. Eng. Res., vol. 3, no. 5, 2012.

[10] Vishnu, B., Manjunath, T. N., & Hamsa, C,”An effective data warehouse security framework” In IJCA Proceedings on National Conference on Recent Advances in Information Technology NCRAIT pp. 33-37, 2014.

[11] Jaiswal, Arvind,“Security Measures For Data warehouse,” Int. J. Sci. Eng. Technol. Res., vol. 3, no. 6, pp. 1729-1733, 2014.

[12] S. Kumar, B. Singh, and G. Kaur, “Data Warehouse Security Issue,” Int. J. Adv. Res. Comput. Sci., vol. 7, no. 6, 2016.

[13] N. Elgamal, A. El-Bastawissy, and G. Galal-Edeen,"An Architecture-Oriented Data Warehouse Testing Approach." In COMAD, pp. 24-34. 2016

[14] J. Feki, S. Taktak, S. Alshomrani, J. Feki, and G. Zurfluh, “The Power of a Model-Driven Approach to Handle Evolving Data Warehouse Requirements,” pp. 169-181 2017.

[15] W. Astriani and R. Trisminingsih, “Extraction, Transformation, and Loading (ETL) Module for Hotspot Spatial Data Warehouse Using Geokettle,” Procedia Environ. Sci., vol. 33, pp. 626–634, Jan. 2016.

[16] B. Oliveira and O. Belo, “A Domain-Specific Language for ETL Patterns Specification in Data Warehousing Systems,” Springer, Cham, 2015, pp. 597–602.

[17] R. Yangui, A. Nabli, and F. Gargouri, “Automatic Transformation of Data Warehouse Schema to NoSQL Data Base: Comparative Study,” Procedia Comput. Sci., vol. 96, pp. 255–264, Jan. 2016.

[18] M. Golfarelli and S. Rizzi, “From Star Schemas to Big Data: 20 $$+$$ Years of Data Warehouse Research,” Springer, Cham, 2018, pp. 93–107.

[19] F. Dehne, D. Robillard, A. Rau-Chaplin, and N. Burke, “VOLAP: A Scalable Distributed System for Real-Time OLAP with High Velocity Data,” in 2016 IEEE International Conference on Cluster Computing (CLUSTER), pp. 354–363, 2016.

[20] Geary, Nigel, et al. "Method and apparatus for automatically creating a data warehouse and OLAP cube." Jun. 2017.

[21] L. Gómez, B. Kuijpers, and A. Vaisman, “Performing OLAP over Graph Data,” in Proceedings of the International Workshop on Real-Time Business Intelligence and Analytics - BIRTE ’17, 2017, pp. 1–8.

[22] Z. Wang, Y. Chu, K.-L. Tan, D. Agrawal, and A. EI Abbadi, “HaCube: Extending MapReduce for Efficient OLAP Cube Materialization and View Maintenance,” Springer, Cham, 2016, pp. 113–129.

[23] R. K. Arora and M. K. Gupta, “e-Governance using Data Warehousing and Data Mining,” Int. J. Comput. Appl., vol. 169, no. 8, pp. 28–31, 2017.

[24] B. Tang, S. Han, M. L. Yiu, R. Ding, and D. Zhang, “Extracting Top-K Insights from Multi-dimensional Data,” in Proceedings of the 2017 ACM International Conference on Management of Data - SIGMOD ’17, 2017, pp. 1509–1524.

[25] K. Zeng, S. Agarwal, and I. Stoica, “iOLAP,” in Proceedings of the 2016 International Conference on Management of Data - SIGMOD ’16, 2016, pp. 1347–1361.

[26] Singh Chaitanya, “Data Abstraction in DBMS.” [Online]. Available: https://beginnersbook.com/2015/04/levels-of-abstraction-in-dbms/. [Accessed: 08-May-2018].

[27] Panoply, “Data Warehouse Architecture: Traditional vs. Cloud” [Online]. Available: https://panoply.io/data-warehouse-guide/data-warehouse-architecture-traditional-vs-cloud/. [Accessed: 08-May-2018].