**Datawarehouse Architecture: Kimball vs. Inmon**

*Abstract: Data Warehouses are important for companies in the field of Business Intelligence. Analyzing data from a data warehouse allows companies to make strategic, conscious decisions going forward. However, architecture of these data warehouses in the world of database professionals remains a controversial subject. In this essay, I attempt to explain what a data warehouse is, articulate what I learned from my research, explore the technical design and pattern differences between the Kimball and Inmon architectures, and finally come up with my own conclusion on when to use either one.*

**Introduction**

We live in a world driven by data. Data is valuable to every industry ranging from tech, education, sports, and even health. However, the data that industries and businesses analyze are not from databases, but rather data warehouses. A data warehouse is a repository where data from one company is kept in a single place. This includes data from different sources as well as current and historical data. Data from this repository can be analyzed, whereas data in a traditional relational database is designed for transaction processing.

**My Research: What Are Data Warehouses?**

From my research, I discovered that a data warehouse requires architecture. According to Lamia Yessad and Aissa Labiod of National High School for Computer Science in Algeria, a data warehouse “… requires a modeling approach that considers all aspects of development such as data modeling, project management, risk management, deployment, and many other essential aspects” (Yessad and Labiod 95). Because of the many aspects that need to be considered, developing architecture for a data warehouse should be carefully constructed and deliberate. We have two different approaches to storing and accessing this data: Kimball Architecture and Inmon Architecture. Professors Jeff Lawyer and Shamsul Chowdhury from Roosevelt University write that “… the Inmon style calls for an atomic-level, third normal form relational format in which to store extracted and transformed data, while the Kimball style calls for a multidimensional style […] in which to store extracted and transformed data” (Laywer and Chowdhury 2). In other words, Inmon architecture stores ‘atomic’ data at the lowest level of detail, while Kimball utilizes multiple tables. However, there is no best architecture. There are certain instances where one is better than the other.

**Kimball Architecture**

First, I will discuss the Kimball architecture. Its approach to data warehousing has been employed in multiple business areas and is highly successful. A key feature of this approach is the star schema. According to Robert Hart and Alex Mu-Hsing Kuo of University of Victoria in Canada, the schema is composed of “… a central database entity known as the fact table surrounded by supporting entities that are referred to as dimension tables” (Hart and Kuo, 697). The fact table would consist of small numbers known as measures and foreign keys to the surrounding dimension tables. Those surrounding tables are a representation of the attributes that define the business process. The purpose of this is to minimize the size of individual fact records in order to optimize performance as it typically contains a large volume of records. The dimension tables are important because it is the backbone of the data warehouse. To demonstrate the Kimball architecture’s usefulness, Hart and Kuo implemented a Kimball Integrated Data Warehouse to meet health care research needs. In their findings, the authors note that their experiment has “… the ability to increase insight into [their] subject areas at minimal effort. [Their] star schemas remain relatively simple and built into constellations of information or rapidly extended with additional information rather than remain as separate areas” (Hart and Kuo 705). Essentially, they applied certain techniques to a Kimball Data Warehouse. They established a unique identifier for every table entity record, and developed a process for dimension table association. This allowed them to create an association between their dimensional models that can be used to capture Emergency Encounters at hospital emergency rooms.

**Inmon Architecture**

Secondly, I will discuss the Inmon architecture. According to Bill Inmon, a warehouse “… is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management’s decision making process” (Inmon). This architecture allows the storage of all events of a company and commits important resources to archive the system. It’s a data-driven approach, meaning the company’s data are loaded without knowing the user requirements. It divides the database into four levels including: operational, Atomic (data warehouse), departmental (data marts) and individual levels. Essentially, data from the database are loaded into the data warehouse using Extract, Transform, and Load. The users create the last level when they analyze and exploit the loaded data into data marts. This architecture can also be used in the health care industry as Dr. The Ying Wah and Ong Suan Sim from University of Malaya used the Data Warehouse for Lymphoma Decision Support System. Their Data Warehouse utilized the Inmon architecture, and in the research the “… Lymphoma or Lymph node cancer is selected as the subject matter. All data that are related […] are stored in the data warehouse. This subject orientation promotes an easy-to-understand data presentation format for doctors and other health professionals” (Wah and Sim, 59). The architecture allowed their study to be reliable and easily evaluated. The basic parameters for its evaluation were Extract, Transform, and Loading process speed, disk space consumption, query performances, and user friendliness. Overall, their development of the data warehouse can serve as a building block of a Clinical DSS and as data input to data mining tools.

**Conclusion**

In conclusion, the two architectures of Data Warehouses have their differences. Kimball architecture implements a unique star schema that utilizes a fact table surrounded by dimension tables. It focuses more on ease of end user accessibility. In contrast, the Inmon architecture stores data at the lowest level of detail. It is subject oriented and data driven. What I noticed from the academia papers I read was that both architectures were implemented in the Health industry. For a company that wants to opt for a decision-making system, they must carefully consider which architectural approach best suits their needs. In one instance, I would choose the Kimball approach if we had to model data marts because the of the query performance. On the other hand, I would recommend the Inmon approach if the requirements are not defined or are very scalable. All in all, both of these approaches can face problems if the data sources change, thus forcing a redesign of the data warehouse.

References

Hart, Robert, and Alex Mu-Hsing Kuo. "Meeting Health Care Research Needs in a Kimball Integrated Data Warehouse." *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (2016): n. pag. Web.

Inmon, Bill. “Building the Warehouse”, 2nd. Newyork: John Wiley & Sons, 1996.

Lawyer, J., and S. Chowdhury. "Best Practices in Data Warehousing to Support Business Initiatives and Needs." *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the* (2004): n. pag. Web.

Wah, Teh Ying, and Ong Suan Sim. "Evaluating a Data Warehouse for Lymphoma Diagnosis and Treatment Decision Support System." *2010 Fourth UKSim European Symposium on Computer Modeling and Simulation* (2010): n. pag. Web.

Yessad, Lamia, and Aissa Labiod. "Comparative Study of Data Warehouses Modeling Approaches: Inmon, Kimball and Data Vault." *2016 International Conference on System Reliability and Science (ICSRS)* (2016): n. pag. Web.