IST722: Unit 02 Participation Questions

**This is an individual assignment.**

**Before you begin, please make sure you’ve read and understand 1) our class honor code, 2) course policies on late work and 3) participation policies as posted on the syllabus. “I didn’t know” is not an excuse.**

**You should cite your sources in a standard format like MPA or APA and include a list of works cited.**

|  |  |
| --- | --- |
| Your Name: | HARSH RAJESH DARJI |
| Your Email: | hrdarji@syr.edu |

# Instructions

Answer each of the following questions as concisely as possible. More is not necessarily better. Please justify your answer by citing your sources from the assigned readings from our textbooks, our class lectures, or online if directed to do so. Be sure to cite in text and include a list of works cited. Place your answer below each question. When you’re finished, print out this document and bring it to class as part of your participation grade.

# Questions

1. What is the difference between a primary key and a business key? What constitutes a good choice for primary key? What is the optimal primary key for an OLTP system and why?

A primary key is special relational database column or combination of columns which is used to uniquely identify all table records. While, business key is a unique, which is formed of attributes that already exists in the real world. Primary key may or may not have logical relationship to attributes within that row. But Business key always has a logical relationship to the attributes within that row. Primary key in relational database can be auto generated depending upon the defined sequence but business key cannot be auto generated. E.g.: For a person attribute, database can generate keys like Person ID:1,2, 3.. which uniquely identifies person. Here, Person ID becomes a primary key. While, combination of Person’s first name, last name and age can be a business key.

Choosing good primary key:

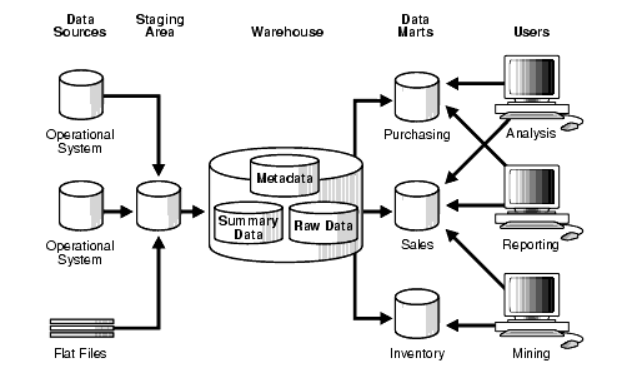
* Never Null – Primary key cannot be null. This is an inviolate rule of the relational model.
* Brevity – Primary key must be one column instead f combination of columns if possible. Because when we want to query our database, the SQL server can speed up computations.
* Simplicity – Primary key should not have embedded spaces and special characters. Because such entries are hard to work with in SQL queries and join operations.
* Data Type – Primary key should be Integer data types if possible, followed by fixed-length character. Because SQL server processes number data type faster than character data type.
* Nonidentifying value – Primary key should not have a meaning because when primary key are is descriptive it might become obsolete.
* Never Change – Primary key must never change.

The optimal primary key for an OLTP system would be a Primary Key with integer types generated sequentially

1. Compare and contrast the Inmon and Kimball definitions of Data Warehouse and Data Mart. How are they similar? Different?

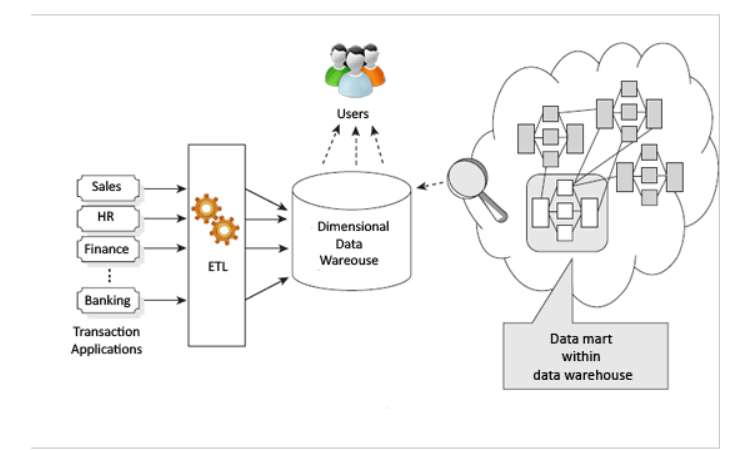
|  |  |  |
| --- | --- | --- |
|  | **Inmon Approach** | **Kimball Approach** |
| **Data Warehouse is:** | Normalised Tables | Dimensional Models |
| **Purpose of Data Warehouse is:** | Data Integration | Query |
| **How is Data Warehouse Built?** | Data first (iterative, bottom-up) | Process first (waterfall, top-down) |
|  | **Inmon Data Warehouse** | **Kimball Data Warehouse** |
|  | Relationship Modeling | Dimension Modeling |
|  | Entity-Relationship Model | Fact Tables and start schema |
|  | Tables in third normal form | Tables are denormalized |
|  | Many Tables using Joins | Easier for end users to understand |
|  | Built for Data Integration | Built for ad hoc queries |
|  | Indirect Access of data by users | Direct Access of data by users |

**Inmon Data Warehouse:**



According to Inmon Approach, data warehouse truly serves as the single source of the truth, as it is the only source for all the data marts. But more ETL work is needed as the data marts are built from the data warehouse. But at the same time data update anomalies are avoided because of very low redundancy.

**Kimball Data Warehouse:**



According to Kimball approach, data marts are within the data warehouse. So, the essence of one truth is lost here. Redundant data can cause data update anomalies. This approach works well for department wise metric as the data marts are geared towards department-wise reporting. But, if new columns are added to fact tables, fact tables will become larger and will not perform well. Also, Integration of legacy data into the data warehouse can e a complex process.

Both Inmon and Kimball approach can successfully deliver data warehouse. The similar variables that are required in both Inmon and Kimball Approach are:

* Source
* Staging
* ETL
* Data Marts
* Business requirements
* Time attribute of Data

WORKS CITED:

Breslin, Mary. 2004. “Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models” Business Intelligence Journal, Winter 2004. Accessed May 22, 2016.

Inmon, W. H. Building the Data Warehouse, Fourth Edition. John Wiley & Sons., 2005.

Marakas, George M. Modern Data Warehousing, Mining, And Visualization. Prentice Hall, 2003.

Inmon, W. H. 2010. “A TALE OF TWO ARCHITECTURES” [InmonCif.com](http://inmoncif.com/). Accessed May 23, 2016.

<http://www.inmoncif.com/products/A%20TALE%20OF%20TWO%20ARCHITECTURES.pdf>

Kimball, Ralph, and Margy Ross. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition. John Wiley & Sons. 2013. Books24x7.

Stanford. 2003. “Data Warehousing Concepts” Stanford.edu. Accessed May 26, 2016.<https://web.stanford.edu/dept/itss/docs/oracle/10g/server.101/b10736/concept.htm#i1006297>

Zentut. 2016. “Ralph Kimball Data Warehouse Architecture” [Zentut.com](http://zentut.com/). Accessed May 25, 2016.  <http://www.zentut.com/data-warehouse/ralph-kimball-data-warehouse-architecture/>