IST722: Class Exercise 2

**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: | Abhijith Anil Vamadev |
| Your Email: | avamadev@syr.edu |

# Instructions (Refer Unit 2)

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 DW Technical Architecture? Give examples.  
- Information was gathered from the assigned lecture videos for assignment 2.

* + 1. Technical Architecture: Logical architecture, discussing how data moves around. Movement of data from one data store ERP to other data store DW.

[2] What is DW System Architecture? Give examples.  
- Information was gathered from the assigned lecture videos for assignment 2.

1. Systems Architecture: Physical configuration of technology and networks that support the technical architecture. Infrastructure that is a combination of hardware, software and networks.

[3] What are the 4 types of data stores found in technical architectures?  
- Information was gathered from the assigned lecture videos for assignment 2.

1. Normalized data store: Internal data store, used as organization single version of the truth for the other systems. Subject oriented, integrated, non-volatile and time-variant. Stored in third normal form, which reduces redundancy. Grows in size due to historical data. Normalized data source is used as a source for data marts and the DDS data store. Common columns in NDS: Created on (when the data was added) and Last update (when the data was last changed).
2. Operational data store: Hybrid data store, parts are internal and user-facing. Integrated, detailed, volatile and current data. Data is removed and updated to reflect current changes, consolidated from different sources. Does not grow over time, references a point in time, which is usually current. Should be stored as a separate DMBS as it is structured differently from NDS or DDS. Common columns in ODS: Last updated.
3. Dimensional Data Store: User-facing data store. Subject-oriented, integrated, non-volatile and time-variant data from source systems. Grows in size over time due to historical data. Data are consolidated and denormalized. It is stored in dimensional format, uses a star-schema for a relational database and a cube for a multi-dimensional database. Common columns in DDS: Effective date, Expiration date and current Row. Reflects the changes to a product over time.
4. MDS: MOLAP/Cube: When the DDS is multidimensional, it is called MOLAP. Multidimensional online analytical processing (MOLAP). Facts are pre-aggregated across all dimensions for improved performance. Is called a cube. Supports semantic meta-data (Sorts the day of the week, using hierarchy or a flow)

[4] Describe the 5 technical architectures discussed throughout the coursework. Be brief.  
- Information was gathered from the assigned lecture videos for assignment 2.

1. Independent data marts: Least complex. Easy to get started with, difficult to scale. Departmentalized, and lacks enterprise focus. Has a source system with one or more DDS. Each DDS is its own system. No data consistency or data integration between data marts. Data marts do not share dimensions.
2. Centralized: Data marts are consolidated into a single a DDS. Lack of integration among the dimensions, no data consistency between data marts. Has copies of the table.
3. Enterprise bus architecture: All data marts in the DDS. Conformed dimensions, they are reused across the data marts. Single dimension for master data. Requires a holistic and enterprise focus to master data. Kimball technical architecture. Requires a lot of foresight.
   * + 1. Enterprises with ODS – Includes an ODS for reporting on current, consolidated data. Conformed dimensions like enterprise bus.
4. Hub and Spoke: Single version of the truth. Inmnon technical architecture. NDS is the single source and data gets send to dimensional data stores as needed. ODS can be added between stage and NDS like enterprise bus architecture. It can be used for other systems.
   * + 1. Hub and Spoke with ODS – ODS is consolidated, current version of data. ODS is sourced from stage or NDS. ODS or stage can populate the NDS.
5. Federated with ETL: Most complex architecture, cases where you have several data warehouse. Common during company takeover, where multiple data warehouses need to be merged. ETL unifies disparate sources into a single federated data warehouse. Used to integrate existing data marts, warehouses and legacy applications into a single federated data warehouse.
   * + 1. Federated with EII – Enterprise application architecture. Software that can be purchased. Federated system is achieved through an EII software. Outputs are aggregated, on the fly.

[5] Discuss the comparative success of the 5 technical architectures. Be brief.  
 - As per the assigned reading - Which Data Warehouse Architecture Is Most Successful?”

1. The predominant architecture was the hub-and-spoke (39%), followed by the bus architecture (26%), centralized (17%), independent data marts (12%), and federated (4%).

2. Independent data marts scored the lowest on all measures, which confirms the conventional wisdom that independent data marts are a poor architectural solution.

3. The bus, hub-and-spoke, and centralized architectures earned similar scores on the success metrics.

4. Overall, the hub-and-spoke architecture was the most expensive and time-consuming to implement. This is not surprising, however, because of the larger domain and size of these warehouses. The architecture also requires a considerable commitment to up-front planning, which takes time and money.

WORKS CITED:

1. Information was gathered from the assigned lecture videos for assignment 2.
2. As per the assigned reading - Which Data Warehouse Architecture Is Most Successful?”