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: | Jaya Varshini Prabakar |
| Your Email: | jprabaka@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.

Technical Architecture, also known as **Logical Architecture,** encompasses a set of procedures and rules that define how the data is **stored** and **moved** from one data store to another data store.

**Some examples of Technical Architecture**: Multidimensional DBMS, Cloudera Hadoop, Star Schemas.

[2] What is DW System Architecture? Give examples.

System Architecture is a **physical** infrastructure containing Servers, Networks and Configurations that implements and supports Technical Architecture.

**Some examples of System architecture**: Mainframe computer, Point of Scale Applications with OLTP as source system, SMP (symmetric multiprocessing), MPP (massively parallel processing) and Hadoop MapReduce.

[3] What are the 4 types of data stores found in technical architectures?

The 4 types of Data Stores in Technical Architectures:

1. **Stage:** This is an Internal Data Store and not user friendly but provides consolidation of data, reduces contention, supports change detection and snapshotting.
2. **NDS (Normalized Data Store):** Used by data warehouses only and not available for end users. Subject oriented, integrated, non-volatile and time variant data from OLTP source.
3. **ODS (Operational Data Store):** Combination of User friendly and internal. The data is integrated, Volatile and current from source system.
4. **DDS (Dimensional Data Store)/MDS:** Not a part of Data warehouse, it is an OLTP user facing data. It is subject oriented, integrated, non-volatile and time variant.

[4] Describe the 5 technical architectures discussed throughout the coursework. Be brief.

The 5 technical architectures discussed are:

1. **Independent Data Mart:** Separate data marts for each department or business unit and lacks enterprise focus.
2. **Centralized:** Data Marts are consolidated into specific DDS but lacks integration among dimensions and there are copies of dimensions for each data mart.
3. **Enterprise Bus Architecture:** Same as centralized but conformed dimensions and are reused across Data Marts. This is a **Kimball technical Architecture**.
4. **Hub and Spoke:** Data is sourced systematically for single version of truth. Dimensional models in data marts are distributed and sourced from NDS. Here NDS acts as a HUB and all other analytical tools and OLTP act as a spoke. This is **Inmon technical Architecture**.
5. **Federated:** Complex technical Architecture in which ETL unifies disparate sources into federated Data Warehouse. Another Federated Architecture is using EII(Enterprise Application Integration) which achieves federation.

[5] Discuss the comparative success of the 5 technical architectures. Be brief.

|  |  |
| --- | --- |
| **Architecture** |  |
| Independent Data Mart | **Success Scenarios:**  **Departmental Analytics** – when different departments have distinct analytical needs.  **Quick And Agile Implementation:** Since they allow departments to setup their own data marts it enables quick data access and analysis.  **Security and Compliance:** Provides data isolation, in scenarios like hacking or data leakage one department might be affected it won’t put others in danger.  **Pros:** Offers flexibility at departmental level, so success depends on specific department needs and the organizational ability to manage and integrate them effectively.  **Cons:** Lack Enterprise focus |
| Centralized | **Success Scenario:**  **Enterprise-wide Reporting and Analytics:** Centralized architecture is more suited for enterprise-wide view analytics and consolidated reporting of departments.  **Pros:** Due to consolidation of data into DDS it is more enterprise focus.  **Cons:** Lacks data consistency and lack of integration as multiple copies of dimensions for each data mart is stored. |
| Enterprise Bus | **Success Scenario:**  Best suited for organization which focus on:  **Service-Oriented Architecture (SOA):** This allows various departments to integrate and interact providing loose coupling and interoperability.  It also gives Real time Data integration and Scalability.  **Pros:** Has conformed dimensions and are reused across data marts and uses single version for master data.  **Cons:** Difficult to achieve because enterprise focus is required when building data marts |
| Hub and spoke | **Success Scenario:**  **Consolidation of Data from Multiple Source Systems:** This architecture is easy to integrate data from disparate source systems.  **Simplified Data Integration and ETL:** Data from various sources are extracted and stored in a central hub (NDW) this gives us easy integration and ensures consistency.  **Pros:** Sourcing from NDS to all DDS department-based data mart provides as ‘single version of truth’ (data sharing across organization). Though there is 3N form it reduces conformed dimensions complexity.  **Cons:** As it has a single hub it might become a single point of failure and might create performance bottleneck |
| Federated ETL | **Success Scenario:**  **Better Data Integration:** Allows for integration of data from multiple databases and eliminated the need for massive data storage.  **Pros:** Several Data warehouses are used which are unified using ETL and sent to Federated Data Warehouse. This is used to integrate existing data mart, legacy applications to single logical data warehouse.  **Con:** Complex Technical Architecture. |