|  |  |
| --- | --- |
| Course: | ISQA 4730/8736 – Decision Support Systems |
| Document: | Video Review Form |

|  |  |
| --- | --- |
| Name: | Brian Detweiler |
| Lectures: | Week 4 - Data Warehouse |
| Date: | September 13th, 2017 |

|  |
| --- |
| Describe the key things you learned from this week’s lectures: |
| 1. Organizations have been collecting data for years, but traditionally, the data collection has been only for transaction processing (OLTP). These systems are optimized for storing individual pieces of facts, not large-scale data analysis. Online Analytical Processing (OLAP), or Data Warehouses, were designed to support large quantities of analysis. |
| 2. Data Warehouses are:  1) Subject-oriented – contains all data about a single subject  2.) Integrated – Contains data from many disparate sources including legacy and DB-systems  3.) Nonvolatile – New data does not overwrite existing data. No updates are done, only reads and writes. Data points are snapshots at that point in time.  4.) Time-variant – Date and time stamps are required.  5.) Aggregated – Data needs to be stored at different levels of granularity. Can pre-store summarized data to make reads more efficient.  They can also be web-based, relational/multidimensional, client-server, real-time, and includes meta-data. |
| 3. There are subcomponents of a Data Warehouse. Data Marts contain a subset of the Data Warehouse and makes it easier for managers to get the information they need. Operational data stores (ODS) is more of a short-term data mart for short-term decision making. Oper marts allow for multidimensional analysis of ODS data. |
| 4. Data Warehouse process flow:  Inflow – Gathering data from various systems is known as ETL (Extract, Transform, Load). This part is usually non-trivial as data can be ugly and transforming it into a uniform format is often difficult.  Upflow – Aggregation or summarization. Various levels of summarization.  Outflow - Most queries will go against highly summarized data.  Downflow – Archiving of older data. Data loses business value over time.  Metaflow – Full path that the metadata takes through the data warehouse. |
| 5. There are various data warehouse architectures. The most typical is the Enterprise Data Warehouse (EDW), which consolidates all of the data sources into a single Data Warehouse and the users connect directly to it. Then there is the Hub-and-Spoke Data Warehouse, which creates data marts as an abstraction layer to the users. This allows for easier customization of UIs and reports. There is added overhead in redundant storage and additional DBAs and operational costs. There is also some data latency, due to the added layer.  Two alternatives to the data warehouse are the Data Mart Centric, and the Virtual, Distributed, Federated architectures. The Data Mart Centric is easy to build since it doesn’t require a monolithic warehouse, but as with the Hub-and-Spoke, there are additional overhead costs. Lastly, the Virtual, Distributed, Federated architecture lets the data live where it was originally created and provides access via middleware. This reduces the overhead of Data Marts, but is really only viable for low volume data. |

|  |  |
| --- | --- |
| Below, provide one substantial online resource related to this week’s lectures. This resource should complement the lectures, aid in our understanding of the lectures’ topics, or illustrate lecture concepts through practical applications in business and industry. | |
| URL: | https://tdwi.org/articles/2015/01/27/hadoop-replace-data-warehouse.aspx |
| Full Citation: | Russom, Philip Ph.D. TDWI. *Can Hadoop Replace a Data Warehouse?* January 27, 2015. |
| Resource Summary: | The answer was nicely summarized by the quote, "Our view is that the data is the warehouse, and our data just happens to be managed with a relational database today. Our data could be managed on a non-relational platform, and it would still be a warehouse." |
| Relevance to the lectures: | This follows up on the slide about Big Data Implementation vs. Data Warehousing. |

|  |  |
| --- | --- |
| Due date: | Fridays at 8:00 a.m. during the week in which lectures were posted. |
| File name: | Format your file name: lastname.doc or lastname.docx |
| Submission: | Submit in Canvas |