DA 320 Assignment 4 M. Blanco

**Part 1: Short Answers (use 3-5 lines to answer the following questions)**

**Problem 1:** Explain why data warehousing is not necessarily a great tool for asking questions quickly.

The goal of data warehousing is to aggregate data from various data silo. Data warehousing it’s ideal but it is often an unattainable goal. The effort that create ETL pipeline is enormous.

First, complete ETL pipeline must be execute before data can be queried. Secondly, changing the data model used might require significant effort. Automated ETL processes could fail if data from any of the disparate sources is interrupted or changed.

**Problem 2:** Explain why data warehousing doesn’t solve every data silo challenge effectively.

Often, we are focus on the data challenge and not the technology. The goal is for the systems used by various organizations (Finance, IT, Statistician) to interact and act more like an analytical database. Some of the data in a data warehouse might be sensitive and would need to be limited. Organization should make every effort to empower employees to ask their own question and invest in technologies such as Hadoop that bridge data silos.

**Problem 3:** Explain the overlapping use cases between data warehousing and distributed computing technologies such as Hadoop.

Hadoop allows for distributed data processing using unstructured data and can be used to execute ad-hoc analysis pipeline. Potentially data can be kept in silos and processed as necessary. An example of an overlapping use cases is the blending of Spark and Shark to provide warehousing and analytics capability.

**Problem 4:**Discuss why there should be consideration of building in the cloud and using cloud computing technologies.

Building data warehousing system in the cloud provides for reduce cost and infrastructure. An added benefit is flexibility for browser-based visualization and reporting tools to access the data source.

**Part 2: Research Analysis**

**Problem 5:** Read and analyze H.P. Luhn’s original business intelligence paper at <http://altaplana.com/ibmrd0204H.pdf>

1. Compare Luhn’s original idea to how today’s business intelligence systems have materialized
2. Do a comparison to determine if Luhn’s idea is or can become a reality or not based on these comparisons. Make sure you use detailed application examples to support your analysis.

In 1958, IBM computer scientist Hans Peter Luhn published a revolutionary article on the subject on Business Intelligence. The article, A Business Intelligence System, described an automatic system that would disseminate information. To stay competitive with increased globalization, competition, and spread of information systems, many organization required a way to organize and simplify the rapidly growth of data. The article defined intelligence as the ability to capture the interrelationships of presented facts in such a way as to guide action towards a desired goal. In its core, BI promotes a way to quickly understand large amounts of information so that informative decisions can be made. Luhn’s contribution not only introduce BI but also expanded the possibilities. His research established methods that are standards for analytical systems used today.

In the 80’s and 90’s we saw advances in data warehouses as way to improve the flow of data from operational systems to support decision. Data access was cut drastically with data warehousing by taking data traditionally stored in multiple places to a single integrated location. With data warehousing, Extract, Transform, and Load (ETL) tools and Online Analytical Processing (OLAP) software were developed. Many BI applications provided functionality to produce data, reports, and visualizations. BI projects are being developed by high skill expert with extensive analytical training, meaning most end users were not capable of executing basic BI tasks on their own. This lead to specialized system and data silos.

At the beginning of the 21st century, many BI technologies were developed to addressed complexity and real-time processing. The growth of Cloud-based programs expanded and simplified the reach of BI platforms. Other technologies introduced self-service access for non-expert users, meaning that users could complete projects without direct support from IT department. BI applications will continue to expand to meet the needs for big data analytics, mobile visualization, and predictive analytics.

Some aspect of the BI system that Luhn envisioned are already available, but not in the way that he imagined it. Currently BI vendors are fine-tuning the functionality of their systems to make them suitable for startups and small businesses.

Alteryx, a self-service data analytics platform, delivers functionality that simplifies access to data and analytics to business analysts and business leaders. The following is a list of features:

* Data Blending
* Predictive Analysis
* Spatial Analysis
* Export Package
* Import Package
* Encrypted Storage
* Interactive Visualizations
* Mapping
* Drag-and-Drop Workflows
* Sharing
* Encryption, Governance & Security

Concepts introduced by Luhn:

* Ability to query and search
* Ability for expert user to carry out difficult queries
* Ability for business users could be allowed to carry out their own queries
* Information on demand
* Security and data ownership