**Week 10 – Self-Service Analytics**

1. What are some of the self-service issues that gave birth to the self-service revolution?
   1. **Delays**  🡺Businesses that use traditional BI tools operate in highly controlled environments where IT maintains access to the data. Users had to request data reports or dashboards, then wait for the analysis to deliver. This process resulted in delays, and finally, when the information has been provided, the data on the report can no longer be valid.
   2. **Accuracy** 🡺Another Self-service analytics common issue is that users either build their solutions and may even choose their tools, with or without permission. This caused some essential processes to be skipped result inaccuracy issues.
   3. **Coupled storage formats** 🡺 Data storage and data processing were tightly connected; this only allowed specific analytical processing tools to process data.
2. How does the creation of a data lake address those problems?
   1. **More Data Available = higher chance of accuracy**🡺One of the main attractions of data lakes is flexibility and getting data into the lake is simple. Data lakes are used as a central repository for raw data, and it serves as the foundation for many "self-service" analytics. This allows for self-service tools to have more data to make decisions.
      1. Data lakes can store: Structured data, semi-structured data, unstructured (raw) data.
   2. **The emergence of standard formats** 🡺 These traditional forms led storage formats to decoupled, which allowed different analytical engines to process the same data without having to replicate the data in repositories.
   3. **Data Independence & flexibility= No more delays when requesting data** 🡺Provides quick answers to user asks no matter where the data resides. Now data consumers can perform their analytics directly against the data lake while all data remains in place.
   4. **Data Permission** 🡺 Sensitive data was masked, row and column-level permission can now be set, which results in role-based control, allowing access to whatever the end-user needs.
   5. **Scalability** 🡺Open format and scales can hold tons of data at a low cost.
3. From what you've learned this term, are the significant challenges in doing so?
   1. **Scattered Data** 🡺 Having more data can be considered an advantage; however, data sometimes becomes scattered in multiple systems, which could cause data analytics to become time-consuming and complex.
      1. Data should be standardized and compressed when loaded into a data lake; this will make it easier for data consumers to work.
   2. **Slow performance & decreased reliability** 🡺 As the data increase in size, it can cause engines to move slower. Over time the data lake can become a data swamp instead.
   3. Lack of security features 🡺 At times, it's hard to secure the data lake's data due to the lack of visibility or the ability to delete or update data.

**References**

Building the enterprise-ready data lake: What it takes to do it right. (2016, July 25). Retrieved May 26, 2021, from <https://www.datanami.com/2016/07/25/building-enterprise-ready-data-lake-takes-right/>.

Dennis, A. (2019, November 11). Using a data lake engine to provide self-service insights. Retrieved May 26, 2021, from <https://www.dataversity.net/using-a-data-lake-engine-to-provide-self-service-insights/>.