Emilija Dimikj

June 19, 2020

BIDD 220 B Sp 20: Data Migration Techniques (ETL Processing)

Lessons Learned

Lessons Learned - ETL Processing

We started covering ETL Basics and Data Warehouse full loading method, this part was combination of SQL programming that we covered in the previous course and basics of the ETL process. After covering the basics, we moved to more complex SQL ETL processes and learned about Slowly Changing Dimensions (SCD) and types of incremental data loading. Special accent was put on MERGE T-SQL statement as the best and most optimal way of handling synchronization. MERGE statement is used to synchronize source and destination tables in one iteration, by inserting, deleting, and updating the target table rows based on the join condition with the source table.

I really enjoyed learning about the SSIS, as this was my first encounter with SSIS. As a visual learner, I liked how data flow is represented graphically and it is easier to define all data paths and transformations. On the other hand, some of the settings are placed on hidden places and windows are not very intuitive to work with. Starting with simple tasks as performing Microsoft’s SSIS tutorial was a good way to learn basic SSIS concepts and understand the principles of SSIS. Moving forward to more complex SSIS projects, we explored the possibilities of this tool. Implementation of a hybrid model give us possibility of using the best of both worlds. Using graphic interface makes it easier to build large, complex, and reliable data flows and is easier to visualize and to control. Overall, the best way to become acquainted with these tools, controls, and features is to use them and that means practice.

As part of learning No-SQL ETL process, we explored C# script, Python and PowerShell script. It was very fun to explore all different ways of making same ETL process with different technologies. I am glad that we got introduced to Python. I never had opportunity to explore it before. It is gaining in popularity so learning it is a good way to go. Python is an elegant, versatile language with lots of powerful modules and code libraries. On other hand, without all available third-party python libraries, it would take a lot more time and effort to create an enterprise level application for the No-SQL ETL process.

Last capstone was supposed to be putting into work all that we learned before but having real world task and encountering real world problems turned out into good learning process itself. I enjoyed doing it, implementing what I know and learning on the way.

Randal’s support and guidance on every step of the way were precious. I really appreciate giving us open hands for the assignments, to explore different techniques and approaches.