Modern Data Architecture: The Role of Data Lakehouse in Data Engineering and Analytics

Giorgi Megeneishvili
e-mail: megeneishvili997@gmail.com
Department of Computer Science,
Faculty of Exact and Natural Sciences,
Iv. Javakhishvili Tbilisi State University
Tbilisi, University St. 13

The evolution of modern data architecture has significantly reshaped the priorities of data engineering and analytics. Traditional Data Warehouses and Data Lakes are progressively merging into a unified concept — the Data Lakehouse, which combines the capabilities of processing both structured and unstructured data to support advanced analytics. This thesis analyzes the stages of data architecture development, explores the core concepts and advantages of the Data Lakehouse model in today's data landscape, and highlights the strategic use of the Medallion Architecture with Bronze, Silver, and Gold layers to ensure data quality and optimize analytical processes. The technical part of the thesis is presented as a practical project, providing detailed descriptions of the implementation phases, the technologies used, and the data engineering strategies employed throughout the process.

References:

1. Martin Kleppmann – Designing Data-Intensive Applications. O'Reilly Media, 2017. https://www.oreilly.com/library/view/designing-data-intensive-applications/9781491903063/
2. Joe Reis, Matt Housley – Fundamentals of Data Engineering. O'Reilly Media, 2022. https://www.oreilly.com/library/view/fundamentals-of-data/9781098108296/
3. Kristin Briney – Data Management for Researchers. Pelagic Publishing, 2015. https://www.springer.com/gp/book/9781498761130
4. Tyler Akidau et al. – Streaming Systems: The What, Where, When, and How of Large-Scale Data Processing. O'Reilly Media, 2018. https://www.oreilly.com/library/view/streaming-systems/9781491983867/
5. Databricks – What is a Data Lakehouse? https://www.databricks.com/solutions/data-lakehouse