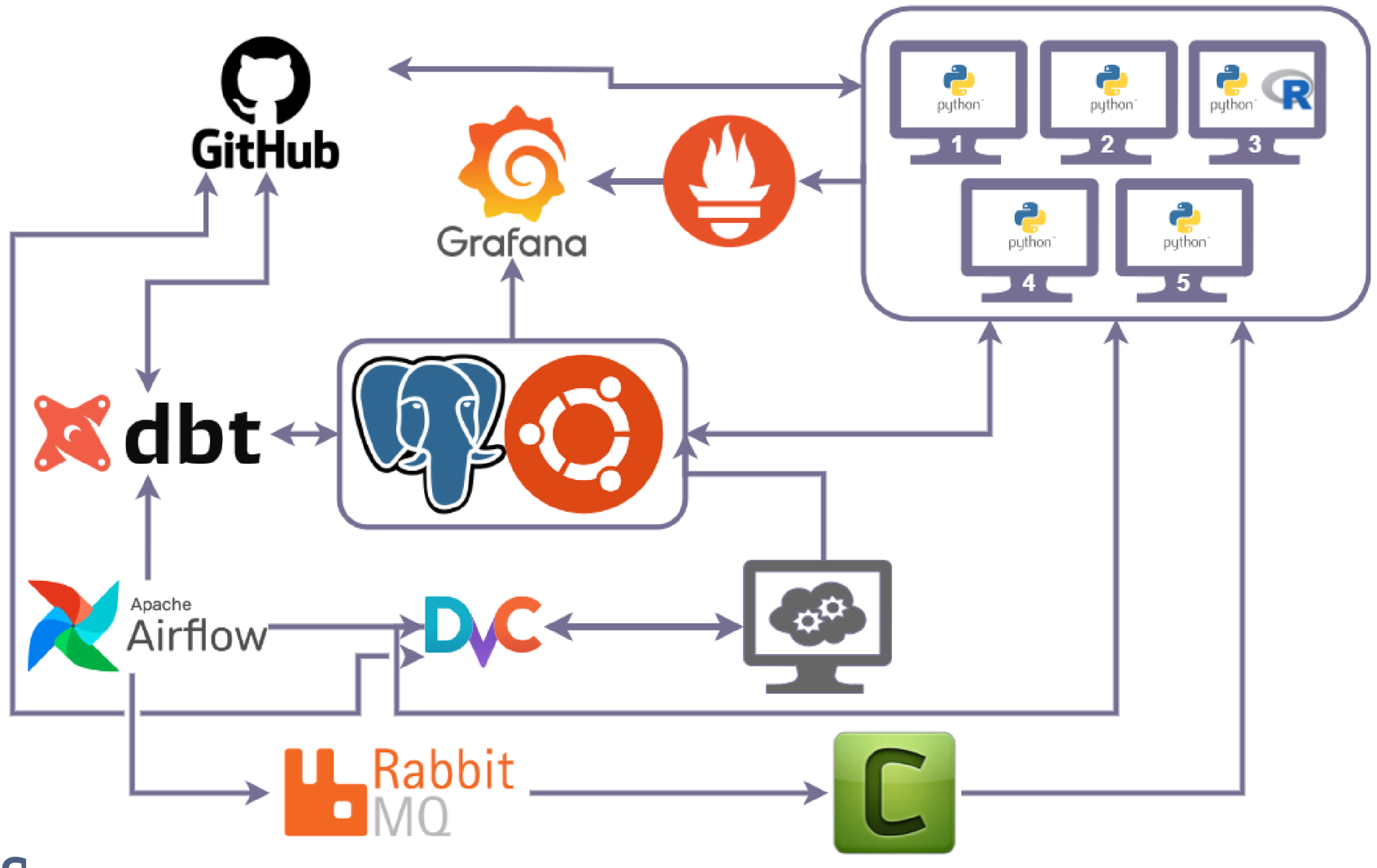
Big Data Processing System for Lithuania Economic Activity Nowcasting

The assessment of economic activity is an important assessment of the state's economy, which allows assessing the current situation, as well as predicting future prospects. The increasing amount of data every year allows this data to be used in the forecasting of economic processes. However, due to the large amount of data, its rapid renewal, and its diversity, it is difficult to evaluate it in traditional ways. Traditional methods of assessing economic activity use monthly or quarterly data, which are no longer appropriate in the face of various economic shocks. A good example of this is the COVID-19 pandemic or the war in Ukraine, which affects the state's economy quite quickly. For this reason, it becomes important to evaluate not only traditional economic indicators, but also various alternative ones collected from various openly available sources. The purpose of this work is to present a possible economic activity assessment system that collects, processes, transforms and visualizes Lithuanian economic activity. The developed economic activity forecasting system allows you to automatically collect text information, prices of products and services, real estate and others. And using machine learn-ing methods, this data is turned into valuable insights, which can be used in state, business and other decision-making.

DATA, SYSTEM AND RESULTS

The main idea of a data analysis system based on big data is the automatic collection of freely available data, their transformation, analysis, and application of machine learning. The goal of this system is to perform all steps in forecasting the country's economic activity with extremely little user intervention. A simplified diagram of the data analysis system is shown in the figure below. Data analysis is performed using distributed computing on several separate servers, each of which has tens of workers. Data collection is controlled using Airflow, which allows you to pre-set data collection intervals and assign data collection to the right server. All the collected data is stored in the PostgresSQL database, and to transform the collected data, the data is used in dbt. Server and data collection worker monitoring uses Prometheus and Grafana, which work to monitor server loads in real-time. The data analysis system is divided into several modules: news module, labour market module, real estate module, etc. Research results obtained by news modules show that news sentiment analysis has a significant impact on forecasting monthly and annual inflation as well as other indicators of economic activity (more: https://doi.org/10.3390/math10193461).



AUTHORS

Mantas Lukauskas*, Vaida Pilinkienė, Jurgita Bruneckienė, Alina Stundžienė, Andrius Grybauskas, Tomas Ruzgas

***mantas.lukauskas@ktu.lt**

Microsoft Teams link: Meeting ID: 383 559 040 865 Passcode: zkoSFb

This project has received funding from European Regional Development Fund (project No. 13.1.1-LMT-K-718-05-0012) under a grant agreement with the Research Council of Lithuania (LMTLT). Funded as European Union's measure in response to the COVID-19 pandemic.