TSEPKOV IAROSLAV

Data Engineer | Analyst
Email: yaroslav.tsepkov@gmail.com
Telegram: Iaroslavtsepkov

Describe

I develop various systems for storing, processing, analyzing and visualizing systems for working with big data. In my career, I have brought several ETL\ELT systems with visualization services and user notifications, as well as monitoring systems for engineers, to the final product. I have experience in implementing systems with machine learning, implemented one video surveillance system and target object detection, as well as a system for searching for anomalies in data.

Education

Master of Science in Samara State Aerospace University

August 2020 - February 2023

Faculty of Computer Science. Applied Mathematics and Computer Science.
Development of intelligent systems. Big data, Machine learning, HPC, statistics, distributed databases.

Bachelor of Science in Samara State University

August 2016 - August 2020

Faculty of Mathematics. Applied Mathematics and Computer Science.
Systems analysis. Include courses statistics, c-lang, python, algorithms and data structs, OS, databases.

Experience

PepsiCo (self employed)
Data Engineer

January 2023 - August 2024

- Implemented and developed a ELT system for collecting and processing data for the e-commerce department using the cloud provider Azure and Databricks on Azure. Configured monitoring and alerting services for the data system. Using the Azure data factory, configured all execution pipelines using DAGs. Together with analysts, I identified requirements for the system, and together we developed a system of requirements for data quality.
- Used stack Azure ecosystem (Azure Databricks, Azure SQL, Azure KeyVault, Azure Data Factory, Azure BlobStorage) and Python + pySpark

Xim inc. (self employed)
Data Engineer

February 2022 - March 2024

- Support of the current ETL system. Updating the API of data collectors, optimizing processes. Implementation of new data collectors for API, email, GCS, S3. Implementation of models and macros for DBT. Formation of data marts for analysts to build reports and execute ad hoc queries. Implemented more than 40 partners with daily data. Set up cloud storage for partners who uploaded their data there. Administered Google cloud, tracked roles and access to data in the cloud. Formed more than 20 new data marts. Executed ad hoc queries to generate certain reports according to requirements. Supported interactive reporting in Redash and Looker Studio.
- Used Google Cloud and S3. Working with BigQuery, ClickHouse, Postgres, Prometheus databases. Using Python and Pandas\Dask + Streamlit for backend and frontend ETL system. I have experience with work many APIs for ad providers like a Google Ad, Apple Search, MyTarget. Working with mediation data and appsflyer. Working with BI tools Grafana, Looker (Data Studio), Redash

Sber
Senior Auditor (Data Scientist)

January 2021 - December 2021

- Working with various data warehouses Oracle, MS SQL Server, Terradata and others. Cleaning and masking data using regex and Levenshtein distance. Conducting analysis using graph theory. Building graphs of connections and processes using Gephi, networkx, SberPM. Research and implemented projects in the field of machine learning: problems of classification, clustering and searching for anomalies in data, document recognition, object detection. Developed internal libraries for future developments.

Integra-S
Software developer

January 2020 - December 2020

- Identifying business requirements for applications. Developed a mobile application for a video surveillance system for Android devices. Together with the designer, we discussed and updated the design of the application. Working with REST API, writing software documentation. Working with RTSP streaming video signal. Reasearch an Optimal Model for Object Detection in Video for the Server Side.

PET projects

BELKA (smart house)

Participation in the smart home project as a mobile developer. Development of a mobile client on iOS, development of database architecture.

High Performance Computing

Solving problems of linear algebra, signal processing using CUDA technology. Writing an article on newtechaudit: Using CUDA technology in solving applied problems.

Image processing

Studying image libraries and implementing image processing algorithms myself. Writing two articles on this topic on newtechaudit: Image analysis and processing using mathematical morphology operations и Developing Analytics Web Applications Using the Streamlit Library

Skills

- programming languages **Python, Bash.**
- databases **BigQuery, ClickHouse, MS SQL server, Oracle, Postgres, Kafka, Prometheus.**
- containers **Docker, Docker-Compose, K8S**
- data formats **Parquet, Avro, ORC, Delta tables, JSON, YAML, MD, Protobuf.**
- data transform **Pandas, Dask, Spark, DBT, Azure Data Factory.**
- machine learning **SciKit-learn, Xgboost, Keras**
- high computing **CUDA, Numba, Multiprocessing, Threads, Concurrent, Async.**
- visualisation **Looker (Data studio), Redash, Yandex Lens, Plotly, Dash, Altair, NetworkX, Gephi.**
- cloud computing **Google Cloud Platform (BigQuery, Cloud Storage, Looker, GooglePlay), Azure (MSSQL, Blob Storage, Data Factory, Databricks, Key Vault), Amazon (S3 bucket)**
- workflow **git, jira, confluence, gitlab, bitbucket, markdown.**
- web **html5, css3, streamlit, fastapi**