

Дунаев Михаил, Data Engineer

E-mail: getroller7@gmail.com

Phone: +7 (914) 090-55-95

Telegram: @getroller

Место проживания: г. Владивосток

Дата рождения: 07.01.2002 (22 года)

Обо мне

Добрый день! Я начинающий Data Engineer. Коммерческий опыт отсутствует. В рамках ВКР строил распределенное КХД на основе реальных данных (Greenplum, Airflow) и моделировал в нем Data Vault. Активно развиваюсь в сфере IT и глубоко интересуюсь информационными технологиями, а также их способностью трансформировать бизнес.

Технологический стек

Python, SQL, ООП, PostgreSQL, Greenplum, Linux, Apache Airflow, Git, Docker, MinIO.

Опыт работы

Выпускная квалификационная работа на тему «Разработка концепции перехода от схем данных к типовым схемам корпоративных хранилищ в ООО «Сервисный Центр ФЕСКО»».

Декабрь, 2023 – Июль, 2024 (8 мес.)

Разворачивал КХД на основе архитектуры заказчика (ООО «Сервисный Центр ФЕСКО»), стек - Greenplum, Airflow. Цель работы – изучение, поиск недостатков и улучшений OLAP-хранилища, обоснование внедрения разработанных решений с точки зрения экономической эффективности.

В процессе прохождения производственной практики изучал и анализировал целевую архитектуру хранилища данных заказчика, устройство MPP-систем и ETL/ELT инструменты.

В процессе работы:

1. Разворачивал кластер Greenplum - 1 master host, 2 segment hosts с репликацией (4 primary и 4 mirror сегмента), PXF, GPFDIST, 3 BM, ОС – RHEL. Настройка резервного копирования, управление доступом к хранилищу.
2. Проектировал и реализовывал слои DWH (RAW, STG, ODS, DDS, DMA).
3. В качестве систем источников использовал базы PostgreSQL, внешние таблицы - PXF, FDW.
4. Моделировал Data Vault на продуктивном датасете, выделял бизнес-сущности, наполнял слои хранилища.
5. Реализовывал и автоматизировал партиционирование таблиц с целью ускорения скорости выполнения запросов, а также с целью эффективного хранения данных (создание слоев хранения данных и вынос холодных партиций в объектное хранилище minIO через PXF).
6. Реализовывал ETL пайплайны (Apache Airflow, Pandas, Requests, Pentaho).
7. Экономически обосновывал то, как разработанные решения генерируют денежные потоки для организации.

Ссылка на проект - <https://github.com/getroller/Final-qualifying-work>

LeetCode - <https://leetcode.com/u/getroller/>

GitHub - <https://github.com/getroller>

Образование

Сентябрь, 2020 – Июль, 2024

Дальневосточный федеральный университет, г. Владивосток
Экономика, Управление фирмой на основе данных (Бакалавр)