



Лагутин Андрей

Мужчина, 27 лет, родился 1 сентября 1997

+7 (916) 8575174 — предпочитаемый способ связи
hofmanez@gmail.com

Проживает: Москва

Гражданство: Россия, есть разрешение на работу: Россия

Готов к переезду, готов к командировкам

Желаемая должность и зарплата

Архитектор информационных систем и баз данных

Специализации:

- Руководитель группы разработки
- Системный администратор
- Системный инженер

Занятость: полная занятость

График работы: полный день

Желательное время в пути до работы: не имеет значения

Опыт работы — 5 лет 8 месяцев

Март 2024 —
настоящее время
1 год 6 месяцев

МегаФон

Москва, www.megafon.ru

Информационные технологии, системная интеграция, интернет

- Разработка программного обеспечения
- Системная интеграция, автоматизации технологических и бизнес-процессов предприятия, ИТ-консалтинг

Телекоммуникации, связь

- Фиксированная связь
- Мобильная связь

Архитектор корпоративного хранилища данных и информационных систем

1) Миграция корпоративного хранилища данных (Exadata → Greenplum/ClickHouse + Hadoop/S3 + Trino):

Проведение миграции с Oracle Exadata на современные OLAP-решения: Greenplum (для аналитики) и ClickHouse (высоконагруженные OLAP-сценарии), охлаждение данных в Hadoop/S3.

Пилотируем связку ИС → Kafka → S3 → Data Lake → Spark → CH.

Data Lake на базе Hadoop с интеграцией Trino (ex-Presto SQL) для высокопроизводительных аналитических запросов.

Реализация архитектуры DWH на основе Data Vault 2.0 с интеграцией модели SID TM Forum (логическая модель, объекты, правила построения).

Автоматизация процессов: генерация логических моделей данных, CI/CD-конвейеры, Change Data Capture (CDC) через IUD-таблицы.

Проведение нагрузочного тестирования, определение метрик RTO/RPO, мониторинг производительности кластеров (Prod/DEV/Pre-prod).

Внедрение Apache Iceberg для управления таблицами в data lake.

Настройка бэкапирования DWH на СХД с протоколом S3.

Разработка концепции Blue/Green Deployments.

Технологии: Trino, S3, Hadoop, Greenplum, ClickHouse, DBT, Airflow, Apache Ni-Fi, Apache Ozone, Iceberg, CI/CD, CDC, Debezium, Delta Lake, LakeHouse, Data LakeHouse, Kafka, Oracle, ODI

2) Проектирование многоуровневой архитектуры хранилища DWH

Создание масштабируемой слоистой структуры с интеграцией 50+ источников данных.

Оптимизация ETL/ELT-процессов через переход с ODI на Airflow и Ni-Fi.

3) Реализация репозитория метаданных для трекинга преобразований данных и сущностей, построения data lineage.

4) Управление командой:

Руководство cross-functional командами (5 человек): постановка задач, оценка трудозатрат, приемочное тестирование.

5) Организация архитектурных комитетов: документирование решений, анализ ПАТ, согласование изменений в соответствии с ISO/TM Forum.

6) Внедрение практик DataOps: мониторинг качества данных, автоматизация тестирования (DBT), рефакторинг legacy-процессов.

Оптимизация и инновации:

Разработка регламентов эксплуатации хранилища: резервное копирование, восстановление, мониторинг SLA.

Основные цели: оптимизация существующих процессов и регламентов, перенос существующих сервисов на новое КХД, решение инцидентов, анализ новых архитектурных решений типа medallion architecture для ClickHouse.

Сентябрь 2023 —
Апрель 2024
8 месяцев

Тера Интегро

Москва

Информационные технологии, системная интеграция, интернет

- Разработка программного обеспечения
- Системная интеграция, автоматизации технологических и бизнес-процессов предприятия, ИТ-консалтинг

Senior Data Engineer

1. Разработка и оптимизация распределенных DWH (Greenplum, ClickHouse)

Архитектура и сопровождение Greenplum: проектирование кластера, тюнинг производительности (партиционирование, индексы), внедрение Prometheus + Grafana для мониторинга метрик.

Создание корпоративного дистрибутива ClickHouse под требования Сбербанка: интеграция шифрования данных (AES-256), настройка репликации (ZooKeeper), разработка мониторинга.

2. Data Lake и гибридные решения

Потоковая обработка данных: реализация Spark Streaming (Python/Scala) для ETL-пайплайнов, интеграция с Kafka для обработки событий в реальном времени.

Управление большими данными: настройка Spark-джобов для batch-обработки, интеграция с Hadoop (HDFS, YARN) и Trino для кросс-платформенных аналитических запросов.

3. Обучение и экспертиза

Разработка корпоративных курсов (Тера Интегро): создание программ по Greenplum (администрирование, оптимизация запросов) и ClickHouse (архитектура, шардирование) на платформе Moodle.

4. Поддержка enterprise-клиентов (Сбербанк, ВТБ, ПСБ)

Премиум-поддержка DWH: аудит производительности кластеров, оптимизация SQL-запросов (EXPLAIN ANALYZE, индексные рекомендации), настройка резервного копирования (pgBackRest, ClickHouse Keeper).

Реализация CI/CD для DWH: деплой скриптов с помощью Airflow и GitLab Pipelines.

5. Инструменты и автоматизация

Разработка кастомных дашбордов в Grafana для мониторинга SLA (запросы/сек, нагрузка на CPU, размер партиций).

6. Создание алерт-систем на базе Zabbix для предупреждения аномалий в работе Greenplum/ClickHouse.

Ключевые технологии и экспертиза:

DWH: Greenplum, ClickHouse, PostgreSQL.

Data Lake/Streaming: Hadoop (HDFS, Hive), Spark (Streaming, SQL), Trino, Kafka.

Инфраструктура: Prometheus, Grafana, Zabbix, Airflow, Docker, Kubernetes.

Январь 2022 —
Август 2023
1 год 8 месяцев

Группа НЛМК

Липецк, nlmk.com/ru

Металлургия, металлообработка

- Продукция черной металлургии (продвижение, оптовая торговля)
- Черная металлургия (производство чугуна, стали, проката)
- Цветная металлургия (выплавка, металлопрокат)
- Металлические изделия, металлоконструкции (производство)

Добывающая отрасль

- Добыча и обогащение угля

Старший разработчик DWH/КХД + DBA

1. Миграция данных:

- Руководство проектами по миграции данных, включая планирование и выполнение переноса данных из различных источников в корпоративное хранилище данных.

(Корпоративное хранилище - PostgreSQL)

- Обеспечение целостности и безопасности данных в процессе миграции, использование инструментов для контроля и проверки данных (Оркестрация - Apache Airflow, Apache Ni-Fi).

2. Оптимизация PostgreSQL:

- Проведение оптимизации производительности баз данных на платформе PostgreSQL, включая анализ и настройку индексов, написание эффективных SQL-запросов. Рефакторинг легаси кода.

- Мониторинг работы баз данных и принятие мер по улучшению производительности.

3. Поддержка проектов и разрешение инцидентов:

- Оказание технической поддержки текущим проектам, выявление и решение инцидентов, связанных с функционированием кластера и инструментов для обработки данных.
- Реализация корректирующих действий и разработка стратегий для предотвращения повторения проблем.

4. ETL-процессы:

- Разработка и внедрение эффективных ETL-процессов для извлечения, преобразования и загрузки данных в хранилище данных, обеспечивая их соответствие требованиям качества данных.
- Оптимизация существующих ETL-процессов для повышения эффективности обработки данных.

5. Построение хранилища данных:

- Участие в проектировании и реализации архитектуры хранилища данных.
- Настройка и оптимизация процессов интеграции с источниками данных. (SAP система)

6. Разработка SQL-скриптов:

- Написание и оптимизация SQL-скриптов на уровне функций для автоматизации процессов обработки и анализа данных, включая создание хранимых процедур и функций.
- Обеспечение их корректности и производительности через тестирование и отладку.

7. Переносы контуров Airflow:

- Осуществление переносов тестовых и производственных контуров с использованием Apache Airflow, включая настройку и управление рабочими процессами ETL.
- Поддержание документации и следование практикам DevOps для обеспечения бесперебойной работы.

8. Документирование процессов:

- Написание и поддержание документации по проектам и разработанным решениям, включая архитектурные схемы, инструкции по использованию и описания процесса ETL.

9. Тестирование моделей Data Vault и Anchor Model:

- Проведение тестирования и оценки моделей Data Vault и Anchor Model в рамках разработки нового корпоративного хранилища данных.

10. Управление командой подрядчиков, написание ТЗ:

- Управлял командой Инженеров Данных (2 человека).

Стек: Hadoop, PostgreSQL, Kafka, Airflow, Zabbix

Январь 2020 —
Декабрь 2021
2 года

IQVIA

Разработчик DWH

1. Разработка и оптимизация хранилищ данных:

- Проектирование, создание и оптимизация хранилищ данных на базе Microsoft SQL Server, Oracle, PostgreSQL.
- Анализ существующих архитектур данных и предложение улучшений для оптимизации хранилищ.

2. Стриминговая обработка данных:

- Организация и внедрение стриминговой обработки данных с использованием ODI, Azure Event Grid.

- Эффективная обработка потоковых данных, включая мониторинг и управление состоянием данных.

3. Модели данных:

- Участие в разработке и внедрении новых моделей данных в компании (Например, Anchor modeling).
- Активное сотрудничество с командами по анализу данных для определения требований к моделям и структурированию данных.

4. Обработка больших данных:

- Применение Apache Spark для выполнения сложных вычислительных задач, анализа и обработки больших объемов данных.
- Работа с экосистемами Hadoop для гарантирования надежного хранения и обработки больших объемов данных.

5. Контейнеризация приложений:

- Реализация контейнеризации приложений с использованием Docker.

6. Программирование на Python:

- Написание скриптов на Python для автоматизации процессов обработки данных, включая ETL-процессы и подготовку отчетности.
- Использование Python-библиотек для анализа и визуализации данных.

Мой карьерный рост в IQVIA:

Вырос от младшего аналитика до владельца продукта: базы данных по суточным продажам (sell out) аптек в России, что позволило развить навыки управления проектами, управления командой (3 человека), обеспечивающей поддержку базы данных, включая планирование разработки и решение проблем, связанных с данными.

Образование

Уровень Среднее образование

Навыки

Знание языков Русский — Родной
Английский — B2 — Средне-продвинутый

Навыки Python PostgreSQL Big Data ETL Git Linux Hadoop Docker
Работа с базами данных Airflow Tableau DWH СУБД Hive Kafka
Postgres Документальное сопровождение Базы данных Azure КХД
Azure Data Lake DAMA-DMBOK Greenplum Clickhouse apache iceberg
Trino ORACLE Системный анализ ERP-системы ESB

Дополнительная информация

Обо мне Experience Summary:

- Programming Languages: Python, SQL, Bash
- SQL Databases: PostgreSQL, Microsoft SQL Server, Greenplum, ClickHouse
- Cloud and Big Data Stack: Apache Airflow, Hadoop (вся экосистема), Spark

- Data Science: PyTorch, Pandas
- BI Tools: Datalens, Tableau, Форсайт, Визиолоджи
- DevOps: Docker, Kubernetes, Ansible
- Operating Systems: Linux/Unix
- Change Data Capture (CDC): Debezium
- Event Streaming Platforms: Apache Kafka, RabbitMQ
- Data Governance: DAMA-DMBOK
- Other Skills: UML
- Object Storage: S3, Minio, Apache Ozone
- Open File Format: Parquet, Avro, ORC
- Open Table Format: Iceberg, Hudi
- Engine: Trino

Professional Skills:

- Опыт выявления узких мест в архитектуре данных и оптимизации интеграционных взаимодействий с различными информационными системами.
- Глубокое понимание интеграционной архитектуры данных и технологий для реализации кластеров высокой доступности и отказоустойчивости.
- Навыки создания схем форматов обмена данными, таких как XSD и JSON-Schema.
- Знание основ мониторинга и лучших практик работы с метриками, включая практики SRE.
- Ознакомлен с этапами разработки в фреймворке TOGAF.
- Развиваюсь как в направлении DWH, так и в гибридных архитектурах типа LakeHouse

Career Goals:

Стремлюсь развивать архитектурные навыки, навыки DevSecOPS, навыки управления командой и проектами.

Hobby:

Шоссейный велоспорт, тяжелая атлетика, учеба (технические навыки, управленческие навыки), походы.

Contact Information:

TG: @xofmanez

Только удаленный формат работы!