

Создание ETL-процесса формирования витрин данных для анализа изменений курса акций

Итоговая аттестационная работа
по курсу «Инженер данных»

Исполнитель: Момотова И.В.

Общее описание проекта

Общая задача: создать ETL-процесс формирования витрин данных для анализа изменений курса

Подробное описание задачи:

- Разработать скрипты загрузки данных в 2-х режимах:
 - Инициализирующий — загрузка полного слепка данных источника
 - Инкрементальный — загрузка дельты данных за прошедшие сутки
- Организовать правильную структуру хранения данных:
 - Сырой слой данных
 - Промежуточный слой
 - Слой витрин

В качестве результата работы программного продукта необходимо написать скрипт, который формирует витрину данных следующего содержания:

- Суррогатный ключ категории
- Название валюты
- Суммарный объем торгов за последние сутки
- Курс валюты на момент открытия торгов для данных суток
- Курс валюты на момент закрытия торгов для данных суток
- Разница (в %) курса с момента открытия до момента закрытия торгов для данных суток
- Минимальный временной интервал, на котором был зафиксирован самый крупный объем торгов для данных суток
- Минимальный временной интервал, на котором был зафиксирован максимальный курс для данных суток
- Минимальный временной интервал, на котором был зафиксирован минимальный курс торгов для данных суток

Дополнение:

В качестве основы витрины необходимо выбрать 3-5 различных акций компаний.

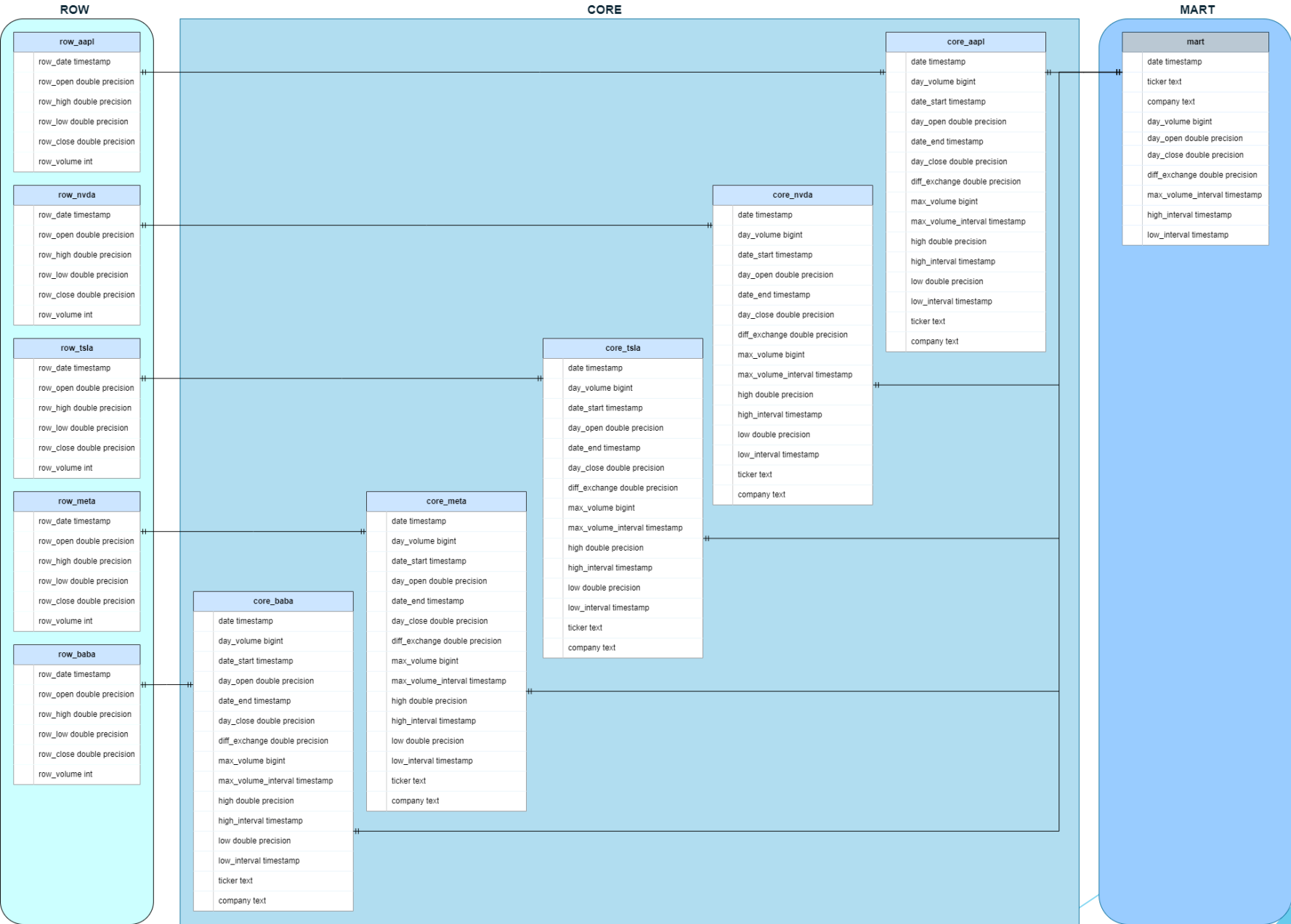
Источники:

[Free Stock APIs in JSON & Excel | Alpha Vantage](#)

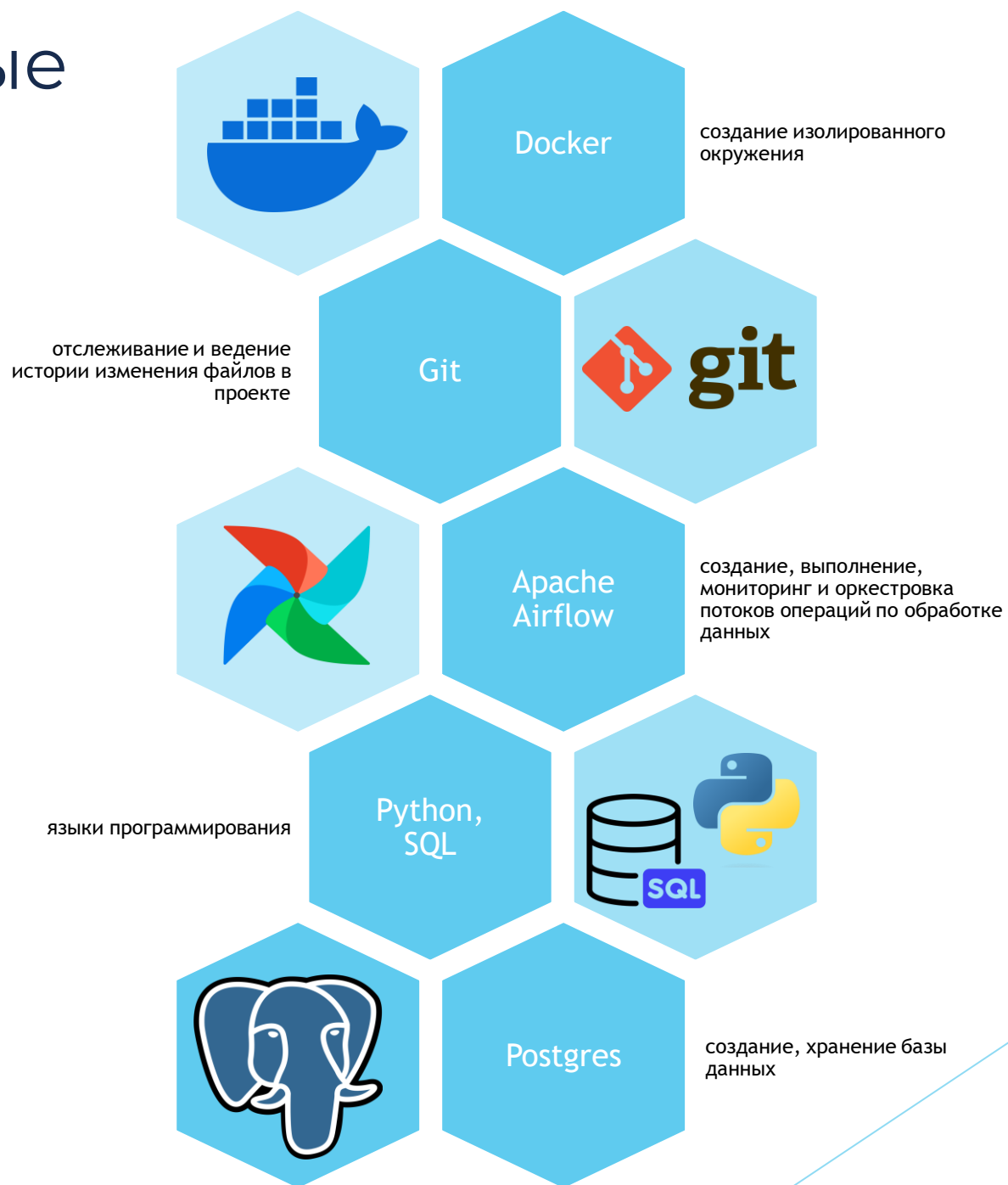
План реализации

1. Анализ источника данных для определения способа сбора информации
2. Проектирование структуры DWH
3. Определение стека используемых при реализации проекта технологий
4. Физическая реализация проекта: написание скриптов для создания БД, настройка соединений

ER-диаграмма

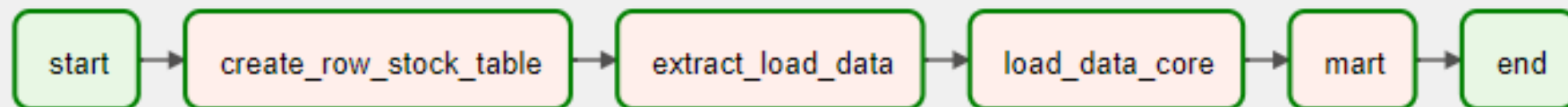


Используемые технологии



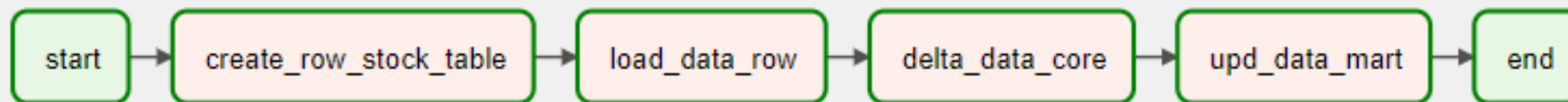
Результаты разработки

Инициализирующий режим загрузки данных



date	ABC ticker	ABC company	123 day_volume	123 day_open	123 day_close	123 diff_exchange	max_volume_interval	high_interval	low_interval
2023-11-30 00:00:00.000	AAPL	APPLE	60 940 414	189,45	190,3	-0,4486671945	2023-11-30 16:00:00.000	2023-11-30 09:33:00.000	2023-11-30 17:33:00.000
2023-11-30 00:00:00.000	NVDA	NVIDIA	60 140 687	481,76	465,98	3,2754898705	2023-11-30 16:00:00.000	2023-11-30 17:15:00.000	2023-11-30 17:34:00.000
2023-11-30 00:00:00.000	TSLA	TESLA	135 270 669	245	235,1	4,0408163265	2023-11-30 16:00:00.000	2023-11-30 09:06:00.000	2023-11-30 17:11:00.000
2023-11-30 00:00:00.000	BABA	ALIBABA ADR	19 758 231	74,67	74,5	0,2276684077	2023-11-30 16:02:00.000	2023-11-30 17:22:00.000	2023-11-30 09:41:00.000
2023-11-30 00:00:00.000	META	META PLATFORMS	26 980 876	332,31	326,32	1,8025337787	2023-11-30 16:00:00.000	2023-11-30 17:33:00.000	2023-11-30 17:10:00.000

Инкрементальный режим загрузки данных



date	ABC ticker	ABC company	123 day_volume	123 day_open	123 day_close	123 diff_exchange	max_volume_interval	high_interval	low_interval
2023-12-01 00:00:00.000	AAPL	APPLE	50 066 616	190,3	191,29	-0,5202312139	2023-12-01 16:00:00.000	2023-12-01 17:21:00.000	2023-12-01 09:22:00.000
2023-12-01 00:00:00.000	NVDA	NVIDIA	37 893 448	472,17	467,55	0,9784611475	2023-12-01 16:00:00.000	2023-12-01 09:22:00.000	2023-12-01 09:35:00.000
2023-12-01 00:00:00.000	TSLA	TESLA	122 516 163	236,5	238,97	-1,044397463	2023-12-01 16:00:00.000	2023-12-01 17:16:00.000	2023-12-01 17:16:00.000
2023-12-01 00:00:00.000	BABA	ALIBABA ADR	32 448 284	73,75	74,14	-0,5288135593	2023-12-01 16:00:00.000	2023-12-01 08:00:00.000	2023-12-01 09:32:00.000
2023-12-01 00:00:00.000	META	META PLATFORMS	16 300 282	327,26	324,81	0,7486402249	2023-12-01 16:00:00.000	2023-12-01 17:12:00.000	2023-12-01 11:00:00.000

Выводы

1. Создан ETL-процесс формирования витрин данных для анализа изменений курса акций
2. В процессе выполнения работы закреплены на практике знания и навыки работы с изученными на курсе технологиями используемыми в инженерии данных