

АРХИТЕКТУРА РЕШЕНИЯ.

Исполнитель: Асет Иманкулов

Должность: разработчик

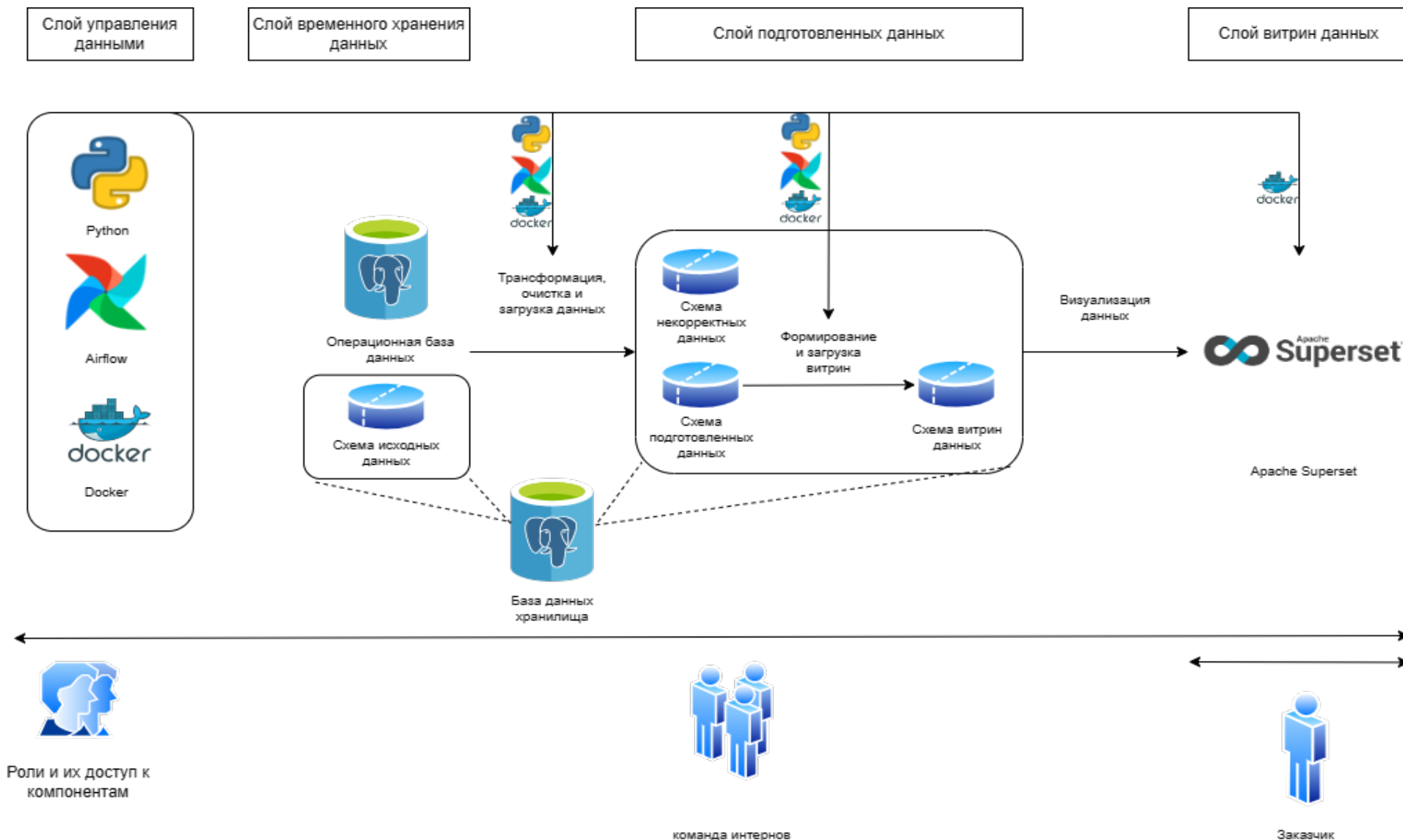
Дата: 31.07.2023

Версия документа: 3.0

Оглавление

1. СХЕМА АРХИТЕКТУРНОГО РЕШЕНИЯ.	3
2. ОПИСАНИЕ КОМПОНЕНТОВ.	4
3. ОПИСАНИЕ СЛОЕВ.	6

1. СХЕМА АРХИТЕКТУРНОГО РЕШЕНИЯ.



2. ОПИСАНИЕ КОМПОНЕНТОВ.

2.1. Наименование баз данных и их элементов в информационной системе

База данных	Наименование
Операционная база данных	internship_sources
База данных хранилища	internship_7_db

База данных	Схема	Наименование
Операционная база данных	Схема исходных данных	sources
База данных хранилища	Схема исходных данных	sources
	Схема подготовленных данных	dds
	Схема некорректных данных	data_quality
	Схема витрин данных	datamarts

2.2. Версии используемого ПО.

Программное обеспечение	Версия программного обеспечения
PostgreSQL	12.10
Python	3.10
Docker (клиент)	24.0.2
Docker (сервер)	24.0.2
Apache Airflow	2.6.2
Apache Superset	2.1.0

2.3. Параметры подключения к компонентам.

Подключение к базам данных:

База данных	Хост	Порт	Пользователь	Схема
internship_sources	10.1.108.29	5432	interns_7	sources
internship_7_db	10.1.108.29	5432	interns_7	sources
				dds
				data_quality
				datamarts

Подключение к Apache Airflow:

Параметр	Значение
URL веб-интерфейса	http://localhost:8080/home
Логин	interns_7
Путь к ДАГам	../docker/airflow/dags/
Путь к исполняемому скрипту	../docker/airflow/docker-compose.yml

Подключение к Apache Superset:

Параметр	Значение
URL веб-интерфейса	http://localhost:9000/superset/welcome/
Логин	interns_7
Путь к исполняемому скрипту	../docker/superset/docker-compose-non-dev.yml

2.4. Используемые сущности.

Слой	База данных	Схема	Сущность	Тип
Слой временного хранения данных	internship_sources	sources	brand	таблица
			category	таблица
			product	таблица
			stock	таблица
			transaction	таблица
	internship_7_db	sources	product_quantity	таблица
			stores	таблица
			stores_emails	таблица
			transaction_stores	таблица
Слой подготовленных данных	internship_7_db	dds	brand	таблица
			category	таблица
			product	таблица
			stock	таблица
			transaction	таблица
			product_quantity	таблица
			stores	таблица
			stores_emails	таблица
			transaction_stores	таблица
		data_quality	brand	таблица
			category	таблица
			product	таблица
			stock	таблица
			transaction	таблица
			product_quantity	таблица
			stores	таблица
			stores_emails	таблица
			transaction_stores	таблица
			upload_tables_tech	таблица
Слой витрин данных	internship_7_db	datamarts	orders_data	таблица
			stock_data	таблица
			stores_data	таблица
			orders_data_mart	представление
			stock_data_mart	представление
			stores_data_mart	представление

3. ОПИСАНИЕ СЛОЕВ.

Архитектурное решение представлено четырьмя слоями: слой управления данными, слой временного хранения данных, слой подготовленных данных, слой витрин данных.

В слое управления данным указаны технологии, осуществляющие обработку данных. Docker осуществляет контейнеризацию Airflow с Python. При этом, Airflow оркестрирует исполнение двух Python-скриптов, содержащих логику обработки данных, и двух python-скриптов, предназначенных для запуска DAG'ов. Первый скрипт предназначен для трансформации и очистки исходных данных и их последующую загрузку в схему подготовленных данных операционной базы данных. Данные, непрошедшие проверку качества попадают в таблицы логирования, размещенные на схеме некорректных данных.

Второй python-скрипт на основе полученных данных формирует и загружает витрины данных. Формирование витрин происходит в два этапа: сперва создаются таблицы соединения, а затем, на их основе, создаются представления. Образованные витрины данных визуализируются с помощью инструмента создания чартов и дашбордов Superset, который, в свою очередь, контейнеризирован с помощью Docker.

Слой временного хранения данных представлен операционной базой (internship_sources), в которой хранятся исходные «сырые» данные и схемой «sources», размещенной на базе данных хранилища.

Слой подготовленных данных содержит базу данных хранилища (internship_7_db) с тремя схемами – схема подготовленных данных (dds), схема некорректных данных (data_quality) и схема витрин данных (datamarts).