



ФГАОУ ВО «Южно-Уральский государственный университет  
(национальный исследовательский университет)»  
Кафедра Электронные вычислительные машины

# Система поддержки аналитического мониторинга профайлов

Автор: студент группы КЭ-222

Топалов М.М.

Руководитель: к.т.н. доцент

Кафтанников И.Л.



# Актуальность работы

## Актуальность:

Обработка и анализ данных стали важным событием последнего десятилетия, многие компании уделяют особое внимание анализу данных. В настоящее время нельзя представить крупную компанию, которая не собирает и не хранит данные. В современном мире данные имеют ценность, которую можно сравнить с нефтью или золотом.

# Цель исследования

## Цель данной работы:

В рамках магистерской диссертации разработать система поддержки аналитического мониторинга профайлов с возможностью оповещения.

В ходе реализации работы были выполнены такие задачи, как анализ предметной области, анализ существующих решений, определение функциональных требований, выбор инструментов и инфраструктуры разрабатываемой системы, проектирование системы, реализация системы и проведение верификации.

# Пример мониторинга профайлов

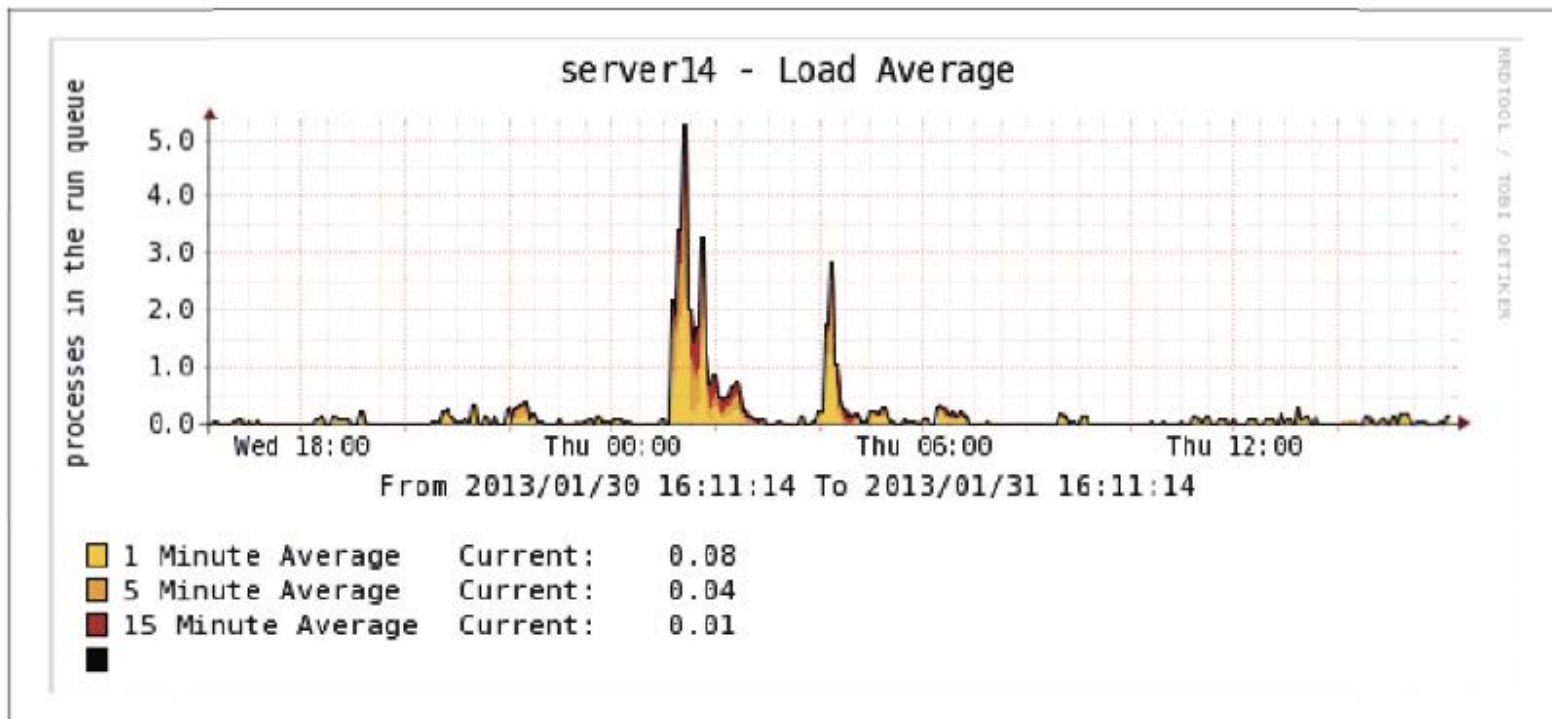


Рисунок 1 – Загрузка сервера

# Пример мониторинга профайлов

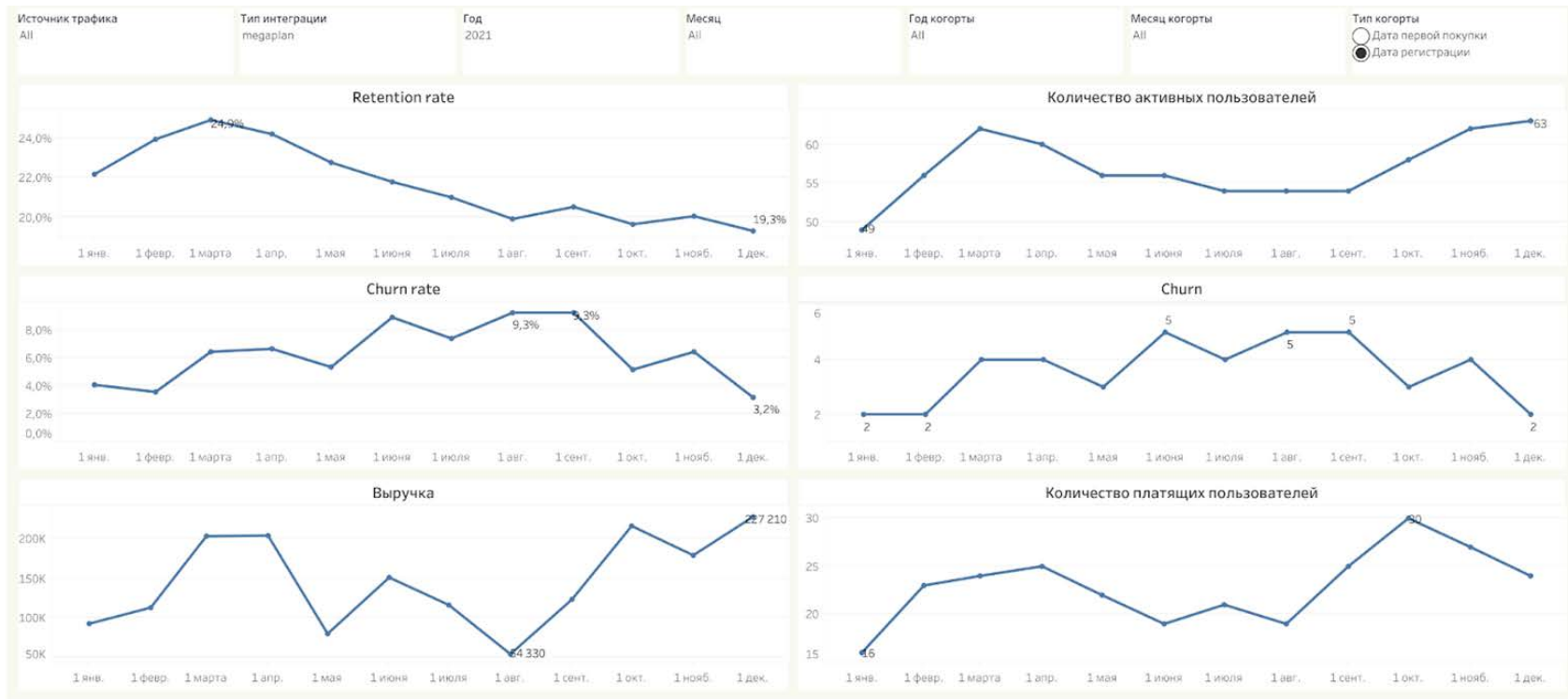


Рисунок 2 – Мониторинг профайлов для верхнеуровневой аналитики

# Пример мониторинга профайлов

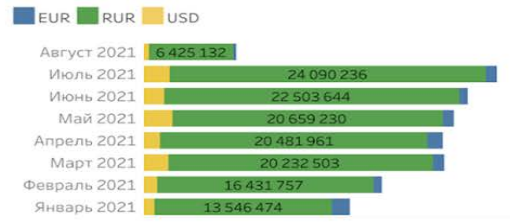
## Выручка по валюте

Укажите валюту  
All

Укажите год  
2021

Укажите месяц  
All

### Структура выручки



### Динамика выручки

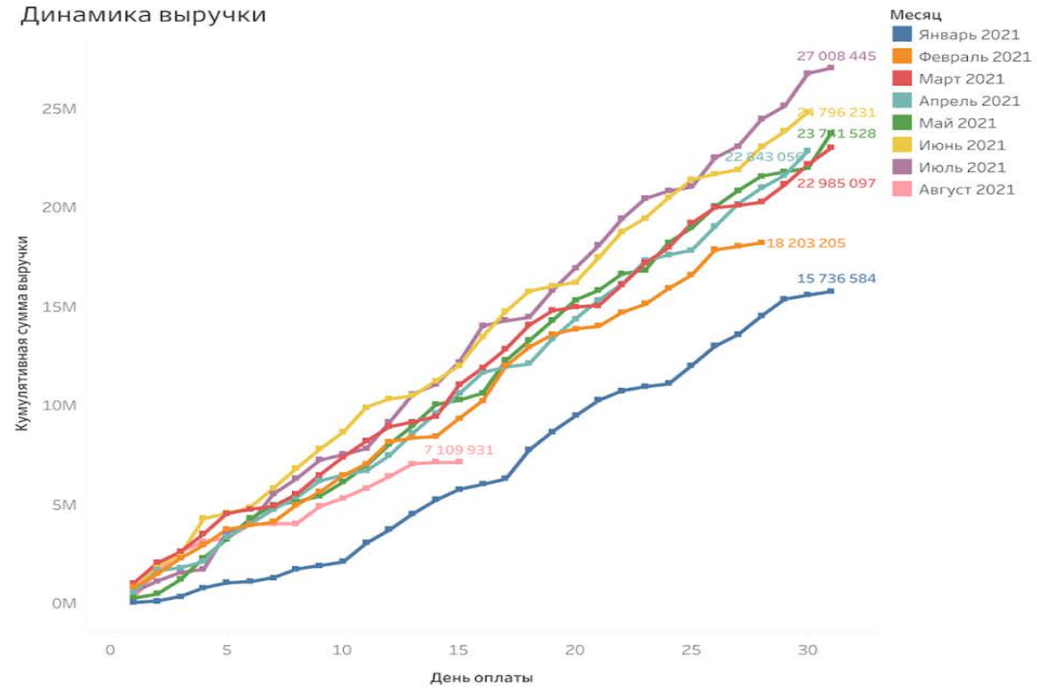


Рисунок 3 – Мониторинг выручки



# Основные этапы разработки мониторинга профайлов

1. Подготовка запроса SQL из сырых данных по рекламным логам
2. Создание витрины DWH
3. Разработка функции в Python для поиска аномальных значений
4. Автоматизация работы в планировщике Apache airflow

# Дата-пайплайн для автоматизации (ETL)

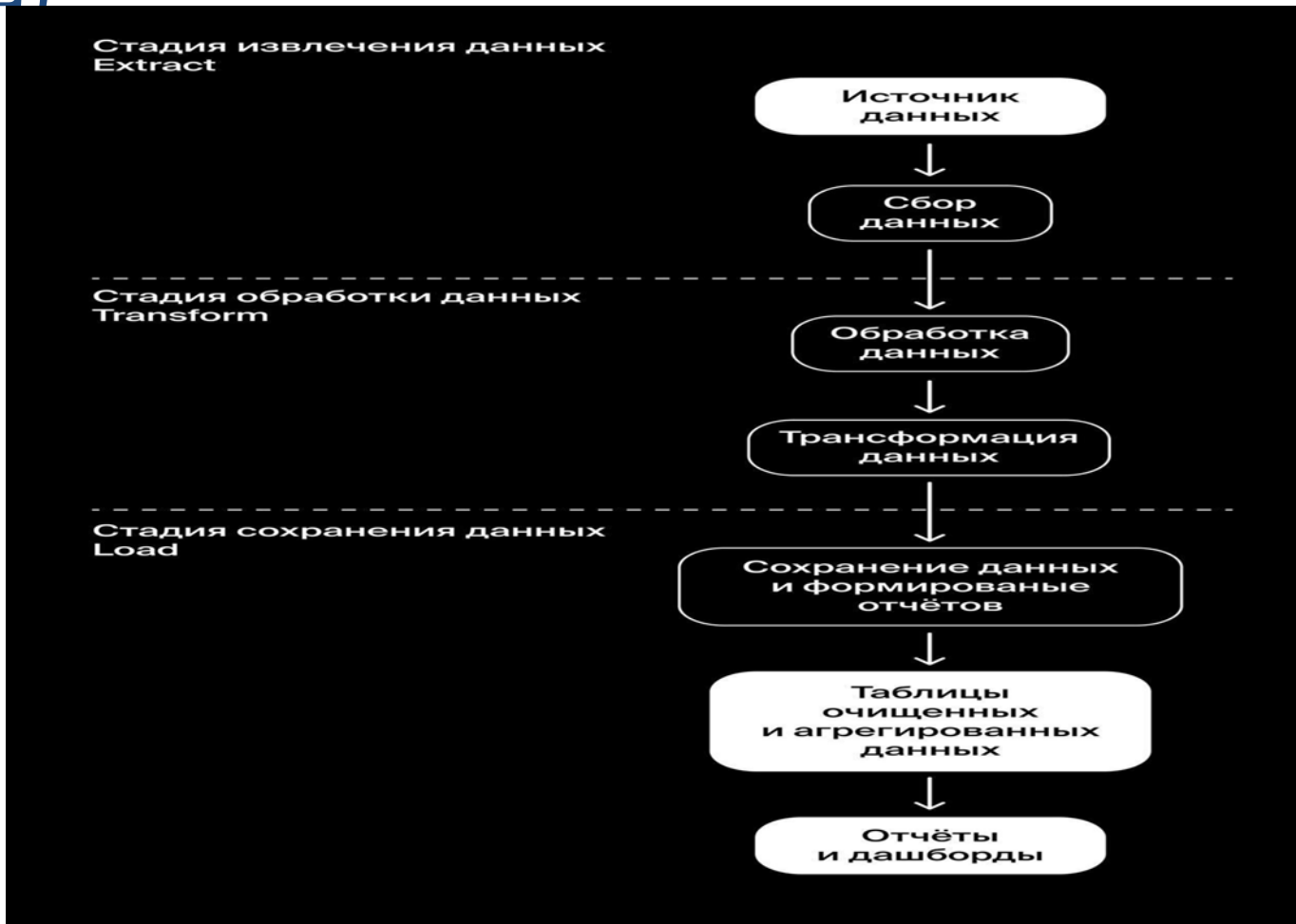


Рисунок 4 – Элементы пайплайна ETL



# Предобработка данных

Название поля	Тип	Описание
id	SERIAL	ID клиента
dateTime	TIMESTAMP	время события
utm	JSON	utm-метка
event	VARCHAR	Событие
campaignId	INT	ID компании
url	VARCHAR	
date	TIMESTAMP	Дата
pageTitle	VARCHAR	Заголовок
agentId	INT	ID агент
path	VARCHAR	Путь
device	VARCHAR	Устройство
details	JSON	
os	VARCHAR	
browser	VARCHAR	Браузер
adId	INT	ID рекламы
cost	VARCHAR	

Рисунок 5 – Структура сырых данных

# Результаты предобработки данных

	client_union_id	campaign_union_id	agency_union_id	ad_id	platform	date_time	event	ad_cost_type
0	client_21074	campaign_27489	no_agency	ad_27489	android	2022-04-08 17:22:25	report	CPM
1	client_22392	campaign_35893	no_agency	ad_35893	android	2022-04-12 21:10:58	click	CPM
2	client_16836	campaign_26799	no_agency	ad_26804	web	2022-04-16 08:16:41	click	CPM
3	client_22920	campaign_37146	no_agency	ad_37146	android	2022-04-09 15:07:59	click	CPM
4	client_645	campaign_15893	no_agency	ad_15944	web	2022-04-11 07:11:57	click	CPM

Рисунок 6 – Готовая витрина данных

# Модель для детектирования аномальных значений

$$EMA_t = \alpha \cdot p_t + (1 - \alpha) \cdot EMA_{t-1}$$

Экспоненциально взвешенное скользящее среднее (EMA), экспоненциальное скользящее среднее — разновидность взвешенной скользящей средней, веса которой убывают экспоненциально и никогда не равны нулю.

Было решено использовать экспоненциальное скользящее среднее, т.к. при использовании скользящего среднего почти всегда возникает лаг, и для его уменьшения необходима экспоненциальная скользящая средняя. Показатель EMA придает последним данным больший вес по сравнению с предыдущими значениями. Этот факт позволяет реагировать на текущие изменения быстрее

# Результат поиска аномальных значений

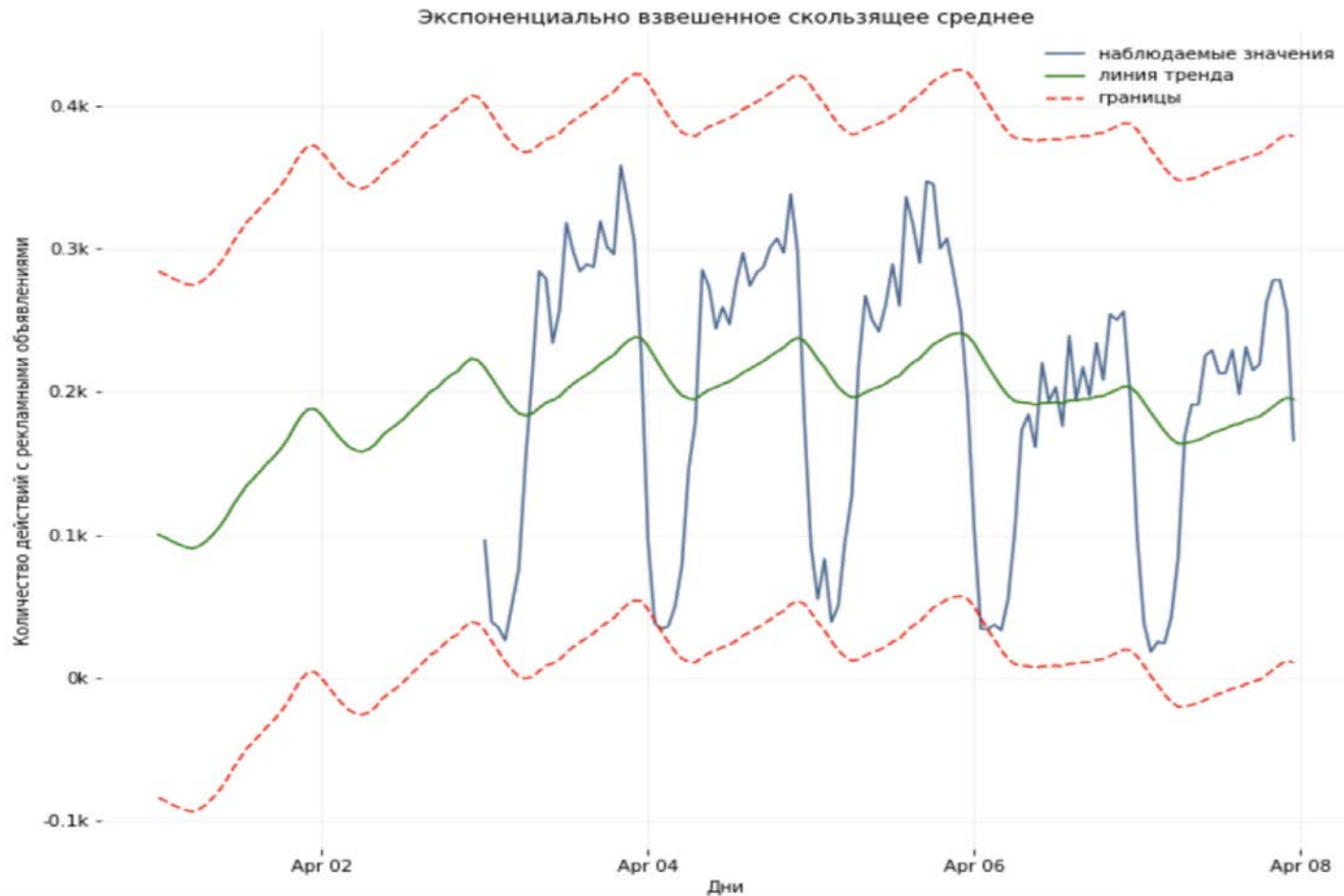


Рисунок 6 – Экспоненциально взвешенное скользящее среднее без аномалий

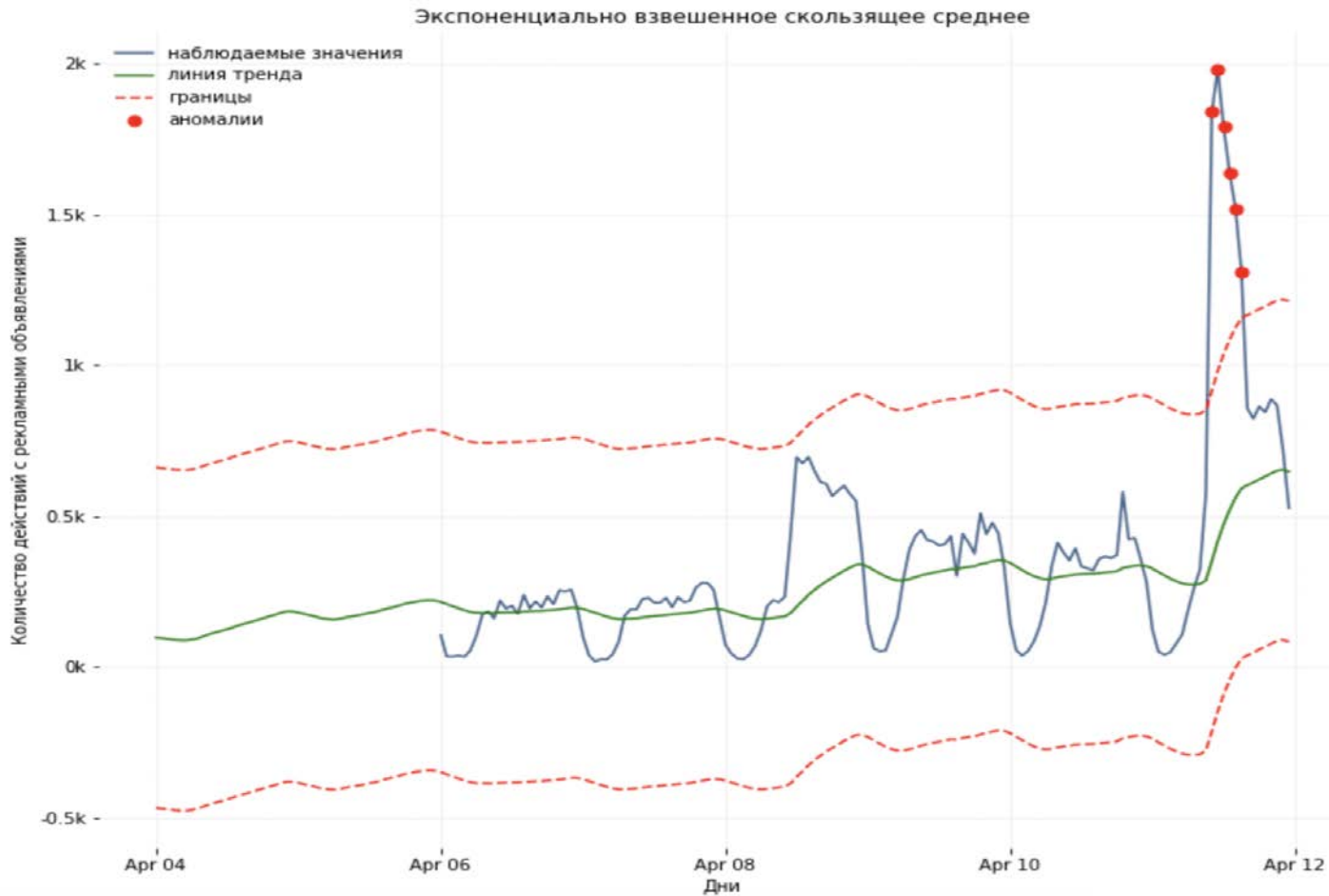


Рисунок 7 – Экспоненциально взвешенное скользящее среднее с аномалиями



# Автоматизация мониторинга профайлов

## DAGs

All 11 Active 7 Paused 4

Filter DAGs by tag

Search DAGs

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks
<input checked="" type="radio"/> all	analytics-team@reg.ru	95 (1)	0 0 * * *	2022-05-07, 00:00:00	2022-05-08, 00:00:00	7
<input checked="" type="radio"/> Anomaly_monitoring	Mikhail Topalov	179	@hourly	2022-05-07, 17:00:00	2022-05-07, 18:00:00	4

Рисунок 8 – DAG в Apache airflow

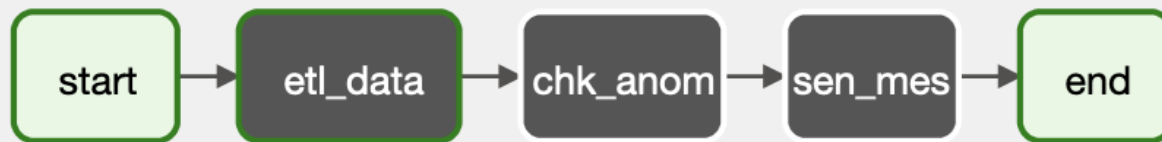


Рисунок 9 – Блок схема DAG

# Заключение

При решении поставленных задач получены результаты, которые заключаются в следующем:

1. Выявлены специфические аспекты мониторинга профайлов, связанные с особенностями его использования и реализации.
2. Проведен анализ методов мониторинга профайлов. Выделены их преимущества и недостатки.
4. Предложены и реализованы модели скользящего среднего, позволяющие выявлять быстрее изменения в данных. Выделены основные преимущества и недостатки методов.
6. Разработана система, позволяющая автоматизировано проводить мониторинг профайлов, искать выбивающиеся значения. Что поможет бизнесу не терять деньги на не выявленных проблемах.





# Вывод

На основании полученных результатов были сделаны следующие выводы:

1. Предложенные методы мониторинга профайлов позволяют выявить основные аномалии в данных.
2. Разработанная система позволяет обрабатывать большой объем данных, а также моментально реагировать на возникающие проблемы в бизнесе.



South Ural  
State University

National Research  
University

Спасибо за  
внимание