**Лабораторная работа (Лекция 54)**

*Предложите архитектурную диаграмму по обнаружению аномалий в значениях потоковых данных, если значение выходит за рамки допустимого интервала по каждому месторождению*

Предлагаемая схема



Данные отправляются по протоколу MQTT на сервер Apache Kafka – Topic 1.Далее информация передается в подсистему анализа данных.

Подсистема анализа данных получает из БД данных адреса параметров и граничные значения для каждого типа параметра. Выполняет проверку полученных из Topic 1 параметров на вхождение в допустимые значения. При определении отклонений отправляет необходимую информацию (параметр, месторождение, значения, тип аларма, время и т.д.) в подсистему отправки уведомлений.

Подсистема отправки уведомлений – отправляет уведомления с определенным сообщением. Контролирует успешность доставки.

Далее подсистемы анализа и отправки уведомлений формируют сообщения в Topic 2, после чего они записываются в БД для дальнейшего хранения.

1. *Нарисуйте и опишите модель данных с хранением обнаруженных аномалий в потоковых данных с указанием того что конкретная аномалия была отправлена как уведомление ответственным лицам*

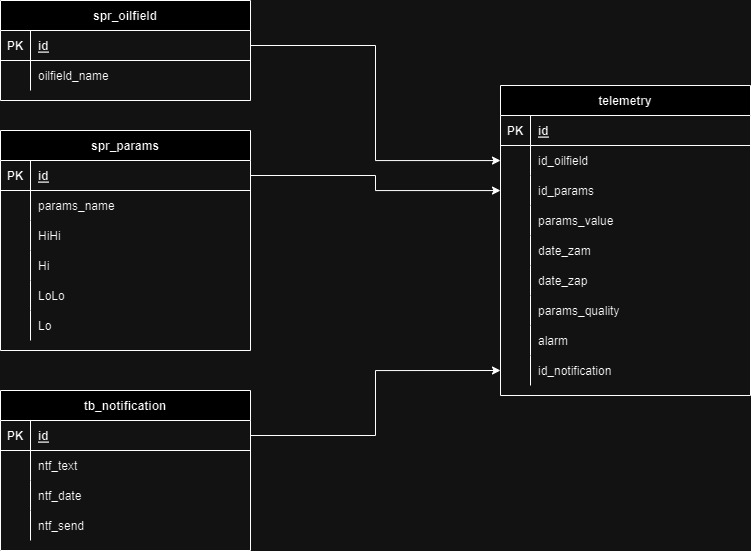
Упрощенная схема предложена на рисунке. Для упрощения принимает, что по всем месторождениям технологические параметры имеют одинаковые граничные значения (на практике это не так).

'spr\_oilfield' – таблица, справочник технологических параметров. Граничные значения по каждому параметру хранятся вместе с параметром.

'spr\_params' – таблица, справочник месторождений.

'tb\_notification' – таблица, хранения сообщений/уведомлений и их статуса (отправлено или нет). Статус 'отправлен' присваивается при получении подтверждения о доставке.

'telemetry' – таблица хранения текущих значений по параметрам, типа аларма и отсылкой на уведомление. Дополнительные поля: дата замера и дата записи в БД. В таблице хранятся все замеры.



1. *Укажите и обоснуйте какие технические компоненты (библиотеки / open source компоненты) вы привлечете для хранения проведенных нотификаций / обнаруженных аномалий и также для хранения пороговых уровней по каждому месторождению.*

Для долговременного хранения всей информации предлагаю использовать БД Greenplum. Распределенная MPP база.

Для аналитики ClickHouse, в которой создавать витрины из Greenplum.