### Полезные ссылки:

<https://habrahabr.ru/post/247609/>

### 1. Автоматизация и коммодизация процессов обработки данных

Большие затраты на техническое обслуживание систем хранения и обработки данных заставляют компании искать более дешевые решения. Одним из таких решений является отказ от собственных дата-центров и переход на облачные технологии хранения и анализа данных такие как [Amazon WS](http://aws.amazon.com/s3/), [Microsoft Azure](http://azure.microsoft.com/ru-ru/).  
  
Острая нехватка специалистов в области анализа данных служит хорошим стимулом для развития автоматизированных инструментов, не требующих глубоких знаний в области технологий и ориентированных на решение бизнес задач. В качестве примера таких инструментов можно привести [Tableau](http://www.tableausoftware.com/), [IBM Watson](http://www.ibm.com/smarterplanet/us/en/ibmwatson/what-is-watson.html) и [Azure Machine Learning](http://azure.microsoft.com/ru-ru/services/machine-learning/).  
  
Отдельно стоит отметить развитие систем автоматической предобработки сырых и неструктурированных данных подобных [OpenRefine](http://openrefine.org/) и [DataWrangler](http://vis.stanford.edu/wrangler/).

<https://aws.amazon.com/ru/s3/>

<https://azure.microsoft.com/ru-ru/>

[https://studio.azureml.net/?selectAccess=true&o=1#](https://studio.azureml.net/?selectAccess=true&o=1)

<https://cloud.google.com/>

**Введение**

В настоящее время получили распространение большое количество разнообразных датчиков потребительского назначения для автоматизированных систем, именуемых как «Умный дом». Для конечного потребителя, интересующегося данной сферой и желающего охватить и контролировать с помощью этих устройств максимальное количество аспектов своей жизнедеятельности, возникает непреодолимая преграда в виде шквала разнообразной информации, поступающей от всевозможных датчиков. Разобраться во всем этом порой не под силу даже опытному инженеру. Помочь конечному потребителю справиться с этим призван проект, который разрабатывается в рамках текущей курсовой работы.

Цели

* Разработать архитектуру облачного анализатора данных для автоматизированных систем домашнего пользования;
* Разработать основной алгоритм, применяемый в анализаторе.

Задачи

* Провести обзор аналогов облачного анализатора данных для автоматизированных систем домашнего пользования;
* Выполнить разработку архитектуры анализатора данных;
* Спроектировать дизайн клиентской части анализатора;
* Изучить алгоритмы, применяемые в аналогах для анализа данных;
* Подготовить данные для разрабатываемого алгоритма;
* Разработать основной алгоритм, применяемый в анализаторе.

**Требования к разрабатываемому облачному анализатору данных автоматизированных систем домашнего пользования**

Облачный анализатор данных автоматизированных систем домашнего пользования представляет из себя средство для сбора, агрегации и анализа данных, поступающих через сеть Интернет с различных датчиков. Он расположен на сервере и взаимодействует с пользователем через браузер или специальное приложение. В нем реализована возможность долговременного хранения информации, ее обработки, отображения в реальном времени поступающей информации с датчиков. Он способен постоянно анализировать информацию и с учетом вновь поступающей производить переоценку сделанных ранее выводов. Стоит отдельно отметить наличие особого алгоритма, который способен обнаруживать в потоке поступающих данных отклонения, не характерные для основной массы поступающей информации и именуемые далее «аномалиями». Так же этот алгоритм может идентифицировать такие аномалии из числа других аномалий по характерным «сигнатурам» – наличию, величине, количеству и продолжительности экстремумов в поступающей информации, а так же после возникновения первых отклонений от нормы в поступающей информации предсказывать определенные аномалии с вычисленной вероятностью их появления.

Функциональные требования:

1. Получение и запись в БД данных датчиков, поступающих через сеть по протоколу HTTP;
2. Добавление нового датчика в анализатор;
3. Информация о датчиках хранится в отдельной таблице БД;
4. Возможность идентифицировать конкретный датчик по данным, поступающим с него;
5. Возможность создавать кластеры данных, поступающих из определенных датчиков;
6. Информация, определяющая кластеры данных, хранится в отдельной таблице БД;
7. Долговременное хранение в БД данных с датчиков;
8. Хранит информацию об ошибках в отдельной таблице;
9. Отображение всех имеющихся в БД данных;
10. Отображение данных из БД за определенный срок;
11. Отображение всех данных из БД одного датчика;
12. Отображение из БД данных определенного датчика или группы датчиков за определенный срок;
13. Отображение всех аномалий за весь период времени на графике;
14. Отображение конкретной аномалии на графике;
15. Информирование об обнаружении аномалии;
16. Отображение вероятности идентификации новой аномалии при ее обнаружении как той или иной уже сохраненной;
17. Автоматическое определение новой аномалии;
18. Аномалия определяется только для определенных кластеров данных;
19. Отображение подробной информации о конкретной аномалии;
20. Отображение графика в выводе;
21. Постоянное обновление данных в выводе;
22. Отображение линии тренда на графике, рассчитанного для данных за определенный период времени;
23. Отображение регрессии на графике, рассчитанного для данных за определенный период времени;
24. Отображение доверительного интервала на графике, рассчитанного для данных за определенный период времени;
25. Возможность предсказывать на ограниченный срок значения, поступающих с датчиков данных;
26. Доступ к функционалу анализатора через авторизацию;
27. Хранение логинов и паролей пользователей в отдельной таблице БД;
28. В анализаторе имеется 2 группы пользователей: администраторы и клиенты;
29. Разные клиенты не могут обратиться к одинаковым таблицам данных БД;
30. Группа «Клиенты»:
    1. Имеет возможность зарегистрироваться;
    2. Может авторизоваться;
    3. Может добавлять и удалять датчики, с которых поступает информация;
    4. Имеет отдельные от других клиентов таблицы данных;
    5. Может делать вышеуказанные выборки данных из БД;
    6. Получать анализ данных;
    7. Имеет возможность описывать обнаруживаемые аномалии;
    8. Имеет возможность удалять сигнатуры аномалий;
31. Группа «Администраторы»:
    1. Может авторизоваться;
    2. Имеет возможность создавать новых администраторов;
    3. Может удалить клиента и привязанные к нему таблицы;
    4. Может удалить администратора;
    5. Может добавить нового клиента;
    6. Имеет возможность просматривать записи из всех таблиц БД;
    7. Имеет возможность удалять записи из всех таблиц БД;
    8. Имеет возможность читать информацию об ошибках;
    9. Имеет возможность добавлять свои сигнатуры аномалий;
    10. Имеет возможность удалять сигнатуры аномалий;
    11. Имеет возможность описывать обнаруживаемые аномалии;
    12. Может вносить изменения в параметры работы алгоритмов анализа;
    13. Может делать вышеуказанные выборки данных из БД;
    14. Получает анализ данных;

Нефункциональные требования:

1. Время отклика анализатора не более 20 с.;
2. Анализатор должен быть отказоустойчивым;
3. Данные в БД хранятся не более 5 лет;
4. Анализатор должен предоставлять удобный и понятный пользовательский интерфейс;
5. Анализатор предоставляет возможность для одновременной работы 20-ти пользователей;

**Обзор аналогов**

Как следует из требований к проектируемому анализатору, основная его идея заключается не просто в отображении всей информации, поступающей из различных датчиков, а в ее агрегации, анализе, предсказывании дальнейших состояний контролируемой системы, представлении в более сжатой и понятной для пользователя форме.   
В текущий момент на рынке нечто подобное в плане применения в автоматизированных системах домашнего пользования найти достаточно трудно. Ниже приведены некоторые из самых распространенных средств:

*MajorDoMo* - представляет собой бесплатную и открытую российскую программную платформу для комплексного управления домашней автоматикой, а так же для информационной поддержки жизнедеятельности. Данная система может быть установлена практически на любой персональный компьютер (на платформе Windows и Linux) и совершенно не требовательна к ресурсам. Даже без привязки к оборудованию она может быть использована для организации персональной инфо-среды. Данный проект является частью эко-системы SmartLiving. Из его возможностей выделяют:

* Кросс-платформенность (Windows/Linux);
* Бесплатная для личного или коммерческого использования;
* Большое и активное сообщество вокруг проекта;
* Поддержка различного оборудования;
* Мульти-язычный интерфейс (Русский/English);
* Веб-доступ с любого устройства;
* Веб-интерфейс с обновлением в реальном времени;
* Push-уведомления;
* Интеграция со сторонними веб-сайтами и сервисами;
* Управление мультимедиа;
* Модель безопасности с разграничением доступа между пользователями;
* CloudSync — облачная синхронизация и простой доступ из любого места;
* Построена на веб-технологиях (PHP/JS/HTML5);
* Программирование с помощью PHP и/или визуальной среды Blockly.

https://majordomo.smartliving.ru/

*DeviceHive* – инструмент, способный любое подключенное устройство сделать частью Интернета вещей. Он обеспечивает коммуникационный уровень, программное обеспечение управления и много платформенные библиотеки для начальной загрузки и разработки умных устройств, домашней автоматизации, дистанционного зондирования, телеметрии, программного обеспечения для дистанционного управления и мониторинга и многое другое. Подключение встраиваемых Linux-систем с использованием Python или C++, протоколов Json или подключения АВР. Разработка клиентских приложений с использованием библиотек HTML5/JavaScript, iOS и Android.

<https://www.devicehive.com/>

*1-M Умный дом* - поддерживает множество датчиков и устройств от разных производителей. Благодаря чему у пользователя появляется свобода создания Умного Дома, не привязываясь к брендам, а выбирая по техническим и ценовым показателям устройства для конкретных задач.  
Поддерживается радиосвязь с датчиками и устройствами фирм: Livolo, Noolite, RAEX, Telis, TKBHOME, GLANZEN, Chuango, Homesecur, Oregon, Nice, Visonic, AnMotors, CAME, Eazylift, EmilLux, FalconEye, Focus, Gant, Ginzzu, Hormann, Kelvin, Baisheng, TechLiCo, Feron, EmilLux, Expert, Fuhaote и др. Среди его возможностей можно выделить:

* голосовое управление без дополнительных программ ( ДУСЯ, Таскер и т.п.);
* оповещение о событиях в системе через SMS или e-mail;
* выполнение сценариев даже без интернет (оффлайн);
* запуск сценариев одной кнопкой или по условиям, событиям, суточным или недельным таймерам, прогнозу погоды и т.п.;
* геолокация - запуск сценариев по положению смартфонов членов семьи;
* ведение журнала работы устройств, просмотр в виде графиков; шифрованная связь.

<https://www.1-m.biz/>

*PRTG Network* – условно-бесплатная программа (пробный период — 30 дней) предназначенная для мониторинга использования сети, работает в семействе операционных систем Windows. Из возможностей данной программы:

* Сбор информации о потоках данных, проходящих через конкретные устройства, с сохранением её в базе данных программы;
* Просмотр статистики в базе данных в виде графиков и таблиц;
* Просмотр результатов в режиме реального времени или за определенный промежуток времени в прошлом на разных устройствах;
* Сбор данных о нагрузке на подсистемы памяти и процессора.

Отметим, что основным предназначением данной программы и приведенных ниже является снятие метрик с серверов и дата центров, нагрузок на всевозможные интернет сервисы. Но порой на практике его применяют и в целях снятия показаний с датчиков, применяемых в системах умного дома.

<https://www.paessler.com/prtg>

*Zabbix* –  свободная система мониторинга и отслеживания статусов разнообразных сервисов компьютерной сети, серверов и сетевого оборудования, написанная Алексеем Владышевым. Для хранения данных используется MySQL, PostgreSQL, SQLite или Oracle Database, веб-интерфейс написан на PHP. Из возможностей данной программы:

* Распределённый мониторинг — до нескольких тысяч узлов. Конфигурация младших узлов полностью контролируется старшими узлами, находящимися на более высоком уровне иерархии;
* Сценарии на основе мониторинга;
* Централизованный мониторинг журналов;
* Веб-интерфейс для администрирования и настройки;
* Отчётность и тенденции;
* Комплексная реакция на события;
* Расширение за счёт выполнения внешних скриптов;
* Гибкая система шаблонов и групп.

<https://www.zabbix.com/>

*Cacti* – open-source веб-приложение, система позволяет строить графики при помощи RRDtool. Cacti собирает статистические данные за определённые временные интервалы и позволяет отобразить их в графическом виде. Преимущественно используются стандартные шаблоны для отображения статистики по загрузке процессора, выделению оперативной памяти, количеству запущенных процессов, использованию входящего/исходящего трафика. Из возможностей Cacti:

* Неограниченное количество элементов графика может быть определено для каждого графика;
* С помощью математических функций CDEF можно управлять данными графа встроенными в RRDTool. Эти функции CDEF можно определить в Cacti и использовать глобально на каждом графике;
* Параметры RRA настраивают для пользователя возможность сбора данных по нестандартным временным интервалам при одновременном хранении различных данных;
* Содержит механизм ввода данных, который позволяет пользователям определять пользовательские сценарии, может быть использовано для сбора данных;
* Шаблоны графиков позволяют группировать общие графики по шаблону.

<https://www.cacti.net/>

Необходимо определить критерии для оценки приложений и сервисов

Для того чтобы провести оценку приведенных ранее приложений сперва выделим те качества, которые нас интересуют, а именно: стоимость приобретения, страну происхождения, расположен ли в облаке, возможность строить графики из поступающей информации, возможность обрабатывать поступающие данные, возможность делать прогнозы, количество подключаемых датчиков. Достоинства и недостатки приведенных выше приложений представлены в таблице 1.

*Таблица 1.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | MajorDoMo | DeviceHive | 1-М Умный дом | PRTG Network | Zabbix | Cacti |
| Стоимость | Бесплатно | Бесплатно | Бесплатно | Бесплатно на 30 дней | Бесплатно | Бесплатно |
| Страна | Россия | США | Россия | Германия | Латвия | США |
| Расположение | ПК | Облако | ПК | ПК | Физический сервер | Физический сервер |
| Наличие графика | Присутствует | Отсутствует | Отсутствует | Присутствует | Присутствует | Присутствует |
| Обработка данных | Отсутствует | Присутствует | Отсутствует | Присутствует | Присутствует | Присутствует |
| Наличие прогноза | Отсутствует | Отсутствует | Отсутствует | Отсутствует | Присутствует | Отсутствует |
| Количество датчиков | Не ограничено | Не ограничено | 50 | 30 | Не ограничено | Не ограничено |

Нетрудно сделать вывод, что большинство из вышеуказанных средств только отображают поступающую информацию с датчиков и отмечают значения, отклоняющиеся от заданных границ. Лишь PRTG Network, Zabbix и Cacti обладают достаточным функционалом, чтобы должным образом анализировать поступающую информацию и даже давать краткосрочные предсказания по значениям поступающих данных (Zabbix). Но их недостатками является то, что компании разработчики расположены за пределами РФ и они располагаются не в облаке, а на физических серверах, следовательно, доступ к их данным возможен через сеть Интернет только после договоренности с провайдером.

**Разработка архитектуры облачного анализатора данных автоматизированных систем домашнего пользования.**

Формирование сущностей и атрибутов, построение схемы модели базы данных.

Согласно требованиям была составлена таблица сущностей и их атрибутов, которые будут применены в БД анализатора (см. табл. 2).

Таблица 2.

|  |  |  |  |
| --- | --- | --- | --- |
| **Название сущности** | **Название атрибута** | **Назначение атрибута** | **Описание атрибута** |
| Data | DataID | Идентификатор записи | Величина типа int. Первичный ключ. Номер записи. Не пустое. |
| Time | Время измерения | Поле типа datetime. Содержит в себе время измерения показаний. Не пустое. |
| Value | Измеренное значение | Поле типа long. Содержит в себе величину, измеренную датчиком. Не пустое. |
| SensorID | Идентификатор датчика | Поле типа int. Внешний ключ. Не пустое. |
| Sensors | SensorID | Идентификатор датчика | Величина типа int. Первичный ключ. Номер датчика. Не пустое. |
| Name | Имя датчика | Величина типа varchar. Содержит не более 50 символов. Не пустое. |
| Description | Информация о датчике | Величина типа varchar. Содержит не более 1000 символов. Может быть пустым. |
| ClusterID | Принадлежность кластеру | Величина типа int. Внешний ключ. Может быть пустым. |
| MaxValue | Максимальная возможная величина, исходящая от датчика | Величина типа long. Может быть пустым. |
| MinValue | Минимальная возможная величина, исходящая от датчика | Величина типа long. Может быть пустым. |
| ClientID | Идентификатор клиента | Поле типа int. Внешний ключ. Не пустое. |
| Clusters | ClusterID | Идентификатор кластера | Величина типа int. Первичный ключ. Номер кластера. Не пустое. |
| Name | Имя кластера | Величина типа varchar. Содержит не более 50 символов. Не пустое. |
| Description | Информация о кластере | Величина типа varchar. Содержит не более 1000 символов. Может быть пустым. |
| Client | ClientID | Идентификатор клиента | Величина типа int. Первичный ключ. Номер клиента. Не пустое. |
| Login | Логин клиента | Величина типа varchar. Содержит не более 50 символов. Не пустое. |
| Password | Пароль клиента | Величина типа varchar. Содержит не более 50 символов. Не пустое. |
| Administrators | AdministratorID | Идентификатор администратора | Величина типа int. Первичный ключ. Номер администратора. Не пустое. |
| Login | Логин администратора | Величина типа varchar. Содержит не более 50 символов. Не пустое. |
| Password | Пароль администратора | Величина типа varchar. Содержит не более 50 символов. Не пустое. |
| Logs | LogID | Идентификатор записи в журнале ошибки | Величина типа int. Первичный ключ. Номер записи в журнале. Не пустое. |
| TimeRegister | Время регистрации ошибки | Поле типа datetime. Содержит в себе время регистрации ошибки. Не пустое. |
| Description | Описание ошибки | Величина типа varchar. Содержит не более 1000 символов. Не пустое. |
| Settings | ClientID | Идентификатор клиента | Величина типа int. Первичный внешний ключ. Номер записи в журнале. Не пустое. |
| IntervalRegr | Интервал вычисляемой регрессии | Величина типа int. Не пустое. |
| IntervalPeace | Интервал измеряемого состояния покоя | Величина типа int. Не пустое. |
| Anomalies | AnomalyID | Идентификатор аномалии | Величина типа int. Первичный ключ. Номер аномалии. Не пустое. |
| TimeRegister | Время регистрации аномалии | Поле типа datetime. Содержит в себе время регистрации аномалии. Не пустое. |
| Name | Название аномалии | Величина типа varchar. Содержит не более 50 символов. Не пустое. |
| Description | Информация об аномалии | Величина типа varchar. Содержит не более 1000 символов. Может быть пустым. |
| ClientID | Идентификатор клиента | Поле типа int. Внешний ключ. Не пустое. |
| Signatures | SignatureID | Идентификатор позиции в сигнатуре | Величина типа int. Первичный ключ. Номер позиции в сигнатуре. Не пустое. |
| Value | Величина позиции сигнатуры | Величина типа long. Не пустое. |
| AnomalyID | Идентификатор аномалии | Поле типа int. Внешний ключ. Не пустое. |
| Labels | LabelID | Идентификатор метки | Величина типа int. Первичный ключ. Номер метки. Не пустое. |
| Start | Время начала метки | Поле типа datetime. Содержит в себе время начала метки. Не пустое. |
| End | Время конца метки | Поле типа datetime. Содержит в себе время окончания метки. Не пустое. |
| AnomalyID | Идентификатор аномалии | Поле типа int. Внешний ключ. Не пустое. |

После построения списка сущностей была построена схема модели базы данных. Сущности были представлены в виде таблиц с полями, а отношения составили связи между ними. Схема модели была построена с помощью Microsoft Visio. Ниже приведена схема модели (см. рис. 1).

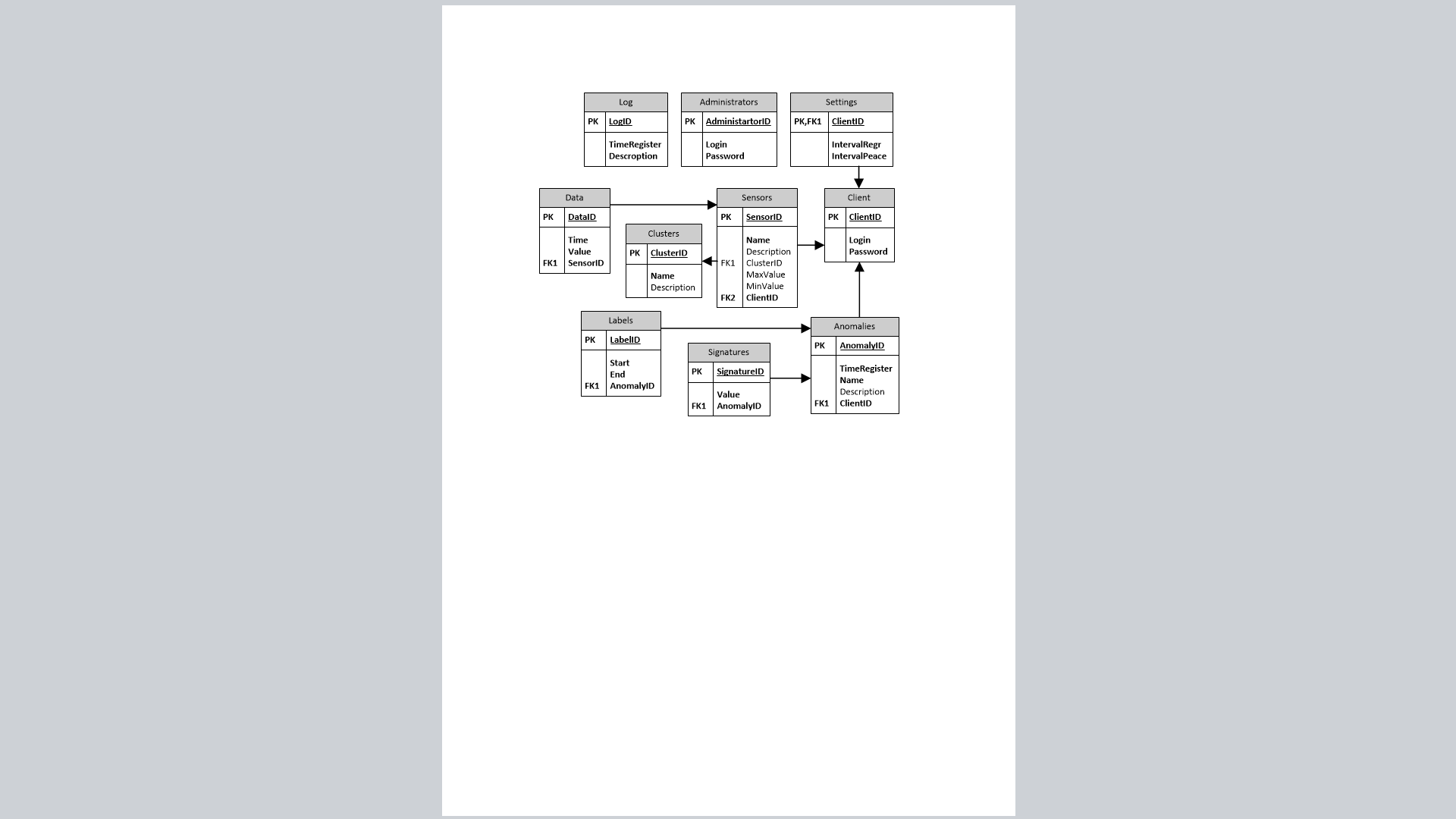


Рисунок 1. – Схема модели базы данных