### 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 системы не более 20 с.;

**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:

MajorDoMo -

https://majordomo.smartliving.ru/

Devicehive -

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

1-M Умный дом -

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

PRTG Network – то, что нужно

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

Zabbix – самое то

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

Cacti – почти то

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

Стоит отметить, что большинство из вышеуказанных средств только отображают поступающую информацию с датчиков и отмечают значения, отклоняющиеся от заданных границ. Лишь PRTG Network и Zabbix обладают достаточным функционалом, чтобы должным образом анализировать поступающую информацию и даже давать краткосрочные предсказания по значениям поступающих данных. Но их недостатками является плата за использование сервиса, ограниченное количество подключаемых источников информации при приобретении за минимальную цену и то, что компании разработчики расположены за пределами РФ.