

ВЗАИМОДЕЙСТВИЕ ТЕХНОЛОГИЙ INTERNET OF THINGS И BIG DATA

М.А. Герасименко

*Московский институт электроники и математики Национального
исследовательского университета «Высшая школа экономики», г. Москва,
Россия Тел. +7(917)529-53-23; E-mail magerasimenko@gmail.com*

Исследуются современные технологии, такие как Big Data и Internet of Things, а также возможности, которые они предоставляют. Проведены обзор и анализ средств, позволяющих работать в данной области: облачные сервисы для хранения больших данных и способ их обработки.

Предложен вариант совместного использования технологий Big Data и Internet of Things.

Ключевые слова: беспроводные технологии, облачные сервисы, сенсорные данные, Internet of Things, big data, управление и обработка данных.

INTERACTION OF TECHNOLOGIES INTERNET OF THINGS AND BIG DATA

M.A. Gerasimenko

*Moscow State the National Research University «Higher school of economics»,
Moscow, Russia*

This research provides the strategy of using two modern directions such as Big Data and Internet of Things and their various opportunities. There are the overview and analysis of tools which helps to work in this area: cloud services for data's storage and its monitoring.

The new using option, linking technologies Big Data and Internet of Things, is represented in this work.

Keywords: wireless technologies, cloud services, internet of things, big data, control and processing data.

Введение

Совместное рассмотрение двух трендов, таких как «Internet of things» и «Big Data», становится все более очевидным в последнее время. Большой объем неструктурированных данных, требующий обработки и стремительно увеличивающийся, является серьезной проблемой, которую необходимо решить. В данной работе обосновываются актуальность и необходимость использования технологий «Internet of things» и «Big Data», а также приводится один из вариантов их использования.

Актуальность темы исследования

По данным компании Intel [1], к 2020 году число устройств, подключенных к сети Интернет, увеличится на 300%, а к 2016 году объем неструктурированных данных будет составлять 90% от всех обрабатываемых. Таким образом, возникает необходимость задуматься над тем, какие из этих данных стоит хранить и обрабатывать, чтобы извлечь из этого пользу, а какие будут только бесполезно тратить вычислительные мощности Data центров.

Сейчас активно ведутся исследования в этой области, и согласно исследованиям компании Cisco, бизнес-возможности рынка, который охватывают «Big Data» и «Internet of things», составляют \$14,3 трлн.

Существуют различные технологии и способы использования облачных баз данных для хранения и обработки информации, полученной от «Internet of things», которые являются компонентами IoT. По мнению специалистов [2], основным компонентом IoT является система RFID (Radio Frequency Identification), которая идентифицирует места назначения и передает данные между ними. Способ автоматической идентификации RFID подразумевает под собой считывание данных для последующей обработки со специальных RFID-меток.

Данные, передаваемые в облачные хранилища, можно разделить на 5 категорий:

1. Данные, полученные от сенсоров.
2. Мультимедийные данные, полученные от приложений.
3. Данные о местоположении, полученные от GPS.
4. Метаданные и дескрипторы, описывающие атрибуты определенного объекта, для того, чтобы определить его тип.
5. Управляющие сигналы.

Сбор данных является важным аспектом работы с информацией. Одним из способов получения данных является использование «Internet of things», которые являются компонентами IoT и представляют собой совокупность датчиков. «Internet of things» не только собирают информацию локально, но и передают ее другим устройствам для анализа и обработки, то есть можно сказать, что именно они создают трафик, который превращается в данные для анализа средствами Big Data.

Инструменты и технологии

Для того чтобы хранить и обрабатывать большие объемы неструктурированных данных, нельзя использовать средства и технологии, которые применяются для работы со стандартными базами данных. Известной технологией для подобных целей является Hadoop, которая использует компьютерные кластеры при работе с данными и парадигму MapReduce. То есть, используется несколько серверов со средними и

малыми мощностями, что позволяет сэкономить на расходах, не потеряв производительности.

В основе технологии лежит распределённая файловая система HDFS (Hadoop Distributed File System), которая обеспечивает хранение данных Hadoop сразу на нескольких узлах кластера [3]. Это позволяет снизить риск потери информации. Фреймворк MapReduce подразумевает обработку данных, состоящую из двух этапов: сначала происходит предварительная параллельная обработка на различных узлах кластера, затем обработанные данные сводятся в единый результат.

Стартапы Big Data

На сегодняшний день существует множество стартапов по Big Data. Например, анализируются поисковые запросы пользователей за большой промежуток времени, что помогает облегчить проведение маркетинговых исследований или успешно распространять рекламу.

В России уже есть несколько стартапов в области Big Data, и их можно разделить на три направления:

1. Стартапы, которые разрабатывают инфраструктуру и базовые технологии.
2. Стартапы, предоставляющие data services.
3. Стартапы, которые работают с потребителями, B2C-сегмент.

На сегодняшний день очень важно анализировать как можно большую часть информации в реальном времени для того, чтобы избежать бесполезного накопления «сырых» данных. В качестве примера такого рода разработок можно рассмотреть команду Cloudbus из Нижнего Новгорода, которая разработала систему интеллектуального управления городским транспортом и сбора информации в реальном времени.

Вариант использования

Исследования и анализ больших данных в настоящее время производятся с информацией, которая получена из социальных сетей, статистики поисковых запросов и посещений web страниц, а также прочих ресурсов, предоставляемых глобальной сетью. Но гораздо интереснее анализировать информацию из реального мира, которая получена от датчиков, которыми и являются «Internet of things» [4].

С помощью датчиков, установленных в автомобили, можно анализировать загруженность дорожных трасс, а также регулировать ее, путем управления светофорами на основе анализа данных о местоположении объектов. Такой сценарий прекрасно подходит для концепции IoT [5], а также является альтернативой работе обычных навигаторов, которые работают со снимками, полученными от спутников.

Установив датчики непосредственно в автомобили, можно следить за количеством аварий и тем, что послужило их причиной, поскольку для этого необходимо обрабатывать большое количество информации.

Для того, чтобы работать с сенсорными данными, хорошим решением будет использование облачных БД. Для успешного хранения и обработки данных, такие системы должны обладать хорошей масштабируемостью, чтобы работа с «Big Data» была успешной. Вот некоторые системы, которые могут использоваться для подобных целей:

- PostgreSQL – объектно-ориентированная open-source система;
- Apache Cassandra – система с индексированием столбцов, которая содержит журнал обновлений, поддерживает денормализацию и представление данных, а также мощный встроенный кэш.
- MongoDB (from «humongous») – документационная open-source система, написанная на C++. 041Видер среди баз данных типа NoSQL;
- IOTMDB – также базируется на принципе NoSQL.

Заключение

В статье рассмотрено взаимодействие двух технологий: Internet Things и Big Data, обоснована актуальность их использования на сегодняшний день.

Предлагается вариант их совместного использования, а также приведены основные средства и инструменты, которые позволяют в дальнейшем развиваться и работать в этом направлении.

Данное научное исследование (проект № 14-05-0064) выполняется при поддержке Программы «Научный фонд ИИУ ВШО» в 2014/2015 гг.

Список литературы

1. Aertebjerg J. Presentation: Business opportunities from IOT and Big Data / J. Aertebjerg // Enterprise Solution Sales Intel EMEA, December 2012. – P. 26.
2. Thi Anh Mai Phan Cloud Databases for Internet-of-Things Data// Aalto University School of Science Degree Programme in Security and Mobile Computing // Master's Thesis Espoo, June 25, 2013.– P. 70
3. Технология Hadoop // Практический опыт и экспертиза DIS-Group [Электронный ресурс] http://www.dis-group.ru/files/hadoop_ot_dis_group.pdf
4. Rolich A. A graphical user interface toolkit for the Web of things / A. Rolich // Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN-2013), Moscow: JSC «TECHNOSPHERA», 2013. – P. 294 – 303.
5. Восков Л. С. Интернет вещей / Л.С. Восков // Новые информационные технологии: тез. докл. XX Междунар. студ. конф. – школы-семинара. – М.: МИЭМ, 2012. – С. 89-94.