

ОБЛАЧНАЯ ПЛАТФОРМА IBM BLUEMIX ДЛЯ ТРЕНИНГА ПО ТЕХНОЛОГИЯМ BIG DATA



И.И. Пилецкий

Научный руководитель совместной лаборатории БГУИР-ИВА и АЦКТ ИВМ, главный специалист по обработке данных ИВА IT Park, доцент кафедры информатики БГУИР, кандидат физико-математических наук



А.Е. Лещев

Заместитель декана по учебной работе факультета компьютерных систем и сетей БГУИР, магистр технических наук



В.И. Козуб

Магистрант кафедры информатики БГУИР

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

E-mail: ianmenski@gmail.com, leschov@bsuir.by, vkozub4@gmail.com

Abstract. The article describes a task of organizing training on Big Data technologies. Two possible ways have been considered: creating a data center infrastructure from scratch and using existing PaaS cloud solutions. As a practical example of the first approach we describe our experience in creating a faculty's data center, our goals and achieved results. To demonstrate the second approach, we describe our experience in using IBM Bluemix cloud platform as a Big Data training platform for our students. Also, the article describes the most relevant Big Data trends and technologies as well as achieved results of the training.

1. ЦОД БГУИР как среда для исследований Big Data. Проблема изучения и освоения новых технологий, связанных с обработкой больших объемов данных, требует нетрадиционного подхода к предоставлению и выделению ИТ ресурсов. Поэтому два года назад было принято решение о создании дата-центра совместной лаборатории БГУИР-ИВА (ЦОД). Этот ЦОД необходим для решения новых задач, связанных с изучением работы с большими и сверхбольшими объемами неструктурированных и структурированных данных. Новая парадигма Big Data, а вместе с ней новые методы обработки данных, не могут быть обеспечены традиционными средствами, применяемыми в университетах на уровне учебных лабораторий.

Одним из решений данной проблемы может быть решение, основанное на объединении существующих в университете достаточно мощных серверов и создании центров обработки данных (ЦОД) горизонтального масштабирования ресурсов, пригодных для проведения исследований Big Data и создания инфраструктуры для моделей IaaS и PaaS.

При проведении исследований были выделены два направления обработки

и анализа Big Data: обработка данных в режиме реального времени и сбор информации о предмете из различных источников, установление связей и анализ такой информации.

Для работы с Big Data мы использовали два семейства продуктов: IBM InfoSphere (InfoSphere BigInsights и InfoSphere Streams) и open source решения: Hadoop, Apache Storm и Apache Spark.

В созданном ЦОД мы провели и продолжаем проводить исследования: получение и анализ данных из социальных сетей, получение и анализ данных из различных Web-сайтов, оптимизация архитектуры, анализ большого объема данных в высоконагруженных системах, апробация методов advanced analytics и их графическая интерпретация. Как правило, эти исследования связаны с магистерскими диссертациями и дипломными работами.

Для проведения анализа полученных данных использовались как собственные разработанные алгоритмы, так и готовые средства типа IBM BigSheets.

Мы использовали IBM InfoSphere BigInsights для анализа профессиональных предпочтений выпускников БГУИР, основываясь на информации, полученной из социальной сети LinkedIn. Также были разработаны приложения для сбора и анализа тональности текстов из социальных сетей с последующей публикацией результатов в облаке в виде веб-сервиса.

Исследовательская и практическая работа по обработке и анализу большого объема данных в высоконагруженных системах позволила реализовать новую архитектуру на базе Apache Storm и Apache Mesos для построения платформы по обработке данных в режиме реального времени для компании «Адформ Бай».

Результаты данной работы были представлены на двух международных конференциях по Big Data в г. Минске (июнь 2015г.) и г. Стамбуле (июль 2015г.) в форме докладов и доступны для более подробного изучения в сборнике докладов [1].

Другое направление связано с обучением студентов для работы с Big Data. Программы для обучения строятся на основе курсов IBM и собственных материалов. Так, в прошлом и текущем учебных годах студенты специализированной лаборатории прошли начальный курс IBM по Big Data и получили сертификаты IBM. Планируется в следующем учебном году начать обучение в магистратуре кафедры информатики по специальности «Обработка больших объемов информации».

В настоящее время облачные технологии и модели вычислений типа IaaS, PaaS, SaaS вошли в повседневный обиход как новая парадигма предоставления ИТ ресурсов. Одним из новейших решений типа PaaS является платформа IBM Bluemix [2], которая поддерживает различные среды исполнения и пользовательские инфраструктуры программирования, веб-сервисы и сервисы приложений, мобильные сервисы, сервисы управления данными, сервисы мониторинга и анализа.

2. *Среда IBM Bluemix.* Bluemix – это новейшее предложение в серии облачных решений IBM. Эта среда позволяет разработчикам и организациям быстро и легко создавать, развёртывать и администрировать приложения в облаке. Bluemix представляет собой реализацию архитектуры IBM Open Cloud Architecture на основе открытого ПО Cloud Foundry, работающего по принципу «платформа как услуга» (Platform as a Service – PaaS). Bluemix предоставляет услуги корпоративного уровня, которые можно легко интегрировать в облачные приложения, не вдаваясь в тонкости их установки и настройки.

Cloud Foundry – это платформа, предоставляемая как услуга (PaaS), с открытым исходным кодом, которая позволяет быстро создавать и развёртывать приложения в облаке. Поскольку Cloud Foundry – ПО с открытым исходным кодом, оно не зависит от поставщика и не привязывает пользователя к какому-либо программному обеспечению и к какой-либо определенной облачной инфраструктуре. Cloud Foundry абстрагирует базовую инфраструктуру, необходимую для работы облака, позволяя сосредоточиться на создании облачных приложений. Важнейшее достоинство Cloud Foundry заключается в широте выбора.

Bluemix – это ключевая технология IBM Cloud, единая среда с мгновенным выделением ресурсов для быстрой разработки и развёртывания приложений, решающих самые разнообразные задачи. Эту открытую, стандартизованную платформу можно использовать для создания, эксплуатации и администрирования приложений, предназначенных для веб-браузеров и мобильных устройств, обработки больших данных и работы с интеллектуальными устройствами интернета вещей. Bluemix поддерживает множество популярных языков программирования и платформ, предоставляет технологии разработки для различных языков программирования, средства поддержки разработки серверной части мобильных приложений, средства мониторинга приложений, технологии с открытым исходным кодом и многое другое.

Стоит отметить, что применение данных технологий существенно снижает трудоемкость разработки приложений. Это достигается благодаря тому, что IBM предоставляет доступ к возможностям ряда своих продуктов, например, для работы с Big Data и аналитики средствами Watson, посредством сервисов IBM Bluemix. Теперь у разработчика есть возможность сосредоточить усилия на решении самой задачи, вместо того чтобы тратить время на установку и настройку громоздкого программного продукта. Благодаря вышеописанным преимуществам IBM Bluemix является отличной средой для быстрого прототипирования приложений.

Использование IBM Bluemix помогает начать обучение студентов и магистрантов университета по специальности «Обработка больших объемов информации», используя новейшие технологии для разработки когнитивных приложений и приложений для работы с Big Data.

3. *Цели и методика тренинга.* Целью тренинга является обучение студентов

современным и востребованным технологиям в следующих областях: Big Data, Cognitive Computing, Internet of Things.

Тренинг проводился в формате лекций и практических занятий. В качестве материалов лекций использовались материалы, предоставляемые IBM в рамках программы академической инициативы, а также собственные наработки.

В рамках курса были затронуты следующие темы: Big Data fundamentals; Hadoop, MapReduce and Cloud Computing; IBM Infosphere BigInsights; IBM Bluemix; Internet of Things; IBM Watson Analytics.

Для студентов данного курса IBM предоставила—**бесплатный доступ к возможностям IBM Bluemix.**

В рамках тренинга мы решили дать общее представление о Big Data, а затем сосредоточиться на работе с Bluemix и решении практических задач. Стоит отметить, что у нас был опыт успешной работы с Bluemix, полученный в том числе в ходе хакатона IBM Moscow Hackathon, прошедшего в декабре 2015 года в Москве. Это хакатон по созданию когнитивных приложений на основе сервисов Watson, размещённых на облачной платформе IBM Bluemix. В ходе хакатона нами было разработано приложение Speciality Helper. По оценке жюри, приложение вошло в тройку лучших проектов, представленных на хакатоне.

Таким образом, Bluemix является мощным инструментом, который при правильном подходе позволит упростить и удешевить процесс разработки облачных приложений в разы. Однако понимание принципов работы с Bluemix требует наработки практического опыта. Такой опыт студенты получили благодаря работе над несколькими практическими задачами.

4. Задачи для тренинга. Для получения практического опыта работы с IBM Bluemix студентам было предложено решить следующие 4 задачи.

4.1. Разработка когнитивного приложения на основе сервисов IBM Watson (суперкомпьютер IBM, часть сервисов которого доступна в рамках Bluemix). В основном сервисы Watson позволяют проводить аналитику больших объёмов данных. Использование этих возможностей позволило нам создать приложение, помогающее абитуриентам выбрать подходящую им специальность в университете, основываясь на составлении и анализе их психологической предрасположенности к тем или иным профессиям. В решении данной задачи использовались следующие сервисы Bluemix: Personality Insights, Dialog, Speech to Text, Question and Answers, Cloudant NoSQL DB, Insights for Twitter и другие.

На рисунке 1 показан интерфейс Bluemix для разработчика приложения. В верхней части расположена панель управления количеством выделяемой приложению памяти и количеством запущенных копий приложения. В нижней части показаны три сервиса Bluemix, используемые в приложении.

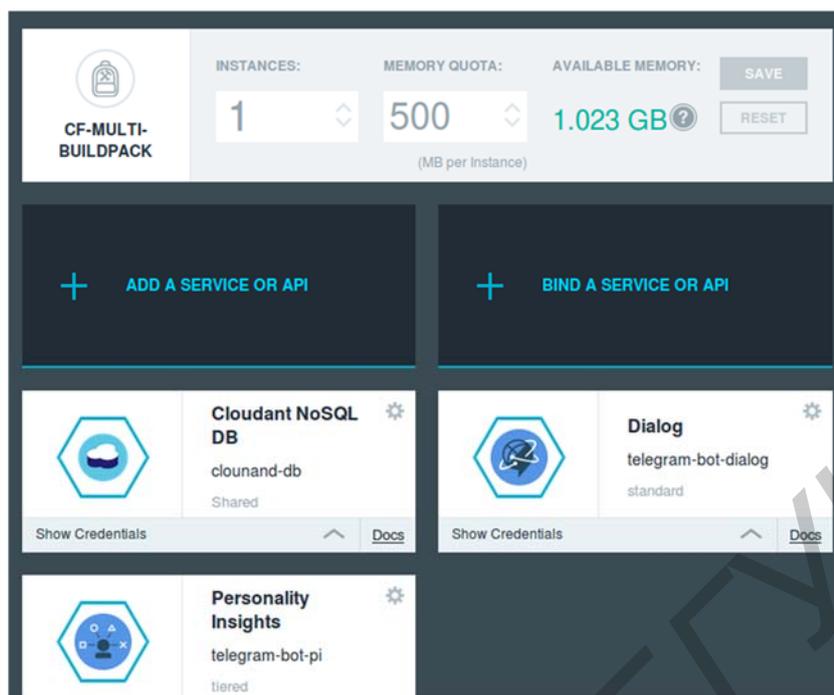


Рис. 1. Интерфейс Bluemix для разработчика приложения

На рисунке 2 показан пользовательский интерфейс приложения. Для взаимодействия с пользователем был разработан чат-бот для мессенджера Telegram, задача которого – получить нужную для анализа информацию о пользователе.

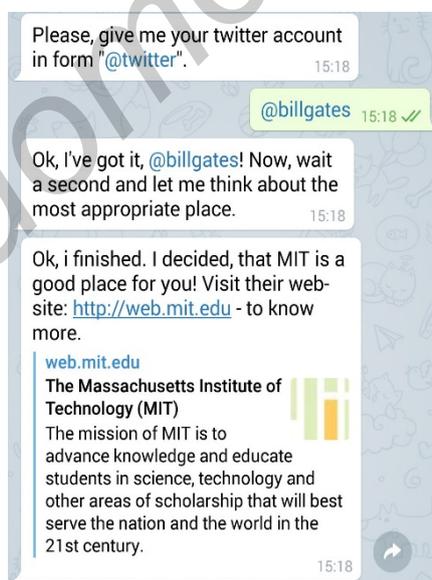


Рис. 2. Пользовательский интерфейс приложения

Стоит отметить, что сфера применения данного приложения может быть значительно расширена и коммерциализирована; например, его можно использовать при поиске кандидатов на определённую должность в компании или при поиске заказчиков.

4.2. Разработать приложение, позволяющее проводить анализ предпочтений людей к определённому бренду, товару, услуге или человеку. Таким образом, появляется возможность сравнения двух продуктов по различным критериям. Например, можно выбрать наиболее популярный продукт, проследить динамику его популярности, прогнозировать спад или подъём покупательской активности, а также выяснить отношение людей к данному продукту. В приложении использовались следующие сервисы IBM Bluemix: AlchemyData News, Tone Analyzer, Concept Insights, Relationship Extraction, Natural Language Classifier. Приложение может быть доработано для предоставления аналитических отчётов, расширения возможностей аналитики и т.д.

4.3. Второе большое направление помимо аналитики, где идёт работа с Big Data, – это интернет вещей (Internet of Things, IoT). Мы предложили студентам несколько задач из области IoT. Так, одна команда студентов занялась разработкой интерактивной карты учебного корпуса БГУИР. Было разработано мобильное приложение, которое содержит карту, отображает текущее положение пользователя на карте и позволяет прокладывать маршруты. На рисунке 3 приложение запущено в режиме измерения расстояний. Видна карта этажа, устройства для позиционирования (круги с буквой В) и маршрут с чекпоинтами. На рисунке 4 приложение запущено в режиме навигации. Движущаяся голубая точка на карте – пользователь приложения.

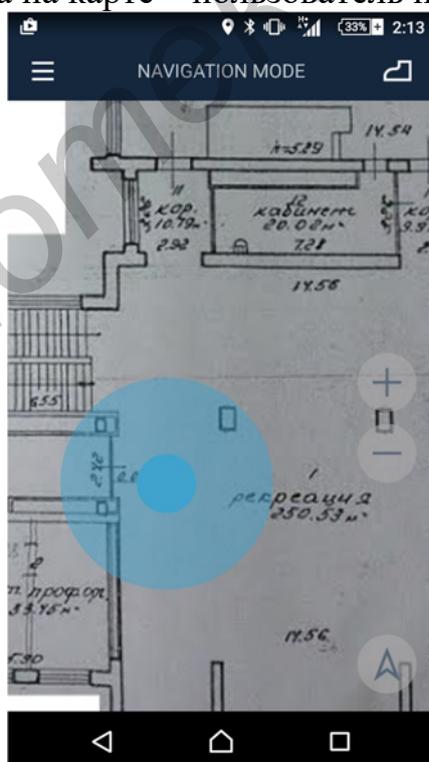


Рис. 3. Режим измерения расстояний

Рис. 4. Режим навигации

Функциональность приложения может быть расширена за счёт средств Bluemix. Например, можно хранить историю передвижений пользователей на

сервере, чтобы затем проводить анализ самых популярных маршрутов и точек скопления людей. Для этих целей подходят следующие сервисы Bluemix: IoT Real-Time Insights, Cloudant NoSQL DB.

4.4. Вторая команда по IoT занялась сбором информации с фитнес-трекеров, набирающих популярность в мире. Стоит отметить, что Bluemix предоставляет удобную платформу разработки приложений для IoT – Internet of Things Platform – что избавляет от необходимости создавать подобную платформу с нуля и позволяет сосредоточиться на решении поставленной задачи. Также в Bluemix доступна среда визуального программирования Node-RED (рис. 5), которая содержит множество шаблонных функций, триггеров и т.п., что значительно ускоряет разработку приложений в области IoT.

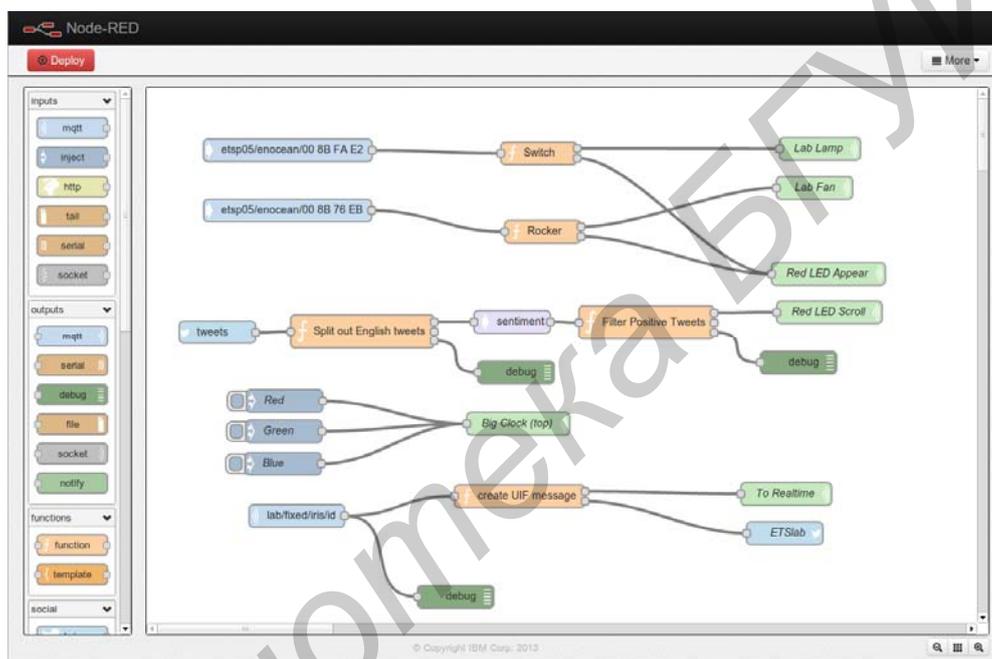


Рис. 5. Среда визуального программирования Node-RED в Bluemix

5. Результаты. В ходе тренинга студенты получили теоретические знания в области Big Data и практические навыки разработки приложений с использованием сервисов IBM Bluemix.

Было разработано несколько приложений, часть из которых планируется применить на факультете.

Кроме того, полученные умения позволят студентам глубже окунуться в Big Data для решения более сложных задач и будут плюсом при трудоустройстве в данной области.

В дальнейшем на факультете планируется проводить углубленные курсы по подготовке специалистов в области Big Data, в том числе с приглашением преподавателей из московского офиса IBM.

Также осенью 2016 года на базе факультета планируется проведение хакатона по Big Data и Bluemix.

Литература

- [1]. И.И. Пилецкий и др. Виртуальная ИТ среда БГУИР для исследования Big Data и VCL, с. 21-32, BIG DATA and Predictive Analytics. Использование BIG DATA для оптимизации бизнеса и информационных технологий : сборник материалов междунар. науч.-практ. конф. / редкол. : М.П. Батура [и др.]. – Минск : БГУИР, 2015. – 220 с.
- [2]. What is Bluemix [Электронный ресурс] / IBM developerWorks. – 2015-2016. – Режим доступа: <https://www.ibm.com/developerworks/cloud/library/cl-bluemixfoundry/>. – Дата доступа: 25.05.2016.

Библиотека БГУИР