

BIG DATA. ТРАНСФОРМАЦИЯ МАГИСТЕРСКИХ ПРОГРАММ



И.И. Пилецкий

Научный руководитель совместной лаборатории БГУИР-ИВА и АЦКТ IBM, архитектор отделения по программному обеспечению 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

Ассистент кафедры информатики БГУИР

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь
E-mail: ianmenski@gmail.com, leschov@bsuir.by, kozub@bsuir.by

Abstract. This paper describes difficulties which we've experienced with training of IT professionals in Big Data field. IT skills with top demand on market are also given. Master degree study programs for speciality 'Big Data Processing' are considered as well as difficulties with training of master degree students in this field. These difficulties are caused by flexible and rapidly changing requirements for this speciality which can't be fulfilled by university itself. Enterprise technologies and solutions can be used in education as a tool for developing necessary skills. Using of these technologies in master degree study programs is also considered.

1. Принципы трансформации. В последние годы мир IT (и не только он) быстро изменяется. Данными является всё то, что мы видим, слышим, что получаем от различных приборов и датчиков. Это огромное количество данных, которые плохо структурированы, например, текстовые данные из Интернет-источников, фотографии, видеозаписи, электронные журналы, геопространственные данные и др.

Так, 90 % данных в мире было создано в последние два года [1-6]:

- каждый день по всему миру продается 2 млн. мобильных устройств;
- по данным GSMA, в 2015 году зафиксировано 7,23 млрд. мобильных соединений;
- в 2016 году более 70 % клиентов зависят от социальных сайтов (от советов до покупки);
- к 2020 году общая рабочая загрузка облачных ЦОД будет приближаться к 90 %, роль традиционных ЦОД будет сокращаться до 8 % (в 2016 году данное соотношение составляло 81 % и 19 % соответственно);
- к 2020 году будет 50 миллиардов IoT-устройств, подключённых к Интернету;
- продолжается рост API экономики, глобальные продажи (E-commerce, M-commerce) неуклонно растут и к 2020 году превысят 600 миллиардов долларов.

По данным корпорации IBM, наблюдается значительный рост использования общественных сетей. Наступает новая постиндустриальная эра, в которой IT займут положение ключевой отрасли в экономике, подобно тому как ткацкое производство занимало эту роль в XVII-XVIII веках, а машиностроение – в XIX-XX веках.

Наиболее востребованными навыками (специальностями) в IT во всём мире в 2016 году по данным LinkedIn являлись следующие (см. таблицу 1). Для получения данных было проанализировано более 330 миллионов LinkedIn профайлов [7-8].

Таблица 1. Наиболее востребованные навыки в IT в 2016 году

#1 Cloud and Distributed Computing	Облачные вычисления и распределённые вычисления
#2 Statistical Analysis and Data Mining	Статистический анализ и интеллектуальный анализ данных
#3 Web Architecture and Development Framework	Архитектура и разработка веб-сайтов
#4 Middleware and Integration Software	Промежуточное и интеграционное ПО
#5 User Interface Design	Дизайн пользовательского интерфейса
#6 Network and Information Security	Сетевая и информационная безопасность
#7 Mobile Development	Разработка ПО для мобильных устройств
#8 Data Presentation	Представление данных (визуализация)
#9 SEO/SEM Marketing	SEO/SEM маркетинг
#10 Storage Systems and Management	Системы хранения данных и управление ими

В табл. 2 приведена тенденция к требуемым специальностям на протяжении 2015 и 2016 годов.

Таблица 2. Тенденция к требуемым специальностям на протяжении 2015 и 2016 годов

The hottest Skills of 2015 on LinkedIn	Top Skills of 2016 on LinkedIn
Cloud and Distributed Computing	Cloud and Distributed Computing
Statistical Analysis and Data Mining	Statistical Analysis and Data Mining
Marketing Campaign Management	Web Architecture and Development Framework
SEO/SEM Marketing	Middleware and Integration Software
Middleware and Integration Software	User Interface Design
Mobile Development	Network and Information Security
Network and Information Security	Mobile Development
Storage Systems and Management	Data Presentation

В IT-сфере появилась новая проблема: адаптация и трансформация образовательных программ к современным требованиям. Основные принципы трансформации программ:

- программы должны отражать тенденции в IT, содержать современный контент;
- помогать студентам в карьерном плане;
- предоставлять новые подходы к образованию, связанные с новыми технологическими возможностями по обучению.

2. *База для трансформации.* Первое направление (на кафедре информатики БГУИР факультета КСиС) связано с обучением студентов в совместной лаборатории БГУИР-ИВА для работы с Big Data. Обучение проводится на облачной платформе ЦОД БГУИР [9-10] и платформе IBM Bluemix [11].

Платформа ЦОД БГУИР обеспечивает сервисные модели IaaS и PaaS. Основу данной платформы составляют два семейства продуктов: IBM InfoSphere (InfoSphere BigInsights и InfoSphere Streams) и Open Source решения: Apache Storm, Apache Spark и др.

Новейшие облачные технологии IBM и сервисные модели типа PaaS и SaaS платформы IBM Bluemix [11] использовались как базис для выполнения практических работ студентами.

Программы для обучения в совместной лаборатории строятся на основе регулярно уточняющихся курсов IBM и материалов, подготовленных преподавателями. Так, на протяжении последних трёх учебных лет студенты факультета КСиС проходят в лаборатории начальный курс IBM по Big Data и получают сертификаты IBM [9].

Второе, основное направление, связано с обучением студентов в магистратуре кафедры информатики по специальности «Обработка больших объёмов информации».

В 2015 году сотрудниками БГУИР был разработан образовательный стандарт для второй

ступени образования (магистратуры) ОСВО 1-40 81 04-2015 «Обработка больших объемов информации», который был утвержден Министерством образования Республики Беларусь. Стандарт обязателен для применения во всех учреждениях высшего образования Республики Беларусь, реализующих образовательные программы магистратуры.

В стандарте предусмотрены *дисциплины государственных компонент*: «Технологическая платформа по управлению большими данными», «Архитектурные решения для обработки больших объемов информации», «Модели и методы обработки и анализа больших объемов информации».

В соответствии с данным стандартом в 2016 году разработаны три учебные программы БГУИР: «Технологическая платформа по управлению большими данными», «Архитектурные решения для обработки больших объемов информации», «Модели и методы обработки и анализа больших объемов информации» – по которым в 2016 году начата подготовка магистров по специальности 1-40 81 04 Обработка больших объемов информации».

Общая тематика этих программ позволяет охватить большинство аспектов «Обработки больших объемов данных». При проведении занятий были рассмотрены такие темы, как: что такое «большие данные», источники больших данных, методы доступа к ним, проблема обработки больших объемов данных, вычислительные модели многопоточной параллельной обработки, файловые системы (HDFS, GPFS, RDD), NoSQL базы данных (типа Riak, Amazon Dynamo, Redis, HBase, CouchDB, MongoDB, Infinite Graph, Neo4J), методы доступа и хранения данных (плоские файлы, JSON, CSV, XML и NoSQL базы данных), платформы, применяемые для работы с большими данными (Open Source, Enterprise Solutions, Cloud Based), обработка данных «в покое», обработка потока данных, архитектурные решения систем вертикального и горизонтального масштабирования, платформы с открытым кодом и реализующие промышленные решения, сравнительный анализ архитектурных решений платформ для обработки больших объемов данных, системы для обработки больших объемов данных, языки параллельного программирования для анализа структурированных и неструктурированных данных больших объемов (Pig, Jaql, Hive, SQL, R), методы, средства и аналитические алгоритмы анализа структурированных и неструктурированных данных, описательная, прогнозная и директивная аналитика, методы нахождения паттернов и аномалий, модели и методы визуального представления данных, примеры графической интерпретации больших объемов информации.

Третье, *научно-исследовательское направление*, связано с исследованиями, которые проводятся при написании диссертаций в магистратуре и аспирантуре.

Здесь выполняются различные работы, связанные с исследованиями методов и технологий обработки естественного языка из Интернет источников, разработкой и реализацией различных IoT-решений с использованием облачных технологий и когнитивных технологий, исследованием методов и средств анализа мультимедийной информации (в том числе и фото), других методов и алгоритмов машинного обучения, методов и средств аналитических решений (в том числе и в сфере транспорта).

В качестве учебной базы для проведения занятий применялись:

- созданный вычислительный кластер в локальном центре обработки данных БГУИР (ЦОД БГУИР) с ПО для обработки больших объемов данных [9];
- платформы Apache Hadoop, Cloudera Hadoop Distribution Platform, Apache Spark, Apache Storm;
- базы данных Riak KV, Amazon Dynamo, Redis, CouchDB, MongoDB, Infinite Graph, Neo4J;
- платформы IBM BigInsights и IBM InfoSphere Streams;
- облачная платформа IBM Bluemix;
- когнитивные сервисы IBM Watson.

В настоящее время, наряду с уже существующей базой, для организации и проведения занятий требуются *индустриальные технологии*, так как решить проблему адаптации и трансформации

ции образовательных программ в сфере ИТ к современным требованиям индустрии силами университетов невозможно.

Индустриальные технологии в образовании для формирования необходимых навыков должны обеспечивать:

- доступ к технологиям и решениям ведущих мировых лидеров в ИТ, предоставляющих как Open Source, так и Enterprise решения;
- свободный доступ для университетов;
- использование облачных ресурсов для выполнения практических и лабораторных работ.

В качестве ресурсов могут быть использованы различные (как платные, так и бесплатные) Интернет ресурсы. Например, на сайте IMS GLOBAL Learning Consortium [12] предлагаются разнообразные сертифицированные курсы (IMS Certified Product Directory), LinkedIn на своем сайте [13] предлагает множество курсов для ИТ-специалистов.

На LinkedIn для каждой специальности предлагаются курсы и открытые рабочие вакансии, так для Cloud and Distributed Computing доступны курсы Cloud Computing, Big Data, Hadoop, Amazon Web Services, а для Statistical Analysis and Data Mining – курсы R, SPSS, Data Analysis. Курсы, как правило, платные.

Мировые лидеры индустрии ИТ также предлагают различные программы для обучения.

Уже длительное время БГУИР плодотворно сотрудничает с компаниями IBM и IBA Group.

В 2008 году была создана научная учебно-производственная лаборатория «Информационных технологий» кафедры Информатики БГУИР и СП ЗАО «Международный деловой альянс» (IBA Group). Благодаря успешной реализации учебно-исследовательских проектов в 2011 году в БГУИР корпорация IBM открыла первый в Республике Академический центр компетенции технологий IBM на базе совместной лаборатории БГУИР и группы компаний IBA.

Кафедра Информатики БГУИР и факультет КСиС сотрудничают с корпорацией IBM по программе IBM Academic Initiative, что позволяет получать бесплатную поддержку по учебным материалам и программным средствам [14].

Корпорация IBM предлагает и обеспечивает актуальные индустриальные технологии в образовании для формирования необходимых навыков у обучаемых, что позволяет обеспечить адаптацию и трансформацию учебных материалов (лекционных, практических) для магистерских программ к современным условиям.

Сотрудничество БГУИР с корпорацией IBM в образовательной области позволяет преподавателям и студентам пользоваться бесплатно такими ресурсами, как: IBM Academic Initiative, IBM Bluemix, IBM Watson, IBM Big Data University, Облачный университет IBM Watson RCIS [15].

По программе Academic Initiative компания IBM предлагает разнообразие тематических учебных материалов как в виде статей и книг, так и в виде on-line занятий и видеоматериалов на You Tube, например:

- Case management educator guide;
- Cloud and Big Data educator guide;
- Cognitive computing educator guide;
- DevOps Services educator guide;
- Enterprise computing educator guide;
- IBM Bluemix™ educator guide;
- Internet of Things educator guide;
- MobileFirst educator guide;
- NoSQL DBaaS with IBM Bluemix educator guide;
- Security educator guide;
- Watson Analytics educator guide;
- Watson Services on Bluemix educator guide.

Разнообразие материалов призвано помочь различным категориям пользователей найти

подходящий способ получения знаний.

Программа IBM Academic Initiative предлагает не только курсы для преподавателей и студентов, но также открытые рабочие вакансии и требования к ним.

Программа IBM Big Data University предлагает разнообразные многоуровневые курсы, после успешного изучения которых присваивается электронный бэйдж трёх уровней в зависимости от сложности курса.

Облачный университет IBM Watson RCIS был создан сотрудниками компании IBM Москва с целью поддержки студентов в выполнении ими различных заданий в процессе обучения.

Облачный университет IBM Watson RCIS – это виртуальный ресурс, использующий инструменты компании IBM, а также продукты с открытым кодом, для поддержки студентов вузов в выполнении ими учебных заданий, выпускных работ и исследовательских проектов, содержание которых связано с созданием веб-приложений и сервисов. Участниками являются преподаватели университетов и студенты, добровольно присоединяющиеся к сообществу, созданному в социальной сети

Университет создается усилиями вузов-организаторов, в число которых входят МГТУ им. Баумана, ВШЭ, БГУИР, Университет «Дубна».

Основными целями проекта являются:

- создание и обновление университетских курсов по информационным технологиям;
- разработка и совершенствование шаблонов курсов и семинарских занятий, создаваемых на основе социальных инструментов;
- совершенствование и аккумулирование знаний, методик и изобретательских подходов для разработки веб-приложений и облачных сервисов; в том числе методов визуального проектирования;
- быстрое прототипирование коммерческих решений для заказчиков;
- создание смотровой площадки для привлечения потенциальных клиентов и заказчиков проектов, размещение решений на IBM Marketplace.

3. Навыки, специальность, карьера

Сотрудничество с корпорацией IBM позволяет БГУИР:

- внедрить индустриальные технологии в образовании для формирования необходимых навыков у студентов;
- обеспечить адаптацию магистерских программ к изменяющимся требованиям IT;
- предоставить ресурсы для обучения в режиме on-line;
- обеспечить выполнение практических и лабораторных работ с помощью облачных ресурсов

В Республике Беларусь, по данным сайта dev.by, зарегистрировано свыше тысячи IT компаний, в которых открыто свыше трёхсот вакансий. Но эти вакансии мало связаны с современными тенденциями в области Big Data.

По данным сайта Парка высоких технологий, в ПВТ зарегистрировано 165 компаний, которые также нуждаются в IT специалистах.

Наиболее перспективными направлениями в настоящее время и недалеком будущем являются: когнитивные системы, приложения и сервисы, аналитика (Watson) + IoT (Internet of Things) – Интернет вещей как сеть взаимодействия M2M (Machine-to-Machine) позволяет создать безлюдное производство, применение беспилотных машин и роботов в армии и на производстве, «Разумные» дома и города, виртуальные глобальные организации, магазины без товаров, логистические предприятия без подвижного состава и т.д. (А. Сорокин, IBM Восточная Европа \ Азия).

Компания IBM предлагает путь карьерного роста через совершенствование навыков и специализации знаний. Требования для занятия вакансии не простые и довольно высокие, но и предложения тоже высокие.

На сайте LinkedIn можно найти около 1300 предложений о найме на работу для наиболее востребованной специальности *Cloud and Distributed Computing*.

Литература

- [1]. Digital-статистика по миру за август 2015 года [Электронный ресурс] / WeAreSocial.sg. – 2015-2017. – Режим доступа: <http://www.likeni.ru/events/digital-statistika-po-miru-za-avgust-2015-goda/>. – Дата доступа: 20.03.2017.
- [2]. Мобильных телефонов – больше, чем людей на планете [Электронный ресурс] / Apps4All. – 2016-2017. – Режим доступа: <http://apps4all.ru/post/10-09-14-mobilnyh-telefonov-bolshe-chem-lyudej-na-planete>. – Дата доступа: 20.03.2017.
- [3]. By 2016 Why 70% of Small Businesses Will Depend On Social Media Tool To Reach New Customers [Электронный ресурс] / Smejoinup. – 2015-2017. – Режим доступа: <http://smejoinup.com/blog/by-2016-why-70-of-small-businesses-will-depend-on-social-media-tool-to-reach-new-customers/>. – Дата доступа: 20.03.2017.
- [4]. Cisco Global Cloud Index: Forecast and Methodology [Электронный ресурс] / Cisco Systems. – 2015-2017. – Режим доступа: <http://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf>. – Дата доступа: 20.03.2017.
- [5]. Popular Internet of Things Forecast of 50 Billion Devices by 2020 Is Outdated [Электронный ресурс] / IEEE Spectrum. – 2016-2017. – Режим доступа: <http://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated>. – Дата доступа: 20.03.2017.
- [6]. The Rise of M-Commerce: Mobile Shopping Stats & Trends [Электронный ресурс] / Business Insider. – 2016-2017. – Режим доступа: <http://www.businessinsider.com/mobile-commerce-shopping-trends-stats-2016-10>. – Дата доступа: 20.03.2017.
- [7]. LinkedIn Unveils The Top Skills That Can Get You Hired In 2017, Offers Free Courses for a Week [Электронный ресурс] / LinkedIn Official Blog. – 2016-2017. – Режим доступа: <https://blog.linkedin.com/2016/10/20/top-skills-2016-week-of-learning-linkedin>. – Дата доступа: 20.03.2017.
- [8]. TOP skills of 2016. Global list [Электронный ресурс] / LinkedIn Week of Learning. – 2016-2017. – Режим доступа: <https://learning.linkedin.com/week-of-learning/top-skills>. – Дата доступа: 20.03.2017.
- [9]. Пилецкий, И. И. Облачная платформа IBM Bluemix для тренинга по технологиям Big Data / И. И. Пилецкий, А. Е. Лещев, В. Н. Козуб // BIG DATA and Advanced Analytics. Использование BIG DATA для оптимизации бизнеса и информационных технологий : сборник материалов II международной научно-практической конференции; Минск, 15-17 июня 2016 г. / редкол. : М. П. Батура [и др.]. – Минск : БГУИР, 2016. – С. 146-153.
- [10]. 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.
- [11]. What is Bluemix [Электронный ресурс] / IBM developerWorks. – 2015-2017. – Режим доступа: <https://www.ibm.com/developerworks/cloud/library/cl-bluemixfoundry/>. – Дата доступа: 20.03.2017.
- [12]. Better Learning From Better Learning Technology [Электронный ресурс] / IMS GLOBAL Learning Consortium. – 2016-2017. – Режим доступа: <https://www.imsglobal.org/>. – Дата доступа: 20.03.2017.
- [13]. Education and Instructional Design [Электронный ресурс] / LinkedIn Learning. – 2016-2017. – Режим доступа: <https://www.linkedin.com/learning/topics/education-and-instructional-design>. – Дата доступа: 20.03.2017.
- [14]. Educator guides [Электронный ресурс] / IBM Academic Initiative. – 2016-2017. – Режим доступа: <https://developer.ibm.com/academic/educator-guides/>. – Дата доступа: 20.03.2017.
- [15]. IBM RCIS Watson Cloud Cognitive University [Электронный ресурс] / IBM Developer Works. – 2016-2017. – Режим доступа: <https://www.ibm.com/developerworks/community/groups/service/html/communitystart?communityUuid=bc004137-b64a-4378-ac02-2caf59c56c2a>