The Government of the Russian Federation

The Federal State Autonomous Institution of Higher Education
"National Research University - Higher School of Economics

Course title: Big Data Collection, Storage&Processing in Heterogeneous Distributed Computer Networks

Area of studies 38.04.05 "Business informatics" Master level

Author:
Komarov Mikhail
mkomarov@hse.ru

Moscow – 2015

1. Course summary

Big data collection, storage&processing in heterogeneous distributed computer networks course is focused on the overview of different methodologies of data collection, storage&processing. This course has been developed to create graduates who can become data scientists capable of understanding and working with massive amounts of data nowadays common to many businesses. It is aimed at students who want to move into this rapidly expanding and exciting area. This course has an aim to deeply involve students into the actual problems connected to the big data collection, storage&processing.

The course covers areas related to massive datasets stored in the cloud or generated from embedded systems and from the Internet of Things (IoT), how data is stored and utilized within distributed systems of enterprise and how organizations can utilize data to change and improve business processes.

2. Area of application and Regulatory References

The Course Program is designed for lecturers, teaching assistants for the Master Program "Big Data Systems", Course "Mobile applications".

The Course Programs has been developed in accordance with:

- National curriculum standard
- Educational Program 38.04.05
- University Academic Plan of the Education Program 38.04.05.

3. Course goals

The general goal of the course is to prepare graduates for effective performance of the managerial role of collection, storage and processing of the big data, work in team and to be able to further commercialize collected data.

Therefore, aims of the course are:

- To deeply involve students into the actual problems connected to the big data collection, storage&processing.
- To present an overview of different sources of big data and how that data is processed and stored.

- To equip students with the knowledge of the major business-models which might be used for big data after processing.
- To equip students with the knowledge of basic principles of developing and managing systems to collect, store and process big data.

4. Students' Competencies to be developed by the Course

The course develops the following competencies:

СК-М1 – Student should estimate and analyze different known scientific methods and approaches in terms of data collection, storage and processing.

СК-М5 – Student should be capable to make managerial decisions, to assess their consequences and to bear responsibility for the outcomes.

These indicated and contributed during the preparation of explanation and analysis of the particular area for data collection, storage and processing, particular market and business-model.

ИК-М1.1ПД_1.1НИД_1.1ПпД_ОУД_4.6_5.2_7.1БИ– Students should identify and make prognoses about modern approaches on increasing business efficiency.

ИК-М1.2_1.3АД_4.5_5.5_7.1_7.2БИ – Students should identify and chose optimal solutions for improving it-infrastructure and business architecture of the company after implementation relevant big data collection, storage and processing processes.

Public presentation of the proposed and developed methods, approaches and architectures during the course time and analytical essay (referat) will indicate the level of efficiency of the students work.

5. How the Course fits in with the Curriculum

The Course is provided as non-obligatory course within the Master Program "Big Data Systems"

The Course is to be based on the acquisition of the following Courses:

- Data Analysis

- Enterprise Architecture Modeling

The Course requires the following students' competencies and knowledge:

- Basic computer science principles and skills
- Basic computer networks principles and skills

6. Course Schedule

№	Topic	Total amount of hours	Classroom Activities		Self-study
			Lectures	Seminars	
1.	Introduction to IoT. Big Data in IoT. Challenges and open issues in IoT.	24	2	4	18
2.	Sensor networks and m2m: standardization and relevant usage models. Sensor Networks and m2m: business use cases and ROI.	14	2	4	8
2.1	Nanocommunications and Internet of Nano Things.	14	2	4	8
3.	Smart Grids.	14	2	4	8
4.	Short-range wireless technologies: data collection, processing and storage	14	2	4	8
5.	Cellular technologies: data collection, processing and storage	14	2	4	8

6.	5G telecommunication networks and systems	12	2	2	8
7.	Use cases, service implementations and business opportunities for operators.	36	2	6	28
8.	Web 2.0, Web 3.0 concept (different definitions and approaches). Emerging Internet of Services, web of services.	12	2	2	8
9.	Data storages. Open Data concept.	12	2	2	8
10.	Mobile applications as a source of data.	12	2	2	8
11.	Enterprise mobility and mobile applications for business. Financial data for business.	12	2	2	8
12.	Essay				
13.	Home task				
14.	Final exam				
	Total:	190	24	40	126

7. Forms and Types of Testing

The course is based on a rather complicated method of grading.

Final score consists of:

• score for home task – O_1;

• score for essay (referat) – O_2;

• score for current work during the classes – O_3;

• final exam presentation – O_4.

Final score O looks as follow:

$O=0,2*O_1+0,2*O_2+0,2*O_3+0,4*O_4$.

Number of credits: 5.

8. Literature

8.1 Core textbooks:

1. Larry L. Peterson, Bruce S. Davie “Computer Networks: A Systems Approach”, Emereo Pty Limited, 2011
2. Terence Craig; Mary E Ludloff “Privacy and Big data”, Sebastopol, CA: O'Reilly, 2011.
3. Kevin Roebuck “Storing and Managing Big Data - Nosql, Hadoop and More”, Emereo Pty Limited, 2011.
4. Ovidiu Vermesan, Peter Friess “Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems”, River Publishers, 2013
5. Gul N. Khan, Krzysztof Iniewski “Embedded and Networking Systems”, CRC Press, 2014
6. Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman “Big Data For Dummies”, John Wiley & Sons, 2013

9. Course content

Topic 1. Introduction to IoT. Big Data in IoT. Challenges and open issues in IoT.

At this topic students would understand what is the Internet of Things, connections between IoT and Big Data. Why IoT is one of the major sources of Big Data. There will be information about the challenges and open issues in IoT especially when we are talking about data storages and when we are talking about real-time processing of the collected data. Examples from the real business and discussions with the students about future of the IoT and Big Data will also be a part of the topic.

Topic 2. Sensor networks and machine-to-machine (m2m). Standardization, relevant usage models, business use cases and ROI.

Within this topic the students will learn the principles of machine-to-machine interaction, correspondent technical challenges, network architectures standardized by ETSI and ITU. The special attention will be given to m2m and conventional operators and service providers, their new demands in m2m business, ways to generate revenues out of m2m. Case studies from different industries will be also provided and analysed.

Textbooks:

1. Vojislav B. Misic, Jelena Misic Machine-to-Machine Communications: Architectures, Technology, Standards, and Applications", CRC Press, 2014.
2. K.C. Chen Machine-to-Machine Communications for Healthcare, Journal of Computing Science and Engineering, Vol. 6, No. 2, June 2012, pp. 119-126
3. Juniper Networks: M2M the rise of machines, http://www.slideshare.net/jpocalles/machine-to-machine-white-paper

Topic 2.1. Nanocommunications and Internet of Nano Things.

Within this topic the students will learn the agenda of nanocommunications and will get acquainted with the paradigm of the Internet of Nano Things. The special attention will be given to the driving forces of the nanocommunications as well as short and long term research agenda. The possible societal impact will also be discussed in detail.

Textbooks:

1. http://en.wikipedia.org/wiki/Nanonetwork, 2014.
2. Akyildiz, I. F., Jornet, J. M., and Pierobon, M. "Nanonetworks: A New Frontier in Communications," Communications of the ACM, vol. 54, no. 11, pp. 84-89, November 2011. http://www.ece.gatech.edu/research/labs/bwn/papers/2011/j12.pdf
3. Akyildiz, I. F. and Jornet, J. M., "The Internet of Nano-Things," IEEE Wireless Communication Magazine, vol. 17, no. 6, pp. 58-63, December 2010 http://www.ece.gatech.edu/research/labs/bwn/surveys/nanothings.pdf

Topic 3. Smart Grids.

Within this topic the students will understand the challenges for Smart Grid and general impact of the technololgy when implemented. The special attention will be given to the challenges of the Internet of Energy, Smart Grid communication standards, interoperability concepts as well as up-to-date status of EU implementation of Smart Grids.

Textbooks:

1. SMART GRID PRIMER (SMART GRID BOOKS), http://energy.gov/oe/technology-development/smart-grid/smart-grid-primer-smart-grid-books
2. James Momoh "Smart Grid: Fundamentals of Design and Analysis", ISBN: 978-0-470-88939-8 232 pages April 2012, Wiley-IEEE Press
3. Stuart Borlase "Smart Grids: Infrastructure, Technology, and Solutions ", CRC Prress, 2012.

Topic 4. Short-range wireless technologies: data collections, processing and storage

Within this topic the students will learn the place of the short-range (capillary) wireless technologies in IoT and role in Big Data collection. The particular attention will be given to the standard technologies such as 6LoWPAN, IEEE 802.11 and .15 and their key features persistent to the Big Data collection and transfer.

Textbooks:

1. Next Gen 802.11ac WiFi for Dummies, Intel, 2014, http://www.intel.com/content/dam/www/public/us/en/documents/pdf/next-gen-80211ac-wifi-for-dummies.pdf

2. Zig Bee Resource Guide, http://www.nxtbook.com/nxtbooks/webcom/zigbee_rg2014/#/0

3. 6LoWPAN presentations, http://www.slideshare.net/mabraham80/6lowpan-final

Topic 5. Cellular technologies: data collections, processing and storage

Within this topic the students will learn the general principles of the cellular technologies and their evolution towards 5G finally allowing implementation of IoT concept and handling of Big Data. A special attention will be given to understanding of network architectures and network capacity problem.

Textbooks:

1. Chilamkurti, Naveen, Zeadally, Sherali, Chaouchi, Hakima (Eds.) "Next-Generation Wireless Technologies", Springer, 2013.

2. 4G Portal, Top 10 books on LTE, 2013. http://4g-portal.com/top10-books-published-or-to-be-published-in-2013

Topic 6. 5G telecommunication networks and systems

Within this topic the students will learn the cellular network concept that is currently under development till 2020. The special attention will be given to challenges for 5G and driving services and applications such as Big Data timely delivery.

Textbooks:

1. LTE-Advanced & 5G Ecosystem 2014 – 2020, http://www.mynewsdesk.com/in/pressreleases/lte-advanced-5g-ecosystem-2014-2020-954972

2. Theodore S. Rappaport "Millimeter Wave Wireless Communications", 2014, ISBN-13: 978-0132172288 ISBN-10: 0132172283.

Topic 7. Use cases, service implementations and business opportunities for operators.

Within this topic the students will learn the challenges network operators have been facing with introduction of Big Data and related infrastructure to collect and store it. The special attention will be given to conventional operator's business models, correspondent CAPEX and OPEX.

Textbooks:

1. M2M service providers and LTE: new revenue opportunities, November 2013, http://www.u-blox.com/images/downloads/Product_Docs/M2M-ServiceProviderLTE_Paper_(UBX-13004668).pdf

2. 4G Portal, business models, http://4g-portal.com/tag/business-models

Topic 8. Web 3.0 concept (different definitions and approaches). Emerging Internet of Services, web of services.

Within this topic the students will learn what is the Web 3.0 concept and differences between different definitions and approaches. What is Web 2.0 and how user generated data is used. Students will learn what is the Internet of Services and web of services how it is possible to use it in business.

Textbooks:

1. In Lee (Ed) "Mobile applications and knowledge advancements in e-business", Idea Group Inc (IGI) , 2013.

2. Brian Ernest Mennecke, Troy J. Strader (eds.), "Mobile Commerce: Technology, Theory, and Applications", Idea Group Inc (IGI), 2003

Topic 9. Data storages. Open Data concept.

Within this topic the students will learn different types of data storages and open data concept, how different open data sources are used in business and what is the key point of open data.
Textbooks:

1. Open government data – USA - www.data.gov

2. Open data hand book - http://opendatahandbook.org/

Topic 10. Mobile applications as a source of data.

Within this topic the students will learn how mobile applications generate data, where they keep it, how it is collected and for which purposes it is used.

Textbooks:

1. Brian Ernest Mennecke, Troy J. Strader (eds.), “Mobile Commerce: Technology, Theory, and Applications”, Idea Group Inc (IGI), 2003
2. In Lee (Ed) “Mobile applications and knowledge advancements in e-business”, Idea Group Inc (IGI) , 2013.

Topic 11. Enterprise mobility and mobile applications for business. Financial data for business.

Within this topic the students will learn what it enterprise mobility, why companies integrate enterprise mobile applications, which financial data could be used for business purposes.
Textbooks:

1. Michael Mordhorst, “How to Help Enterprises Going Mobile: Investigation on Influences and Requirements of Business Apps Within Enterprise Mobility”, Anchor Academic Publishing, 2014

2. Lyndsay Wise, “Using Open Source Platforms for Business Intelligence: Avoid Pitfalls and Maximize ROI”, Newnes, 2012.

10. Seminars content for topic 1 – 11:

The main goal of all seminars – to practice using the knowledge gained during the lectures, to propose and develop new business processes, it-infrastructures e.g.

11. Forms of current control

Essay (referat) expects from the students to explain and analyze in 10-15 pages particular topic from the course as business prospective.

Home task expects from the students to prepare technical documentation of the developed mobile application with the explanation (from the essay) of the area for the mobile application and conclusions.

12. Educational technologies

The course aims to bring student close to the real practices of big data collections and processing and business usage, where ugly bosses give unclear order in impolite manner with tight deadlines

and absence of resources, where colleagues are not supportive and subordinates are not motivated.

Subsequently the course will use some elements of "stress ignition":

- Dialog with the students – the lecture may be interrupted at any moment by a question of the lecturer either addressed to the whole audience or addressed to a particular student.
- Analysis and harsh critique of individual and collective presentations.
- Collective responsibility – all members of a project group share the same grade.
- Computer-assisted simulations and presentations are an integrated part of the course.

13. Additional information (examples of lectures)

Example of part of the lecture on topic 8:

The Web 3.0 concept has many different meanings, but all of them consider using information by the machines Web 1.0, or the information web, was straightforward enough. IT was full of static content and could be seen as an extension of offline media, such as print and TV. This version of the web was able to provide information to users in a broadcast model of information distribution. The next evolution of the web brought about web 2.0 or the social web which is characterised by users communications, contributing and collaborating. Web 2.0 has empowered users and consumers of the web to shift from being passive consumers of content and information into active producers of content and information. It allows users to equally participate in the production of content creation and in sharing that content with a wider audience online. Web 3.0 means that our things, our belongings will have the power to learn, intuit and decide.

It is our future in terms of our life, it is our tomorrow in terms of how fast is technological progress integrating new devices into our life.

Technological concept by Marcelo Dias de Amorim was proposed during the FP6 project and includes the description of the Web 3.0 as Internet of Things – when all the things communicate with each other and with the user.

Another technological concept in terms of software methods – is the concept by Tim Berners-Lee (director of W3C) – where he defines the Web 3.0 as the Internet where user can read-write and execute – in terms of execution at the internet he means semantic web and web-services.

Concept by Sramana Mitra is something between technology and philosophy – her formula is this: Web 3.0 results from combining content, commerce, community and context, with personalization and vertical search. Or, to put in a handy phrase: Web 3.0 = (4C + P + VS). By "context," She means the reason you're surfing the Web. Looking for a job is "context," as is planning a trip or shopping for clothes. Fundamental to context is the user. And when you fuse a specific user with genuine context, you wind up with truly personalized service .

And very philosophic concept is the concept by Ted Hoy (former vice-president of the Digital river company). As part of cloud computing, user applications are being transformed into rich Internet applications. Instead of pedestrian computing functionality — like conducting online searches with words — Web 3.0 offers users a richer application tier with far more logic. The breakthrough is whenthe technology itself is aggregated from the client and put on the Web. In this environment, software-as-a-service models will be able to be developed and deployed much more quickly and cost effectively because the infrastructure for development (particular software code) will live in the clouds.

In a Web 3.0 world, the way users experience the Internet will once again be different. Because content and applications will live up in the "cloud," users will be able to experience the Web on a phone, or move from device to device, instead of being limited to a PC. Advancements in usability will be key point and will need to catch up in order for this new multiplatform paradigm to succeed - same hardware parts, but different software parts, which will be in the clouds.

Over the last decade, the service sector has become the biggest and fastest-growing business sector in the world. Our economy is service-oriented and to continue growth services should be easily available and should be more productive. There are different descriptions of different types of services in terms of software engineering, in terms of business, in terms of web and the Internet which are described. As it was shown during the TEXO project which main goal was to develop new business models for the Web and which targeted on the development of an open platform for the development, distribution and provision of services, there was need for the marketplace of services like goods and during that project was proposed special TEXO marketplace as an essential place to go for finding services and solutions. There are special roles in global service delivery: service gateway which makes services available; service aggregator

which composes services into value added services; service broker which bills value added services; service channel which delivers value added service using customer specific channels. But there is a need to make all this processes automated and that is where the Internet of Things concept should help. With the development and integration of the new IPv.6 we have got possibility to integrate all of our devices and things into the same global network. Thus we will have companies with different business-models, based on the data from the Internet for their production facilities with the use of Internet of Things concept.

As it is written with the development of cloud-oriented architecture there will be big change in interaction between users and machines. A cloud-oriented architecture (COA) is a conceptual model encompassing all elements in a cloud environment. According to Jason Bloomberg, of ZapThink, the development of a global cloud-oriented architecture is an essential block of the Internet of Things, in which anything that can be identifies – including people, coffee machines, park benches and just about any other random item you can think of – can be tagged and connected through the Internet or a similar wide-area network. A cloud-oriented architecture is related to both service-oriented architectures (SOA) and event-driven architectures (EDA) and is a combination of two other architectural models: the resource-oriented architecture (ROA) and the hypermedia-oriented architecture. A ROA is based on the idea that any entity can be assigned a uniform resource identifier (URI) is a resource. Thus it includes not only infrastructure elements such as servers, computers and other devices, but also WEB pages, scripts, multimedia information and virtual reality all together. We should consider resource as any component of an application that needs to be used or addressed (physical objects and WEB ―data‖).

As the vision of Internet of Things can be seen from two perspectives – ―Internet‖ centric and ―Thing‖ centric. The Internet centric architecture involved Internet services being the main focus while data is contributed by the objects. In the object centric architecture [9], the smart objects take the center stage. In order to implement the full potential of Cloud computing and cloud-oriented architecture as well as ubiquitous sensing, Cloud at the center seems suitable. This not only gives the flexibility of dividing associated costs in the most logical manner but it is also highly scalable which is important in terms of global services support. Service aggregators can offer their services with the use of data from the devices using a storage Cloud; service brokers can provide their software tools. The main idea of the Cloud ―Internet‖ centric perspective is quite simple – user or customer should not be involved in a process of data analysis, data distribution or data visualization. The same should apply to application developers,

as they should easily connect resources from the Cloud (devices, data, visualization etc.) and provide at the end complete application. Everything will be done inside the Cloud and customers will get only services they ordered. A key concept of having standard algorithms and protocols at the Cloud for all the devices from the Internet of Things (like standards for widgets on the WEB) will definitely make new era at the Internet of Services implementation."