

[Обзорная панель](#)[Мои курсы](#)[Английский язык для профессиональных целей. Весна. _1](#)[Unit 2. Big Data \(Большие данные\)](#)[Class Work 2](#)Вопрос **Инфо**

1. Warming-up questions.

1. What springs to mind when you hear 'Big Data'?
2. How would you define Big Data?
3. Can you give an example of businesses and industries that deal with Big Data?
4. Do you think we need huge and powerful computers to process big data?

Do it there <https://e.sfu-kras.ru/mod/assign/view.php?id=1572162>

Вопрос **1**

Осталось
попыток: 1

Балл:
10,00

2. Watch this video to learn about Big Data and fill in the gaps.



1. We constantly produce a lot of data, for example, via (1) , public transport and GPS.
2. Daily we (2) 55 million pictures, 340 million tweets and one (3) documents.
3. To (4) big data you don't need huge computers.
4. People work with a (5) and endless network of normal servers, and (6) algorithms.
5. Right now the big data of (7) is being analyzed to (8) a car that can drive completely accident-free all by itself.
6. In the future we can even use the big data of (9) to (10) the perfect treatment.

[Проверить](#)

Вопрос **Инфо**

3. Answer the questions related to the video in pairs and in class.

1. What is an amount of data produced every day?
2. What do we call Big Data?
3. What do we need to process Big Data?
4. What is a web site mentioned in the video?
5. What Big Data is analyzed by the site? Why?
6. What is the Big Data of traffic being analyzed for?

Вопрос **Инфо**

4. Read and translate the text.

Big Data (What it is and Why it Matters)

Big data is a popular term used to describe the exponential growth and availability of data, both structured and unstructured. And big data may be as important to business – and society – as the Internet has become. Why? More data may lead to more accurate analyses. More accurate analyses may lead to more confident decision making. And better decisions can mean greater operational efficiencies, cost reductions and reduced risk.

As far back as 2001, industry analyst Doug Laney (currently with Gartner) articulated the now mainstream definition of big data as the three Vs of big data: volume, velocity and variety.

Volume. Many factors contribute to the increase in data volume. Transaction-based data stored through the years. Unstructured data streaming in from social media. Increasing amounts of sensor and machine-to-machine data being collected. In the past, excessive data volume was a storage issue. But with decreasing storage costs, other issues emerge, including how to determine relevance within large data volumes and how to use analytics to create value from relevant data.

Velocity. Data is streaming in at unprecedented speed and must be dealt with in a timely manner. RFID tags, sensors and smart metering are driving the need to deal with torrents of data in near-real time. Reacting quickly enough to deal with data velocity is a challenge for most organizations.

Variety. Data today comes in all types of formats. Structured, numeric data in traditional databases. Information created from line-of-business applications. Unstructured text documents, email, video, audio, stock ticker data and financial transactions. Managing, merging and governing different varieties of data is something many organizations still grapple with.

The real issue is not that you are acquiring large amounts of data. It's what you do with the data that counts. The hopeful vision is that organizations will be able to take data from any source, harness relevant data and analyze it to find answers that enable 1) cost reductions, 2) time reductions, 3) new product development and optimized offerings, and 4) smarter business decision making.

from http://www.sas.com/en_us/insights/big-data/what-is-big-data.html

Вопрос **Инфо**

5. Answer the questions related to the text.

1. [What is Big Data?](#)
2. Why is Big Data so important to business?
3. What is the mainstream definition of Big Data according to analyst Doug Laney?
4. What factors influence on increasing in data volume?
5. Why is velocity important to deal with Big data?
6. What varieties of data are mentioned in the text?
7. What is the point of dealing with Big data?

Do it there <https://e.sfu-kras.ru/mod/assign/view.php?id=1572162>

6. Render the text. Do it there <https://e.sfu-kras.ru/mod/assign/view.php?id=1572162>



Hadoop — проект фонда [Apache Software Foundation](#). Hadoop – [свободно распространяемый](#) набор [утилит](#), [библиотек](#) и [фреймворк](#).

Этот продукт предназначен для разработки и выполнения [распределённых программ](#) (distributed programs), работающих на [кластерах](#) из сотен и тысяч узлов.

Hadoop используется для реализации поисковых и контекстных механизмов многих [высоконагруженных](#) ([high-loaded](#)) веб-сайтов, например, для [Yahoo!](#) и [Facebook](#). Разработан на [Java](#) в рамках вычислительной парадигмы [MapReduce](#).

Согласно этой парадигме, приложение разделяется на большое количество одинаковых элементарных заданий, выполнимых на узлах кластера и [сводимых](#) ([reducible to](#)) в конечный результат естественным образом.

В 2014 год проект состоял из четырёх модулей:

- **Hadoop Common** ([связующее программное обеспечение](#) или промежуточное [программное обеспечение](#));
- **HDFS** ([распределённая файловая система](#));
- **YARN** (система для планирования заданий и управления кластером);
- **Hadoop MapReduce** (платформа программирования и выполнения распределённых MapReduce-вычислений).

Раньше в Hadoop входил целый ряд других проектов, ставших самостоятельными в рамках системы проектов Apache Software Foundation. Разработка **Hadoop** была инициирована в начале 2005 года [Дугом Каттингом](#) ([Doug Cutting](#)) с целью построения программной инфраструктуры распределённых вычислений для проекта [Nutch](#) — свободной [программной поисковой машины](#) (Open-source web-search engine) на [Java](#).

Идейной основой этой разработки стала публикация сотрудников [Google](#) Джеффри Дина и Санжая Гемавата о вычислительной концепции [MapReduce](#). Новый проект был назван в честь *игрушечного слонёнка* ребёнка основателя проекта.

from <https://ru.wikipedia.org/wiki/Hadoop>

7. Invent questions so that the sentences below could be the answers to them.

1. ...? - In the information economy big data will likely be the platform that supports a new business initiative, new competitive advantage, new source of revenue, new way of doing research, and so forth.
2. ...? - We as a society are creating unprecedented amounts of information, and it shows no signs of slowing down anytime soon. But the "big data" perspective sees this as an opportunity, rather than a problem.
3. ...? - Not all big data applications involve analytical analysis of structured data.
4. ...? - Organizations are still unsure how to use Big Data in their situation. Perhaps it is so as they feel their organization is not ready. Companies may have a lot of data but they fail to read or decipher it correctly.
5. ...? - The authors suggest three mutually supportive capabilities; (a) Ability to identify/combine/manage data from multiple sources (b) Ability to build predictive and optimization models using this data (c) Willingness of the organization to make transformations based on the findings from this data.
6. ...? - The inherent risk with modeling exercise is that the designed model may turn out to be so complex that it is no longer practical to be used.
7. ...? - Many implementations of big data and analytics failed because they weren't in sync with the company's day-to-day processes and decision making norms.
8. ...? - Big Data offers the humanistic disciplines a new way to claim the status of quantitative science and objective method.
9. ...? - As a large mass of raw information, Big Data is not self-explanatory. And yet the specific methodologies for interpreting the data are open to all sorts of philosophical debate.
10. ...? - There are three types of technologies currently utilized for big data analytics: software database appliances, hardware database appliances, and distributed databases.

ПРЕДЫДУЩИЙ ЭЛЕМЕНТ КУРСА

[◀ E-workbook. Unit 2](#)

Перейти на...

СЛЕДУЮЩИЙ ЭЛЕМЕНТ КУРСА

[Homework 2 ▶](#)

© 2010–2024 Центр обучающих систем
[Сибирского федерального университета, sfu-kras.ru](http://sfu-kras.ru)

Разработано на платформе moodle
Beta-version (3.9.1.5.w3)

[Политика конфиденциальности](#)

[Соглашение о Персональных данных](#)

[Политика допустимого использования](#)

Контакты +7(391) 206-27-05
info-ms@sfu-kras.ru

[Скачать мобильное приложение](#)



[Инструкции по работе в системе](#)