INTRO 4.0

BIG DATA

INTRO 4.0

INTRODUCTION TO THE INDUSTRIAL REVOLUTION 4.0

These didactical materials, which have been developed in the framework of the European project 'Industry 4.0 - INTRO 4.0', funded by the European Commission aims to come up with an overview of what has been done in the European Industry in terms of Industry 4.0.

The content of these didactical materials provides the most relevant and useful information on Industry 4.0 to a target group that includes: adults, educators (VET & Higher Education), teachers, trainers, coaches, employers, employees, the general public, and suppliers of innovative solutions.

This information is rooted within the report 'Current Status Of The Industry 4.0' and the report 'Summary Report of the expert interviews/questionnaires and the specific research on the field of manufacturing companies", both developed by the partners of this project.

INTRO 4.0

BIG DATA

INDEX

THIS CONTENT MAY BE OF GREATER INTEREST TO THE COMPANIES

THIS CONTENT MAY BE OF GREATER INTEREST TO THE GENERAL PUBLIC

🎯 LEARNING OBJECTIVES

❖ Understand Big Data.

❖ Identify Big Data opportunities.

❖ Recognize and improve most value skills.

❖ Monitor organizational challenges and good practices.

❖ Increase company resources and benefits.

❖ Perform useful strategies.

❖ Set practical uses.

❖ Provide useful information about courses and certificates

INTRO 4.0

INTRODUCTION

Big Data refers to large amounts of data produced very quickly by a high number of diverse sources.

Learning Objectives

- Understand Big Data
- Identify Big Data opportunities
- Recognize and improve most value skills
- Monitor organizational challenges and good practices
- Increase company resources and benefits
- Perform useful strategies
- Set practical uses
- Provide useful information about courses and certificates

Taking decisions

Identification of risk to the product/services

What is this for?

Customer service

Operational efficiency

Volume | Variety | Velocity | Veracity | Value | Virtue

6 Vs of Big Data

Designed to extract value economically from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis.

FUTURE APPLICATIONS
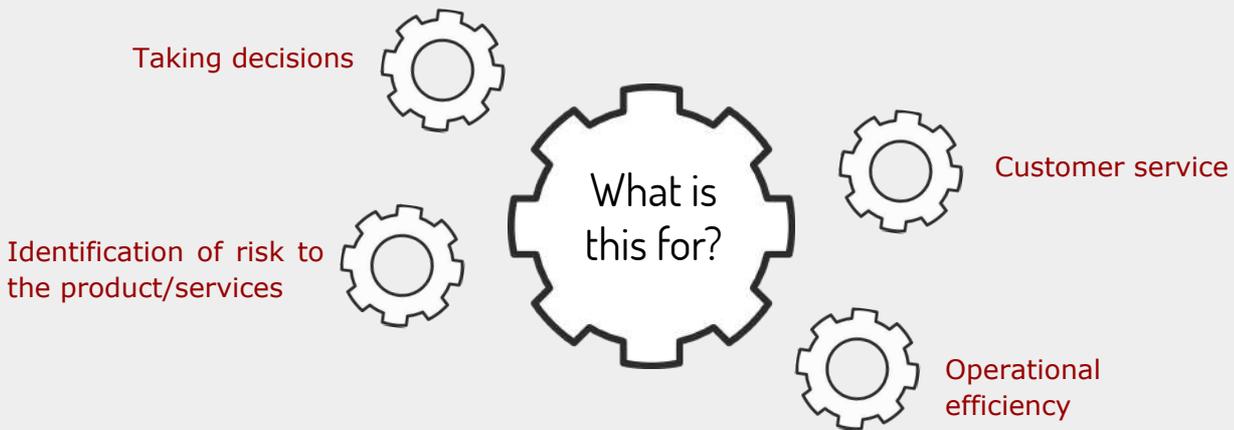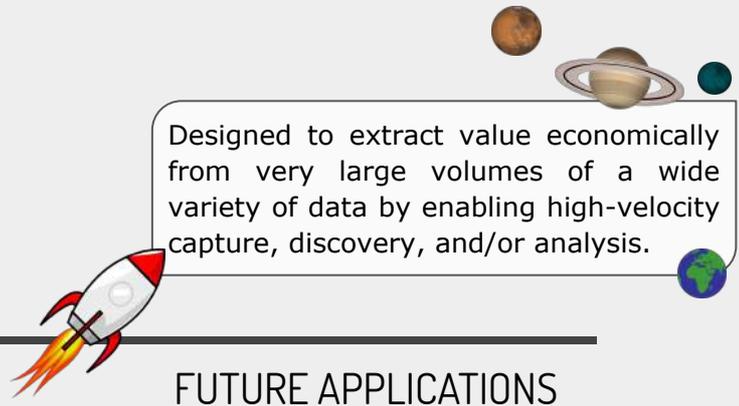
WHAT IS IT?

Across society, from health to agriculture and transport, from energy to climate change and security, practitioners in every discipline recognise the potential of the enormous amounts of data being created every day. The challenge is to capture, manage and process that information to derive meaningful results and make a difference to people's lives. Data can either be created by people or generated by machines, such as sensors gathering climate information, satellite imagery, digital pictures and videos, purchase transaction records, GPS signals, etc. It covers many sectors, from healthcare to transport and energy, communications and retail.

Figure 1. Big Data uses. Source: www.edureka.com

Generating value at the different stages of the data value chain will be at the centre of the future knowledge economy. Good use of data can also bring opportunities to more traditional sectors such as transport, health or manufacturing.

INTRO 4.0

WHAT IS IT?

- Transform service industries by generating a wide range of innovative information products and services

- Increase the productivity of all sectors of the economy through improved business intelligence

- Better address many of the challenges that face our societies

- Improve research and speed up innovation

- Achieve cost reductions through more personalised services

- Increase efficiency in the public sector

- Digitising European Industry

Figure 2. Predictive analytics. Source: www.dreamstime.com

XV-XIX

- The Emergence of Statistics
- First recorded experiment in statistical data analysis.
- The Hollerith Tabulating Machine used punch cards reducing 10 years' work to three months

XX

- The Early Days of Modern Data Storage
- The Beginnings of Business Intelligence
- The Start of Large Data Centers
- The Emergence of the Internet

XXI

- Early Ideas of Big Data
- "Internet of Things"
- Web 2.0 Increases Data Volumes
- Today's Use of the Term 'Big Data' Emerges
- Final Thought

INTRO 4.0

WHAT IS THIS FOR?

Big Data presents great opportunities as they help us develop new creative products and services, for example apps on mobile phones or business intelligence products for companies. It can boost growth and jobs in Europe, but also improve the quality of life of Europeans.

Healthcare

Enhancing diagnosis and treatment while preserving privacy, Big Data offers solutions for improved efficiency in healthcare information processing which in turn creates value for businesses, public sector and citizens. The analysis of large clinical datasets can result in the optimisation of the clinical and cost effectiveness of new drugs and treatments and patients can benefit from more timely and appropriate care. Data interoperability is of utmost importance since the data is derived from diverse and heterogeneous sources such as bio-signal streams, health records, genomics and clinical lab tests. Privacy-preserving technologies aim at providing access to health data for patients, healthcare professionals and clinical researchers in a uniform way and in an anonymized and aggregated form to develop better prevention or treatment options.

Data Markets

Information technology has driven, directly or indirectly, much of Europe's economic growth during the last decades as the role of data transitioned from the support of business decisions to becoming a good in itself. An open approach towards data value creation has become critical in the new networked economy, with Europe well placed to nurture this new revolution.

Transport: fewer accidents and traffic jams

The transport sector can clearly benefit from Big Data collected through sensors, GPS data and social media in particular. A smart use of Big Data supports governments in optimising multimodal transport and managing traffic flows, making our cities smarter. Citizens and companies can save time through the use of route planning support systems.

INTRO 4.0

WHAT IS THIS FOR?

Environment: reduced energy consumption

The Big Data revolution brings about novel ways of understanding and addressing environmental challenges. A better use of globally available national and local datasets helps scientists in their research and enables policy-makers to make informed and evidence-based decisions related to natural disasters like flooding, to fight against climate change and reduce costs. Smart cities also host data centres adapting the power consumption of public buildings to the availability of renewable energy and other useful indicators. At the same time, our mobiles devices become smarter by integrating analytical tools to reduce our energy consumption and save money.

Open Data

Open Data refers to the information collected, produced or paid for by the public bodies and made freely available for re-use for any purpose. Public sector information is information held by the public sector. The Directive on the re-use of public sector information provides a common legal framework for a European market for government-held data. It is built around the key pillars of the internal market: free flow of data, transparency and fair competition.

Agriculture: safer food and increased productivity

A smart use of Big Data in agriculture can increase productivity, food security and farmer incomes at the same time. Through an intelligent and widespread use of data coming from sensors and Earth observations such as the open data from the Copernicus Programme the ways we are farming today can be changed entirely for the better. This can lead to a more efficient use of natural resources (including water or sunlight) in our farming practices. With advanced technologies farmers can have access to data in real time on how their farm machinery is working as well as to historic weather patterns, topography and crop performance.

Industrial impact / Big Data access technologies / Research

Maximally exploiting available data is increasingly critical to industrial competitiveness. Accessing the relevant data is becoming progressively difficult due to the explosion in the size and complexity of data sets. Maximally exploiting data requires flexible access and engineers need to explore the data in ways not supported by current applications. Engineers spend up to 80% of their time on data access problems. Apart from the enormous direct cost, freeing up expert time would lead to even greater value creation through deeper analysis and improved decision making.

INTRO 4.0

## WHAT IS THIS FOR?

TOP 8 BIG DATA SKILLS FOR WORKERS
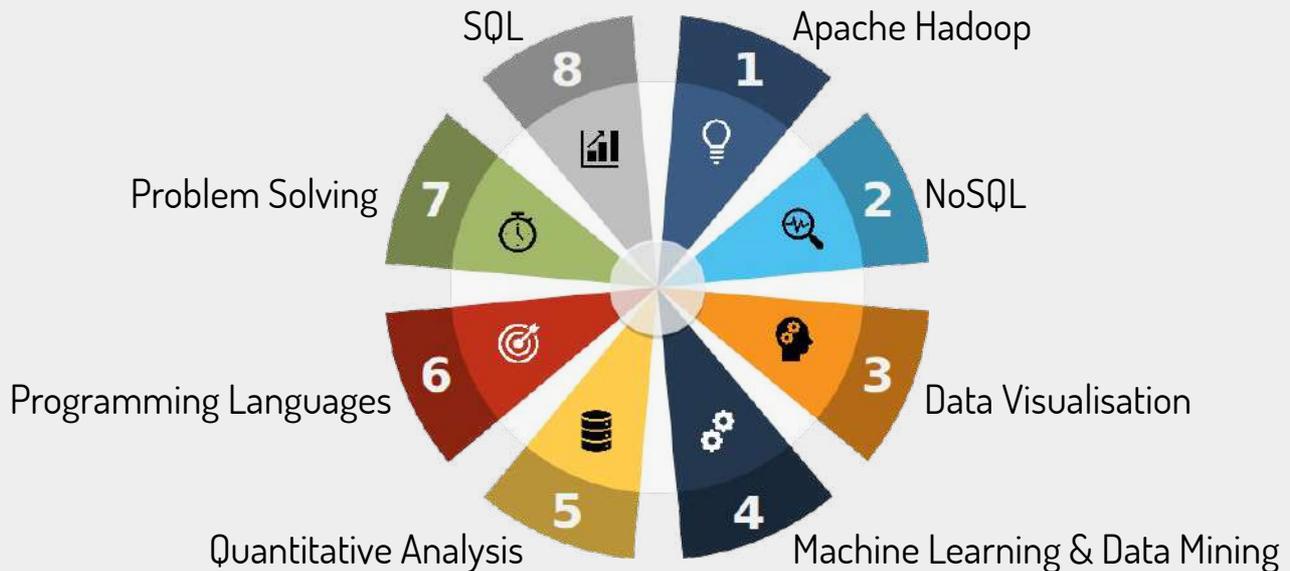
Figure 3. Top 8 Big Data skills for workers.
Source: Self made

Apache Hadoop: Hadoop is an open-source, a Java-based programming framework that continues the processing of large data sets in a distributed computing environment. It runs few applications on distributed systems with thousands of nodes involving petabytes of information. It has a distributed file system, called Hadoop Distributed File System or HDFS, which enables fast data transfer among the nodes. A modern implementation of Hadoop now features an ecosystem of related projects that provide a rich set of big data services:

➢ Apache Spark is a distributed processing engine that performs high performance, in-memory processing of large data sets.

➢ Apache Hive provides built-in data warehousing capabilities to the Hadoop system using a SQL-like access methods for querying data and analytics.

➢ Apache HBase is a scalable, distributed NoSQL wide column database built on top of HDFS.

➢ Apache Zeppelin is a web-based, multi-purpose notebook that enables interactive data processing including ingestion, exploration, visualization, and collaboration features for Hadoop and Spark.

INTRO 4.0

BIG DATA

WHAT IS THIS FOR?

NoSQL: The NoSQL databases including Couchbase, MongoDB, etc. are replacing the traditional SQL databases like DB2, Oracle, etc. These distributed NoSQL databases help in meeting the Big Data storage and access needs. This complements the expertise of Hadoop with its data crunching ability. The professionals with NoSQL expertise can find opportunities everywhere.

Data Visualisation: The data visualization tools like QlikView, Tableau can help in understanding the analysis performed by the analytics tools. The complex Big Data technologies and processes carried out are tough to grasp, and this is where the role of professionals come into the picture. A professional well versed with data visualization tools can get a chance to grow in their career with big organizations.

Machine Learning & Data Mining: Data mining and Machine Learning are the two hot fields of Big Data. Though the landscape of Big Data is vast, these two make an important contribution to the field. The professionals that can use machine learning for carrying out predictive and prescriptive analysis are scarce. These fields can help in developing recommendation, classification and personalization systems. The professionals with the knowledge of data mining and machine learning are heavily paid as well.

Quantitative Analysis: Quantitative and Statistical analysis is a significant part of Big Data as it is all about numbers. The background in statistics and mathematics helps a lot. The knowledge of tools like SAS, SPSS, R, etc. help in adding to your skills as well. Hence, the industry required professionals with the quantitative background in large numbers.

Programming Languages: Certain general-purpose programming languages can help you a great deal in gaining a competitive edge over others. These programming languages include Java, Python, C, Scala, etc. Even the programmers with experience in data analytics are in great demand.

INTRO 4.0

BIG DATA

WHAT IS THIS FOR?

Problem Solving: Even if you carry the knowledge of all tools and technologies in the Big Data field, the ability of problem-solving and creativity will help you to perform your tasks well. Implementation of Big Data techniques for efficient solutions will require both these qualities in a professional.

SQL: SQL is the data centered language that works as a base for the Big Data era. The knowledge of Structured Query Language will essentially be an added advantage to the programmers while working on Big Data technologies like NoSQL. It is also an important part of the Hadoop Scala warehouses.
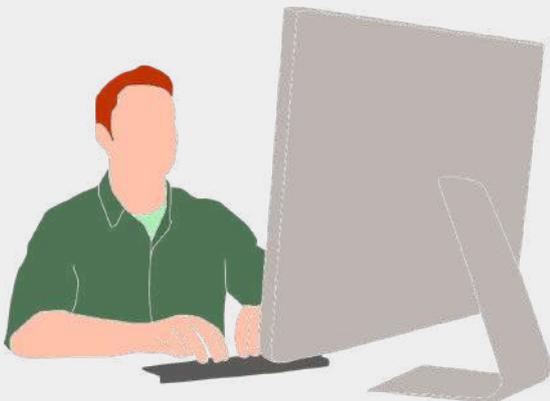
Most valued skills:

1. Teamwork
2. Business acumen
3. Intellectual curiosity
4. Problem solving
5. Communication skills

Not all applications of Big Data technologies are for analysis of data. Some are used for deploying Web sites for social media or gaming applications, and others are used for large content stores that provide information access to massive amounts of documents. Examples include:

Analytics (e.g., data mining, multi-dimensional analysis, data visualization)

Operations (e.g., running a Web site, processing online orders)

Information access (e.g., search-based access to information, normalization, and access across content and data sources)

INTRO 4.0

WHAT IS THIS FOR?

Improving your Big Data skills you will be able to...

Provide ideas from huge amounts of data from multiple sources

Mitigate risks by optimizing complex decisions

Improve client commitment and loyalty.

Monitor the performance or operations of the businesses

Identify the causes of failures and problems in real time

Evaluate the portfolio risk quickly.

Locate, get, extract, change, analyze, and blend data with different tools.

Understand the potential of data-driven marketing.

Customize the client experience.

Improve the quality of decision making.

Generate customers based on their purchasing habits.

Add value to the interactions with online and offline customers.

Figure 4. Capabilities when improving Big Data skills. Source: Self made

INTRO 4.0

BIG DATA

WHAT IS THIS FOR?

How to Develop Your Big Data Skills?

Data Visualisation and Analytical Skills

Big Data tools essentially carry out data analysis to derive important insights from the large datasets. Being familiar with the business domain can help you to understand the data for which the analysis is conducted.

The data professionals shall have an ability to interpret the data by visualizing it. This needs a specific science and mathematics edge to understand the complex data with creativity and imagination. Learning the analytics tools can help you to develop your data visualization and analytics skills.

Programming Ability

The ability to code and conduct statistical and quantitative analysis is a significant requirement in Big Data market. The background in mathematics will help greatly. Knowledge of Object-Oriented Languages and the fundamental knowledge of data structures and algorithms can go a long way. It is important to be familiar with sorting algorithms, data types, etc.

Familiarity with Technologies

It is important for a Big Data professional to be familiar with a range of tools and technologies that the industry uses. The number of tools you can work with, the better it is. These tools help in performing research analysis and reaching conclusions.

These technologies include SPSS, Excel, SQL, SAS, R, MatLab, Python, Linux, Hadoop, Scala, etc. There are many open source technologies written in other languages giving an edge to the technical experts. The demand for people with both statistical and programming skills is even higher.

Hands on experience and developing yourself

Intenta obtener experiencia práctica con las herramientas de Big Data que estás aprendiendo. Dado que la tecnología cambia muy rápidamente, estudiar algunos cursos puede ayudar significativamente. La interacción con las bases de datos puede ayudar a comprender mejor las herramientas de datos. El aprendizaje automático y la minería de datos pueden ayudarte a obtener una mejor experiencia en herramientas Big Data. Puedes buscar cursos en línea para aprender más sobre estas tecnologías.

BIG DATA

GOOD PRACTICES

Today, almost any interaction made over the Internet or through the consumption of goods and services is being tracked, stored, and used in targeted ways. This has led to the notion of Big Data massive amounts of data that reflect the behavior and actions of various people. Data scientists and data collection platforms are now able to computationally organize petabytes and exabytes of data so that it is easy to analyze and identify patterns that may have otherwise gone undetected.

Big Data technologies are being deployed in support of processes within commercial, nonprofit, or government organizations. The challenges and problems organizations face are not Big Data challenges but rather business or organizational challenges that are impacted by Big Data. Big Data technology deployment cases can be found across business processes such as:

- Customer relationship management (sales, marketing, customer service, etc.)
- Supply chain and operations
- Administration (focused on finance and accounting, human resources, legal, etc.)
- Research and development
- Information technology management
- Risk management

GOOD PRACTICES

BBVA

The history of BBVA is the history of the many different people who have been a part of the more than one hundred financial institutions that have joined our corporate endeavor since it originated in the mid-19th Century. During the economic development of the 1960s until now , BBVA has expanded, acquiring other banks and creating a solvent financial group. Prestigious financial publications recognized the efficiency of BBVA's integration, naming it the best bank in the world (Forbes) and in Spain (The Banker) and in the year 2000, the best bank in Latin America (Forbes) and the best bank in Europe (Lafferty) in 2001.

Corporate responsibility is at the core of its business model. BBVA fosters financial education and inclusion, and supports scientific research and culture. It operates with the highest integrity, a long-term vision and applies the best practices. They really believe the knowledge derived from financial data can transform the banking industry. Thus, they implement and use the most advanced analytics and artificial intelligence to offer the best digital interaction with the customer. The challenge, and more accurately the opportunity, comes not from the gathering and storing, but rather from how insight is derived from that data – how it is put to good use, and it's an area where BBVA is gaining recognition as a leader and expert in the field,transforming Big Data into financial intelligence for large institutions as well as smaller companies and individuals that in the past had not access to such advantages.

Some leading companies:

vmware ESXi

Microsoft

IBM

G

Hewlett Packard Enterprise

alteryx

ORACLE®

TERADATA

SAP

cogito

amazon web services™

splunk>

DELL EMC²

INTRO 4.0

BENEFITS FOR THE COMPANY

Big Data Management solutions:

❏ Provide companies the ability to add a variety of data from hundreds of different sources in real time. This means that you can increase the client's commitment since you can have more effective interactions with them and better marketing proposals, which ultimately lead the company to achieve a longer and more profitable relationship with the client.

❏ Eliminate data niches, so that organizations can obtain a unique view of the customers that include countless descriptive, calculated and industry-specific metrics that allow for the construction of a detailed record of the behavior of each client. These profiles provide organizations with a global understanding of their clients through in-depth knowledge of the client and its operations.

❏ Provide organizations with complete customers' profiles, which allows for more personalized customer experiences at each point where contact is made throughout the entire journey of the company.

❏ Your organization can address the data it needs to obtain actionable information and increase the value of the entire relationship with the client. Apps developed by a smartphone app development company can be used to sustain a good relationship with your clients.

The reality is that as data volumes continue to increase, its promise for companies also appears to exponentially grow. This allows companies to convert raw data into relevant projections, predictions, and trends with accuracy.

Big Data is Timely: 60 percent of each work day, knowledge workers spend it trying to find and manage the data. Big Data can provide timely reports immediately.

Big Data is Accessible: Half of the senior executives indicate that access to correct data is often difficult.

Big Data is Holistic: The information is currently stored in silos within many organizations. Marketing data, for example, can be found in web analytics, mobile analysis, social analysis, CRM systems, A/B testing tools, email marketing systems, and many other sites each with its focus on its silo.

INTRO 4.0

BENEFITS FOR THE COMPANY

Big Data is Reliable: Things as simple as securing the correct contact data of customers through the review of multiple systems can save thousands of Dollars in incorrectly sent communications.

Big Data is Relevant: 43 percent of companies are not satisfied with the ability of their tools to filter irrelevant data.

Big Data is Safe: A breach of data security costs hundreds of dollars per customer.

Big Data is Precise: Businesses have difficulty with multiple versions of the fact based on the supply of their info. Combining multiple reliable sources, businesses can create accurate correct of intelligence.

Big Data is Usable: Many companies make bad decisions due to obsolete or bad data. Big Data can ensure that the data is usable without fear of mistakes.

Advantages

- Better decision-making
- Increased productivity
- Reduce costs
- Improved customer service
- Increased revenue
- Increased agility
- Greater innovation
- Faster speed to market
- Fraud detection

Disadvantages

- Need for talent
- Address data quality issues
- Need for cultural change
- Compliance with regulations
- Cybersecurity risks
- Technology changes
- Hardware needs
- Difficulty integrating legacy systems

Figure 5. Advantages and Disadvantages of Big Data.
Source: Self made

INTRO 4.0

FUTURE APPLICATIONS

The importance of Big Data does not revolve around how much data a company has but how a company utilizes the collected data. Every company uses data in its own way; the more efficiently a company uses its data, the more potential it has to grow. The company can take data from any source and analyze it to find answers.

New Product Development

By knowing the trends of customer needs and satisfaction through analytics you can create products according to the wants of customers.

Cost savings

Some tools of Big Data like Hadoop and Cloud-Based Analytics can bring cost advantages to business when large amounts of data are to be stored and these tools also help in identifying more efficient ways of doing business.

Time Reductions

The high speed of tools like Hadoop and in-memory analytics can easily identify new sources of data which helps businesses analyzing data immediately and make quick decisions based on the learnings.

Understand the market

By analyzing customers' purchasing behaviors, a company can find out the products that are sold the most and produce products according to this trend. By this, it can get ahead of its competitors.

Control online reputation

Big Data tools can do sentiment analysis by monitoring and improving the online presence of your business.

Figure 8. Benefits of Big Data for companies. Source: Self made

The use of Big Data is becoming common these days by the companies to outperform their peers. In most industries, existing competitors and new entrants alike will use the strategies resulting from the analyzed data to compete, innovate and capture value.

Big Data helps the organizations to create new growth opportunities and entirely new categories of companies that can combine and analyze industry data. These companies have ample information about the products and services, buyers and suppliers, consumer preferences that can be captured and analyzed.

🚀 FUTURE APPLICATIONS

Practical Uses of Big Data:

Different industries are using Big Data in different ways. In our list, we have compiled the uses of Big Data and what industries are using them.

Location Tracking: Logistic companies have been using location analytics to track and report orders for quite some time. With Big Data in the picture, it is now possible to track the condition of the good in transit and estimate the losses. It is now possible to gather real-time data about traffic and weather conditions and define routes for transportation. This will help logistic companies to mitigate risks in transport, improve speed and reliability in delivery.

Precision Medicine: With Big Data, hospitals can improve the level of patient care they provide. 24x7 monitoring can be provided to intensive care patients without the need of direct supervision. On top of that, the efficiency of medication can be improved by analyzing the past records of the patients and the medicines provided to them. The need for guesswork can be significantly reduced. In the case of certain biopharmaceuticals, there are many variables that impact the final product. For example, while manufacturing insulin intense care needs to be taken to ensure the product of desired quality. By analyzing all the factors impacting the final drug Big Data analysis can point out key factors that might result in incompetence in production.

INTRO 4.0

🚀 FUTURE APPLICATIONS

Fraud Detection & Handling: Banking and finance sector is using Big Data to predict and prevent cyber crimes, card fraud detection, archival of audit trails, etc. By analyzing the past data of their customers and the data on previous brute force attacks banks can predict future attempts. Not just Big Data helps in predicting cyber crimes but it also helps in handling issues related to miss transactions and failures in net banking. It can even predict possible spikes on servers so that banks can manage transactions accordingly. The Securities Exchange Commission (SEC) is using Big Data to monitor financial markets for possible illegal trades and suspicious activities. The SEC is using network analytics and natural language processors to identify possible frauds in the financial markets.

Advertising: Advertisers are one of the biggest players in Big Data. Be it Facebook, Google, Twitter or any other online giant, all keep a track of the user behavior and transactions. These internet giants provide a great deal of data about people to the advertisers so that they can run targeted campaigns. Take Facebook, for example, here you can target people based on buying intent, website visits, interests, job role, demographics and what not. All this data is collected by Facebook algorithms using Big Data analysis techniques. The same goes for Google, when you target people based on clicks you will get different results and when you create a campaign for leads that you will get different results. All this is made possible using Big Data.

Entertainment & Media: It is focused on targeting people with the right content at the right time. Based on your past views and your behavior online you will be shown different recommendations. This technique is popularly used by Netflix and Youtube to increase engagement and drive more revenues. This will allow in better revenue from ads and will provide a more engaging user experience.

INTRO 4.0

🚀 FUTURE APPLICATIONS

TRENDS:

❖ **Rapidly Growing IoT Networks**

It is becoming quite common that our smartphones are being used to control our home appliances, thanks to the technology called the Internet of Things (IoT). With smart devices such as Google Assistant and Microsoft Cortana trending in homes to automate specific tasks, the growing IoT craze is drawing companies to invest in the technology's development.More organizations will jump on the opportunity in providing better IoT solutions. This will lead to more ways to collect vast amounts of data, and along with it the means to manage and analyze it. The industry response is to push for more new devices that are more capable of collecting, analyzing and processing data.

INTERNET OF THINGS (IOT)

The Internet of Things, or IoT, is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.

❖ **Accessible Artificial Intelligence**

Artificial intelligence is now more commonly utilized to help both big and small companies improve their business processes. AI programs can now execute tasks that make it faster and more precise than humans, cutting down errors along the way and improving the overall flow. This allows humans to focus better on more critical tasks and further enhance the quality of service.The good news is everybody can have access to pre-built machines that run AI applications to address the growing demand, which levels the playing field for companies in the same industry. Individual organizations may gain an advantage if they find the most efficient way to integrate this into their business process.

FUTURE APPLICATIONS

❖ The Rise of Predictive Analytics

Big Data analytics has always been a key strategy for businesses to have a competitive edge and achieve their goals. They use the necessary analytics tools to process Big Data and determine the reasons why certain events happen. Now, predictive analysis through Big Data can help predict what may occur in the future. There is no doubt this kind of strategy is highly effective in helping analyze gathered information to predict consumer behavior. This allows companies to determine the measures they have to take by knowing a customer's next action before they even do it. Analytics can also provide more context on data to help understand the reasons behind them.

❖ Dark Data Migration to the Cloud

Information that is yet to be transformed into digital format is called dark data, and it is a huge reservoir that is currently untapped. These analog databases are expected to be digitized and migrated to the cloud, so they can be used for predictive analytics that benefits businesses.

❖ Chief Data Officers Will Have Bigger Roles

Now that Big Data is increasingly becoming an essential part of executing business strategies, chief data officers are adopting a more critical role in their organization. They are expected to take a more active position in steering the company towards the right direction. This trend opens doors for data marketers who are looking for career growth.

🚀 FUTURE APPLICATIONS

❖ Quantum Computing

Getting to analyze and interpret massive amounts of data can take a lot of time with the current technology we are using. If only we can crunch billions of data at once in just a few minutes, we can cut processing time immensely, giving companies the opportunity to make timely decisions to achieve more desired results. This huge undertaking can only be possible through quantum computing. Despite being in its infancy, experiments are currently being carried out on quantum computers in an effort to help in practical and theoretical research across different industries. Pretty soon, large tech companies such as Google, IBM, and Microsoft will start testing quantum computers to integrate them into their business processes.

❖ Smarter and Tighter Cybersecurity

Organizations have grown paranoid over the past scandals that involved hacking and system breaches. This has prompted them to focus on strengthening information confidentiality. IoT is also being a cause for concern with all the data being collected; cybersecurity is an issue. To address this perpetually impending threat, Big Data companies pitch in to help organizations use data analytics as a tool to predict and detect cybersecurity threats. Big Data can be integrated into a cybersecurity strategy through security log data where it can be used to provide information about past threats. This can help companies prevent and mitigate the impact of future hacks and data breaches.

🚀 FUTURE APPLICATIONS

❖ Open Source Solutions

There are many public data solutions available, such as open source software, that have been making considerable improvements to speed up data processing. They now have features that allow access and response to data in real time. For this reason, they are expected to flourish and will be in high demand from 2019. There is no doubt open source software is cheaper, in that it can help your business cut operations costs.However, there are some downsides that you need to know if you are willing to give them a shot.

❖ Edge Computing

Edge computing is set to leave the cloud in the dust when it comes to processing data. It delivers a better performance since there is less data flowing in and out of the network, with less cloud computing costs. The company can also benefit from storage and infrastructure costs if they choose to delete unnecessary data collected from IoT. Additionally, edge computing can speed up data analysis, giving companies ample time to react.

❖ Smarter Chatbots

Powered by smarter AI, chatbots are now being deployed by companies to handle customer queries to deliver more personalized interactions while eliminating the need for actual human personnel. Big Data has a lot to do with delivering a more pleasant customer experience as bots process large amounts of data to provide relevant answers based on the entered keywords by customers in their queries. During interactions, they are also able to collect and analyze information about customers from conversations. This process can help marketers develop a more streamlined strategy to achieve better conversions.

ADVANCED CONTENT

The vast collections of information available on the web and in the cloud could help prevent the next financial crisis, or even tell you exactly when your bus is due. The key lies in giving everything (whether it's a person, business or product) a unique identifier.

The EU data market has been analysed in the past years by several studies and reports. Despite of this growing bottom-line market, there are some barriers:

- Europe has been slow to adopt data technologies compared to the US.

- Data skills gap.

- Standardization. Increasing complexity and variety on standards can slow innovation.

- Privacy and data protection. A reliable legal framework, like GDPR, is complex but can guarantee success of the companies.

- Reaching all sorts of SMEs and start-ups. Companies emerging from entrepreneurial ecosystems like accelerators or incubators are usually not so linked to EU initiatives.

Data Market Services is born to overcome the barriers of data-based SMEs and start-ups in Europe in data skills, entrepreneurial opportunities, legal issues and standardization, thanks to the provision of free support services for them.

INTRO 4.0

ADVANCED CONTENT

To learn Big Data, it is important to get some hands-on experience apart from the theoretical knowledge. Organisations looking for data analysts and data scientists often prefer the talent with special Big Data certifications in the field. The candidates can gain an edge over others by having certain certifications on their resume.

There are some important, helpful certifications in building a career in Big Data:

Cloudera Certifications

built around the Apache Hadoop platform and provide the tools to extract the most value from your customer data.

MongoDB

is a document database with the scalability and flexibility that you want with the querying and indexing that you need

Apache Storm

is a free and open source distributed realtime computation system. Storm makes it easy to reliably process unbounded streams of data

Scala

is a programming language used for functional programming and strong static systems. It is object-oriented and it runs on JVM

Hortonworks Certifications

help Big Data Hadoop professionals to establish valid and important Hadoop credentials. you over the rest.

R software

is a language and environment for statistical computing and graphics. R is an integrated suite of software facilities for data manipulation, calculation and graphical display.

Python

is an object-oriented, high-level programming language with integrated dynamic semantics primarily for web and app development.

MapReduce

is a processing technique and a program model for distributed computing based on java.
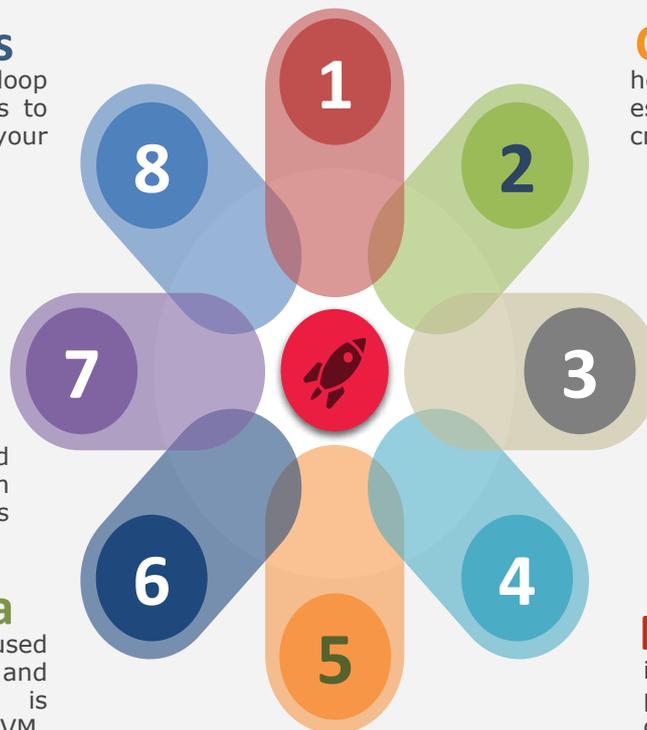
Figure 9. Certifications in building a career in Big Data.
Source: Self made

ADVANCED CONTENT

Creating a strategy

1   **Identify What You Want:**

Your end goal has the biggest impact on the shape of your overall strategy. You need to decide whether you want to increase the efficiency of customer reps, improve operational efficiency, increase revenues, provide better customer experience or improve marketing. The goal you have should be precise, certain and direct. Any strategy with just the sole purpose of exploring possibilities is likely to end up in confusion. Based on your goal you can choose a methodology, hire employees and select the right sources of data. So create SMART (Specific, Measurable, Attainable, Relevant and Timely) goals and make plans accordingly.

2   **Leverage a Proven Big Data Strategy:**

There are 4 proven ways to create a working Big Data strategy. Based on your end goal and availability of data you can choose either of the below Big Data strategies to attain successful results:

A.   Performance Management: It involves using transactional data like customer purchase history, turnover and inventory levels to make decisions relating to store management and operational supremacy. This data is available within the organization and gives insights into subjects relating to short term decision making and long term planning. It works well with companies with large historical databases that can be leveraged without much pain. It can also help with better customer segmentation and targeting.

ADVANCED CONTENT

B.   Data Exploration: This approach makes heavy use of data mining and research to find solutions and correlations that are not easily discoverable with in-house data. Currently, it is used by companies focusing on robust inbound marketing to generate insight on prospects behavior on the website. It helps you identify new segments of data and bring out insights regarding customer's behavior and preferences.

C.   Social Analytics: Social analytics measures the non-transactional data on various social mediums and review sites like Facebook, Twitter and Google+. It is based on the analysis of conversations and reviews that come up on these platforms. It brings out three primary analytics viz. awareness, engagement, and word-of-mouth. In-stream data analysis techniques like sentiment analysis prove very effective in these cases. It gives insights on the brand identity and customer's opinions on new offerings and services. The social analysis also proves effective in predicting spikes in demand for certain products.

D.   Decision Science: Decision science refers to the experiments and analysis on non-transactional data, such as consumer-generated content, ideas, and reviews. Decision science is more about exploring possibilities than measuring known objectives. Unlike social analysis, that is based on engagement analytics, decision science focuses on hypothesis testing and ideation process. This involves extensive use of text and sentiment analysis to understand customer's opinions about new services and schemes.

INTRO 4.0

BIG DATA

ADVANCED CONTENT

3 Identify Infrastructural Changes:

To leverage Big Data particularly historical databases you might need to create many infrastructural changes in the company. If the old company data was stored in traditional formats it might not facilitate the running of complex algorithms and analysis. Moreover, different departments may need integration to collect and streamline data to put it to more usable format. Integration between different departments is key to bringing and implementing changes at scale. If your existing infrastructure is not interlinked properly then you will need to prepare for big changes.

4 Establish Talent Pool:

Human Resources is one of the most critical aspects of creating a Big Data strategy. Your Big Data team must have statisticians to make sense out of data, business analysts to communicate insights to the decision makers and key decision makers themselves who are capable to lead the team. Without a proper team, the discussions on Big Data may revolve around jargons that are not clear to either of the teams. A proper language needs to created to facilitate discussions between the business leaders and the technical team. If this is not done properly then no side will be able to understand the insights and the entire execution will end up with regrets and blame games.

INTRO 4.0

ADVANCED CONTENT

5 Obsess Over Customer Satisfaction:

The key use of Big Data is to generate insights that can help companies serve their customers in a better way. Customer oriented marketing is the new way of approaching the market and making revenues. At the end of the day, you need to communicate to your customer that you are there to solve a problem and not just to make money. Big Data provides such insights into the customer mind set that can be used to improve and even alter the current marketing practices. Another thing you need to focus on is to create a fine line between data gathering and privacy abuse. Your customers should not feel like they are spied.

6 Ensure Usability:

Many times it happens that the insights created by the statisticians are beyond comprehension for staff. The data, analytics, and insights that are collected by the analysts needs to be communicated precisely to the implementation team. The information should be comprehended and represented in a way that its value is identified by people who are not from a statistical background. This can be done by using graphical representation and by communicating direct instructions to the teams involved.

7 Be Agile:

This goes without saying. While implementing disruptive technologies many hurdles might come up that no one initially thought about. You need to adjust your budget, people, and ideologies based on the circumstances and insights you gather. It is best to start with a high-level plan and make changes as the need be. You might come up with an action plan that is nowhere close to the initial idea but it will be worth the toil.

INTRO 4.0

ADVANCED CONTENT

SOME BIG DATA TOOLS:

FusionCharts Suite XT

Interactive charts for reports, dashboards, analytics, monitors, and surveys.

QlikView

Software for business intelligence & data visualization. You can analyze data and use your data discoveries to support decision making.

Tibco Spotfire

Is a very powerful data visualization tool that allows users to access and combine data in a single analysis. It gives you a quick insight into your data.

Watson Analytics

Is a smart data analysis and visualization service on the cloud that helps just about anyone quickly discover patterns and meanings in their data.

Sisense

It is a platform that gives analytic builders the tools to simplify complex data and provide insights to everyone.
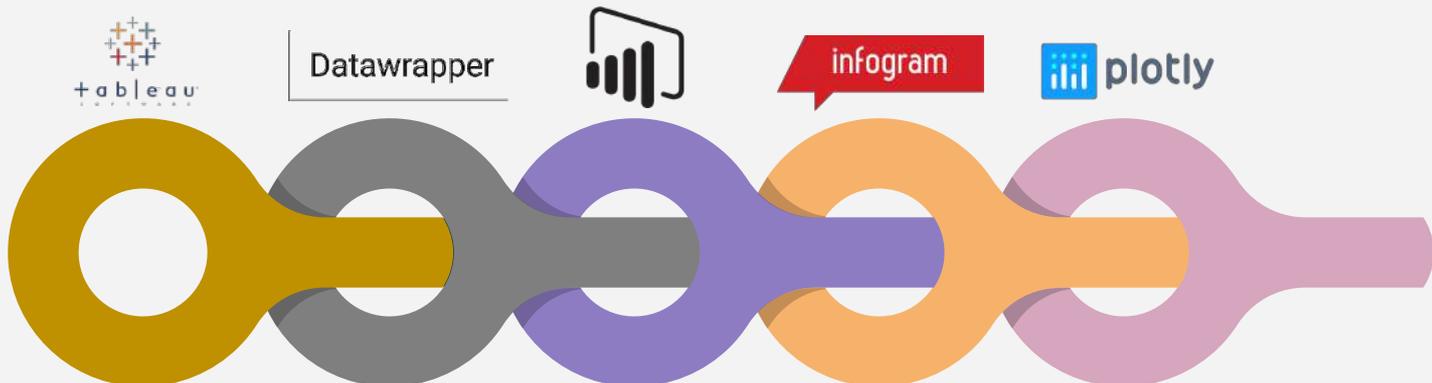
Tableau

You can create interactive and flexible dashboards making use of custom filters and its drag and drop functions.

Datawrapper

Is an easy to use tool for creating visualizations like infographics, maps, data tables and responsive charts like line, bar, stacked bar, donut, etc.

Microsoft Power BI

Is a business analytical tool which makes it easy for businessmen to visually analyze their data and form strategies based on it.

Infogram

It makes accessing data easy by letting you edit the data in the editor and connect to your desirable cloud service.

Plot.ly

Has a graphical user interface for importing and analyzing data into a grid and using stats tools.

Figure 10. Some Big Data tools. Source: Self made

INTRO 4.0

EDUCATION

Test your knowledge of Big Data tools in the cloud with this quiz:

Self-assessment tool:

https://searchcloudcomputing.techtarget
.com/quiz/Test-your-knowledge-of-big-d
ata-cloud-services

Degrees/Masters

❏ MSIT: BUSINESS INTELLIGENCE & DATA ANALYTICS (BIDA) - Carnegie Mellon University's Heinz College

❏ M.S. in Statistics: Data Science - Stanford University

❏ Big Data, Strategic Decisions: Analysis to Action - Stanford Graduate School of Business

❏ Master of Science in Data Science – ETH Zurich

MOOCs

❏ Big Data Analysis: Hive, Spark SQL, DataFrames and GraphFrames -Coursera

❏ Big Data Applications: Machine Learning at Scale -Coursera

❏ Managing Big Data with MySQL - Coursera

❏ Intro to Machine Learning - Udacity

INTRO 4.0

EDUCATION

External manuals & tutorials for more information

❏ Data Analytics Made Accessible, by A. Maheshwari

❏ Lean Analytics: Use Data to Build a Better Startup Faster, by A. Croll and B. Yoskovitz

❏ Big Data and Hadoop Tutorial - Intellipaat

❏ Introduction to Machine Learning

❏ Introduction to Data Science

❏ Big Data Quick Exploratory Self-Assessment Guide

Certifications

❏ Cloudera Certified Professional

❏ Intellipaat Big Data Hadoop Certification

❏ Microsoft's MCSE : Data Management and Analytics

❏ Hortonworks Hadoop Certification

INTRO 4.0

BIBLIOGRAPHY

❖ European Commission (2018). *Big Data*. Digital Single Market. Retrieved from
https://ec.europa.eu/digital-single-market/en/big-data

❖ European Commission (2014). *Worldwide Big Data Technology and Services - 2012-2015 Forecast*. Digital Single Market. Retrieved from
https://ec.europa.eu/digital-single-market/en/news/worldwide-big-data-technology-and-services-2012-2015-forecast

❖ European Commission (2013). *FACTSHEET: What is Big Data?*. Retrieved from
http://europa.eu/rapid/press-release_MEMO-13-965_en.htm

❖ Gaitho, M. (2018). *How Applications of Big Data Drive Industries*. Retrieved from https://www.simplilearn.com/big-data-applications-in-industries-article

❖ Simplilearn. (2018). *9 Must-have skills you need to become a Data Scientist, updated*. Retrieved from
https://www.kdnuggets.com/2018/05/simplilearn-9-must-have-skills-data-scientist.html

❖ Burtch, L. (2014). T*he Must-Have Skills You Need to Become a Data Scientist*. Burtch Works. Retrieved from
https://www.burtchworks.com/2014/11/17/must-have-skills-to-become-a-data-scientist/

❖ SAS. (2013). *Big Data Analytics An assessment of demand for labour and skills, 2012-2017*. E-skills UK. Retrieved from
https://ec.europa.eu/digital-single-market/en/news/big-data-analytics-assessment-demand-labour-and-skills-2012-2017

❖ *5 Practical Uses of Big Data*. (2017). Retrieved from
https://www.newgenapps.com/blog/5-practical-uses-of-big-data

❖ *Top 13 Best Big Data Companies of 2019.* (2019). Retrieved from
https://www.softwaretestinghelp.com/big-data-companies/

❖ Vesset, D., Morris, H.D., Little, G., Borovick, L., Feldman, S., Eastwood, M., ... Yezhkova, N. (2012).*Worldwide Big Data Technology and Services 2012 – 2015 Forecast*. Framingham, USA: IDC. Retrieved from
https://www.researchgate.net/profile/Shafagat_Mahmudova/post/How_can_big_data_analytics_and_AI_apply_to_risk_and_contingency_management/attachment/59d6525979197b80779aa96a/AS%3A511969745489920%401499074507185/download/Big_Data_Analytics_as_a_Service_for_Business_Intelligence1.pdf

INTRO 4.0

SELF APPRAISAL

★ After having read this text do I have a clear idea of what Big Data is?

★ What tools am I used to?

★ Do I know the benefits that Big Data can bring to my company?

★ Can I recognize the advantages and disadvantages of its implementation for my company?

INTRODUCTION TO THE INDUSTRIAL REVOLUTION 4.0