Big Data Skillset and Sources

Tushar B. Kute,
http://tusharkute.com

Analytical Skills

- Analytical skills are one of the most prominent Big Data Skills required to become the right expert in Big Data.
- To Understand the complex data, One should have useful mathematics and specific science skills in Big Data.
- Analytics tools in Big Data can help one to learn the analytical skills required to solve the problem in Big Data.

Data Visualization Skills

- An individual who wants to become a Big Data professional should work on their Data Visualization Skills.
- Data has to be adequately presented to convey the specific message. This makes visualization skills are essential in this area.
- One can start by learning the Data Visualization options in the Big Data Tools and software to improve their Data Visualization skills.
- It will also help them to increase their imagination and creativity, which is a handy skill in the Big Data field. The ability to interpret the data visually is a must for data professionals.

Data Visualization Skills

tusharkute.com

Business Domain and Big Data Tools

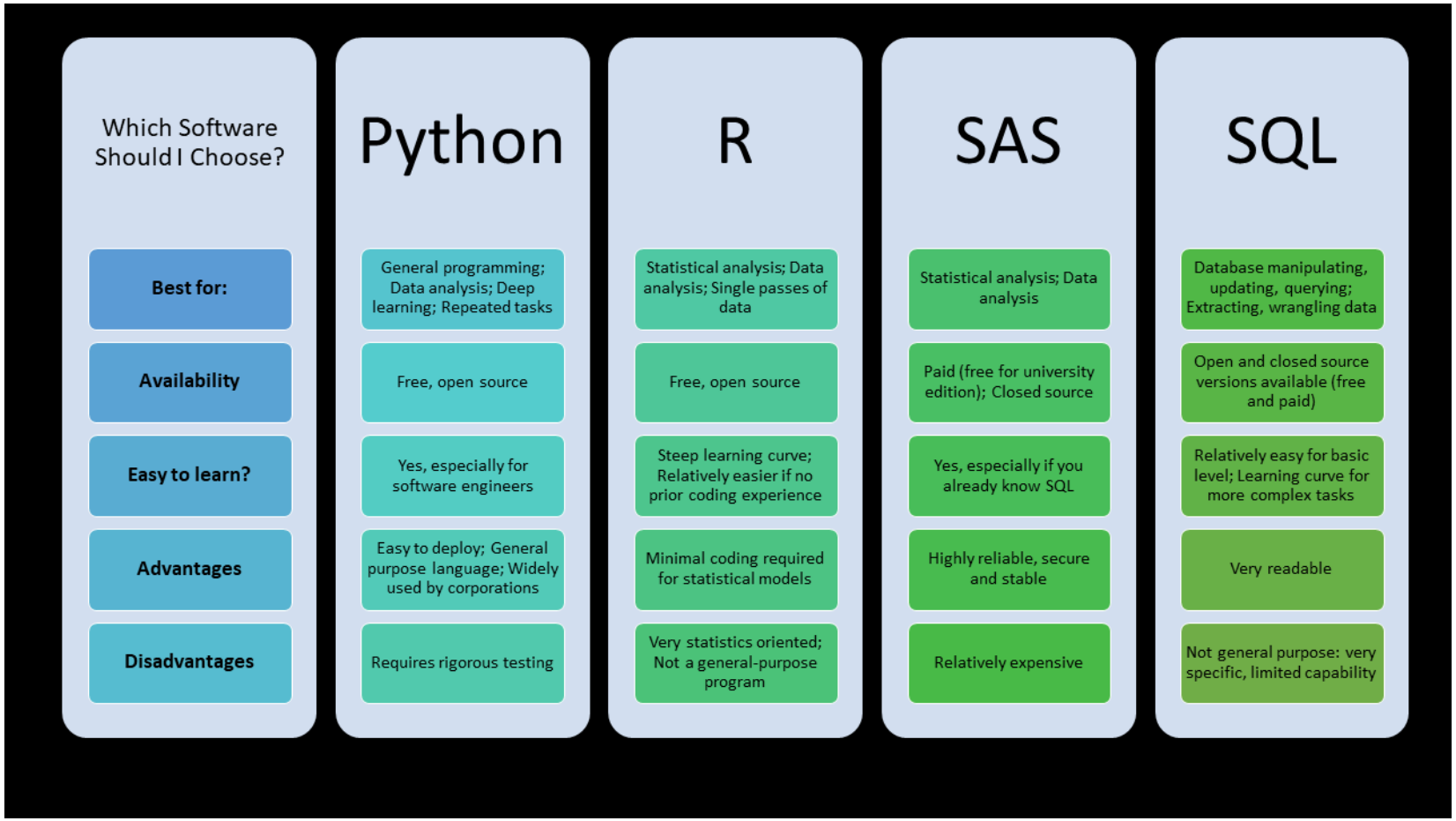
- Insights from massive datasets are derived and analyzed by using Big data tools.
- To understand the data in a better way by Big Data professionals, they will need to become more familiar with the business domain, especially with the business domain of the data they are working on.

Skills of Programming

- Having knowledge and expertise in Scala, C, Python, Java and many more programming languages are added advantages to Big Data Professional.
- There is a high demand for programmers who are experienced in Data analytics.
- To become an excellent Big Data Professional, one should also have good knowledge of fundamentals of Algorithms, Data Structures and Object-Oriented Languages.
- In Big Data Market, a professional should be able to conduct and code Quantitative and Statistical Analysis.

miTu skillologies
tusharkute.com

Skills of Programming



Skills of Programming

- One should also have a sound knowledge of mathematics and logical thinking.
- Big Data Professional should have familiarity with sorting of data types, algorithms and many more.
- Database skills are required to deal with a significantly massive volume of data.
- One will grow very far if they have an excellent technical and analytical perspective.

Problem Solving Skills

- The ability to solve a problem can go a long way in the field of Big Data. Big Data is considered to be a problem because of its unstructured data in nature.
- The one who has an interest in solving problems is the best person to work in this field of Big Data.
- Their creativity will help them to come out with a better solution to a problem. Knowledge and skills are only good up to a limit.
- Creativity and problem-solving skills are even more essential to become a competent professional in Big Data.

SQL

- In this era of Big Data, SQL work like a base. Structured Query Language is a data centred language.
- It will be beneficial for a programmer while working on Big data technologies such as NoSQL to know SQL.

SQL

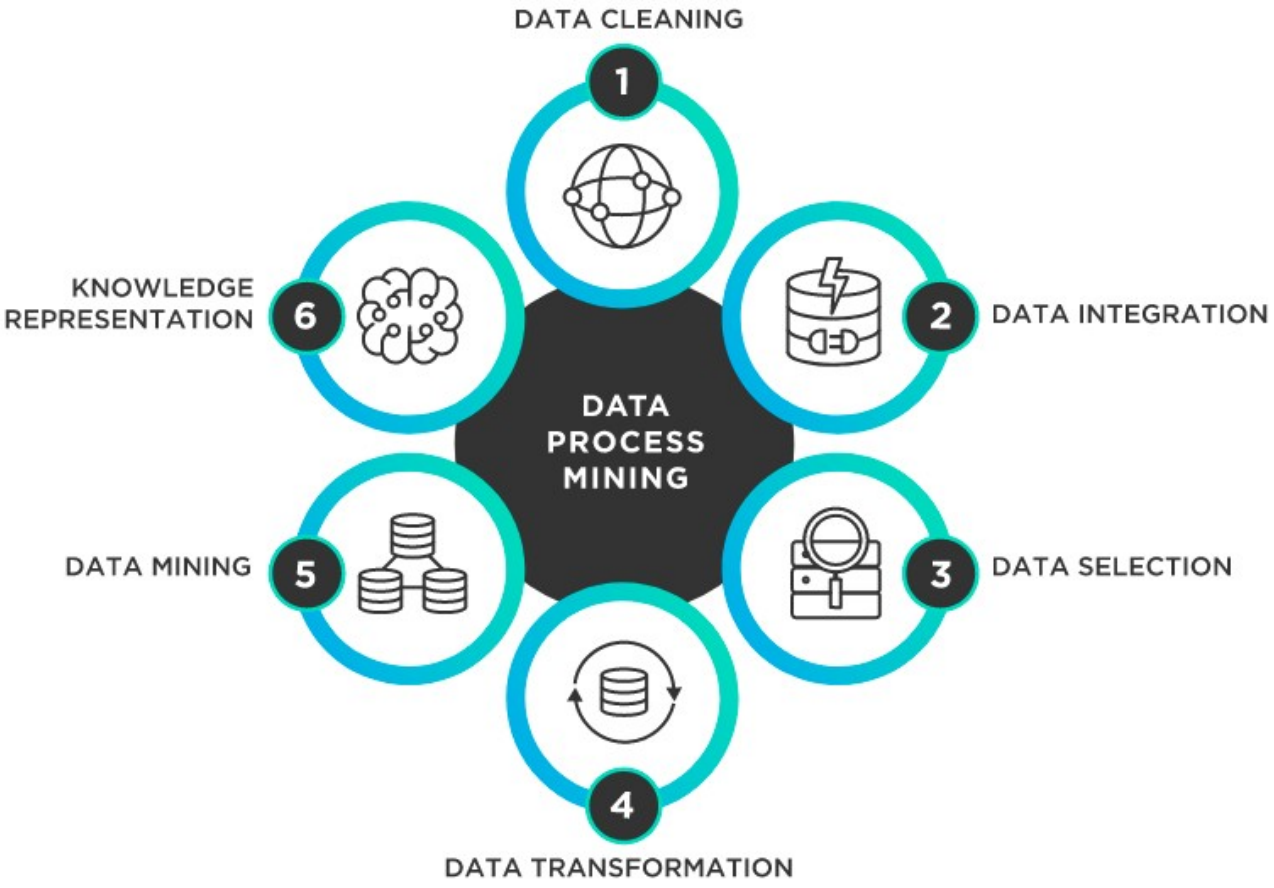
SQL	NOSQL
Relational Database management system	Distributed Database management system
Vertically Scalable	Horizontally Scalable
Fixed or predifined Schema	Dynamic Schema
Not suitable for hierarchical data storage	Best suitable for hierarchical data storage
Can be used for complex queries	Not good for complex queries

miTu skillologies
tusharkute.com

Data Mining

- Experienced Data mining professionals are in high demand.
- One should gain skills and experiences in technologies and tools of data mining to grow in their careers.
- Professionals should develop most-sought data mining skills by learning from top data mining tools such as KNIME, Apache Mahout, Rapid Miner and many more.

Data Mining


DATA CLEANING
1
2
DATA INTEGRATION
3
DATA SELECTION
4
DATA TRANSFORMATION
5
DATA MINING
6
KNOWLEDGE REPRESENTATION
DATA PROCESS MINING

Familiarity with Technologies

- Professionals of Big Data Field should be familiar with a range of technologies and tools that are used by the Big Data Industry.
- Big Data tools help in conducting research analysis and to conclude.
- It is always better to work with a maximum number of big data tools and technologies such as Scala, Hadoop, Linux, MatLab, R, SAS, SQL, Excel, SPSS and many more.
- There is a higher demand for professional have excellent skills and knowledge in programming and statistics.

tusharkute.com

Familiarity With Public Cloud and Hybrid Clouds

- Most Big Data teams will use a cloud set up to store data and ensure the high availability of Data.
- Organisations prefer cloud storage as it is cheaper to store large volumes of data when compared to building an in-house storage infrastructures.
- Many organizations even have a hybrid cloud implementation where in data can be stored in-house or on public cloud as per the requirements and organisation policies.

tusharkute.com

Familiarity With Public Cloud and Hybrid Clouds

- Some of the public clouds that one must know are Amazon Web Services (AWS), Microsoft Azure, Alibaba Cloud etc.
- The in-house cloud technologies include OpenStack, Vagrant, Openshift, Docker, Kubernetes etc.

tusharkute.com

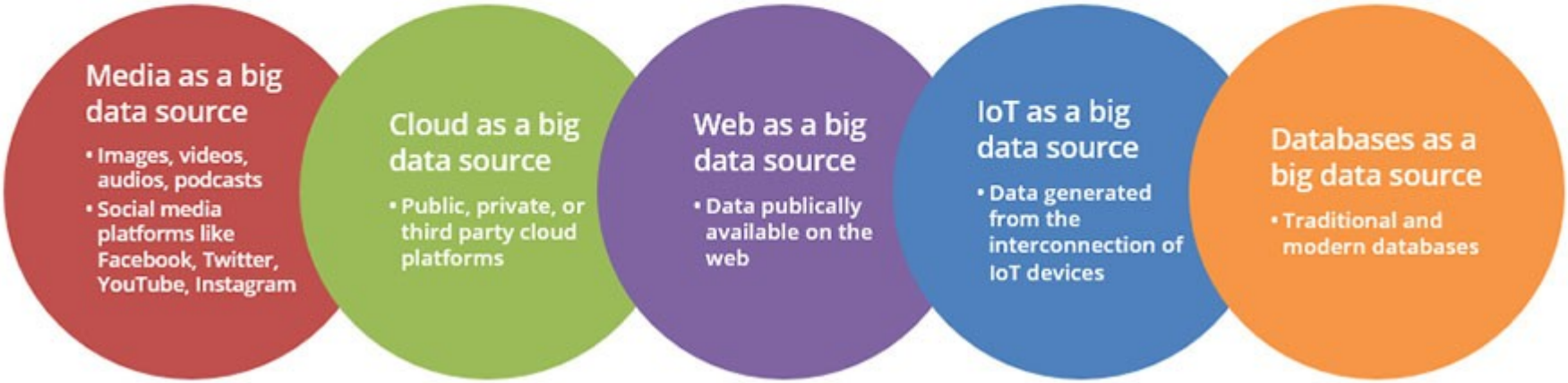
Skills from Hands-on experience

- An aspiring Big Data Professional should gain hands-on experience to learn the Big data tools.
- One can also go for short-term courses to learn the technology faster.
- If one has good knowledge about newer technologies, then it will help them in understanding the data better by using modern tools.
- Their interaction with the data will improve give them an edge over the others by bringing out better results.

Big Data Sources

- Data, as we know, is massive and exists in various forms. If it is not classified or sourced well, it can end up wasting precious time and resources.
- In order to achieve success with big data, it is important that companies have the know-how to sift between the various data sources available and accordingly classify its usability and relevance.

Big Data Sources



Media

- Media is the most popular source of big data, as it provides valuable insights on consumer preferences and changing trends.
- Since it is self-broadcasted and crosses all physical and demographical barriers, it is the fastest way for businesses to get an in-depth overview of their target audience, draw patterns and conclusions, and enhance their decision-making.
- Media includes social media and interactive platforms, like Google, Facebook, Twitter, YouTube, Instagram, as well as generic media like images, videos, audios, and podcasts that provide quantitative and qualitative insights on every aspect of user interaction.

Cloud

- Today, companies have moved ahead of traditional data sources by shifting their data on the cloud.
- Cloud storage accommodates structured and unstructured data and provides business with real-time information and on-demand insights.
- The main attribute of cloud computing is its flexibility and scalability.
- As big data can be stored and sourced on public or private clouds, via networks and servers, cloud makes for an efficient and economical data source.

IOT

- Machine-generated content or data created from IoT constitute a valuable source of big data. This data is usually generated from the sensors that are connected to electronic devices.
- The sourcing capacity depends on the ability of the sensors to provide real-time accurate information.
- IoT is now gaining momentum and includes big data generated, not only from computers and smartphones, but also possibly from every device that can emit data.
- With IoT, data can now be sourced from medical devices, vehicular processes, video games, meters, cameras, household appliances, and the like.

miTu skillologies
tusharkute.com

Web

- The public web constitutes big data that is widespread and easily accessible. Data on the Web or ‘Internet’ is commonly available to individuals and companies alike.
- Moreover, web services such as Wikipedia provide free and quick informational insights to everyone.
- The enormity of the Web ensures for its diverse usability and is especially beneficial to start-ups and SME’s, as they don’t have to wait to develop their own big data infrastructure and repositories before they can leverage big data.

Databases

- Businesses today prefer to use an amalgamation of traditional and modern databases to acquire relevant big data.
- This integration paves the way for a hybrid data model and requires low investment and IT infrastructural costs.
- Furthermore, these databases are deployed for several business intelligence purposes as well. These databases can then provide for the extraction of insights that are used to drive business profits.
- Popular databases include a variety of data sources, such as MS Access, DB2, Oracle, SQL, and Amazon Simple, among others.

Summary

- The process of extracting and analyzing data amongst extensive big data sources is a complex process and can be frustrating and time-consuming.
- These complications can be resolved if organizations encompass all the necessary considerations of big data, take into account relevant data sources, and deploy them in a manner which is well tuned to their organizational goals.

Thank you

This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com