**Instructor**

Dr. Ayaz ul Hassan Khan

|  |  |
| --- | --- |
| **Group Members** |  |
|  |  |
| Jibran Rasheed Khan | IS-010/2018 |
| Muhammad Waqar | IS-011/2018 |
| Hafiz Idrees Riaz | IS-025/2018 |

Distributed Systems (CT-509)

Project proposal

**Mega Tool-Set Comparison and Data Science Framework**

**Abstract**

Data Science is a scientific technique to unfold the hidden mystery of data and extract the valuable insight that can boom the business needs, beware of what probably be happened in near future and most importantly make human to act that react on things. Nowadays, data is the key element that use to classify and aids in to foretell each and every aspect of human; behavior, nature, standards, lifestyle, passion, desire, business needs and lot more uncountable. But, data is not so simple nor resides in any specified form. This usually be available in any of 4Vs; Volume, Velocity, Verity and Veracity. Due to complex nature of data, many organization, scientist and practitioners suffer trouble to incorporate them in a suitable platform, desirable tools, libraries and other supportive algorithms or calculations. Also, data science itself not a separate field of study, but it’s a composition of several incorporating fields – Information Technology, Statistics and Science. The core propose of this study presents the generic platform model for data science practitioner and professionals. In this work, we will explore each aspect and compositor of data Science, drill down to identify individual field hierarchy, tools, and platform available. Furthermore, it covers the concise comparison study renewed and most preferable element in each filed. The study, will beneficial in building a generic platform and composite tool that will ease the practitioners to learn more, building better and flexible system and more compatible integrating features.

# **Objectives**

To list down the technology, algorithms, tools and their comparison for

1. Big data
2. Data Analytics
3. Algorithms/Methods
4. Visualization

# Targeting Conference

### **DataEngConf’19**

DataEngConf focuses on technical aspects of data science. It aims to remove the gap between data scientists, data engineers and data analysts. Main focus of the conference will be on topics related to the real world application of data science. Conference will also discuss data pipeline architectures. Among the panel of speakers would be data scientists and engineers from top companies such as Facebook, Lyft, Airbnb and Instacart among others. Attendees will include data scientists and engineers from from a variety of different fields and startups including media, finance and technology.

|  |  |  |
| --- | --- | --- |
| Location | San Francisco |  |
| Date | April 17-18, 2019 |  |

<https://www.dataengconf.com/>

# **Important URLs:**

https://www.ngdata.com/top-tools-for-data-scientists/

# **References**

If anyone would like to add online PDF/source file links add here. Otherwise shared via following google drive

<https://drive.google.com/drive/folders/1kN4d49MhpUI217pM57NGI8N50xwKLpC6?usp=sharing>

Following paper uploaded to google drive.

[1] I. Yaqoob *et al.*, “International Journal of Information Management Big data : From beginning to future,” *Int. J. Inf. Manage.*, vol. 36, no. 6, pp. 1231–1247, 2016.

[2] B. M. S. Thillaieswari, M. Phil, and B. Ed, “Comparative Study on Tools and Techniques of Big Data Analysis,” *Int. J. Adv. Netw. Appl.*, vol. 66, no. 61, pp. 61–66, 2017.

[3] D. Singh and C. K. Reddy, “A survey on platforms for big data analytics,” *J. Big Data 2014,* vol. 1, no. 8, pp. 1–20, 2014.

[4] A. Siddiqa, I. A. Targiohashem, I. Yaqoob, M. Marjani, A. Gani, and F. Nasaruddin, “A Survey of Big Data Management: Taxonomy and State-of-the-Art,” *J. Netw. Comput. Appl.*, vol. 71, no. August, pp. 151–166, 2016.

[5] T. L. Coelho, R. P. Magalh, and D. Ara, “Big Data Analytics Technologies and Platforms : a brief review,” in *CEUR Workshop Proceedings*, 2018, pp. 25–32.

[6] A. Oussous, F. Benjelloun, A. A. Lahcen, and S. Belfkih, “Big Data Technologies : A Survey,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 30, no. 4, pp. 431–448, 2018.

[7] R. Kune, P. K. Konugurthi, and A. Agarwal, “The anatomy of big data computing,” *Softw. Pract. Exp.*, vol. 46, no. October 2015, pp. 79–105, 2016.

[8] J. Williamson, *Getting a Big Data Job For Dummies*. John Wiley & Sons, Inc., 2015.

[9] F. Tekiner and J. A. Keane, “Big Data Framework,” in *IEEE International Conference on Systems, Man, and Cybernetics*, 2013, pp. 1494–1499.

[10] C. W. Tsai, C. F. Lai, H. C. Chao, and A. V Vasilakos, “Big data analytics : a survey,” *J. Big Data*, vol. 2, no. 21, pp. 1–32, 2015.

[11] C. L. P. Chen and C. Zhang, “Data-intensive applications , challenges , techniques and technologies : A survey on Big Data,” *Inf. Sci. (Ny).*, vol. 275, no. August, pp. 314–347, 2014.

[12] P. Pääkkönen and D. Pakkala, “Reference Architecture and Classification of Technologies , Products and Services for Big Data Systems,” *Big Data Res.*, vol. 2, no. 4, pp. 166–186, 2015.

[13] A. Gandomi and M. Haider, “Beyond the hype : Big data concepts , methods , and analytics,” *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, 2015.