**Prepared By**

Jibran Rasheed Khan

jibran\_rasheed@hotmail.com

**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

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 supportives. 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.