Research Data Centre to Improve Learning and Research in Higher Education Institutions

P. Bavithra Matharasi
Dept. of MCA Mount Carmel College Bengaluru

ABSTRACT: Data centre is the heart for any functionality of an organization or Institution. As the data has become big data, a data centre will require huge storages, racks of servers and networks. When the cost overruns or maintenance hitches, cloud is the saviour. In this cloud era, implementing a Research Data Centre (RDC) in Mount Carmel College (MCC) is unique and need of the hour. Though there are rich source of dataset, tools and techniques to access the dataset that are available in internet, accessing them all under a single umbrella is not possible. An initiative is taken to establish a RDC through research department at MCC, keeping the goal of providing dataset that caters all domains. This will include both public and private dataset. These datasets will be used to obtain rich knowledge and wisdom through projects and research. Many industry tie-up will help to grow data and research. RDC will be a hub to ensure high quality service to academics and researchers, providing an interconnection facility for global research network in near future.

KEYWORDS: research dataset, private data, public data

I.   INTRODUCTION

The quantity of the data in the world is growing exponentially. Number of resources also grow faster in generating dataset. With the invention of web and mobile, the world itself has gone online, digitally interconnected. This factor has increased the growth of data rapidly and rigorously. Thus the term 'data' takes a new avatar 'Big Data'. The major sources of Big Data are social media sites, eCommerce, digital images, digital videos, data from cell phones,   web logs, medical records, military surveillance, complex scientific research and so on. All these data amountsto around some Quintillion bytes of data. By 2020, this voluminous data will be increased to 40 Zettabytes. The data is big data for which, special tools, techniques and procedures are used for analysing and manipulating the data. Big data is rich with information, from which one can infer knowledge and wisdom by analysing and manipulating.

It is a common practice to have a data centre in any organization [2]. Data Centres are the pivotal infrastructure of any organization of any size. It is a form of silos that offers value-added service to the resources for processing and storing data. Universities and Colleges are not less in usage and generation of data. Having a data centre in Institutions are not uncommon but started few decades. Owning and running a data center is a costly, cumbersome, and time-consuming process [3]. The biggest expenditures of data centers are not the one-time purchases of hardware and software. The fact is that the, building an infrastructure is the smallest expenditures.

The total required investment is evident only when the data begins to grow adding up the recurring costs. This recurring cost includes server hardware and refreshes, networking, software, bandwidth, power, cooling, disaster recovery, security and staffing are just a few of the ongoing and recurring costs. The solution for academia when it comes to implement a data centers, is the same as it is for nearly every organization is "the cloud". Cloud could be picturized as "infrastructure on demand," meaning that paying only based on usage [4]. It is a trend and worthy for any organization to migrate towards cloud for such data centres. Organizations is benefit by both owning and hiring a cloud.

Further, the growth of higher education is increased both in the aspect of quality and quantity. There are number of institutions, colleges and universities opened. These institutions provide the best curriculum in a state of art infrastructure with highly qualified teachers. The growth is tremendous that they train and launch industry-ready professionals to the market. Similarly, the growth of industry is also increased both in quantity and quality. Though there is a tremendous growth, there is always a void in the need of industry-ready professionals for the industry. A

broad gap is created between the academia and industry. This is due to the facts that
1. The requirement of an industry is vast and unique
2. Students are not trained for any specific industry

Instituting a dedicated data centre for only research and project is a new of its kind available today. The data centre will be called as Research Data Center (RDC) that is used only for research and learning. It is a dedicated data centre with state-of-art infrastructure and big data of dataset. This new RDC will support digital learning through projects, improve research curve and outreach society. This will be a bridge in collaborating with number of industries and other academic institutions. This will provide a platform for students to give projects and internship. This will be a hub for academic and research units providing connectivity across the world.

1.1.1 Categories of data
This era is also known as 'era of data' as data is freely and voluminously available anywhere. anytime. Use of latest gadgets and social media has tremendously changed the way of human lifestyle. Researchers have predicted that 40 Zettabytes (40,000 Exabytes) will be generated by 2020, which is an increase of 300 times from 2005 [1]. The data centre in any University, College or Institution will include both the data created from scratch and data generated. The created data from the scratch will be the data pertaining to that particular institution like mark details, study materials or stakeholder information. The data generated could be statistical data, experimental results, observational data or interview recordings. The former data that applies to a particular institute is confidential and private data.
The data available in any data centre could be categorized into two broad classifications.
1.   Public Data   2. Private Data

- **Public Data**
Public data are readily available at free of cost in the internet. There are number of organizations providing their dataset freely available for research purpose. There are number of websites are available from where we can download those datasets.

- **Private Data**
Private dataset are confidential, secured and private data, belonging to that particular organization only. The private data will be created and curated by the organization itself. These data

are more secured. These private datasets need special permission to access from the data generator or creator. These private datasets could be acquired with agreement and licensing.

II. OBJECTIVES
1.   Establish Research Data Centre with state-of-art infrastructure
2.   Create private and public dataset for research and projects
3.   MoUs or tie-up with industries to improve learning, research and job opportunities
4.   Create a platform for Students and Staff to pursue research and projects
5.   Improve the learning and research curve

III. LITERATURE SURVEY
Higher education has made a lot of significance in shaping the world by the outcome from research. Educators engaged in research are updated with latest information and knowledge. When the students are expected to involve in research, Teachers should be involved in research to come out with latest and original information must have familiarity with various aspects to the concept. Students when engaged in research-based learning are bound to learn more and better than they would without the integration of research. This highlights the importance of hiring professionals from industry who are familiar with research-based practise. To integrate the practice of research it is important to redesign the curriculum aligned to the practice of research [5].

Higher education can be seen as a focal point of knowledge and its application, an institution which makes a great contribution to the economic growth and development through fostering innovation and increasing higher skills. Higher education is broadly defined as one of key drivers of growth performance, prosperity and competitiveness. UNESCO says its social role provides the link between the intellectual and educational role of universities on one hand and the development of society on the other [6]. Universities and higher education intuitions are keen about research and carrying complex and essential work [7]. For any research, data vital. A research is data centric. Research data are a valuable resource that often requires a great deal of time and money to create. Research data should be managed in an appropriate and timely manner. Research data should possess few characteristics. They should ensure research integrity and validation of results. They should increase research efficiency. They should facilitate data security and minimize the risk of data loss. They should ensure

wider dissemination and increased impact. They should enable research continuity through secondary data use [8].

There are so many tools that are available for data cleaning, convert structured data to unstructured data, building a model with the data and visualizing the data. To name few OpenRefine, WolframAlpha, Impoer.ioPlot.ly and BigML [9]. Research data management covers the planning, creating, storing, organising, accessing, sharing, describing, publishing and curating of data [10]. There are some of the factors to be considered while effectively managing the research data. The impact of research could be increased by maximizing the visibility of the data and promoting transparency in research. The accessibility of research work could be improved by ensuring the quality and integrity of the data is maintained during and beyond the life of the research project. research data could be prevented from unauthorized use by addressing privacy and confidentiality issues throughout and beyond the research project. research data should be safeguarded by establishing appropriate storage, back-up and management.
The idea of establishing a data centre takes new dimensions and more innovations are added in all aspects. Microsoft has launched a watertight data centre, comprising of 864 servers across 12 racks, under the ocean [11].

IV. METHODOLOGY
In order to develop a project, choosing a project development methodology is most important. A project has a goal to achieve and plans to work within a defined schedule and budget. A methodology helps at every stage of project from initial start-up to implementation. According to the Project Management Institute (PMI), a methodology is defined as 'a system of practices, techniques, procedures, and rules used by those who work in a discipline. There are ample number of methodologies available. Selecting the most suitable project management methodology could be a tricky task. Different project management methodologies have their own advantages and disadvantages for different types of project. when choosing an appropriate methodology, there are number of factors to be considered. This project uses an dynamic an iterative methodology called 'Agile'. This methodology tries to provide rapid, continuous delivery of products to the stakeholder.

4.1. Agile Project Development Methodology
Agile is an iterative approach to develop any project. This agile project management helps teams deliver value to their stakeholder faster and with lesser nuisances. Instead of delivering everything on a single final launch, an agile team delivers work in small, but consumable, increments. It is a methodology to build products incrementally using iterations of 1 to 4 weeks, so that the development process is aligned with the stakeholder's need. Instead of a single-pass development of 6 to 18 months where all the requirements and risks are predicted upfront, this methodology adopts a process of frequent feedback where a workable product is delivered after 1 to 4 week iteration [12]. The time limit for each iteration can be set by the team according to the need of the project development.

This project of establishing RDC also works on the Agile project development methodology. Since it requires data growth, the development process is monitored, delivered and discussed with the stakeholders frequently. This is a systematic and structured process that enables to improve practices and decisions gradually. We can learn from the outcomes of the decisions that were made during prior stages of the project and from previous projects. It is all about adapting to the needs of the stakeholder so that we can ultimately add value with project deliverables.

Figure 1 Agile Project Development Methodology

V. ARCHITECTURE OF RDC
RDC will be the central hub for creation and curating data for research or project. The data center infrastructure is central to the IT architecture, from which all content is sourced or passes through. Proper planning of the data center infrastructure design is critical, and performance, resiliency, and scalability need to be carefully considered [13]. Another important aspect of the data center design is flexibility in quickly deploying and supporting new services.

Designing a flexible architecture that has the ability to support new applications in a short time frame can result in a significant competitive advantage. Such a design requires solid initial planning and thoughtful consideration in the areas of tools and techniques to be used [14]. Since this is a first of its kind, which is to cater the need of various departments, the infrastructure is incorporated into the already built strong network of the college.
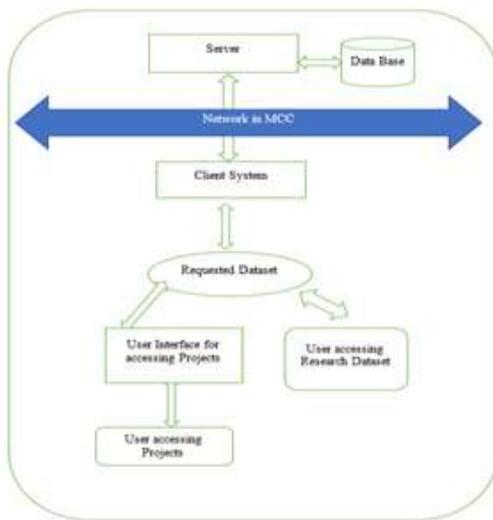

Figure 2 Architecture of RDC

VI. BASIC SERVICES IN THE RESEARCH DATA CENTER

The RDC will be a lighthouse in two aspects, namely Projects and Research. Data could be created, utilized and generated through projects and research. But the usage and generation of dataset will be different in projects and research.

6.1 Projects

Students and Staff can involve in project work and access dataset to improve their learning curve. Students also has the proximity to get internships and placements through RDC. A web portal is available through which the students and staff can register for project works. To register in this project portal, first a login should be acquired by registering to the Innovation cell manually. Once the login is obtained, students and staff can login and choose among the list of projects offered by list of companies.

6.2 Research

Staff and Students can undergo research work on any topic. The researchers should manually register to the Innovation cell and request for the dataset. Once the manual registration is over, a login id will be sent the respective researchers email and a space will be provided in the server. The researchers can login with this id and access both private dataset and public dataset.

VII. FEATURES OF RDC

Research is the heart of any higher education institution.

7.1 Consultation Service:RDC will provide advises and consultation to both Staff and Students in how best could dataset be used or to choose right tools and techniques for either research or project. Support is provided for all skill levels and for all areas of research within the College. The data set will be available and caters the need of every department in the college.

7.2 Best Training: The stakeholders will not be refrained after choosing research or project. They are offered best training from industry and bundled with latest materials for learning.

7.3 Improve Learning: Learning through research and project will improve the learning ability of any individual.

7.4 Spur Research: Whether a simple analysis or a discovery of a new component, dataset is vital for the research. RDC will readily offer created and curated dataset for the purpose. Research could be taken in any field of interest.

7.5 Industry Linkage: The main goal of this RDC is shaking hands with many industries.

7.6 Create Jobs: Since RDC has linkage with more industries, job opportunities will grow and both Students and Industry will get benefits.

VIII. CONCLUSION

RDC is the need of the hour. It is implemented in the right time. With the right tools, techniques and infrastructure to deliver leading technology, more importantly to students and researchers, RDC will yield better success rate. The RDC will offer Students and Staff with plethora of opportunity in projects and research.

IX. ACKNOWLEDGEMENT

to establish a data centre for the purpose of research and projects.

REFERENCES

[1]. https://www.edureka.co/blog/big-data-tutorial "Big Data Tutorial: All You Need To Know About Big Data!", Awanish, Published on Dec 24,2018

[2]. https://msutoday.msu.edu/news/2018/msu-announces-new-it-strategy-and-opening-of-new-data-center/, "MSU announces new IT strategy and opening of new data center", Emily Guerrant, Jessi AdlerPublished: April 12, 2018

[3]. https://www.networkworld.com/article/3223692/data-center/what-is-a-data-centerhow-its-changed-and-what-you-need-to-know.html, "How a data center works [and how they're changing]",Zeus Kerravala, Network World, SEP 25, 2017

[4]. https://msft.engineer/why-academia-shouldnt-build-data-centers-eb8486519a9ccial Media, "Why Academia Shouldn't Build Data Centers", Brad Groux, May 16, 2017

[5]. http://edtechreview.in/trends-insights/insights/2741-research-in-higher-education," Role and Need of Research in Higher Education", Priyanka Gupta, 14 April 2017

[6]. https://qswownews.com/importance-of-higher-education/, "Importance of higher education for today's economy", QS Asia News Network, March 23, 2018

[7]. https://www.kaplanpathways.com/about/news/university-research-important-students/, "Why is university research important for students?", BarniHomden, 09 March 2017

[8]. http://libguides.ucd.ie/data/why_manage

[9]. https://www.import.io/post/what-is-data-and-why-is-it-important/, "What is data, and why is it important?", June 28, 2018

[10]. https://libguides.library.curtin.edu.au/research-data-management, "Research data management: Introduction", Oct 16, 2018

[11]. https://www.wired.co.uk/article/microsoft-data-centre-orkney-sinks-project-natick

[12]. https://tutorialspoint.com/management_concepts/agile_project_management.htm

[13]. https://cisco.com/c/en/us/td/docs/solutions/Enterprise/Data_Center/DC_Infra2_5/DCInfra_1.html

[14]. https://www.myvirtualjourney.com/data-center-architecture-an-overview/,"Data Center Architecture: An Overview", Nisar Ahmad, November 25, 2017