Technofist

BIG DATA PROJECT LIST 2018 -2019

At work as usual: 080-40969981 | Write to me: technofist.projects@gmail.com, |
when u need us the most: **+91-9008001602, 080-40969981**| On the
Web:www.technofist.com www.technofist.in , www.itcdp.in

TECHNOFIST a leading student's project solution providing company established in Bangalore since 2007. With perfect infrastructure, lab set up, Work shop, Expertise faculties make us competitive service providers.

Here is the list of project titles 2018 and 2019.

DOORS OF TECHNOLOGY:

- EMBEDDED SYSTEMS

- MICROCONTROLLERS / ARM /PIC / AVR

- WIRELESS TECHNOLOGIES

- ROBOTICS

- ARDUINO

- GSM & GPS/ ZIGBEE

- MATLAB / VLSI

- IEEE PROJECTS ON JAVA / DOT NET

- INTERNET OF THINGS

- ANDROID BASED PROJECTS

- PHP

- AND COMPLETE MECHANICAL FABRICATIONS

- MECHANICAL DESIGN AND ANALYSIS

Projects are available for all branches of **ENGINEERING, DIPLOMA, MCA/BCA, and MSc/BSc.**

CORPORATEOFFICE:

Technofist

TECHNOFIST
YES Complex,
#19/3&4, 2nd Floor,
Dinnur Main Road, R T Nagar Post
Opposite to Navodaya Vidyaneketan School
Bangalore – 32

Here we provided a **latest Big Data 2018 project lis**t with abstracts. we do train a student from basic level of software which includes basic java Classes, projects implementation, final project demo and final code explanations. If you have questions regarding these projects feel free to contact us. You may also ask for abstract of a project idea that you have or want to work on. The **own projects idea** for diploma and Engineering students can also encouraged here.

IEEE BIG DATA PROJECT LIST 2018 AND 2019

2018 – 19 IEEE PROJECT TITLES ON BIG DATA
TEB001
TITLE - SOCIALQ&A: AN ONLINE SOCIAL NETWORK BASED QUESTION AND ANSWER SYSTEM **ABSTRACT** - Question and Answer (Q&A) systems play a vital role in our daily life for information and knowledge sharing. Users post questions and pick questions to answer in the system. Due to the rapidly growing user population and the number of questions, it is unlikely for a user to stumble upon a question by chance that (s) he can answer. Also, altruism does not encourage all users to provide answers, not to mention high quality answers with a short answer wait

	time. The primary objective of this paper is to improve the performance of Q&A systems by actively forwarding questions to users who are capable and willing to answer the questions. To this end, we have designed and implemented SocialQ&A, an online social network based Q&A system.
TEB002	**TITLE** - **PRIVACY-PRESERVING DATA ENCRYPTION STRATEGY FOR BIG DATA IN MOBILE CLOUD COMPUTING** **ABSTRACT** - Privacy has become a considerable issue when the applications of big data are dramatically growing in cloud computing. The benefits of the implementation for these emerging technologies have improved or changed service models and improve application performances in various perspectives. The execution time of the data encryption is one of the serious issues during the data processing and transmissions. Many current applications abandon data encryptions in order to reach an adoptive performance level companioning with privacy concerns. In this paper, we concentrate on privacy and propose a novel data encryption approach, which is called Dynamic Data Encryption Strategy (D2ES).
TEB003	**TITLE** - **EFFICIENT PROCESSING OF SKYLINE QUERIES USING MAPREDUCE** **ABSTRACT** - The skyline operator has attracted considerable attention recently due to its broad applications. However, computing a skyline is challenging today since we have to deal with big data. For data-intensive applications, the MapReduce framework has been widely used recently. In this paper, we propose the efficient parallel algorithm SKY-MR+ for processing skyline queries using MapReduce. We first build a quadtree-based histogram for space partitioning by deciding whether to split each leaf node judiciously based on the benefit of splitting in terms of the estimated execution time. In addition, we apply the dominance power filtering method to effectively prune non-skyline points in advance.

At work as usual: 080-40969981 | Write to me: technofist.projects@gmail.com, |
when u need us the most: **+91-9008001602, 080-40969981**| On the
Web:www.technofist.com www.technofist.in , www.itcdp.in

TEB004	**TITLE - FIDOOP-DP: DATA PARTITIONING IN FREQUENT ITEMSET MINING ON HADOOP CLUSTERS** **ABSTRACT** - Traditional parallel algorithms for mining frequent itemsets aim to balance load by equally partitioning data among a group of computing nodes. We start this study by discovering a serious performance problem of the existing parallel Frequent Itemset Mining algorithms. Given a large dataset, data partitioning strategies in the existing solutions suffer high communication and mining overhead induced by redundant transactions transmitted among computing nodes. We address this problem by developing a data partitioning approach called FiDoop-DP using the MapReduce programming model. The overarching goal of FiDoop-DP is to boost the performance of parallel Frequent Itemset Mining on Hadoop clusters.
TEB005	**TITLE - USER-CENTRIC SIMILARITY SEARCH** **ABSTRACT** - User preferences play a significant role in market analysis. In the database literature there has been extensive work on query primitives, such as the well known top-k query that can be used for the ranking of products based on the preferences customers have expressed. Still, the fundamental operation that evaluates the similarity between products is typically done ignoring these preferences. Instead products are depicted in a feature space based on their attributes and similarity is computed via traditional distance metrics on that space. In this work we utilize the rankings of the products based on the opinions of their customers in order to map the products in a user-centric space where similarity calculations are performed.
TEB006	**TITLE - PRACTICAL PRIVACY-PRESERVING MAPREDUCE BASED K-MEANS CLUSTERING OVER LARGE-SCALE DATASET** **ABSTRACT** - Clustering techniques have been widely adopted in many real world data analysis applications, such as customer behavior analysis, medical data Analysis, digital forensics, etc. With the explosion of data in today's big data era, a major trend to handle a clustering over large-scale datasets is outsourcing it to HDFS platforms. This is because cloud computing offers not only reliable services with performance guarantees, but also savings on in-house IT infrastructures. However, as datasets used for clustering may contain sensitive information, e.g., patient health information, commercial data, and behavioral data, etc, directly outsourcing them to any Distributed servers inevitably raise privacy concerns.

TEB007	**TITLE - SECURE BIG DATA STORAGE AND SHARING SCHEME FOR CLOUD TENANTS** **ABSTRACT** - The Cloud is increasingly being used to store and process big data for its tenants and classical security mechanisms using encryption are neither sufficiently efficient nor suited to the task of protecting big data in the Cloud. In this paper, we present an alternative approach which divides big data into sequenced parts and stores them among multiple Cloud storage service providers. Instead of protecting the big data itself, the proposed scheme protects the mapping of the various data elements to each provider using a trapdoor function.
TEB008	**TITLE - SENTIMENT ANALYSIS OF TOP COLLEGES USING TWITTER DATA** **ABSTRACT** - In today's world, opinions and reviews accessible to us are one of the most critical factors in formulating our views and influencing the success of a brand, product or service. With the advent and growth of social media in the world, stakeholders often take to expressing their opinions on popular social media, namely twitter. While Twitter data is extremely informative, it presents a challenge for analysis because of its humongous and disorganized nature. This paper is a thorough effort to dive into the novel domain of performing sentiment analysis of people's opinions regarding top colleges in India. Besides taking additional preprocessing measures like the expansion of net lingo and removal of duplicate tweets
TEB009	**TITLE - ON TRAFFIC-AWARE PARTITION AND AGGREGATION IN MAPREDUCE FOR BIG DATA APPLICATIONS** **ABSTRACT** - The MapReduce programming model simplifies large-scale data processing on commodity cluster by exploiting parallel map tasks and reduce tasks. Although many efforts have been made to improve the performance of MapReduce jobs, they ignore the network traffic generated in the shuffle phase, which plays a critical role in performance enhancement. Traditionally, a hash function is used to partition intermediate data among reduce tasks, which, however, is not traffic-efficient because network topology and data size associated with each key are not taken into consideration. In this paper, we study to reduce network traffic cost for a MapReduce job by designing a novel intermediate data partition scheme.
TEB010	

	TITLE - A PARALLEL PATIENT TREATMENT TIME PREDICTION ALGORITHM AND ITS APPLICATIONS IN HOSPITAL QUEUING-RECOMMENDATION IN A BIG DATA ENVIRONMENT **ABSTRACT** - Effective patient queue management to minimize patient wait delays and patient overcrowding is one of the major challenges faced by hospitals. Unnecessary and annoying waits for long periods result in substantial human resource and time wastage and increase the frustration endured by patients. For each patient in the queue, the total treatment time of all the patients before him is the time that he must wait. It would be convenient and preferable if the patients could receive the most efficient treatment plan and know the predicted waiting time through a mobile application that updates in real time. Therefore, we propose a Patient Treatment Time Prediction (PTTP) algorithm to predict the waiting time for each treatment task for a patient.
TEB011	**TITLE - PROTECTION OF BIG DATA PRIVACY ENCRYPTED CLOUD DATA** **ABSTRACT** - In recent years, big data have become a hot research topic. The increasing amount of big data also increases the chance of breaching the privacy of individuals. Since big data require high computational power and large storage, distributed systems are used. As multiple parties are involved in these systems, the risk of privacy violation is increased. There have been a number of privacy-preserving mechanisms developed for privacy protection at different stages (e.g., data generation, data storage, and data processing) of a big data life cycle. The goal of this paper is to provide a comprehensive overview of the privacy preservation mechanisms in big data and present the challenges for existing mechanisms.
TEB012	**TITLE - NETSPAM: A NETWORK-BASED SPAM DETECTION FRAMEWORK FOR REVIEWS IN ONLINE SOCIAL MEDIA** **ABSTRACT** - Nowadays, a big part of people rely on available con-tent in social media in their decisions (e.g. reviews and feedback on a topic or product). The possibility that anybody can leave a review provide a golden opportunity for spammers to write spam reviews about products and services for different interests. Identifying these spammers and the spam content is a hot topic of research and although a considerable number of studies have been done recently toward this end, but so far the methodologies put forth still barely detect spam reviews, and none of them show the importance of each extracted feature type. In this study, we propose a novel framework, named NetSpam, which utilizes spam features for modeling review datasets as heterogeneous information networks to map spam detection procedure into a classification problem in such networks.

At work as usual: 080-40969981 | Write to me: technofist.projects@gmail.com, |
when u need us the most: **+91-9008001602, 080-40969981**| On the
Web:www.technofist.com www.technofist.in , www.itcdp.in

TEB013	**TITLE - EFFICIENT RECOMMENDATION OF DE-IDENTIFICATION POLICIES USING MAPREDUCE** **ABSTRACT** - Many data owners are required to release the data in a variety of real world application, since it is of vital importance to discovery valuable information stay behind the data. However, existing re-identification attacks on the AOL and ADULTS datasets have shown that publish such data directly may cause tremendous threads to the individual privacy. Thus, it is urgent to resolve all kinds of re-identification risks by recommending effective de-identification policies to guarantee both privacy and utility of the data.De-identification policies is one of the models that can be used to achieve such requirements, however, the number of de-identification policies is exponentially large due to the broad domain of quasi-identifier attributes.
TEB014	**TITLE - A SECURE AND VERIFIABLE ACCESS CONTROL SCHEME FOR BIG DATA STORAGE IN CLOUDS** **ABSTRACT** - Due to the complexity and volume, outsourcing ciphertexts to a cloud is deemed to be one of the most effective approaches for big data storage and access. Nevertheless, verifying the access legitimacy of a user and securely updating a ciphertext in the cloud based on a new access policy designated by the data owner are two critical challenges to make cloud-based big data storage practical and effective. Traditional approaches either completely ignore the issue of access policy update or delegate the update to a third party authority; but in practice, access policy update is important for enhancing security and dealing with the dynamism caused by user join and leave activities.

At work as usual: 080-40969981 | Write to me: technofist.projects@gmail.com, |
when u need us the most: **+91-9008001602, 080-40969981**| On the
Web:www.technofist.com www.technofist.in , www.itcdp.in