ABSTRACT

Physical deployment of privacy preserving data mining system is a great challenge for organizations nowadays. What are different hardware and software files needed for deployment? How different software files across the system should interlinked together for functioning correctly. What execution environment should be provided for different platforms on the same hardware? What are various protocol needed for hardware communication. Its prior knowledge will assist the developers for successful implementation of the entire framework. This paper deals with UML pictorial model for enabling the developers in successful installation of Privacy Preserving Data Mining Systems during run time environment. This would enable for the engineers in developing the software projects within optimum time, within budget, reduce the chances of errors and in turn minimizes the development and maintenance effort.

Keywords: unified modeling language, privacy preserving data mining systems, component diagrams, deployment diagrams, centralize PPDM systems, distributed PPDM systems, object oriented software engineering, query language.

1. INTRODUCTION

Software programs are often too big and complex. It is a great challenge for programmers to develop the error free software. It often involves different stakeholders to coordinate together for writing codes in order to function correctly. Object oriented approach has been evolved with the aim to simulate real world applications. It deals with breaking of programs into objects. One can models the real world and translates the created objects from analysis into design. The first step in Object Oriented analysis is create a precise, concise, understandable and correct model of real worlds. The software components being develops should exhibit the properties of Correctness, reusability, extensibility, compatibility, portability and friendliness [16] [17]. It has many advantages such as protecting the other resources in the operating system, support encapsulating properties and Abstract data types. It abstraction property helps in making the design flexible. Class inheritance and object composition supports reusing functionality, which facilitates newer requirements to be compatible with existing requirements and leads the system to evolve and makes the system robust to the particular types of change. New functionality could be added without altering the existing functions [16]. UML is used to model the full range of practical system needed to be built. Model is very essential for the software development process. It is the visual description of the necessary details of project and guides the software developers built the error free software applications as any flaw in the design could produce the catastrophic result, and disturb the overall budget [18]. Computations mainly involves three ingredients-processors (or thread of control), actions (or functions) and data (also called objects). A system architecture could be obtained from function or from objects. It mainly concern with what the object does, rather than concerning what the system do[19]. Object Oriented Paradigm solve all the problems of classical paradigm and is the best approach available today. It makes the task of development and maintenance easier. It exhibits the property of encapsulation which implies object's interdependence amongst each other [20]. Object-Oriented Design is the process of design encompassing object-oriented decomposition and a notation of representing both the logical and physical, as well as static and dynamic model of the system under design [21]. UML is standard diagrammatic notation for specifying, visualizing, constructing and documenting the artifacts of software system, as well as non-software system. It is the de-facto standard for Object Oriented Modeling and its 13 different diagrams are helpful for visualizing the complex software easier before actual implementation [22]. It provides the software professionals a stable and common design language that could be used to built complex applications for creating and discriminating design plans [23].

Concealing sensitive data from the outside world during mining process and simultaneously preserving the underlying data patterns so as to retain data utility which in turn could be exploited for gaining trade benefits has been major concern in the field of healthcare [2], business, web-usage mining [24], market basket analysis [25] and biometric [26] to name a few. Organizations mainly apply data mining for gaining trade benefits for extracting useful patterns while hiding personal information from competitive organizations. In the previous paper, authors have discuss the static modeling of various approach of Privacy Preserving Data Mining Systems by various UML diagrams such as Use Case diagram, Class diagram, Activity diagram and sequence diagram to name a few. Authors in this paper will discuss components and deployment diagrams for its physical deployment during run time. To the best of our knowledge, this is the first work which concern with UML Modeling of generalized PPDM system.
2. RELATED WORK


3. UML MODELING OF THE PROPOSED FRAMEWORK

Authors in this paper have discussed both centralized and distributed framework for achieving privacy preserving data mining tasks. Let us first discuss the centralized approach.

Runtime Environment, a software component containing source classes, object classes and all the necessary packages providing environment suitable for java programs execution, is installed in each sets of the hardware components. Figure below represents how each of these package and object files interact together for achieving the desired operations. Each Data Provider's information is stored in the tables and java object class fetch data from the table and applies privacy preserving operations on it. The data are outsourced to the Data Warehouse Server. The clean and integrated data from multiple Data Provider's datasets are stored in Data Warehouse database, which also include other information such as login credentials. As the operation on Data Mining Server involves any Object Oriented Program in the front-end and Query Language in backend, both Object Oriented Programming and Query Package environment are needed to installed in the Data Mining Server. Queries are sent to the Data Warehouse Server. Username and Password are needed which are stored in login credential database, and connection took place after successful authentication. Data is fetch from the Data warehouse database and privacy preserving operations are applied on it. Figure below represents the entire operations.
Figure-2 represents the component diagram of distributed Privacy Preserving Data Mining Systems. Different PPDM frameworks are connected via TCP/IP protocol suite and shares local data mining results to perform global data mining.

Figure-3 represents the deployment diagram of centralize Privacy Preserving Data Mining Systems. Various hardware components (or nodes) - Data Provider, Data Warehouse Server and Data Mining Server are installed in the systems. Data Provider and Data Warehouse Server interact together by TCP/IP protocol, while communication between Data Mining Server and Data Warehouse Server takes place connectivity. Data Provider gives input for performing the entire operations while Data Miners access the data mining output after PPDM operations.
Figure-3 depicts the deployment diagram of centralized PPDM systems. Each PPDM framework is connected via TCP/IP Protocol suite. Each Data Mining Server across different frameworks share the local mining result with other framework by TCP/IP protocol suite to perform global data mining operations. Data Provider input the data and mining output is accessed by the each data miners of the corresponding framework.
CONCLUSIONS
Complexity is inherent in the development of software codes. The building of complex software often results in late, over budget and deficient in needed requirements, often called as software crisis. So, it is often required to decompose the software into objects. Deployment of Privacy Preserving systems is one of the biggest challenges nowadays. This could only be successfully achieved if developers have prior knowledge of complete hardware and software artifacts and its mutual interaction during deployments. Authors have explained step by step pictorial representation of various diagrams for both centralize and distributed data mining systems. This enables the software technicians to develop error free and robust application which in turn reduce further development and maintenance effort.

REFERENCES


[3] Lin Zhang; Yan Liu; Ruchuan Wang; Xiong Fu; Qiaomin Lin. 2017. Efficient privacy-preserving classification construction model with differential


