Weill Cornell Medicine

2018 - SUMMER

CURRICULAR PRACTICAL TRAINING

CPT – REPORT

SIDDHESH SAHADEO POWAR

Student ID- 10424918

SOFTWARE ENGINEER INTERN

RESEARCH INFORMATICS GROUP

WEILL CORNELL MEDICINE

DATE: 9TH AUGUST 2018

1) Organization: Describe the organization you interned for.

I am Siddhesh Powar, pursuing Masters in Information Systems at Stevens Institute of Technology. I have completed two semesters at Stevens and I started my Summer Internship on 13th July 2018 at Weill Cornell Medicine. Weill Cornell Medicine is committed to excellence in patient care, scientific discovery and the education of future physicians in New York City and around the world. Since 1898, doctors and scientists at Weill Cornell have been engaged in world-class clinical care and cutting-edge research that connect patients to the latest treatment innovations and prevention strategies. Weill Cornell Medicine's powerful network of collaborators extends to its parent university Cornell University; to Qatar, where Weill Cornell Medicine-Qatar offers a Cornell University medical degree; and to programs in Tanzania, Haiti, Brazil, Austria and Turkey. Weill Cornell Medicine faculty provide comprehensive patient care at New York-Presbyterian/Weill Cornell Medical Center, New York-Presbyterian Lower Manhattan Hospital, New York-Presbyterian Queens and New York-Presbyterian Brooklyn Methodist Hospital. Weill Cornell Medicine is also affiliated with Houston Methodist.

At Weill Cornell Medicine, I am working in the Information Technology and Services Department. The Information Technologies and Services Department provides IT infrastructure, management, service and support for the entire institution. The department supports more than 10,000 computer users across 1,000 servers, and numerous clinical, research and administrative software applications. The staff ensures

institutional security standards, manages high speed networking systems, develops websites, creates innovative computer systems and applications for special projects, and assists with hardware purchases.

In the Information Technology and Services Department, I work in the Research Informatics Group. Weill Cornell Medical College is dedicated to supporting its researchers with comprehensive, state-of-the-art computing resources. Faculty members have online access to cluster computing services, data center facilities, Internet 2 connectivity to the main campus, and many other resources. Research Informatics facilitates the conduct and administration of clinical and translational research through its Architecture for Research Computing in Health (ARCH) and Data Integration teams. It is focused on combining and integrating new and existing datasets to enable access to clinical data for research as well as creating custom research repositories for hypothesis-driven research.

2) Role: Describe the role you played as an intern in the organization.

I work as a Software Engineer Intern in the Research Informatics Group at Weill Cornell Medicine. Under the direction of the Manager of Research Informatics Software Engineering, I assist in the design, implementation and maintenance of software to support the scientific workflows. As a Research Informatics Software Engineer Intern my

job is to primarily focus on the back-end and front-end application components as well as the integration of the data from multiple source systems. Through these activities, I work with the other team members to provide robust, scalable software solutions to the research enterprise. This internship has given me an experience in the field of clinical and translational research informatics.

As an Intern, I coordinate with the other senior team members to get as much information as possible on the projects which we are working on. Some of the projects and modules which we are working on are:

Architecture for Research Computing in Health (ARCH)

Architecture for Research Computing in Health (ARCH) matches Weill Cornell Medical College investigators with tools and services for obtaining electronic patient data.

Biobank Informatics

Biobank informatics supports workflows for sample handling and annotation of biospecimens as well as data integration with clinical and research information systems.

Clinical Trial Monitor Support

Investigators can request access to electronic health record system data to support clinical trial monitors.

Cohort Discovery (i2b2)

Informatics for Integrating Biology and the Bedside (i2b2) enables researchers to query de-identified data from electronic health record (EHR) and research systems of Weill Cornell Medical College to discover cohorts of patients preparatory to research.

Electronic Case Report Forms (eCRFs)

RED Cap is an electronic data capture tool that enables investigators to build case report forms that support prospective studies and retrospective chart reviews.

Electronic Health Record Interventions

EHR systems provide a platform for implementing new clinical interventions developed from analysis of previously captured EHR data and evaluating effects using prospectively collected EHR data.

Electronic Health Record Reporting

Weill Cornell Medical College investigators can request inpatient electronic health record data via DISCOVERY/TRAC and outpatient data via Physician Organization Information Services.

High Performance Computing

ITS can provide infrastructure and management for high performance computing needs, including specialized setups.

Imaging Informatics

The Weill Cornell Department of Radiology's Imaging Data Evaluation and Analytics Lab (IDEAL) provides research support for imaging data management services and image processing and analysis.

Multi-Institutional Data Sharing

By combining electronic patient data across institutions, investigators may increase the completeness and comprehensiveness of clinical data and improve clinical trial accrual and comparative effectiveness research.

Online Surveys

Survey solutions for research studies and administrative data collection.

Red Cloud Secure

Red Cloud Secure offers dedicated hosts only for the time needed and can accommodate non-public data.

Research Data Repositories

Research data repositories (RDRs) enable investigators to aggregate and transform data from multiple systems into a single secure resource customized to their needs.

Subject Enrollment & Billing Compliance

Electronic health record and billing systems from the Weill Cornell Physician Organization and New York-Presbyterian Hospital support subject enrollment and billing compliance activities.

WCM Data Core

Data Core is a secure, scalable computing and storage environment within which users can share access to a collection of data sets and process the data with a variety of software tools, while meeting appropriate regulatory requirements to protect data. I am also given several tasks which have already been worked upon, to find a better solution to the problem with a different approach, by putting my knowledge into practice.

3) Tasks: Describe what tasks you were responsible for during your internship

My role is to develop solutions to the research problems by putting into practice my skills in managing and analyzing large databases and querying them to create several tables which give the information related to specific research study or disease. With my knowledge in Python Coding I play a role of developing solutions to several research problems. I am also responsible in automating several tasks through the codes and making the processes faster and efficient.

Our company deals with huge datasets which have hundreds of tables related to the patient information, diseases and allergies information, age groups, regional data, etc. These tables have records of patients all over the world and the diseases and the treatments they are undergoing. This data is highly confidential as it has the patient's health information. So, in my first week I underwent a HIPAA (Health Insurance Portability and Accountability Act) training wherein I learned about how the patient data needs to be secured and used for a rightful purpose.

Later on, I was given some tasks where I used Microsoft SQL Server to query the database. Performing data analysis on several databases by using several queries and merging different tables and records to obtain a solution and after analyzing this data creating reports in SQL Server Reporting Services (SSRS) to get a better understanding of the analysis, were some of the tasks which I worked upon.

I have used Python 3.7 to develop some scripts which were able to successfully automate certain processes. In one task, I had to work on pymssql module of Python to establish connection to our SQL Servers and then querying the databases through Python by executing several stored procedures. In this task I successfully extracted the contents of a CSV file into Python and then imported it into the database by querying the database by using several SELECT statements and Clauses and several stored procedures to update the database. Then, I committed the changes to the server connection in Python and after these databases were updated the contents of the updated table was exported into another CSV file automatically.

Previously, this whole process was manually performed by the software engineers. But using the above approach and code I was successfully able to implement an automated process which was quicker and efficient in producing the results. In Python I have worked on pyodbc and pymssql environments.

I also used Subversion (SVN) to keep track of the files in our servers used SVN repositories and PyDev to keep track of the scripts and run them. Currently, I am working on parsing XML documents to databases in Python. Thus, overall, I am working on tools such as Python 3.7, Microsoft SQL Server Management Studio, SSRS, JSON, XML and Excel.

4) Relationship to Curriculum: How did your internship relate to the different courses you took during the curriculum? What useful skills did you bring from the classroom to your workplace? What skills did you acquire during the internship that will help you during the remainder of your studies?

I one of my courses at Stevens MIS-630 (Data and Knowledge Management), wherein I learned about SQL and how it is used for Database Management. During the course of my Internship I have consistently worked on SQL. The knowledge which I gained during this course helped me work on huge databases and produce SQL scripts to query the databases. While I was studying this course, I had worked on a project to create a Library Database Management System. This was not a big project and it had only a few records and tables. But, at my internship I am dealing with the tables which have millions of records and thus, I have learned to optimize the SQL scripts to produce faster results for larger datasets. In this way, my knowledge of dealing with complex systems has increased and I have acquired the skills to optimize the code, which will help me work on my future projects at Stevens in a more efficient way.

I took up BIA-652 (Multivariate Data Analysis) where I learned to code in R. This made it easier for me to code in Python at my workplace. In this course, I learned about developing programming logic to analyze the data. Using this knowledge, I am working on several projects which run in Python.

I also took up (MIS-636) Business Intelligence and Data Warehousing where I learned about the Data Warehouses and the Architecture. How the data is extracted from several sources and then using the ETL tools this data is stored in the Data Warehouses to perform Data Analysis and after analyzing this data is moved into Data Marts and can be used for reporting services and by end-users. I used this knowledge to understand the workflow of several projects during my internship and how this data is analyzed and then used for Business Intelligence and Research Purposes.

5) Feedback: What worked, what could be improved?

Working at Weill Cornell Medicine has indeed been a great learning experience for me and working under the guidance of Senior Team Members and the Manager, I really found it exciting to work on several challenging projects. With the knowledge which I gained during the courses I took at Stevens, I was able to work in this fast-paced and challenging environment at my workplace.

I am privileged to study at such a reputed university and obtain an internship at such a big company. Still, I always look forward to making improvements by studying hard and working more harder and reach further heights. The area of improvement I would work upon is learning more languages and tools in my next semesters at Stevens which will enhance my knowledge in the Information Systems domain. Thank You.

REFERENCES:

https://weill.cornell.edu/