

October 20, 2022

## Plan B: Massively Learning Activities

*M.S. Computer Science: In Woo Park*

### Abstract

This document serves as a general-purpose outline of my Plan B: Capstone Project for M.S. Computer Science. My project will be conducted with TASI/PHIDC and the project will be advised under Director and Faculty Specialist, **Norman Okamura**. My expected graduation date is Spring 2023.

## 1 Introduction

The Telecommunications and Social Informatics Research Program / Pacific Health Informatics and Data Center (TASI/PHIDC), formerly TASI/PEACESAT, is part of the Social Science Research Institute (SSRI) of the College of Social Sciences (CSS) at the University of Hawai'i at Manoa. TASI/PHIDC programs incorporate an interdisciplinary approach to education and research, and work with partners from across the University of Hawai'i system, State of Hawai'i and other government and academic institutions from the Asia and Pacific Islands region.

SAS Viya, or SAS Visual Analytics, is a cloud-enabled\*, in-memory analytics engine that provides elastic, scalable, and fault-tolerant processing. Cloud Analytics Services (CAS) is the in-memory analytics engine SAS Viya uses for both on-premise as well as cloud-service environments (i.e., AWS, Azure, GCP). CAS uses a combination of hardware and software where data management and analytics take place on either a single machine or as a distributed server across multiple machines. There are two types of CAS server configurations: a Symmetric Multiprocessing (SMP) environment where a CAS server consists of a controller and runs on a single machine, or a Massively Parallel Processing (MPP) environment where a distributed CAS server consists of one controller, one or more workers, and one backup controller (optional), each running on a separate machine. Although many configuration alternatives exist, a software-based hyper-converged infrastructure or a virtualized multi-server hyper-converged infrastructure is preferred for SAS Viya to perform data analytics in parallel.

Hyper-Converged Infrastructure, or HCI, is a software-defined IT infrastructure that virtualizes all of the elements of conventional "hardware-defined" systems. In a traditional IT infrastructure, compute, networking, management, and storage exists as separate components. HCI combines these elements with virtualization in order to dynamically allocate resources to virtual machines or containers. This reduces overall latency of the infrastructure and enables parallel processing.

TASI/PHIDC is a Technical Assistance and Research Partner or "TARP" who has an Intergovernmental Cooperative Agreement (ICA) with the Commonwealth of the Northern Mariana Islands (CNMI) State Medicaid Agency (SMA) to design an infrastructure that would allow advanced data analytics and parallel processing of Protected Health Information. After careful consideration, TASI/PHIDC has opted for SAS technologies in a hyper-converged infrastructure.

---

\*Cloud-Enabled: Applications built traditionally and then migrated to the cloud.

## 2 Purpose

The purpose of this project is to assist in the planning, acquisition, and configuration of hardware and software to support advanced data analytics using SAS Viya technologies in a parallel processing environment.

### 2.1 Design a Master Plan with SDLC Framework

Assist in designing a master plan for the acquisition of hardware and configuration of SAS software using VMware and multiple physical servers that will be capable of being upgraded as new servers are acquired and installed.

The project will document the plan within a Systems Development life cycle (SDLC) framework to describe the technical hardware and software configuration on existing servers and ultimate plan to acquire, migrate, and reconfigure the VMs (and perhaps other SAS configurations) to support the CAS functionality in the new server hardware environment. The new environment will consist of new servers with increased in-memory capacity and optimizes on hard disks as well as a Storage Area Network (SAN) capability to implement CAS functionality.

#### **Configured Architecture**

A key issue that will need to be understood and documented will be the steps involved in re-configuring the VMs in the new server hardware and to optimize the blend of in-memory, SAN, and hard drives. Other processes such as backup of original and copies of data loaded into the CAS systems environment that will support the worker functions will also need to be clearly articulated. The reason is that SAS corporation is being engaged (and paid) for the installation of the software and re-configurations will result in additional expenses. So, it will be essential that these configuration and re-configuration processes (especially the steps and order of operations of configuration) be understood and documented since it is not yet determined how many servers and the exact configuration of the servers have been conclusively made.

The SDLC framework uses life cycle phases to plan and implement systems development. These life cycle phases are planning, system analysis, system design, development, implementation, integration, and testing. The SDLC framework is required when designing a master plan as TASI/PHIDC must follow the same regulations and compliances as the Centers for Medicare & Medicaid Services (CMS).

#### **Data Access, Compliance, Risk Adjustment**

CMS data will be requested through the Research Data Assistance Center (ResDAC), a CMS contractor that provides assistance to researches interested in CMS data. ResDAC is able to distribute CMS data as a SAS read-in file for smooth integration with SAS technologies.

The project's design must comply with the same regulations and compliances as the Centers for Medicare & Medicaid Services (CMS). The Health Insurance Portability and Accountability Act (HIPPA), requires the protection of a patient's electronically stored, protected health information (ePHI), by using appropriate administrative, physical and technical safeguards to ensure confidentiality, integrity, and security of this information.

In addition, the CMS Hierarchical Condition Category (HCC) risk adjustment model is used to calculate risk scores, which will adjust capitated payments made for aged and disabled beneficia-

ries enrolled in Medicare Advantage (MA) and other plans. The project will use the same risk adjustment algorithms when performing data analytics.

### **Learning Objectives**

Understand the intricacies and nuances required when designing a master plan for federal agencies by following the SDLC framework. Participate in multiple life cycle phases to ensure a deep understanding of the project's plan, design, implementation, deployment, and testing. Understand the concept of virtualization, SANs, and CAS, in a hyper-converged infrastructure to enable parallel processing data analytics for SAS technologies. Understand necessary regulations, compliance, and security measures when dealing with Protected Health Information.

## **2.2 Data Migration (Tentative)**

If time permits, assist in the full-scale data migration of CNMI State Medicaid Agency PHI records.

### **Learning Objectives**

Understand the concept of HIPPA, and how to protect sensitive patient health information. Understand the concept of data-at-rest and data-in-transit encryption when conducting data migration of PHI records.

## **2.3 Data Analytics (Tentative)**

If time permits, assist in the analysis of CNMI State Medicaid Agency PHI records the Agency for Healthcare Research and Quality (AHRQ) Quality Indicators. The AHRQ Quality Indicators (QIs) comprise four measure areas: inpatient, prevention, patient safety, and pediatric care. The Agency for Healthcare Research and Quality is one of twelve agencies within the United States Department of Health and Human Services.

### **Learning Objectives**

Understand the relationship between SAS and AHRQ when conducting data analysis on PHI. Data analytics training on SAS Viya through AHRQ QI Modules (PQI, IQI, PSI, PDI) at the enterprise level.

## **3 Conclusion**

Understand non-trivial database and data analytics topics by assisting in the acquisition of hardware and software to configure SAS technologies in a virtualized infrastructure for CNMI State Medicaid Agency PHI records, using the SDLC framework to respect time constraints, compliance, and risk adjustments at the enterprise level.

Building and deploying this architecture with SAS Technologies is not primarily for the benefit of PHIDC/TASI. UHM provides PHIDC/TASI the SAS software license by sponsoring the annual cost. The consequence of this project is to create a framework for a non-secure, hyper-converged, SAS environment that can be used by the UHM faculty members to train undergraduate and graduate students in data analytics. PHIDC/TASI will create this framework by designing for PHI and then scaling out to build an infrastructure that can be used by UHM faculty and students.