TASI
TELECOMMUNICATIONS AND SOCIAL INFORMATICS
RESEARCH PROGRAM

Massively Learning Activities

Full Name	Student ID
In Woo Park	25141090

April 15, 2023

Table of Contents

1 | Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

2 | Introduction

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.1 | TASI/PHIDC

ABOUT US
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. Program and research focus areas include policy, planning, information and communications technologies and systems, health information technology, health informatics in Hawai'i and the Pacific Islands region.

MISSION
The TASI/PHIDC Research Program missions are to: (1) Provide technical assistance in policy, program planning and evaluation; (2) Facilitate public and private sector collaboration to improve community resiliency, sustainability, and health system performance; and (3) Build capacity in information technology, health data management, analytics, and data sciences.

FACULTY RESEARCH
TASI/PHIDC conducts interdisciplinary and applied research and provides policy, program, technical assistance, education, and training in Hawai'i and the Pacific Islands Region related to:

- Accessible and affordable Information and Communication Technology (ICT)
- Health Information Technology (HIT)
- Electronic Health Record (EHR)
- Healthcare and claims data management, analytics, and programs
- Telehealth
- Meteorological and disaster communications

2.2 | Contract Explained (TBD TITLE)

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.

- Modernize data archive and storage (paper to electronic) of PHI data.
- Want to perform data analytics and machine learning.
- Used RCUH funds to purchase SAS license.
- Therefore, SAS needs to be accessible to multi-tenants and UH themselves.

3 | VMware

VMware is a company that specializes in developing technologies for virtualization and cloud computing. Its software products and services enable organizations to efficiently manage their IT infrastructure, improve performance, and reduce costs. VMware offers solutions for network virtualization, cloud management, digital workspace solutions, and security solutions.

3.1 | vSphere 6.5

vSphere is VMware's virtualization software suite that allows you to create and manage virtual machines and computing environments, using a set of software tools and services. With vSphere, you can run multiple virtual machines on the same physical server, each running its own operating system and applications. vSphere includes many features and capabilities that help make virtualized environments more reliable, scalable, and performant, such as:

- **vSphere Web Client**: A web-based management interface.
- **ESXi**: The bare metal hypervisor installed on your machines.
- **vCenter**: A centralized management system for your vSphere environment.
- **vSAN**: A software-defined storage solution to create a distributed storage platform in vSphere.
- **NSX**: A software-defined networking solution for your vSphere environment.
- **VMotion:** Software to migrate VMs between servers without interruption of service.

3.2 | vSphere Web Client

The vSphere Client is an application that enables administrators to manage and monitor VMware vSphere environments. It comes with a graphical user interface (GUI) and allows users to connect to VMware vCenter Server, which serves as a central management console for multiple VMware vSphere hosts.

Through the vSphere Client, administrators can create and modify virtual machines, manage storage, configure networking, and monitor system performance, among other things. Essentially, it provides a range of tools that enable users to manage virtual infrastructure components effectively. In addition to the traditional Windows-based vSphere Client, there's also a web-based version called the vSphere Client (HTML5), which is designed to work seamlessly across different operating systems and devices, including desktops, laptops, and mobile devices. This new client offers a simplified interface, improved performance, and support for new features introduced in vSphere 6.5 and later versions.

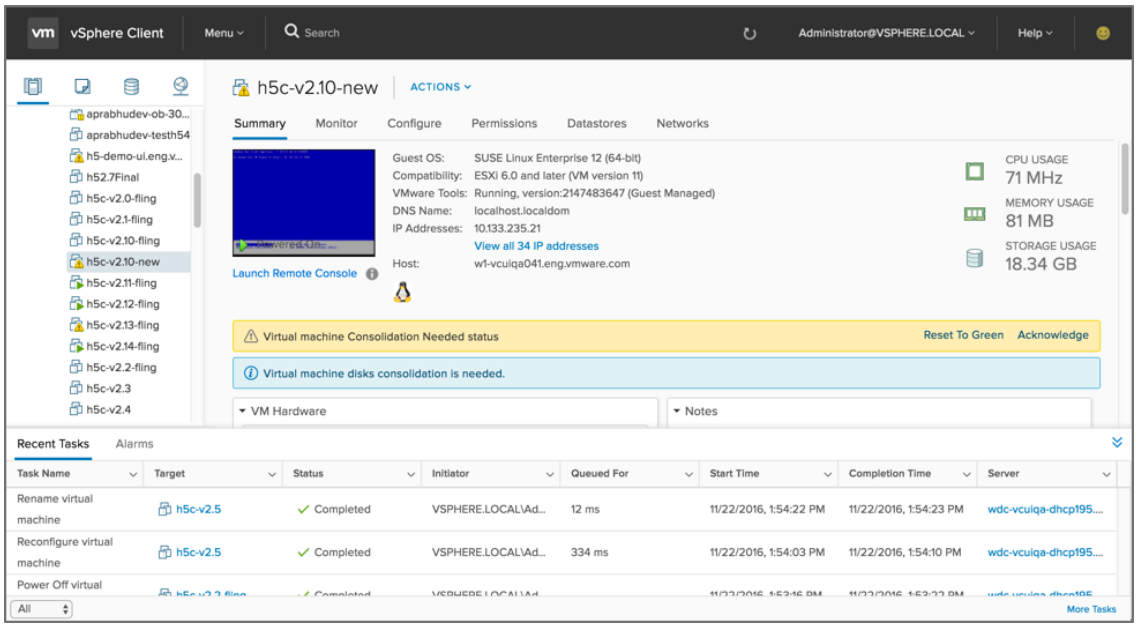

Figure 3.1: vSphere Client (STOLEN EXAMPLE)

3.3 | ESXi

VMware ESXi formerly known as ESX is a bare metal hypervisor that is installed directly on the physical server hardware and provides the ability to create, run, and manage virtual machines.

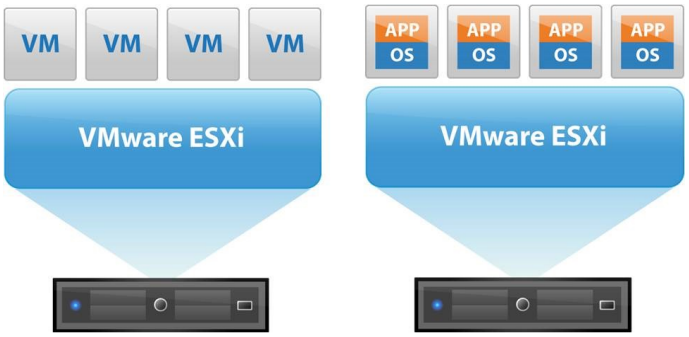

Figure 3.2: ESXi (STOLEN EXAMPLE)

3.4 | vCenter

vCenter is a software platform that provides centralized management and control for their suite of virtualization products, including vSphere. By providing a single point of control, it simplifies management and reduces complexity, making it easier to manage many virtual machines and components. With vCenter, you can manage hosts, clusters, virtual machines, networks, and storage resources to support a virtualized environment with high availability, disaster recovery, and workload balancing.

In addition, vCenter provides advanced capabilities like automation, orchestration, and policy-based management. These features allow you to automate routine tasks, streamline operations, and enforce policies across your virtualized environment. Examples of automated tasks include: provisioning VMs [1], patches and updates [2], backup and recovery [3], monitor and reports [4], and resource allocation [5].

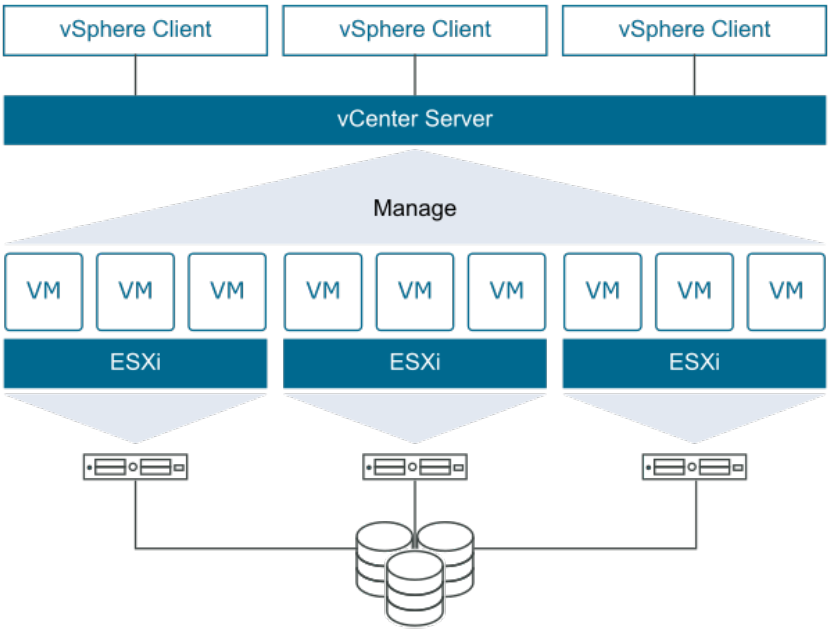

Figure 3.3: vCenter (STOLEN EXAMPLE)

[1] Create new virtual machines, configure virtual hardware, and install operating systems and applications.
[2] Deploy software updates, apply security patches, and performing maintenance tasks.
[3] Create backup schedules, perform backup and restore operations, and monitor backup performance.
[4] Generate reports on virtual machine performance, track resource usage, and monitor system health.
[5] Adjust CPU and memory resources, configure storage allocations, and manage network bandwidth.

3.4.1 | vCenter Security and Risks

Security is a critical aspect of virtualized environments, and vCenter provides a range of security features to protect against unauthorized access, data theft, and data manipulation. These security features include: role-based access control[6], auditing[7], encryption[8], secure communication[9], integration[10], and two-factor authentication[11]. These security features help to ensure confidentiality, integrity, and availability of the virtualized infrastructure, a requirement when working with PHI data.

3.5 | vSAN

vSAN is a software-defined storage solution developed by VMware, which allows organizations to create a distributed storage platform that is integrated with vSphere. This provides a highly scalable and available storage infrastructure, using standard hardware.

By creating a shared data store using the internal disks of ESXi hosts in a vSphere cluster, vSAN allows organizations to pool their storage capacity and performance into a single datastore, scaling it easily by adding more hosts to the cluster. vSAN features data replication, erasure coding, and automatic data rebalancing. Additionally, it offers advanced storage services such as deduplication, compression, and encryption, ensuring optimal storage efficiency and security which streamlines storage management, automates routine tasks, and helps to optimize storage utilization and cost savings.

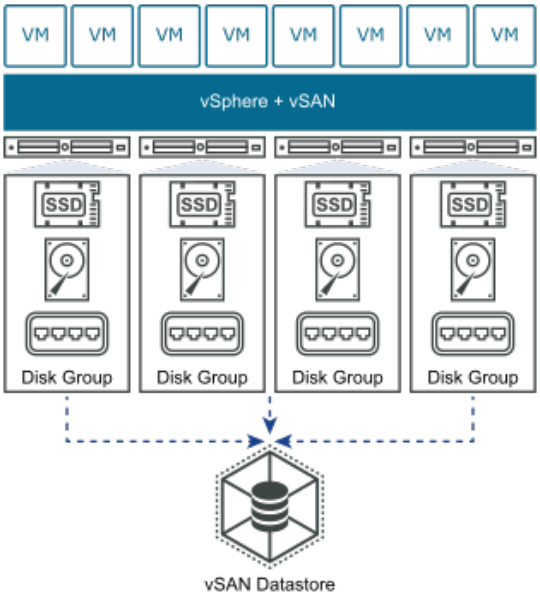

Figure 3.4: Standard vSAN Cluster (STOLEN EXAMPLE)

3.6 | NSX

NSX is a network virtualization and security platform created by VMware that provides a software-defined networking (SDN) solution that enables organizations to virtualize their network infrastructure, creating a more flexible, scalable, and manageable network.

NSX allows for all network components in your infrastructure to be virtualized, decoupling your network from existing hardware. This abstraction enables organizations to pool and automate network resources, which can reduce the time and cost of deploying and managing network infrastructure. NSX also offers advanced security features and networking capabilities which allows administrators to apply precise

[6] Define roles and permissions to users based on their roles to prevent unauthorized access.
[7] Track user activity and changes to identify security issues and log actions taken within the virtualized environment.
[8] Encrypt VM data, configuration files, and communication between hosts.
[9] Supports SSL/TLS encryption to secure communication between hosts and the vCenter server.
[10] Integrate with third-party security products (e.g., antivirus, IDS) to provide additional layers of security
[11] Provide two forms of identification before accessing the VM to prevent unauthorized access.

policies to specific workloads or applications. For example, NSX provides: network automation, multi-cloud and on-premises support, network segmentation, minimal cost and resource overhead, switching and routing, and load balancing features.

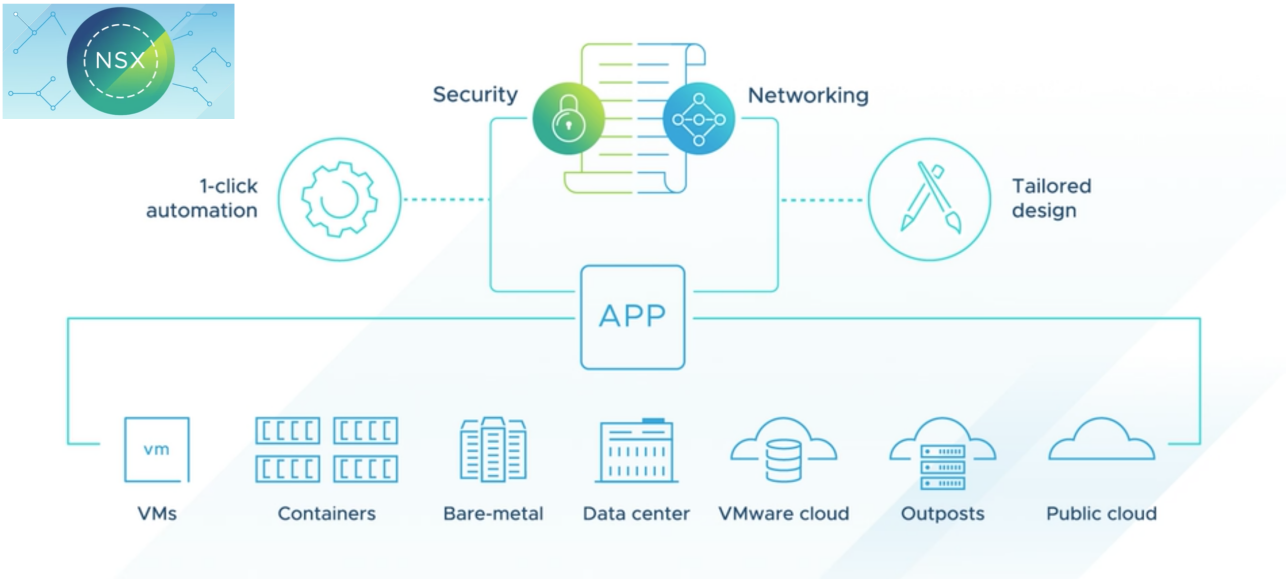

Figure 3.5: NSX Infrastructure (STOLEN EXAMPLE)

3.7 | VMotion

VMotion is virtualization software that enables IT administrators to move VMs between physical servers or hosts without disrupting service. The process involves copying the entire state of the VM, including memory, CPU state, and network connections, from one host to another. The benefits of VMotion include increased availability and uptime, improved hardware utilization, workload balancing, and reduced downtime for maintenance and upgrades. However, the feature also requires specialized hardware and software, increasing the complexity of virtualized environments. VMotion uses shared storage, high-speed networking, and specialized software to ensure a seamless migration.

The main use case for VMotion is to provide high availability and workload balancing for virtualized environments by optimizing resource usage, improving performance, and avoiding downtime during maintenance or upgrades. For example, an IT administrator can use VMotion to move running VMs to another host during server maintenance, ensuring uninterrupted service for end-users. Once the maintenance is complete, the VMs can be moved back to the original host. VMotion also allows for the consolidation of workloads and the migration of VMs to new hosts for improved hardware utilization and cost savings.

When migrating VMs with sensitive data, such as protected health information (PHI), there may be compliance issues with regulations like HIPAA. To ensure compliance, virtualization infrastructure and VMotion must be configured to meet data protection, access control, and auditability requirements. Encryption and other security measures should also be implemented to protect the confidentiality, integrity, and availability of PHI during migration. IT administrators should ensure that host servers and network connections used for VMotion are secure and protected from unauthorized access.

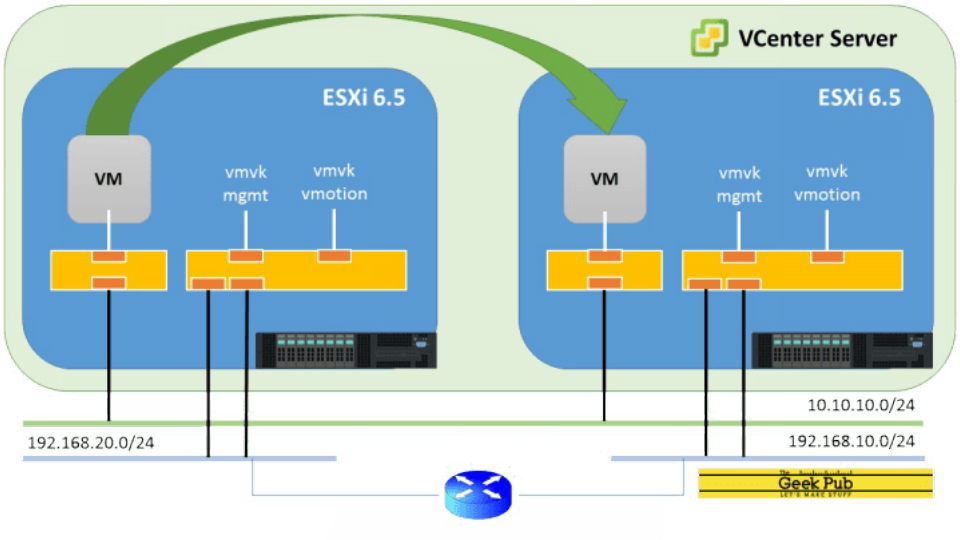

Figure 3.6: VMotion (STOLEN EXAMPLE)

4 | Statistical Analysis System (SAS)

SAS, or Statistical Analysis System, is a software suite that has been used for advanced analytics, business intelligence, data management, and predictive analytics since it was first released in 1976. Developed by the SAS Institute, it offers a range of statistical and data analysis tools, which are suitable for many applications including data mining, forecasting, econometrics, quality control, and statistical analysis.

The software provides a user-friendly graphical interface for data analysis and reporting, as well as a powerful programming language that allows users to customize their analysis and automate repetitive tasks. Its ability to handle large and complex datasets and perform advanced statistical analyses make it popular in various industries, including finance, healthcare, government, and academia. SAS is widely used for purposes such as fraud detection, risk management, clinical research, and marketing analysis, and is a popular choice among data scientists and statisticians.

SAS offers multiple product **suites**. The SAS Enterprise Suite is a collection of SAS products designed for enterprise-level data management and analysis. The SAS Platform provides two engines for managing foundational capabilities such as distributed processing, security, administration, program development and execution, resource management, user interfaces, cloud integration, operating systems and third-party software. These engines are SAS 9.4 and SAS Visual Analytics.

In addition to the SAS Enterprise Suite and the SAS Platform, SAS also offers SAS Data Management Advanced which is a powerful ETL solution for preparing data for both analytic engines.

4.1 | SAS Data Management Advanced (SAS DMA)

SAS Data Management Advanced (SAS DMA), is a software suite that provides a comprehensive set of tools for data integration and data quality. The primary purpose of SAS DMA is to support data ETL (Extract, Transform, Load) processes, which involve extracting data from multiple sources, transforming it to meet specific requirements, and loading it into a target system for analysis and reporting. SAS DMA is a stand-alone software suite and can be deployed on-premises or in the cloud. It is not part of SAS 9.4 (i.e. SAS Data Management and Analytics), which is a separate software suite that provides a wide range of tools for data analysis, reporting, and visualization.

To support these functions, SAS DMA relies on three different types of servers:

- The **Mid-Tier** server is a web-based interface that provides access to SAS DMA workflows. This server is responsible for user authentication and authorization, job scheduling and monitoring, and other functions that are necessary for effective workflow management. It acts as a gateway for users to interact with the SAS DMA system.
- The **Metadata** server is responsible for managing information about data sources and workflows. It provides a central repository for storing metadata, which enables efficient management of SAS objects, definition of relationships between objects, and tracking of changes to data. In the case of SAS DMA, the metadata server manages information about data integration workflows and data quality rules.
- The **Compute** server provides the processing power and resources necessary to run data integration and data quality jobs. This server is responsible for executing the actual data integration and ETL tasks defined in the workflows created in SAS DMA. It ensures that the workflows are run efficiently and effectively, regardless of the size or complexity of the data being processed.

4.2 | SAS 9.4

SAS 9.4 is a software suite that provides tools for data management, statistical analysis, business intelligence, and predictive modeling. SAS 9.4 can handle large datasets and complex analyses by using a wide range of built-in functions and procedures that can save time and effort when working with data.

For example, a pharmaceutical company might use SAS 9.4 to analyze clinical trial data to determine the efficacy and safety of a new drug. A bank might use SAS 9.4 to perform risk analysis on its loan portfolio. A retail company might use SAS 9.4 to analyze customer data to better understand buying patterns and preferences.

SAS 9.4 is composed of several modules that provide a wide range of functionalities:

- Base SAS - The basic programming language, data access, and management capabilities of SAS.
- SAS/STAT - A comprehensive set of statistical analysis procedures for data exploration and modeling.
- SAS/GRAPH - A set of tools for creating high-quality graphical output from SAS data.
- SAS/ETS - A set of time series analysis and forecasting procedures.
- SAS/IML - An interactive matrix language for matrix manipulation, data analysis, and numerical optimization.
- SAS/ACCESS - Connectivity to data sources such as relational databases and spreadsheets.
- SAS Enterprise Guide - A graphical user interface (GUI) for SAS programming, data management, and reporting.

4.3 | SAS Visual Analytics (SAS Viya)

SAS Visual Analytics (SAS Viya), is a cloud-based analytics platform that provides a suite of tools and services for elastic, scalable, and fault-tolerant data analytics, data processing, and machine learning for enterprise environments. It allows organizations to store, manage, analyze, and share large volumes of data across different sources and formats, all within a single platform.

SAS Viya is composed of several software that provide a wide range of functionalities:

- SAS Visual Analytics: A tool for creating interactive reports and dashboards to explore and visualize data.
- SAS Visual Statistics: A tool for performing statistical analysis and building predictive models on large data sets.
- SAS Visual Data Mining and Machine Learning: A tool for exploring and analyzing large data sets using advanced analytics techniques such as clustering, decision trees, and neural networks.
- SAS Visual Forecasting: A tool for creating accurate and reliable forecasts using time series data.
- SAS In-Memory Statistics: A tool for performing high-performance analytics and modeling on large data sets using in-memory processing.

When performing analytics on large datasets, SAS Viya uses Cloud Analytic Services.

4.4 | Cloud Analytics Services (CAS)

Cloud Analytics Services (CAS) is the in-memory analytics engine SAS Viya uses for both on-premise as well as cloud-service environments (e.g., 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. In either single or distributed deployment, each machine (host, node, etc) will be assigned one of three roles: CAS Controller, CAS Backup Controller, CAS Worker.

Analogy
In a restaurant kitchen, there exists three primary chefs. They are the (1) executive chef, (2) sous chef, and (3) station chef(s). The executive chef's primary role is to manage the kitchen and its staff whilst doing very little cooking. The sous chef's primary role is to be the right-hand to the executive chef, ready to manage the kitchen, share, or take over the responsibility of the executive chef at a moments notice. The station chef(s) merely wait for instructions from the executive chef, then executes the job they are given.

This is the relationship of each CAS node with each other:

- The CAS Controller is the executive chef managing the kitchen and its staff, delegating work.
- The CAS Backup Controller is the sous chef ready to take over the responsibility of the executive chef.
- The CAS Worker(s) are the station chefs cooking what they are assigned to by the executive chef.

4.4.1 | Role 1: CAS Controller

Controller is the first role that can be assigned to a host for SAS Cloud Analytic Services. For both server architectures, single-machine and distributed, one machine must be designated as the Controller. The role of the Controller is to parse out work to each Worker host available. In other words, the Controller manages and controls the overall operation of the CAS environment. As the master node, the Controller is responsible for distributing workload among available CAS Workers, managing user sessions, and providing a secure environment for data retrieval and data storage.

In a single-machine environment, the CAS Controller and CAS Worker roles can be performed by different processes or threads within the same operating system instance. However, we are not limited to this deployment method as it is also possible to have the CAS Controller and CAS Worker(s) virtually separated (on the same hardware) to increase the scalability of the deployment. The configuration of your architecture depends on what you need out of CAS.

In a distributed environment, the CAS Controller is responsible for managing and controlling the CAS environment whilst the actual data processing and data analytics are performed by the CAS Worker(s).

4.4.2 | Role 2: CAS Backup Controller

Backup Controller is the second role that can be assigned to a host for SAS Cloud Analytic Services. Although optional, the CAS Backup Controller is highly recommended in a distributed server environment. The role of the CAS Backup Controller is to act as a standby or hot-backup for the primary CAS Controller in case of a failure. Its primary purpose is to ensure that the system can continue to function in the event of a failure of the primary controller. The Backup Controller is typically set to passively monitor the primary controller for any signs of failure, such as a loss of connectivity or failure to respond to heartbeat messages. It does not actively participate in task scheduling or job execution while the primary controller is running normally.

If the primary CAS Controller fails, the Backup Controller will take over as the primary controller and assume responsibility for managing the CAS worker nodes and scheduling tasks. In this scenario, the CAS worker nodes will send their status updates and job results to the Backup Controller instead of the failed primary controller.[12]

In some systems, the Backup Controller can also be given jobs to execute as a CAS worker node. This can help to improve the system's overall performance by increasing the number of available processing resources. In this scenario, the Backup Controller can perform both the role of a CAS Controller and a CAS worker node.[13]

4.4.3 | Role 3: CAS Workers

Worker is the third role that can be assigned to a host for SAS Cloud Analytic Services. The CAS Worker is responsible for performing data processes and data analytics sent from the CAS Controller. For example, CAS Workers can perform data manipulations, transformations or computations on large/complex datasets. These computations are but not limited to: statistical analysis, machine learning models, text analysis, time series analysis, optimization, etc. Workers execute these computations using data stored on disk, in-memory, or in a distributed file system.

In a distributed environment, one host will be assigned as your controller and any additional hosts are considered workers (optional CAS Backup Controller). Workers increase the overall computing power of your distributed-server and provides a solution for a scalable (up/down), distributed, and fault-tolerant environment for data storage and data analysis because the worker manages the storage of data/metadata across multiple nodes. The amount of CAS Workers needed to create an optimized distributed environment is highly dependant on data size, computation type, and workload.

We can create two types of CAS configurations: a single-machine environment using symmetric multiprocessing **(SMP)**, or distributed server environment using massively parallel processing **(MPP)**.

[12] If the main CAS Controller fails, how does each CAS Worker respond to the Backup Controller with their completed jobs?
[13] Can the CAS Backup Controller be assigned work as well as passively monitor the main CAS Controller?

4.4.4 | Symmetric Multiprocessing (SMP)

The Symmetric multiprocessing (SMP) architecture is used when you want to run CAS on a single server or virtual machine (VM) with multiple CPU cores. This is called shared-memory architecture because all the CPUs share the same memory. When a job is submitted to CAS in an SMP architecture, it is processed by the worker node(s) in parallel using the shared memory. The results are returned to the controller node, which sends them back to the user.

A typical SMP architecture for CAS might consist of a single VM that serves as both the controller and worker node. The number of VMs required will depend on the size of your data and the processing requirements of your workload. For example, if you have a large dataset or complex analytical workloads, you might deploy CAS on a VM with 8 or 16 CPU cores.

Some examples of use cases for CAS on an SMP architecture include:

- Exploratory data analysis
- Statistical modeling and regression analysis
- Predictive analytics
- Machine learning and deep learning

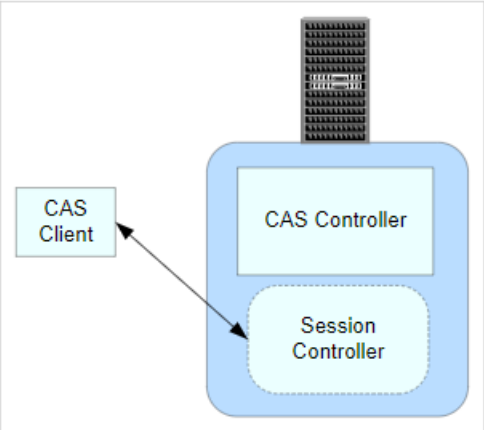

Figure 4.1: Single-machine CAS Server (STOLEN EXAMPLE)

4.4.5 | Massively Parallel Processing (MPP)

The Massively Parallel Processing (MPP) architecture is used when you want to run CAS on a cluster of multiple servers or VMs. This is called distributed-memory architecture because the data is partitioned and stored across multiple servers or nodes. When a job is submitted to CAS in an MPP architecture, it is distributed across the worker nodes in parallel. Each worker node processes its own subset of the data and returns the results to the controller node. The controller node then aggregates the results from all worker nodes and sends them back to the user.

A typical MPP architecture for CAS might consist of multiple VMs or servers, with some dedicated as controller nodes and others as worker nodes. The number of VMs or servers required will depend on the size of your data and the processing requirements of your workload. For example, you might choose to deploy CAS on a cluster of 10 or more VMs or servers to handle large-scale data processing tasks.

Some examples of use cases for CAS on an MPP architecture include:

- Big data processing and analysis
- High-performance computing
- Large-scale machine learning and deep learning
- High-throughput data processing, such as in genomics or drug discovery

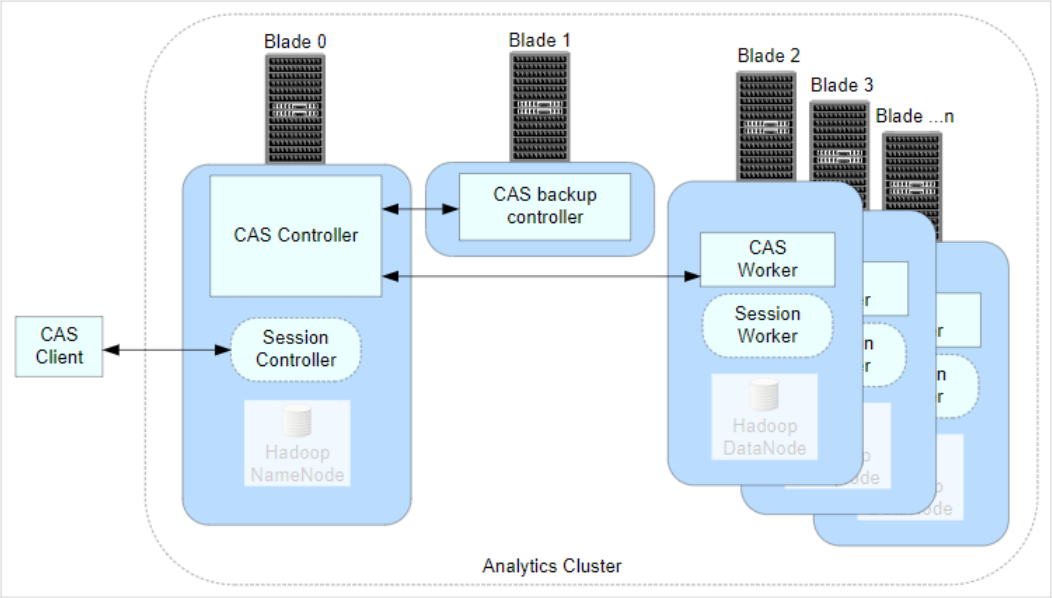

Figure 4.2: Distributed CAS Server (STOLEN EXAMPLE)

5 | Security and Risk Management

This chapter provides an introduction to security and risk management by covering key concepts such as data compliance, identity and access management (IAM), data governance, data encryption, and data backups. This chapter is not intended to be a comprehensive handbook for implementing proper security measures, but rather as an overview of the security measures to consider when developing a strategy for storing and accessing sensitive information.

5.1 | Data Compliance

- HIPAA Compliance
- UHM Compliance
- RCUH Compliance
- TASI Compliance
- State of Hawaii Compliance

5.2 | Identity and Access Management (IAM)

Identity and Access Management (IAM) is a security practice that safeguards sensitive information by allowing only authorized individuals to access confidential resources and data.

Identity management looks to confirm that an accessing user is who they say they are, whilst access management uses a users identity to determine which resource they are allowed to access.

IAM components can be classified into four major categories: authentication, authorisation, user management, and central user repository.

5.2.1 | Authentication

Authentication is a component of IAM in which a user is required to provide sufficient credentials to gain access to an application system.

Sufficient credentials for accessing sensitive healthcare information are defined as authentication methods that comply with the HIPAA Security Rule (Section 5.1). The HIPAA Security Rule requires covered entities to implement multi-factor authentication or an equivalent authentication method for accessing ePHI.

According to HIPAA, the multi-factor authentication method must use two of the following three elements:

- Something you know (Password or PIN)
- Something you have (Smart Card or Security Token)
- Something you are (Fingerprint or Facial Recognition)

Two new additional standards are not required but provide additional authentication methods:

- Somewhere you are (IP Address or Geo-location)
- Something you do (Signature or Gesture)

Once a user is authenticated, a session is created to allow the user to interact with the application system. The session will remain open until the user's task is completed or through termination by other means (e.g., timeout). By centrally maintaining the session of a user, the authentication module can provide single sign-on services.

Single sign-on (SSO) is a mechanism that allows users to authenticate once and access multiple systems or applications without having to re-enter their credentials. SSO simplifies access control and user permissions by providing a centrally managed solution for user authentication policies across all systems. There are several options when deciding on a SSO solution. (e.g., LDAP, OAuth, SAML, RADIUS, PKI, etc).

5.2.2 | Authorization

Authorization is a component of IAM in which a user is given permission to access a particular resource.

This component comes after a user has successfully authenticated to an application system with sufficient credentials. Authorization is performed by checking the resource access request (e.g., web-based application URL), against an IAM policy store and is the core module that implements Role-Based/Attribute-based, access control.

- Role-Based Access Control (RBAC) is a method of access control that assigns roles to users and access permissions to those roles in order to provide a centrally managed solution for authorization.
- Attribute-Based Access Control (ABAC) is a method of access control that assigns permissions based on a user's attributes (e.g., job title, location, department).

The authorization model can provide more complex access control policies other than user role/groups and user attributes (e.g., access channels, time, resource requested, external data, business rules).

5.2.3 | Central User Repository

The Central User Repository (CUR) stores and delivers identity information in order to verify credentials submitted from clients. Identity information is equivalent to user account information (e.g., usernames, passwords, etc). The most common CUR protocol is the Lightweight Directory Access Protocol (LDAP).

LDAP is a protocol for accessing and maintaining distributed directory information services over an Internet Protocol network in order to provide a centrally managed authentication and authorization solution for application systems.

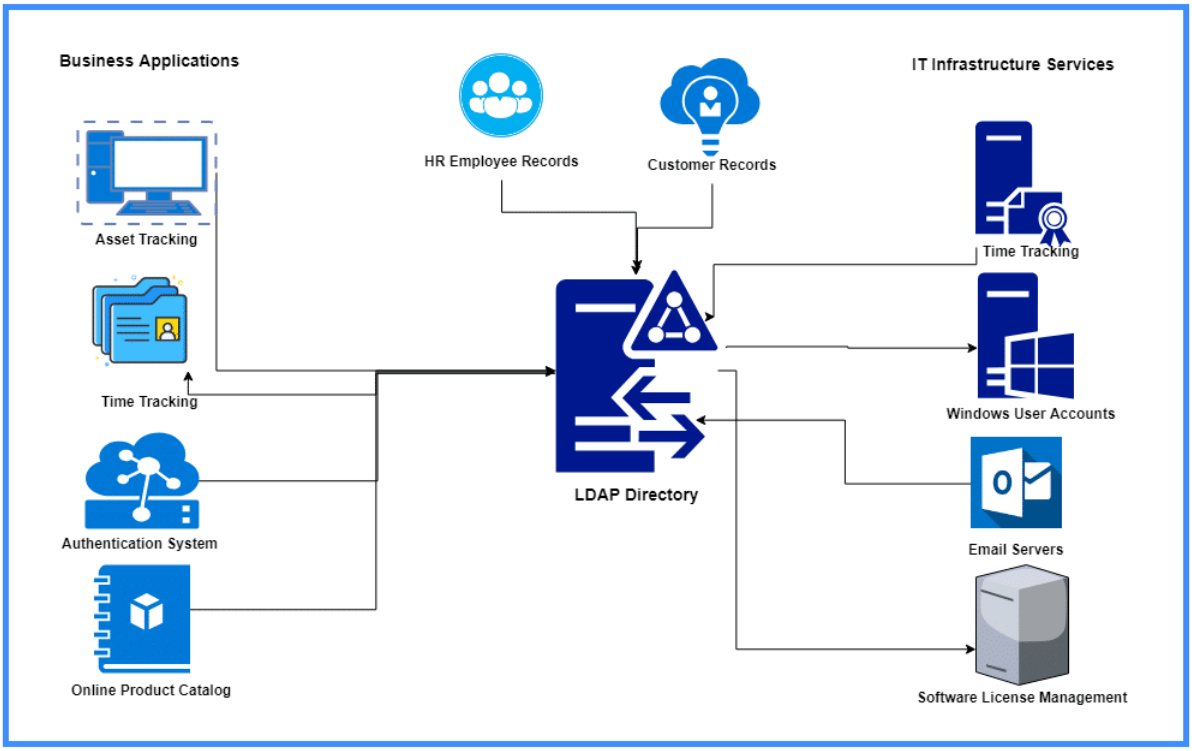

Figure 5.1: Lightweight Directory Access Protocol (STOLEN EXAMPLE)

LDAP allows system administrators to manage user accounts, configure access and permissions, and monitor and audit user activity.

5.2.4 | User Management

- Onboarding Process: [Making new accounts, role generation]

- Offboarding Process: [Removing account permissions and archival/auditing]

5.3 | Data Governance

To be completed during the Data Governance Seminar in early May 2023.

5.4 | Data Encryption

SSL & TLS protocols for data encryption.

- Data At-Rest (Encrypted)
- Data In-Transit (Encrypted)
- Data In-Memory (Unencrypted)

5.5 | Data Backups

- Data In-Memory:
- Data Main Storage:
- Data Cold Site

6 | Massively Learning Activities I - Initial Deployment

TASI has been contracted by CNMI to create an infrastructure that allows for data analytics on Protected Health Information (PHI). This infrastructure will initially be hosted on-premises, with plans to move towards a hybrid solution in the future. To achieve this, we will be providing a Platform as a Service (PaaS) solution, by hosting SAS Viya services on our own hardware and allowing tenants to access and utilize the platform for their own analytics applications.

The tenants, including APCD, CMNI, CMA, Criminal Justice, and several Education environments, will provide the necessary data, which will be submitted to an ETL data pipeline for processing before being sent to SAS on-prem servers. Once the data has been processed, tenants may perform data analytics using advanced algorithms in SAS programming language.

To ensure secure operations, we will configure the security relationships between the software, hardware, and tenants using LDAP, security groups, encryption and other related tools. Our goal is to architect a high-performance infrastructure that allows for advanced data analytics while maintaining the confidentiality and security of PHI.

Due to SAS being a time sensitive project, the initial deployment will have SAS suites and VMs installed on existing hardware, with plans to migrate the infrastructure to newly acquired hardware in the future.

6.1 | Planning

The System Development Life-cycle (SDLC) is a project management model that defines different stages that are necessary to bring a project from conception to deployment and later maintenance. The SDLC model consists of several phases, which typically include requirements gathering, design, development, testing, deployment, and maintenance. The specific activities within each phase may vary depending on the project and the organization, but the basic principles are the same. The SDLC model is a flexible framework that can be adapted to suit the needs of different projects and organizations. It provides a systematic approach to software development that helps ensure that software is built efficiently, effectively, and with minimal risk.

Massively Learning Activities will follow a similar variation to the SDLC project management model where each SDLC stage will correspond to a subsection in this chapter.

6.2 | Requirements of Analysis

6.3 | Security and Risks

Security In-Depth

- Identity and Access Management single sign on, guest, and user /security groups
- Liability % of TASI fault, % of customer fault
- Data Encryption data in transit and data at rest
- Compliance HIPAA/HITECH, other
-

6.4 | Design and Protoyping I

Massively Learning Activities (MLA) is divided into two phases, (1) the initial deployment of SAS on existing infrastructure and later (2) the migration of SAS onto scaled infrastructure.

In either deployment stage, SAS Viya will be deployed in a multi-tenant environment. A multi-tenant deployment of SAS Viya allows for a single deployment to serve multiple customers[14]. These customers can share some physical resources while remaining logically separated. A multi-tenant deployment allows for these distinct groups to share IT resources in a secure and cost-effect manner. Multi-tenancy deploys into a Kubernetes namespace. The deployment includes a provider tenant, shared mid-tier services, application-specific database schemas, shared applications, and a designated SAS administrator for the

[14]Customers are tenants (etc: CNMI, APCD, CMA, Med-Quest, UH Education) but each tenant has its own set of users and groups.

provider tenant. Administrators with elevated Kubernetes privileges onboard one or more tenants. After tenant on-boarding, Kubernetes administrators onboard one or more Cloud Analytic Services into each new tenant, then each CAS server is uniquely configured during the on-boarding process to meet the specific tenant requirements.

The final and completed deployment of SAS Viya will expect a total of 8 tenants:

- **Tenant 1**: Commonwealth of the Northern Mariana Islands (CNMI)
- **Tenant 2**: All-Payer Claims Database (APCD)
- **Tenant 3**: Centers for Medicare & Medicaid Services (CMA)
- **Tenant 4**: Med-Quest
- **Tenant 5**: UH Education 1
- **Tenant 6**: UH Education 2
- **Tenant 7**: UH Education 3
- **Tenant 8**: UH Education 4

6.4.1 | Initial Deployment

The initial deployment of MLA will involve installing SAS Viya and SAS DMA on existing infrastructure[15]. The existing infrastructure is an available Dell PowerEdge FX2 Enclosure located in TASI's NOC. The PowerEdge FX2 Enclosure is a 2U hybrid rack-based computing platform that combines multiple blades[16] into a single enclosure to increase the efficiency and reduce the cost of rack-based systems. This multi-blade enclosure has a XXGB connection to a storage pool (describe me).

Each blade is installed with **ESXi 6.7** as the host operating system and is equipped with **12 CPU(s) w/ XX cores**, **256GB** of RAM, and **XXGB** of personal storage. Theses resources will be logically separated to configure a multi-tenant environment. These virtual machines (VMs) will be created and managed through vCenter (and later through SAS), which is running on a separate server. The VMs responsible for a CAS role will use RHEL 7.X as the host operating system.

CNMI, APCD, CMA, and Med-Quest will be the initial four tenants with each tenant having their own deployment configuration based on their respective requirements. CNMI and APDC will be configured in a 5-server environment [17] and every other tenant will be configured in a 3-server environment[18]. The other tenants that are yet to be added will be considered during the migration stage of MLA.

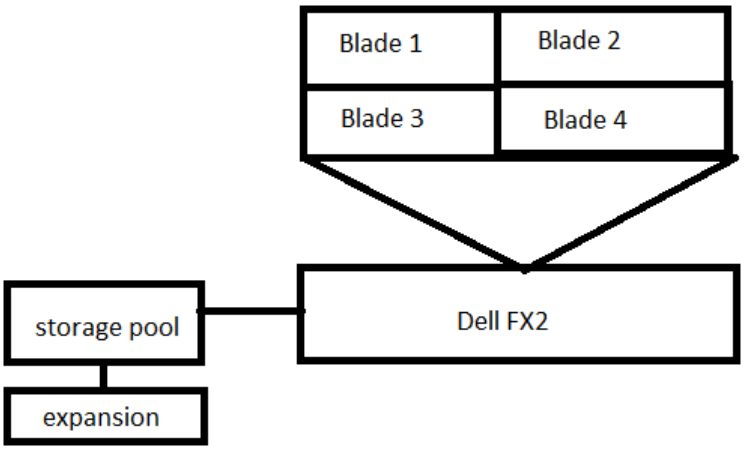

Figure 6.1: TASI On-Premise Environment (needs VISIO)

[15] Infrastructure is used to describe existing hardware available to use on-premises at TASI/PHIDC.
[16] Blades are complete servers in a smaller form factor that have their own CPU(s), memory, storage, and networking components.
[17] 5 Server: (1) Primary CAS Controller, (2) Backup CAS Controller, (3) CAS Worker 1, (4) CAS Worker 2, (5) CAS Worker 3
[18] 3 Server: (1) Primary CAS Controller, (2) Backup CAS Controller, (3) CAS Worker 1

There are several factors that impact the configuration of SAS technologies in a virtualized environment: **(1) high availability and redundancy**; **(2) optimization**; and **(3) security and compliance**.

(1) Consider that although virtualizing the CAS Controller and Backup Controller on the same hardware can offer several benefits, such as cost savings, simplified management, and easier backup and recovery processes, it will be better to virtualize them on separate hardware to provide high availability and redundancy of CAS controllers (Figure 6.2), in the case of unexpected downtime. As for CAS Workers, it is recommended to virtualize them on separate hardware to reduce resource conflicts but it is not necessary.

(2) The performance overhead of using SAS Viya and CAS on VMs instead of dedicated hardware can depend on several factors including workload characteristics, available hardware resources, and the virtualization technology used.

In section 6.3.1, if the advantage of distributing the SAS components across blades is for load balancing, it would be good to include a comment to that effect – otherwise, it would probably be more efficient to have them all on the same blade.

(3) In a logically separated multi-tenant environment, MLA must ensure network and data isolation between each tenant, data encryption for data at rest and in transit, a well configured access control list for hardware and software, and compliance certification for handling PHI data (i.e., HIPAA, PCI DSS, SOC 2, HITECH, Hawaii Information Privacy Act, etc).

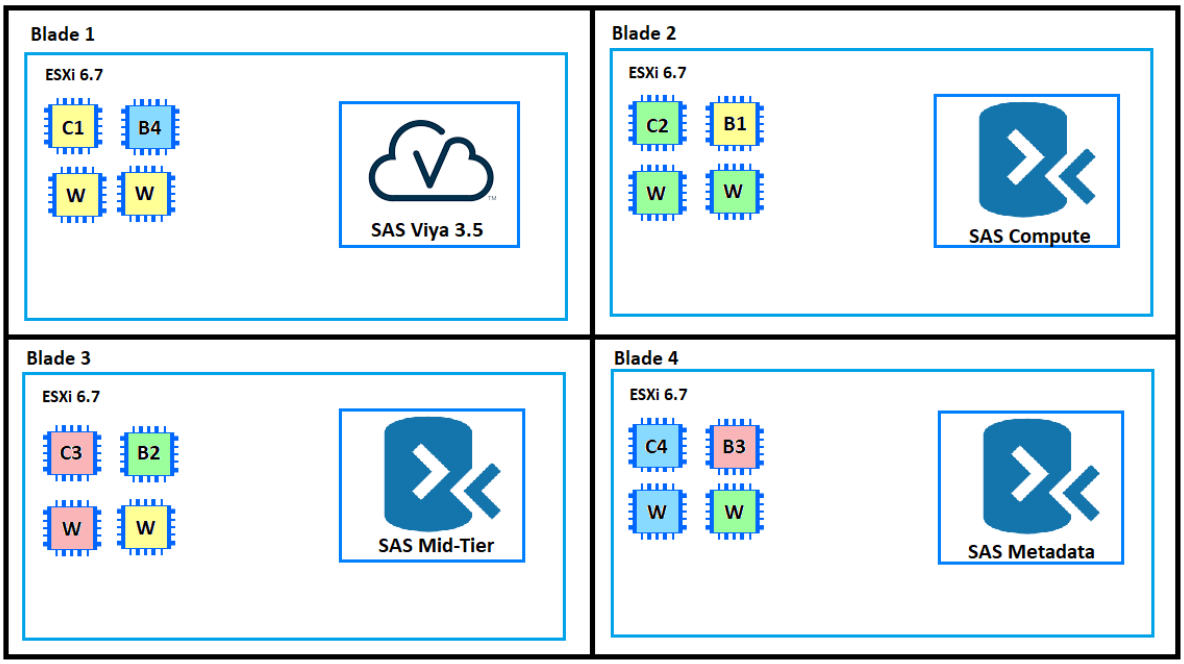

Figure 6.2: Initial Multi-Tenant Deployment (needs VISIO)

To maximize resource efficiency, CAS nodes will be evenly distributed across each blade, where a blade will consist of one controller, one backup controller, and two workers. The controller and backup controller, configured on the same system, will belong to separate tenants. The workers will also belong to separate tenants but each blade will have at least one related controller and worker per system.

Subsequently, four additional VMs will be created to support the installation of SAS Viya and SAS DMA. SAS Viya will be installed as software on top of a RHEL 3.7X VM instance, in Blade 1. SAS DMA consists of three software components that will installed as software on top of Windows Server 2019 VM instances, in Blades' 2, 3, and 4.

6.5 | Deployment and Prototyping I (Initial)

6.6 | Testing & Integration I

6.7 | Operations and Maintenance I

7 | Hyper-Converged Infrastructure (HCI)

HCI, or Hyper-Converged Infrastructure, is a software-defined, unified system that combines the traditional elements of IT infrastructure (e.g., compute, networking, management, storage) with virtualization, simplifying infrastructure, reducing costs, and increasing scalability and flexibility. In a traditional IT Infrastructure, servers, storage networks, and storage systems are physically separated as stand alone hardware devices (e.g., servers, network switches, disk arrays). Consolidating these components into a single, integrated system simplifies the management, deployment, configuration, and maintenance of your IT Infrastructure.

The benefits of an HCI environment include:

- Scalability: Designed to scale out by adding additional nodes on-demand to your system.
- Efficiency: Improve resource utilization by using or eliminating idle storage capacity.
- Agility: Quickly deploy new applications and workloads without extensive planning across systems.
- Data Protection: Integrated backup and disaster recovery.
- Reduced Hardware Costs: Reduce the amount of hardware required reducing CAPEX[19]/OPEX[20] costs.

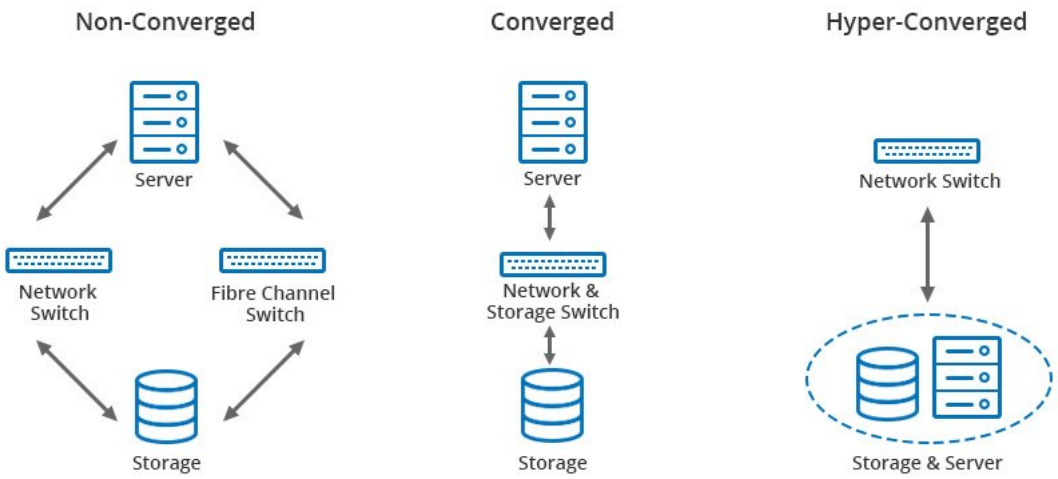

Figure 7.1: Types of IT Infrastructures (STOLEN EXAMPLE)

In HCI, multiple servers or nodes are combined to create a cluster. These nodes share their computing and storage resources with each other to create a multi-purpose integrated system. The design of your HCI cluster will depend on your specific needs and requirements.

The software that powers HCI also includes a management layer, which automates tasks like resource provisioning, data migration, and load balancing. This layer abstracts the hardware, making it easier to manage and deploy your IT infrastructure. Overall, HCI is a powerful and flexible solution that can help organizations streamline their IT operations, reduce costs, and improve efficiency.

[19] Capital expenditure is the cost a business incurs to acquire assets that will provide benefits beyond the current year.
[20] Operating expenses refer to the money a company spends to run day-to-day operations.

8 | Massively Learning Activities II - Migration Deployment

TASI has been contracted by CNMI to create an infrastructure that allows for data analytics on Protected Health Information (PHI). This infrastructure will initially be hosted on-premises, with plans to move towards a hybrid solution in the future. To achieve this, we will be providing a Platform as a Service (PaaS) solution, by hosting SAS Viya services on our own hardware and allowing tenants to access and utilize the platform for their own analytics applications.

The tenants, including APCD, CMNI, CMA, Criminal Justice, and several Education environments, will provide the necessary data, which will be submitted to an ETL data pipeline for processing before being sent to SAS on-prem servers. Once the data has been processed, tenants may perform data analytics using advanced algorithms in SAS programming language.

To ensure secure operations, we will configure the security relationships between the software, hardware, and tenants using LDAP, security groups, encryption and other related tools. Our goal is to architect a high-performance infrastructure that allows for advanced data analytics while maintaining the confidentiality and security of PHI.

Due to SAS being a time sensitive project, the initial deployment will have SAS suites and VMs installed on existing hardware, with plans to migrate the infrastructure to newly acquired hardware in the future.

8.1 | Planning

The System Development Lifecycle (SDLC) is a project management model that defines different stages that are necessary to bring a project from conception to deployment and later maintenance. The SDLC model consists of several phases, which typically include requirements gathering, design, development, testing, deployment, and maintenance. The specific activities within each phase may vary depending on the project and the organization, but the basic principles are the same. The SDLC model is a flexible framework that can be adapted to suit the needs of different projects and organizations. It provides a systematic approach to software development that helps ensure that software is built efficiently, effectively, and with minimal risk.

Massively Learning Activities will follow a similar variation to the SDLC project management model where each SDLC stage will correspond to a subsection in this chapter.

8.2 | Requirements of Analysis

Refer to the sizing documents (2) and the current resources document comparison to see what we are missing.

8.3 | Security and Risks

Security In-Depth

- LDAP
- VMware Security Policies
- SAS Security Policies
- HIPPA, other Federal Laws

8.4 | Deployment and Prototyping II (Migration)

VMotion in action (See 6.1.2).

8.5 | Testing & Integration II

8.6 | Operations and Maintenance II

A | Appendix A title

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.