

# **REPORT ON INTERNSHIP AT COCHIN SHIPYARD LIMITED**



**MASTER OF COMPUTER APPLICATIONS  
(2023-2025)**

**SUBMITTED BY  
MERIN JAI**

May 2024

## ACKNOWLEDGEMENT

I am delighted to express my heartfelt gratitude to the esteemed professionals who made my internship experience at Cochin Shipyard Limited truly enriching and rewarding. I am deeply grateful to **Mr. Harikrishnan M K, Assistant General Manager (IT)** at Cochin Shipyard Limited, for his exceptional mentorship and insightful guidance, which were pivotal in shaping my learning experience.

I express my deepest thanks to **Mrs. Bini Joseph, Senior Manager (IT)** at Cochin Shipyard Limited for your support and guidance throughout my internship. Your patience and willingness to share your knowledge have greatly enhanced my understanding of IT operations and project management.

I would also like to extend my thanks **Mr. Jeevan Raveendran and Mr. Vinayan M T, Training Institute**, Cochin Shipyard Limited for their support and coordination in facilitating this internship program. The training and exposure provided have been invaluable in developing my skills and professional growth.

I am also thankful to all the staff members of the IT department and the Training Institute for their assistance and cooperation during my internship.

I am indebted to **Fr. Dr. Saju M. D. CMI Principal**, Rajagiri College of Social Sciences for his unwavering support and positive energy throughout the duration of the internship.

I would like to convey my sincere gratitude to **Dr. Bindiya M Varghese**, the highly esteemed Dean of the Computer Science Department, for her unwavering dedication in serving as my mentor throughout the entirety of my internship.

Furthermore, I would like to express my sincere gratitude to all the faculty members of the department and acquaintances who provided invaluable support and guidance throughout my internship, contributing significantly to my successful completion of the program.

Merin Jai

## **ABSTRACT**

Cochin Shipyard Limited is one of the largest shipbuilding and maintenance facilities in India. Located in Kochi, Kerala, it was established in the year 1972 as a government-owned company under the Ministry of Shipping, Government of India. Initially, it was primarily focused on ship repair and gradually expanded into shipbuilding. Cochin Shipyard is known for its commitment to quality and safety standards. It follows international best practices in shipbuilding and repair, ensuring reliable and durable vessels. The shipyard also focuses on training and skill development programs for its employees and offers training to marine engineering students through its training institute.

During my internship at Cochin Shipyard Limited's Information Systems Department (ISD), I had the opportunity to work on two significant automation projects using leading robotic process automation (RPA) platforms. The first project involved leveraging OpenRPA to develop a task bot that automated the processing of Excel invoices. This bot efficiently extracted data, generated daily reports, and seamlessly emailed them to the respective manager, streamlining the invoicing process and enhancing reporting accuracy.

In the second project, I utilized UiPath's capabilities to design an intelligent bot focused on automating the conversion of PDF invoices into Excel format. This intelligent bot incorporated advanced data extraction techniques, allowing for precise and error-free invoice processing. The automation of PDF-to-Excel conversions significantly reduced manual effort and improved workflow efficiency within the organization.

Alongside project work, I gained invaluable insights into the functioning of ISD departments, their various web applications, intranet, servers, and data centers. I witnessed firsthand how ISD collaborates with departments across the organization to streamline processes, improve efficiency, and enhance overall business operations.

My internship at ISD was a transformative experience that not only allowed me to apply cutting-edge RPA technologies but also provided me with insights into the seamless integration of IT solutions across departments within a dynamic organization like Cochin Shipyard Limited.

# TABLE OF CONTENTS

<b>CHAPTER-1 INTRODUCTION.....</b>	<b>Error! Bookmark not defined.</b>
<b>1.1 INTRODUCTION.....</b>	<b>Error! Bookmark not defined.</b>
<b>1.2 ORGANIZATION PROFILE.....</b>	<b>Error! Bookmark not defined.</b>
<b>CHAPTER-2 METHODS AND DISCUSSION .....</b>	<b>Error! Bookmark not defined.</b>
<b>2.1 Intranet at Cochin Shipyard: Synergy Portal .....</b>	<b>Error! Bookmark not defined.</b>
<b>2.2 SAP .....</b>	<b>Error! Bookmark not defined.</b>
<b>2.2.1 FLM (File Lifecycle Management).....</b>	<b>Error! Bookmark not defined.</b>
<b>2.3 Security, Networking and Data Centers.....</b>	<b>Error! Bookmark not defined.</b>
<b>2.4 Active Directory .....</b>	<b>Error! Bookmark not defined.</b>
<b>2.4.1 Integration with Computer Login.....</b>	<b>Error! Bookmark not defined.</b>
<b>2.4.2 Integration with VPN Access .....</b>	<b>Error! Bookmark not defined.</b>
<b>2.4.3 Integration with Applications .....</b>	<b>Error! Bookmark not defined.</b>
<b>CHAPTER-3 PROJECT .....</b>	<b>Error! Bookmark not defined.</b>
<b>3.1 Problem Definition .....</b>	<b>Error! Bookmark not defined.</b>
<b>3.2 Proposed Solution: Automation through Task Bot Development.....</b>	<b>Error! Bookmark not defined.</b>
<b>3.3 Scope of work .....</b>	<b>Error! Bookmark not defined.</b>
<b>3.4 Methodology and Implementation .....</b>	<b>Error! Bookmark not defined.</b>
<b>3.4.1 Automation using OpenRPA.....</b>	<b>Error! Bookmark not defined.</b>
<b>3.4.2 Automation with UiPath .....</b>	<b>Error! Bookmark not defined.</b>
<b>3.5 Conclusion .....</b>	<b>Error! Bookmark not defined.</b>
<b>WEEKLY OVERVIEW .....</b>	<b>6</b>

# **Chapter-1**

## **Introduction**

## 1.1 Introduction

The internship at Cochin Shipyard Limited provided a valuable opportunity to gain practical insights into the shipbuilding and repair industry. Cochin Shipyard, established in 1972, is a renowned shipyard in India known for its expertise in constructing various vessels, providing repair services, and offering marine engineering training through its Marine Engineering Training Institute (METI).

In the last five decades CSL has emerged as a forerunner in the Indian Shipbuilding & Ship repair Industry and also a well-known player on the global shipbuilding front. CSL can build ships up to 1,10,000 DWT and repair ships upto 1,25,000 DWT. CSL has many a feather on its cap like it has built and delivered the largest warship ever to be built in India – INS Vikrant to the Indian Navy, has delivered two of India’s largest double hull Aframax tankers each of 95,000 DWT, has built various types of vessels including Tankers, Bulk Carriers, Port Crafts, Passenger Vessels etc. CSL has secured shipbuilding orders from internationally renowned companies and has exported around 47 ships to various commercial clients outside India such as National Petroleum Construction Company (Abu Dhabi), the Clipper Group (Bahamas), Vroon Offshore (Netherlands) and SIGBA AS (Norway).

## 1.2 Organization Profile

Cochin Shipyard is structured into three main divisions:

**Ship Repair Division:** This division specializes in the repair and maintenance of various types of vessels, including tankers, bulk carriers, and offshore vessels. It offers services such as hull repairs, machinery overhauls, and painting.

**Shipbuilding Division:** The shipbuilding division is responsible for constructing new vessels, including cargo ships, passenger ferries, and offshore support vessels. It encompasses various departments, including Information Systems Department (ISD), which supports the shipbuilding activities through technology and information management.

**Marine Engineering Training Institute (METI):** METI is the training arm of Cochin Shipyard, providing specialized courses in marine engineering and naval architecture to aspiring professionals in the maritime industry.

### Role of Information Systems Department (ISD) within Shipbuilding

The Information Systems Department (ISD) plays a critical role within the Shipbuilding Division of Cochin Shipyard. It is responsible for:

---

- Managing and maintaining the shipyard's IT infrastructure, including network systems, servers, and software applications used in ship design and production.
- Developing and implementing information systems to optimize project management, procurement processes, and quality control within the shipbuilding projects.

Supporting digital transformation initiatives by integrating technology into shipyard operations, enhancing efficiency and productivity.

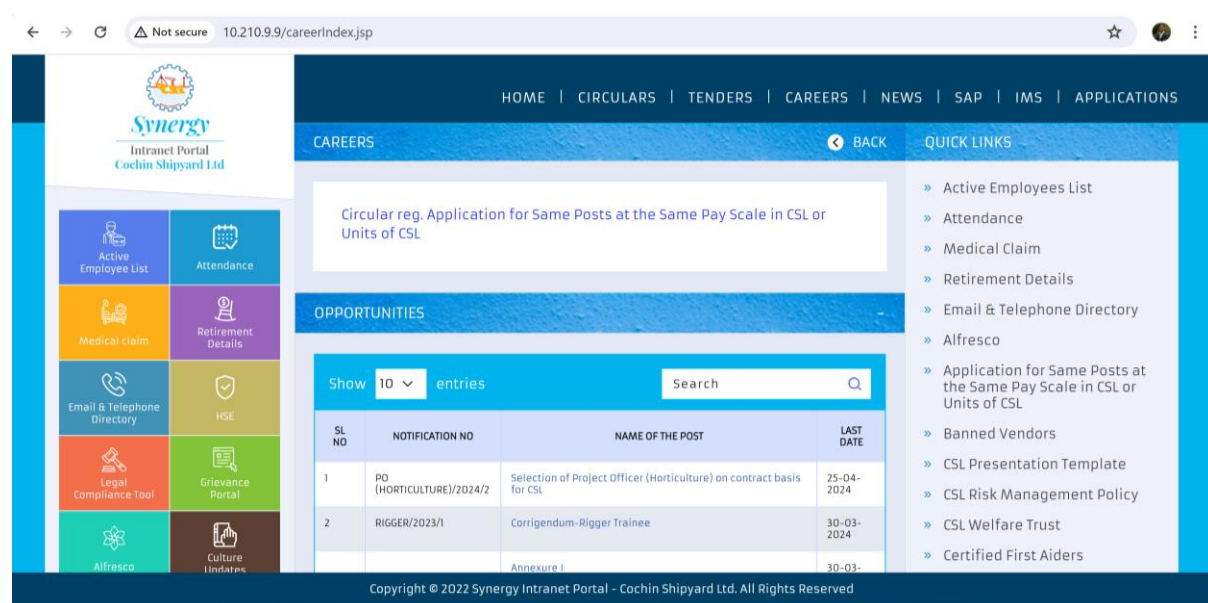
## **Chapter-2**

### **Methods and Discussions**



## 2.1 Intranet at Cochin Shipyard: Synergy Portal

Cochin Shipyard employs an Intranet system known as the Synergy portal, serving as a centralized platform for internal communication and collaboration among employees. The Synergy portal offers a range of features and functionalities designed to enhance employee engagement, information dissemination, and workflow efficiency.



It is a Content management-based application which is managed by the ISD. It was developed in Java and uses SQLServer Enterprise Edition as its backend. It is having various functionalities such as Active Employee List, Attendance details, Retirement details, circulars, grievance portal and so.

It also has the SAP (ESS) portal. In the Employee Self Service portal, each employee can view their payroll, leave requests and so on

Some of the web applications in Intranet which are managed by ISD are:

### i. SMAX(Service Management Automation)

It is a ticketing tool which was developed for Disaster recovery. Suppose we need to connect a system to a network then we must raise a ticket through SMAX.

### ii. Fluence DMS

It is Document Management system hosted on Next cloud. Nextcloud is an open-source, self-hosted cloud storage and collaboration platform that allows

individuals and organizations to store, share, and sync files securely across devices. It offers a range of features similar to popular cloud services like Dropbox, Google Drive, and Microsoft OneDrive, but with a focus on privacy, security, and customization.

**iii. NOVEX (Noive to Expert)**

It is a web app that provide a E-learning platform for the employees for the training which is hosted in cloud. There will be training materials provided by the HR department.

## **2.2 SAP (ERP)**

SAP (Systems, Applications, and Products) is a German multinational leading enterprise resource planning (ERP) software suite used by organizations worldwide, including Cochin Shipyard Limited. SAP offers a wide range of modules that cater to different business functions and processes. The ERP (Enterprise Resource Planning) system used in Cochin Shipyard is SAP. It was customized as per the business needs.

Initially SAP was based on SAP ECC, or SAP ERP Central Component SAP ECC is scalable, supporting small to large enterprises and adapting to changing business needs and growth. It was migrated to S/4HANA. SAP S/4HANA is the successor to SAP ECC and offers enhanced features such as in-memory computing, simplified data models, advanced analytics, and a modern user experience.

There are various Functional modules in SAP. Some of which are used in Cochin shipyard is:

- i. Financial Accounting Controlling (FICO):** Manages financial transactions, accounting, reporting, and financial planning and focuses on cost controlling, budgeting, and profitability analysis.
- ii. Materials Management (MM):** Manages procurement processes, inventory management, purchasing, and vendor/supplier relationships.
- iii. Sales and Distribution (SD):** Handles sales order processing, pricing, billing, shipping, and customer relationship management.
- iv. Quality Management (QM):** Ensures quality control and compliance with quality standards throughout the production process.

- v. **Plant Maintenance (PM):** Manages equipment, maintenance planning, work orders, and asset management.
- vi. **Human Capital Management (HCM):** Handles human resources, payroll, talent management, and workforce planning.
- vii. **Project System (PS):** Manages project planning, execution, and monitoring.
- viii. **Supplier Relationship Management (SRM):** Manages supplier relationships, procurement processes, sourcing, and contract management.
- ix. **Production Planning (PP):** Deals with production planning, scheduling, and manufacturing processes.
- x. **Enterprise Asset Management (EAM):** Manages assets, maintenance activities, work orders, and asset lifecycle management.
- xi. **Document Management System (DMS):** The Document Management System (DMS) module in SAP is designed to manage electronic documents and files within an organization.

## T- Codes

In SAP, T-codes (Transaction Codes) play a significant role in navigating and accessing various functionalities within the system. Transaction Codes, often abbreviated as T-codes, are shortcuts or codes used to access specific transactions or functions within SAP. Each T-code corresponds to a particular task or process, making it easier for users to navigate through the system and perform their daily activities efficiently.

## ABAP in SAP S/4HANA

ABAP (Advanced Business Application Programming) is a programming language used in SAP's software systems, including SAP S/4HANA. ABAP is primarily used for developing and customizing applications within the SAP environment. It allows developers to create reports, forms, interfaces, and workflows tailored to the specific needs of an organization.

In SAP S/4HANA, ABAP continues to play a crucial role in extending and enhancing the functionality of the ERP system. It provides developers with tools and frameworks to build innovative solutions, integrate with external systems, and optimize business processes.

## Management by Wipro at Cochin Shipyard

At Cochin Shipyard, the management of SAP S/4HANA, including ABAP development, was handled by Wipro. Wipro, a leading global IT consulting and services company, specializes in SAP implementation, customization, and support services. Their expertise in managing SAP environments ensured the smooth operation and customization of SAP S/4HANA at Cochin Shipyard, meeting the organization's business requirements effectively.

### **Current AMC by EY**

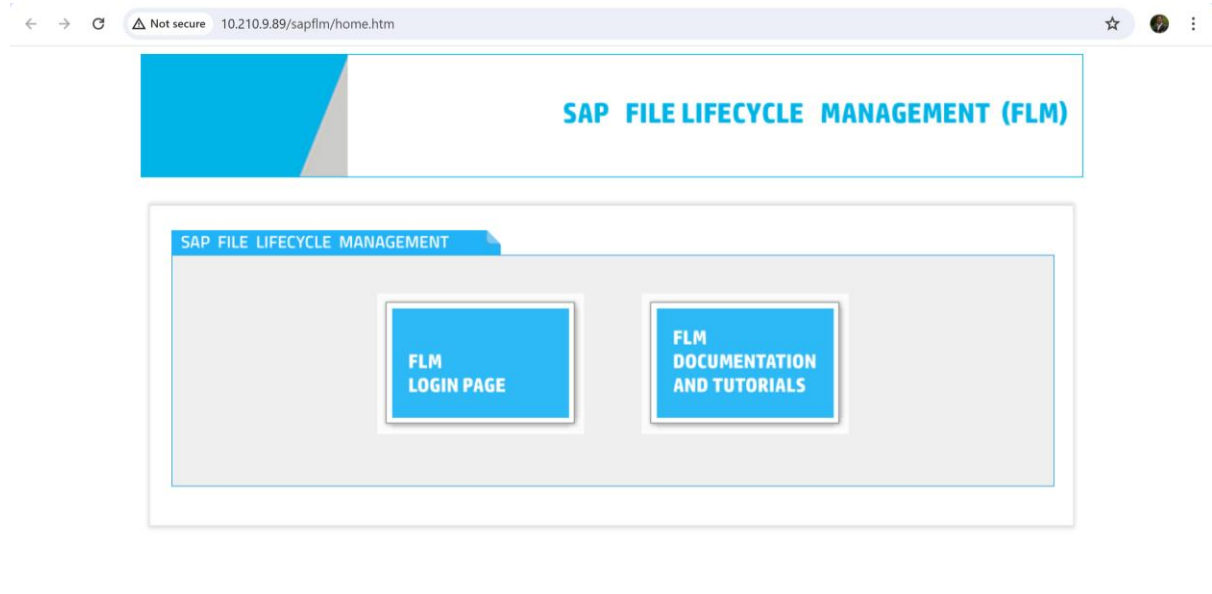
The Annual Maintenance Contract (AMC) for SAP S/4HANA, including ABAP services, is now being managed by EY (Ernst & Young) at Cochin Shipyard. EY is a multinational professional services firm known for its consulting, auditing, and advisory services. With EY's involvement in the AMC, Cochin Shipyard can benefit from EY's expertise in SAP solutions, ensuring continued support, maintenance, and optimization of their SAP S/4HANA environment.

This collaborative effort between Wipro and EY reflects Cochin Shipyard's commitment to leveraging top-tier IT services providers for the management and enhancement of their SAP S/4HANA system, including ABAP development and ongoing support.

The SAP

### **2.2.1 FLM (File Lifecycle Management)**

File Lifecycle Management (FLM) at Cochin Shipyard represents a significant shift in document handling practices, moving from physical files to a digitalized approach. This transition is governed by the Delegation of Power (DOP) framework, ensuring efficient and structured management of documents throughout their lifecycle.



Implemented as a product of SAP, FLM reflects Cochin Shipyard's commitment to leveraging cutting-edge technology for improved operational efficiency and productivity. The technical aspects of FLM, including system configuration and maintenance, are overseen by the proficient Information Systems Department (ISD) at Cochin Shipyard, ensuring seamless integration and optimization of FLM to meet the organization's unique requirements.

The adoption of FLM brings numerous benefits to Cochin Shipyard, including streamlined document access and retrieval, enhanced collaboration among teams, improved compliance with regulatory standards and internal policies, and significant cost and time savings. By centralizing document storage and enforcing standardized practices, FLM promotes a digital transformation within the organization, reducing reliance on physical paperwork and enabling faster and more efficient workflows. This strategic initiative aligns Cochin Shipyard with industry best practices and underscores its commitment to embracing digital innovation for enhanced operational excellence and competitiveness.

## 2.3 Security, Networking and Data Centers

Networking infrastructure plays a crucial role in facilitating seamless communication and data exchange within Cochin Shipyard Limited (CSL). Some of the networking and security practices at CSL:

- 
- **SMAX Ticketing System:** CSL utilizes the SMAX (Service Management Automation X) ticketing system for network-related requests and issues. Employees can raise tickets through SMAX to request network access, troubleshoot connectivity issues, or request network configuration changes.
  - **VLAN Segmentation:** CSL employs VLAN (Virtual Local Area Network) segmentation to logically divide the network into separate segments. Each department within CSL is assigned a specific VLAN to ensure network isolation and optimize network performance. For instance, the Ship Building (SB) department may be allocated the 90 series of IP addresses, while the IT department may have the 99 series.
  - **IP Address Assignment:** Departments at CSL are assigned IP addresses based on the VLAN segmentation. This IP addressing scheme helps in identifying and routing traffic within the network efficiently. It also allows for easier management and troubleshooting of network devices and systems.
  - **MAC Address Security:** CSL implements MAC address binding for enhanced network security. MAC addresses, which are unique identifiers assigned to network devices, are bound to specific ports or interfaces. This binding ensures that only authorized devices with registered MAC addresses can connect to the network, enhancing security against unauthorized access.
  - **Data Centers:** CSL's data centers serve as the backbone of its IT infrastructure, hosting critical applications, databases, and services. The data centers are equipped with robust networking equipment, including routers, switches, firewalls, and load balancers, to ensure high availability, scalability, and security of IT resources.
  - **Redundancy and Failover:** CSL's networking architecture includes redundancy and failover mechanisms to minimize downtime and ensure business continuity. Redundant network links, power sources, and data center components are implemented to mitigate the risk of service interruptions due to hardware failures or network issues.
-

- 
- **Network Monitoring and Management:** CSL employs network monitoring tools and protocols to proactively monitor network performance, detect anomalies, and troubleshoot issues in real time. Network administrators utilize these tools to optimize network performance, enforce security policies, and ensure compliance with industry standards.
  - **Future-Ready Infrastructure:** CSL continually invests in upgrading and modernizing its networking infrastructure to adapt to evolving technology trends, such as cloud computing, IoT (Internet of Things), and digital transformation initiatives. This ensures that the network infrastructure remains scalable, resilient, and capable of supporting future business requirements.

### 2.3.1 Data Centers and Servers

At Cochin Shipyard Limited (CSL), the infrastructure supporting critical applications and services is housed within the in-house data center. Here's an overview of CSL's data centers and servers based on the provided points:

- **Server Placement and Resources:** The servers at CSL are located within the in-house data center, ensuring centralized management, security, and reliability. The resources required for these servers include processor speed, memory (RAM), and storage capacity. CSL's IT infrastructure is designed to meet the computational and storage demands of its operations and business applications effectively.
- **Storage and RAID Levels:** Storage within CSL's servers is organized using various RAID (Redundant Array of Independent Disks) levels. RAID configurations provide data redundancy, fault tolerance, and performance optimization, ensuring data integrity and availability across CSL's systems.
- **Server Environments:** CSL follows a structured approach to server environments for system development, testing, and production:
  - **Development Server:** This server is dedicated to modifications, changes, and software development activities. It provides a sandbox environment for developers to test and implement new features without affecting live data.

- 
- **Quality Servers:** Quality servers are utilized for testing purposes, primarily conducted by users to validate system functionality, performance, and compatibility before deployment.
  - **Production Server:** The production server hosts the live system and is accessible to users for day-to-day operations and transactions. It ensures reliability, uptime, and data integrity for critical business processes.
- **SAP ERP System Servers:** CSL's SAP ERP system, including SAP S/4HANA and SAP portal - ESS (Employee Self-Service), requires dedicated servers for development, testing (quality), and production environments. Additionally, CSL integrates seamlessly with databases using data synchronization (datasync) and process orchestration to ensure data consistency and workflow efficiency.
- **SAP S/4HANA Servers:** These servers host various modules written in ABAP and require separate instances for development, quality assurance, and production to facilitate software development lifecycle processes.
  - **SAP Portal - ESS Servers:** The ESS (Employee Self-Service) portal, developed in WebDynpro, also follows a similar three-server setup for development, quality testing, and production environments.
  - **SRM Servers:** For tender procurement processes, CSL utilizes SRM (Supplier Relationship Management) servers dedicated to handling procurement activities and vendor interactions.
- **Total SAP Servers:** Considering the complexity and scale of CSL's SAP landscape, the total number of servers dedicated to SAP operations could range between 60 to 70 servers. This includes servers for SAP S/4HANA, SAP portal - ESS, SRM, database synchronization, and other related components.

### 2.3.2 Structure of Data centers

The structure of a data center typically involves several key components and considerations to ensure reliable and efficient operations.

- **Physical Facility:**
-



- 
- **Building:** CSL's data center is housed in a dedicated facility within the company's main block. This building is designed to meet the specific requirements for power, cooling, security, and space needed for data center operations.
  - **Power Infrastructure:** CSL's data center incorporates robust electrical systems, including power distribution units (PDUs), uninterruptible power supplies (UPS), backup generators, and power management systems. These systems ensure continuous and reliable power supply to the equipment.
  - **Cooling Systems:** CSL's data center is equipped with advanced cooling systems such as air conditioning units, precision cooling systems, or liquid cooling solutions. These systems maintain optimal operating temperatures for servers, networking equipment, and other critical components.
  - **Security:** Physical security measures are paramount at CSL's data center. Access control systems, surveillance cameras, biometric scanners, and security personnel are deployed to protect the data center from unauthorized access and potential threats.

➤ **Networking Infrastructure:**

- **Network Equipment:** CSL's data center features a comprehensive suite of network equipment, including routers, switches, firewalls, load balancers, and other networking devices. These devices manage data traffic within the data center and facilitate secure communication with external networks.
- **Cabling:** Structured cabling systems are meticulously installed to interconnect servers, storage devices, networking equipment, and other components within CSL's data center. This ensures efficient data transmission and connectivity.

➤ **Server and Storage Infrastructure:**

- **Servers:** CSL utilizes rack-mounted servers or blade servers to host various applications, databases, virtual machines (VMs), and other workloads.
-

.....

These servers are organized within server racks, typically in a standard 42U configuration.

- **Storage:** CSL employs storage arrays or storage area networks (SANs) to provide centralized storage for data, files, and backups. Technologies like RAID are implemented for data redundancy and performance optimization.

➤ **Data Center Services:**

- **Virtualization:** CSL leverages virtualization technologies such as VMware vSphere or Microsoft Hyper-V to create and manage VMs effectively. Virtualization optimizes resource utilization and scalability within the data center.
- **Monitoring and Management:** CSL's data center is equipped with robust data center management software and monitoring tools. These tools track performance metrics, manage resources, automate tasks, and provide insights for optimization and proactive maintenance.
- **Backup and Disaster Recovery:** CSL implements comprehensive backup solutions and disaster recovery plans to safeguard data and ensure business continuity. These measures mitigate risks associated with hardware failures, cyberattacks, or natural disasters.

➤ **Environmental Controls:**

- **Fire Suppression:** Fire detection systems and suppression mechanisms are strategically deployed throughout CSL's data center to prevent and contain fires effectively.
  - **Environmental Monitoring:** Sensors continuously monitor environmental factors such as temperature, humidity, air quality, and other parameters. This monitoring ensures optimal conditions for equipment operation and reliability.
  - **Physical Layout:** CSL's data center is organized into structured areas, including server racks, hot and cold aisles, and equipment zones. This
- .....

layout optimizes airflow, cooling efficiency, and overall operational effectiveness.

## **2.4 Active Directory**

Active Directory (AD) is a directory service developed by Microsoft that provides centralized management of network resources, including user accounts, computers, groups, and applications. It plays a critical role in identity and access management at CSL.

### **2.4.1 Integration with Computer Login**

#### **Windows Domain Join**

At Cochin Shipyard, all corporate computers are joined to the Active Directory domain. This process allows IT administrators to centrally manage and secure these computers. Domain-joined computers leverage AD for user authentication and apply group policies defined in Active Directory.

#### **User Authentication**

When users log in to their domain-joined computers, they authenticate against Active Directory. AD verifies the username and password provided by the user, granting access based on the user's permissions and group memberships stored in AD.

### **2.4.2 Integration with VPN Access**

#### **VPN Configuration**

Cochin Shipyard uses a VPN solution for secure remote access to the corporate network. The VPN server is configured to authenticate users against Active Directory. This ensures that only authorized users with valid AD credentials can establish a VPN connection.

#### **Authentication and Authorization**

During VPN authentication, the VPN server queries Active Directory to validate user credentials. AD also enforces access policies based on user attributes and group memberships. This ensures that remote users have appropriate access levels as defined in AD.

### **2.4.3 Integration with Applications**

---

## **Fluence Application**

Fluence, a key application used at Cochin Shipyard, integrates with Active Directory for user authentication. Users log in to Fluence using their AD credentials, enabling single sign-on (SSO) and simplifying access management for IT administrators.

## **Smax Application**

Smax, another critical application, leverages Active Directory for user management and access control. AD groups are used to define user roles and permissions within Smax, ensuring that users have the right level of access based on their AD group memberships.

## **Benefits and Impact of AD Integration**

The integration of Active Directory with computer login, VPN access, and applications at Cochin Shipyard offers several benefits:

**Centralized Management:** AD provides a single source of truth for user accounts and access policies.

**Enhanced Security:** AD enforces strong authentication and access controls, reducing security risks.

**Improved User Experience:** Single sign-on (SSO) simplifies user access to multiple applications.

**Efficient IT Management:** Centralized user provisioning and policy enforcement streamline IT operations.

## **Chapter-3**

### **Project**

---

### 3.1 Problem Definition

The current manual process at CSL involves a dedicated team responsible for checking and generating daily summary reports of invoices. This manual process is highly time-consuming, requiring significant manpower to extract data from invoices, perform calculations for generating summaries, and verifying the accuracy of the information. Additionally, the manual nature of the process increases the risk of human errors, leading to potential discrepancies in the generated reports. As a result, CSL faces challenges in terms of operational efficiency, resource utilization, and the accuracy of financial reporting due to the limitations and inefficiencies of the manual summarization and calculation tasks.

### 3.2 Proposed Solution: Automation through Task Bot Development

#### *Task Bot Development:*

In the era of digital transformation, organizations like Cochin Shipyard Limited (CSL) are increasingly turning to intelligent automation solutions to streamline their operations. One such solution is the deployment of task bots or intelligent bots, powered by Robotic Process Automation (RPA) and artificial intelligence (AI), to automate repetitive and rule-based tasks. In the context of CSL's manual process of invoice summarization, task bots offer a transformative approach to accelerate workflows, enhance accuracy, and optimize resource utilization.

Some of the Automation Benefits:

- **Time Savings:** Automating repetitive tasks reduces processing time and frees up manpower for more strategic activities.
- **Accuracy:** The bot can perform calculations with precision, minimizing the risk of human errors and ensuring accurate report generation.
- **Scalability:** The automated process can handle a large volume of invoices efficiently, scaling to meet growing demands without significant resource constraints.

### 3.3 Scope of work

---

- 
- Develop and implement task bots or intelligent bots to automate the invoice summarization process at CSL.Ensure accurate extraction of data from various invoice formats, calculation of totals, and generation of daily summary reports.
  - Task Bot Development: Design, develop, and test task bots capable of extracting data fields (e.g., supplier information, invoice numbers, line items, totals) from diverse invoice layouts.
  - Data Training: Train the task bots using machine learning algorithms to recognize and process different invoice formats accurately.
  - Calculation Algorithms: Develop algorithms within the bots for performing calculations such as subtotalling, tax computations, discounts, and total amounts.
  - Integration: Integrate the task bots seamlessly into CSL's existing systems, databases, and workflows for data retrieval, processing, and reporting.
  - Testing and Validation: Conduct rigorous testing and validation of the automated process to ensure accuracy, reliability, and compliance with business rules.
  - Deployment: Deploy the automated solution into production, ensuring smooth transition and user acceptance.

### **3.4 Methodology and Implementation**

#### **RPA (Robotic Process Automation)**

RPA is a technology that uses software robots or "bots" to automate repetitive, rule-based tasks previously performed by humans. These bots can mimic human interactions with digital systems, such as logging into applications, navigating interfaces, and performing data entry tasks. RPA aims to streamline workflows, increase efficiency, and reduce errors in business processes across various industries.

#### **OpenRPA**

OpenRPA is an open-source RPA platform designed for automating business processes. It provides a user-friendly interface for creating and managing automation workflows without the need for extensive coding knowledge. OpenRPA supports integration with

.....

various applications and systems, enabling organizations to automate tasks across different platforms.

## **UiPath**

UiPath is a prominent name in the realm of Robotic Process Automation (RPA), offering a powerful platform that enables organizations to streamline their operations through automation. With its intuitive user interface and visual design tools, UiPath makes it accessible for both technical and non-technical users to create automation workflows effortlessly. Its capabilities span across various automation tasks, including data entry, document processing, and application integration, allowing businesses to free up human resources from mundane tasks and focus on higher-value activities. UiPath's scalability enables the deployment of multiple bots to handle complex processes, improving efficiency and productivity. Moreover, its integration capabilities with diverse systems, coupled with AI and machine learning features, empower organizations to achieve advanced automation, enhance decision-making, and drive innovation. With robust analytics, monitoring tools, and stringent security measures, UiPath ensures that automation processes are optimized, compliant, and secure, contributing significantly to digital transformation initiatives.

### **3.4.1 Automation using OpenRPA**

#### **Workflow Explanation:**

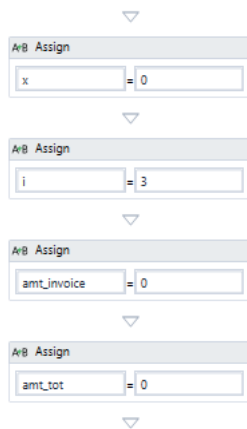
This workflow is designed to automate the processing of data from an Excel file of invoices. It reads specific data, performs calculations, and writes back processed information into a different Excel file. It uses a sequence of activities within a loop (For Each) to iterate over rows of data, updating values, and then writing back results. Additionally, it integrates with external systems by sending an email notification with an attachment. The use of InvokeCode activities allows for the integration of custom Python scripts to extend the workflow's capabilities beyond what's available with standard activities in the workflow designer.

#### **➤ Variables:**

This workflow defines several variables (input, i, x, filename, amt\_invoice, amt\_tot, etc.) to store values and perform calculations during execution.

.....

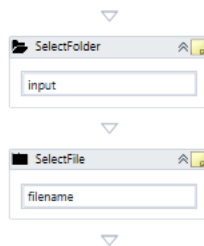




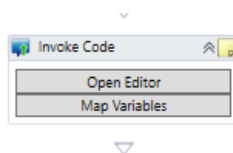
## 2. File Handling:

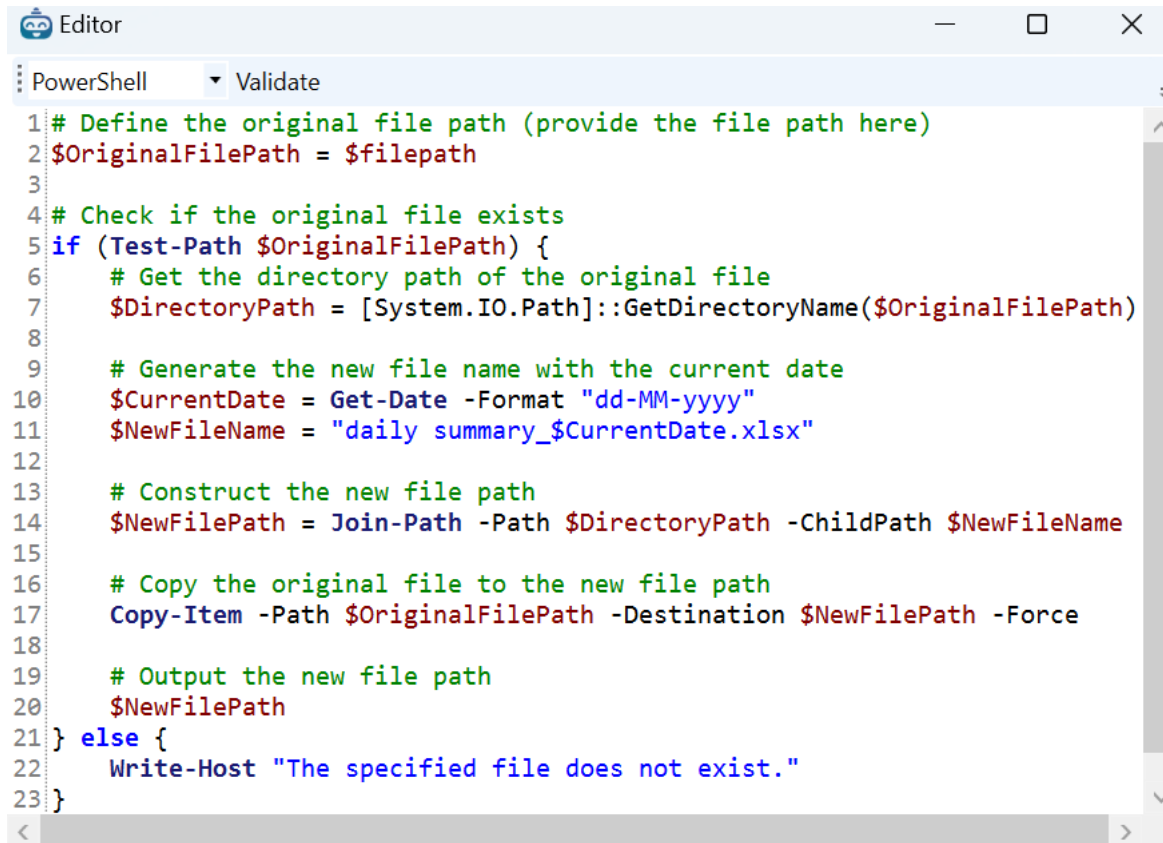
**SelectFolder:** Prompts the user to select an input folder.

**SelectFile:** Prompts the user to select an output destination file.



**InvokeCode (PowerShell):** Saves a file using the current date and handles file operations like checking file existence, copying, and moving files.





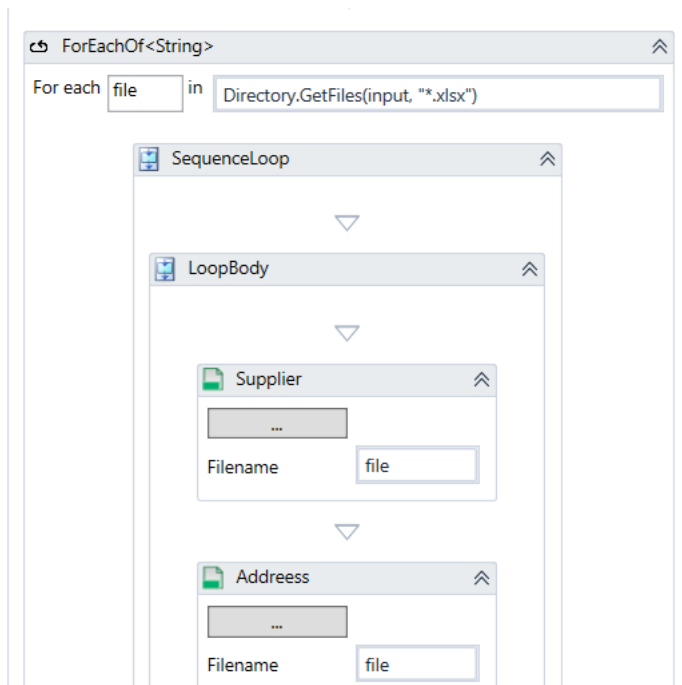
The screenshot shows a PowerShell script in an Editor window. The script defines a function to process a file path. It checks if the file exists, gets the directory path, generates a new file name with the current date, constructs the new file path, copies the original file to the new path, and outputs the new file path. If the file does not exist, it writes a message to the host.

```

1 # Define the original file path (provide the file path here)
2 $OriginalFilePath = $filepath
3
4 # Check if the original file exists
5 if (Test-Path $OriginalFilePath) {
6     # Get the directory path of the original file
7     $DirectoryPath = [System.IO.Path]::GetDirectoryName($OriginalFilePath)
8
9     # Generate the new file name with the current date
10    $CurrentDate = Get-Date -Format "dd-MM-yyyy"
11    $NewFileName = "daily summary_{$CurrentDate}.xlsx"
12
13    # Construct the new file path
14    $NewFilePath = Join-Path -Path $DirectoryPath -ChildPath $NewFileName
15
16    # Copy the original file to the new file path
17    Copy-Item -Path $OriginalFilePath -Destination $NewFilePath -Force
18
19    # Output the new file path
20    $NewFilePath
21 } else {
22     Write-Host "The specified file does not exist."
23 }

```

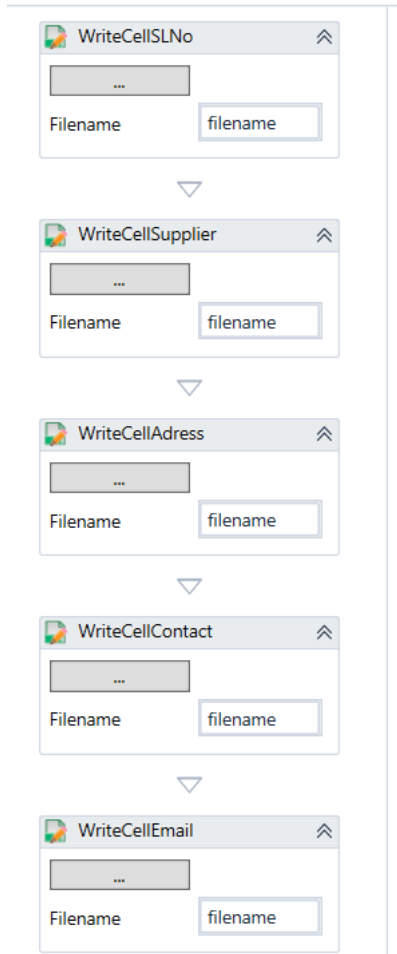
**ForEachOf:** Iterates over each Excel file (\*.xlsx) in the selected input folder.



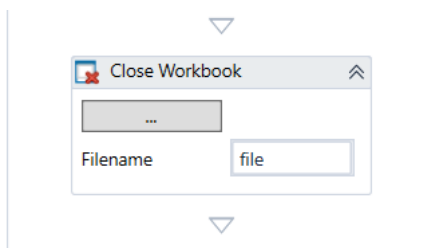
**ReadCell:** Reads specific cells from the Excel file ([file]) and stores the values in workflow variables (supplier, address, contact, email, etc.).

The screenshot displays a vertical sequence of workflow actions in a drag-and-drop interface. The first four actions are 'ReadCell' components, each with a document icon and a title: 'Contact', 'Email', 'ItemSummary', and 'InvoiceDate'. Each 'ReadCell' action contains a 'Filename' input field with the value 'file'. Below these is an 'Assign' action, indicated by a small 'A/B' icon, which contains an assignment: 'date\_in' followed by an equals sign and the expression 'invoicedate.ToStrir'. The actions are connected by downward-pointing arrows, and each has a collapse/expand icon in its top right corner.

**WriteCell:** Writes values to specific cells in the Excel file ([filename]).



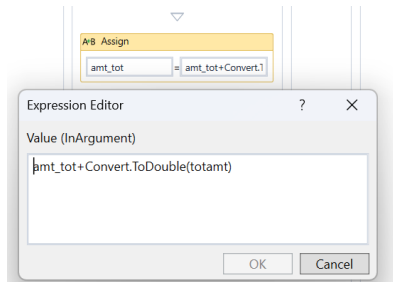
**CloseWorkbook:** Closes the Excel workbook after operations.



### Data Processing:

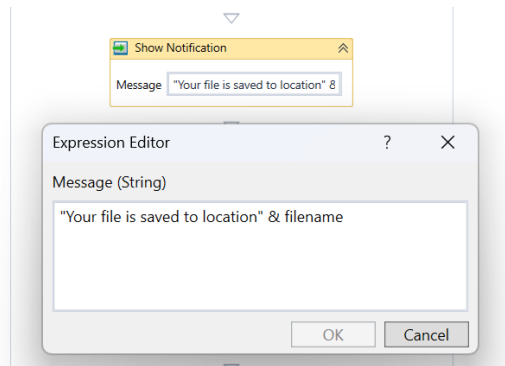
Data extracted from Excel files is processed (e.g., calculating invoice amounts, tax percentages, total amounts, formatting dates) within nested sequences and assigned to variables.

Accumulated values (amt\_invoice, amt\_tot) are updated iteratively.



### User Interaction:

ShowNotification: Displays a notification after file processing is complete.



**InvokeCode (Python):** Sends an email using SMTP to notify the user and potentially attach the processed Excel file.

```

1 import smtplib
2 from email.mime.multipart import MIMEMultipart
3 from email.mime.text import MIMEText
4 from email.mime.base import MIMEBase
5 from email import encoders
6
7 def send_email(sender_email, sender_password, recipient_email, subject, body, attachment_path):
8     # Setup the MIME
9     message = MIMEMultipart()
10    message['From'] = sender_email
11    message['To'] = recipient_email
12    message['Subject'] = subject
13
14    # Attach the body to the email
15    message.attach(MIMEText(body, 'plain'))
16
17    # Attach the file
18    if attachment_path:
19        attachment = open(attachment_path, "rb")
20        part = MIMEBase('application', 'octet-stream')
21        part.set_payload(attachment.read())
22        encoders.encode_base64(part)
23        part.add_header('Content-Disposition', f"attachment; filename= {attachment_path}")
24        message.attach(part)
25
26    # Connect to the SMTP server
27    with smtplib.SMTP('smtp.gmail.com', 587) as smtp_server:
28        smtp_server.starttls()
29
30        # Login to the SMTP server using sender's credentials
31        smtp_server.login(sender_email, sender_password)
32
33        # Send the email
34        smtp_server.send_message(message)
35
36 # Example usage
37 sender_email = 'openrpai@gmail.com' # Update with your Gmail address
38 sender_password = 'rgavvalbxavsguu'
39 recipient_email = 'leothomas8095@gmail.com' # Update with recipient's email address
40 subject = 'Daily summary'
41 body = 'Hi, Attaching the daily summary'
42 attachment_path = path # Replace with the path to your Excel file
43 send_email(sender_email, sender_password, recipient_email, subject, body, attachment_path)
44

```

## Functionality:

The workflow reads data from Excel files, performs calculations, updates the Excel file with processed data, saves the updated file, and then sends a notification email with an option to open the saved file.

## Findings and Results

The bot was tested with 8 datasets and the accuracy rate was 100% with a completion time of 0.50 mins. The input invoices was in excel format.

daily summary 08-05-2024 - Excel

FileHomeInsertPage LayoutFormulasDataReviewViewHelpTell me what you want to do

Calibri11A A

</

### 3.4.2 Automation with UiPath

The bot was trained with 50 datasets and test dataset of 500 invoices were given. We did the testing with about 60 invoices and the accuracy rate was 60%.

#### Sequence Activity

Purpose: Acts as the main container for a sequence of activities.

Variables:

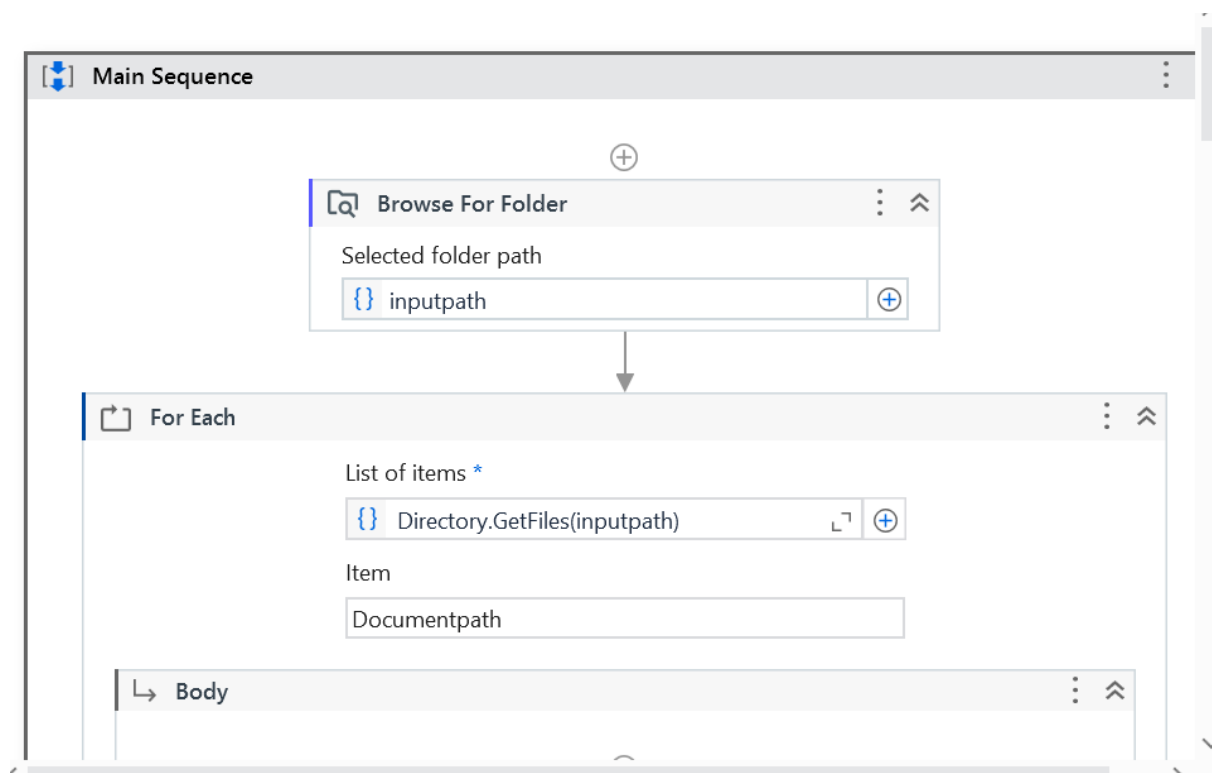
- dt1: Variable of type System.Data.DataTable.
- inputpath: Variable of type String.

#### SelectFolder Activity

Purpose: Displays a dialog box to select a folder interactively.

Property:

- SelectedFolder: Binds to the inputpath variable to store the selected folder path.

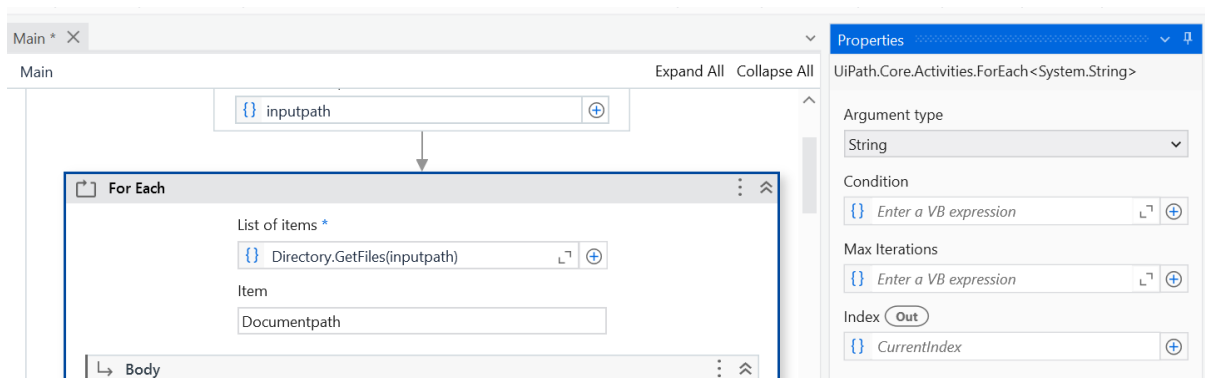


## ForEach Activity

Purpose: Iterates over a collection of file paths obtained from a directory.

Properties:

- Values: Binds to Directory.GetFiles(inputpath) to retrieve file paths.
- Child Activities: Contains a nested sequence of activities (ActivityAction) for each file path.

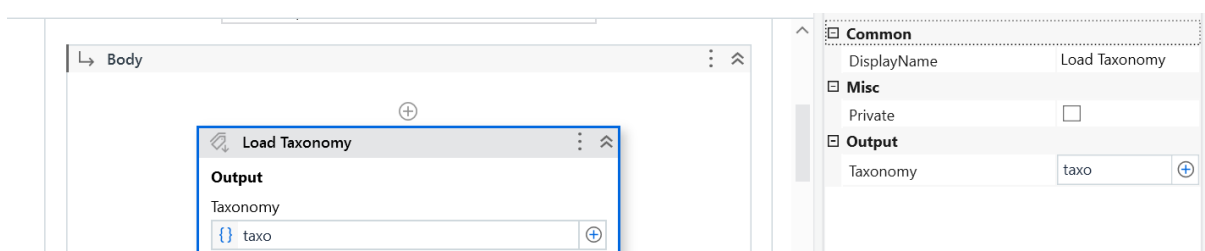


## LoadTaxonomy Activity

Purpose: Loads a taxonomy definition.

Properties:

- Taxonomy: Binds to the taxo variable of type DocumentTaxonomy.



## 5. DigitizeDocument Activity

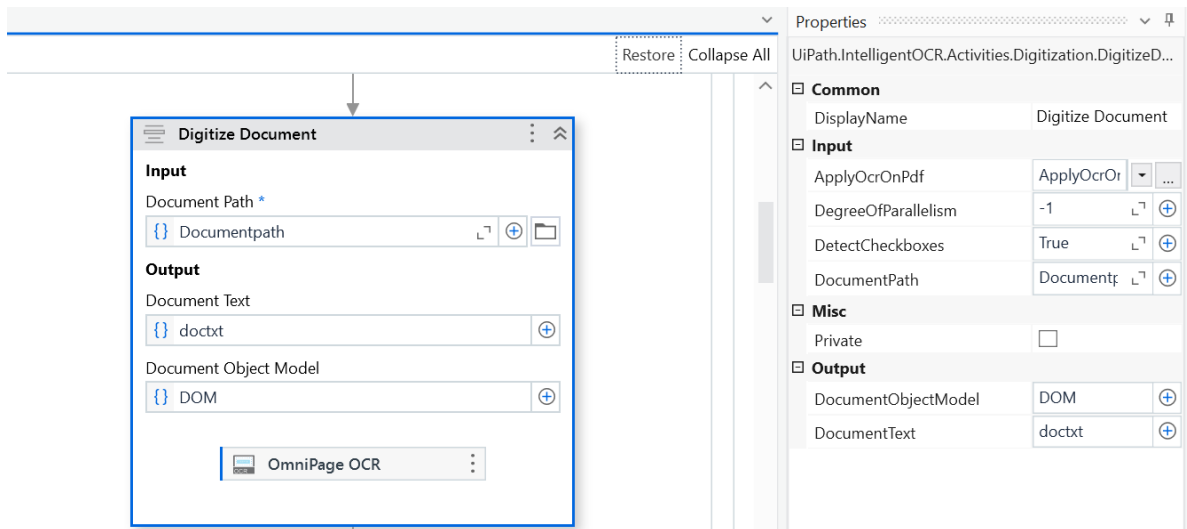
**Purpose:** Converts a document into text using OCR.

**Properties:**

- DocumentPath: Binds to the current file path (Documentpath) from the ForEach loop.



- DocumentText: Outputs the extracted text to the doctxt variable.

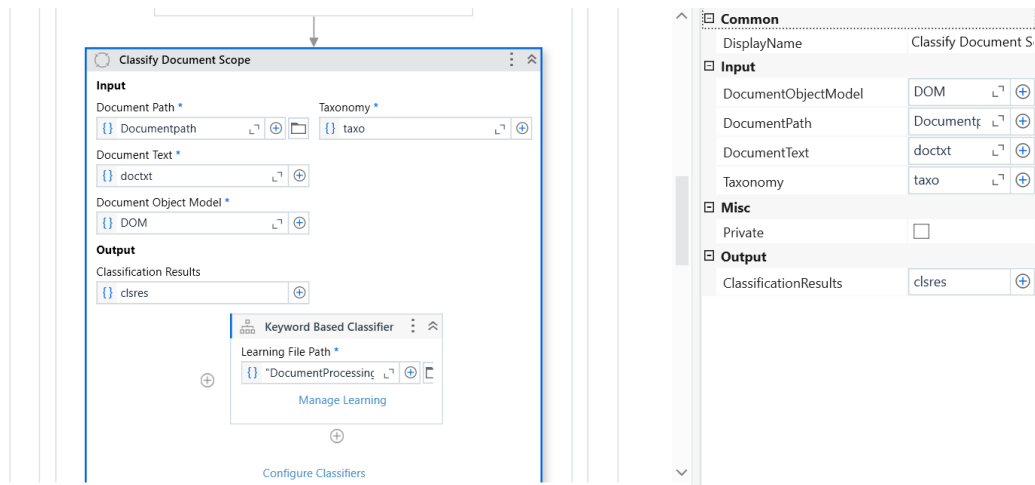


### ClassifyDocumentScope Activity

Purpose: Classifies a document based on a taxonomy.

Properties:

- DocumentPath: Binds to the current file path (Documentpath) from the ForEach loop.
- DocumentText: Binds to the extracted text (doctxt).
- DocumentObjectModel: Outputs the structured document model to the DOM variable.
- ClassificationResults: Outputs classification results to the clsres variable.
- Child Activities: Contains a KeywordBasedClassifier for document classification.

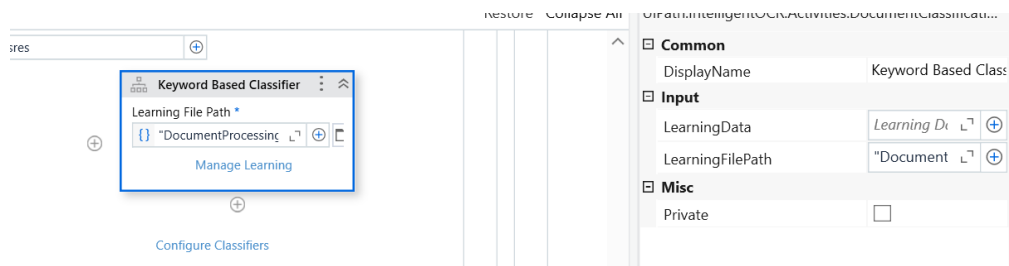


## KeywordBasedClassifier Activity

Purpose: Classifies documents based on keyword matching.

Properties:

- **LearningFilePath:** Specifies the path to the file containing learning data for the classifier.
- **PersistenceId:** Identifier for the classifier's persistence.



## DataExtractionScope Activity

Purpose: Extracts structured data from a classified document.

Properties:

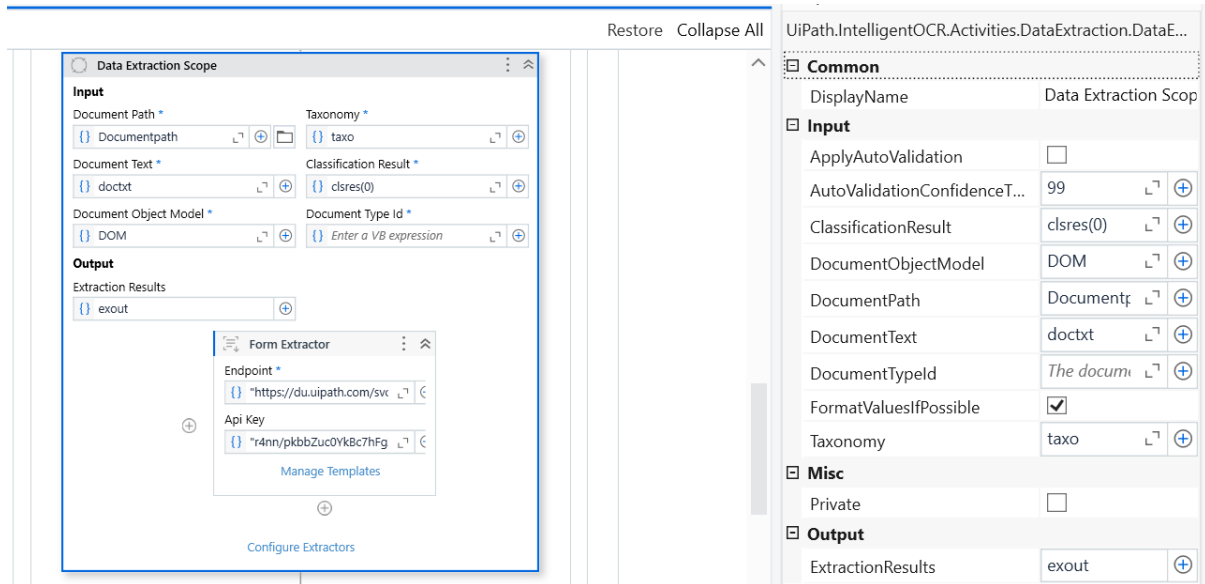
**DocumentPath:** Binds to the current file path (Documentpath) from the ForEach loop.

**DocumentText:** Binds to the extracted text (doctxt).

**DocumentObjectModel:** Outputs the structured document model to the DOM variable.

**ExtractionResult:** Outputs extraction results to the exout variable.

**Child Activities:** Contains a FormExtractor for data extraction.

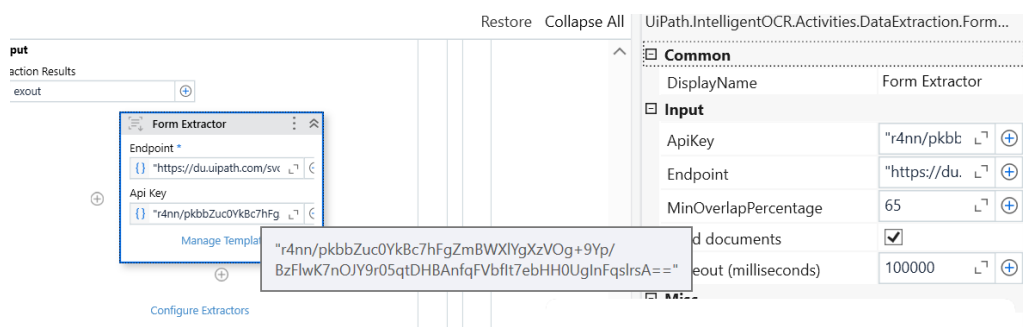


## FormExtractor Activity

**Purpose:** Extracts data fields from a document using predefined templates.

**Properties:**

- **Endpoint:** Specifies the URL endpoint for the form extraction service.
- **SerializedTemplates:** Defines templates for extracting specific data fields.
- **PersistenceId:** Identifier for the form extractor's persistence.



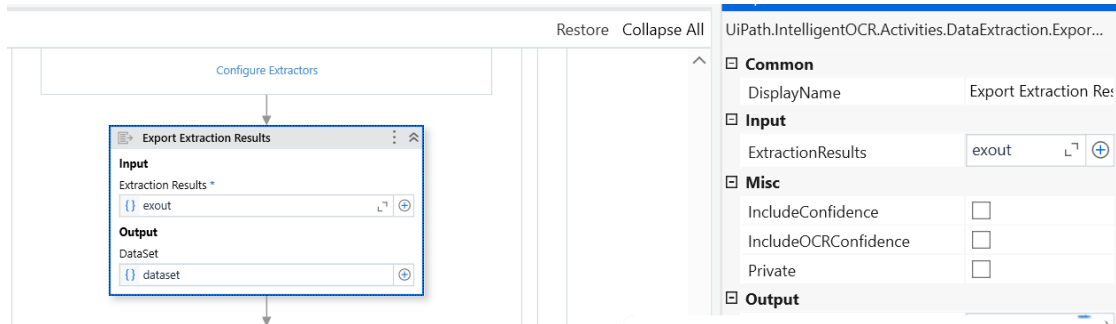
## ExportExtractionResults Activity

**Purpose:** Exports extraction results to a dataset.

Properties:

DataSet: Binds to the dataset variable of type System.Data.DataSet.

ExtractionResult: Binds to the extraction results (exout).

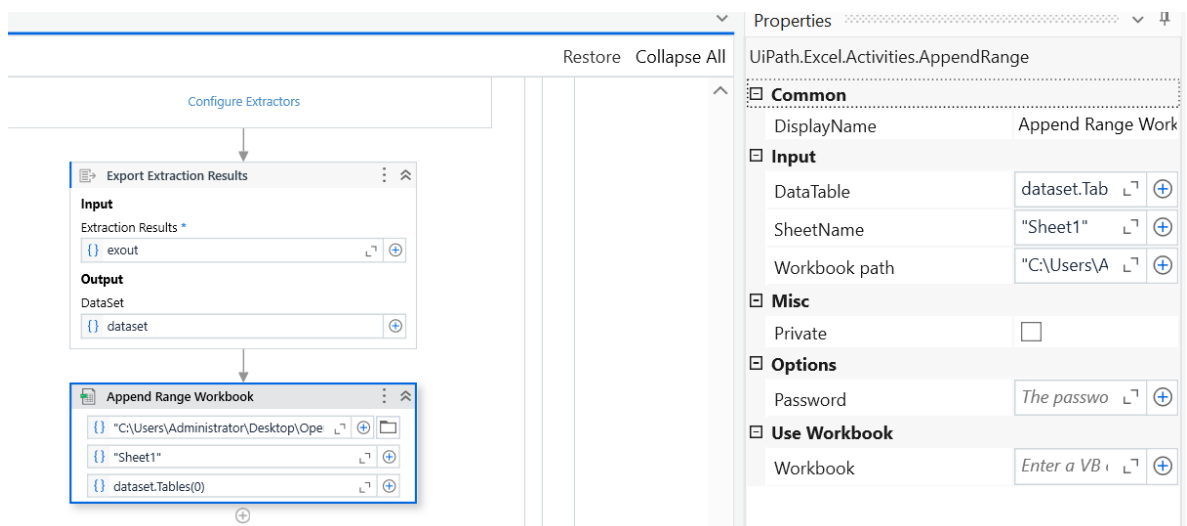


## AppendRange Activity

Purpose: Appends data to an Excel workbook.

Properties:

- DataTable: Binds to the first table in the dataset (dataset.Tables(0)).
- SheetName: Specifies the name of the Excel sheet to append data to.
- WorkbookPath: Specifies the path to the Excel workbook (sample.xlsx).



## Overall Workflow Purpose:

This workflow automates the processing of documents stored in a specified folder. It iterates over each document, extracts text using OCR, classifies the document based on a

.....

predefined taxonomy, and then extracts structured data using specific extraction templates. Finally, it exports the extracted data to an Excel workbook for further analysis or processing.

Each activity plays a crucial role in different stages of the document processing workflow, from data extraction to result export. Adjustments and configuration within each activity ensure accurate and efficient automation of document-related tasks. If you have more specific questions about any activity or component, feel free to ask!

## WEEKLY OVERVIEW OF INTERNSHIP ACTIVITIES

WEEK	DAY	DATE	DESCRIPTION OF WORK
Week 1	1	22/04/2024	Introduction to the organization, Introduction to METI, Safety Class
Week 1	2	23/04/2024	Introduction to ISD, its structure and working.
Week 1	3	24/04/2024	Introduction and a session on Robotic Process Automation and OpenRPA. Installation of OpenRPA.
Week 1	4	25/04/2024	Project 1 assigned. Read data from a excel file and assign to another excel file. Started with basics of OpenRPA and read data from excel file and displayed to output console. Creation and usage of variables.  READ EXCEL, READ CELL, READ RANGE, WRITE LINE
Week 2	5	29/04/2024	Successfully extracted the data from input excel invoices and write into output excel.  FOR EACH, READ CELL, WRITE CELL, SELECT FOLDER, SELECT FILE
Week 2	6	30/04/2024	Creation of new Workflow, implementation of features like saving and closing all the opened excel files, saving the output file(report file) with current date.  Calculated the total amount and total tax amount and SL no in the output excel  ASSIGN, INVOKE CODE (PowerShell for saving file with current date), CLOSE WORKBOOK
Week 2	7	02/05/2024	Added extra feature to save and close the output excel and display a prompt saying

			the file saved at "path". Give OK/CANCEL to open the saved excel. SHOW NOTIFICATION, INVOKE CODE (PowerShell)
Week 2	8	03/05/2024	Work 2 assigned. Introduction to UiPath. Installation of UiPath.
Week 2	9	04/05/2024	Visit to Ship building dock, ship repair dock, Quay 1 and Quay 3 UiPath –Read from pdf invoices
Week 3	10	06/05/2024	Introduction to Taxonomy manager, OmniPage OCR. Extracted text from pdf which are of same types
Week 3	11	07/05/2024	Learned about AI center, Form extractor, Classify Document Scope – Keyword Based Classifier, Data Extraction scope, Export extraction result. Extracted data from pdf by training 5 pdf formats. Accuracy rate was 60%
Week 3	12	08/05/2024	Trained about 40-50 pdf invoices. Extracted the test, checked for accuracy and did the project presentation
Week 3	13	09/05/2024	Session on SAP and shipyard Business scope. Worked on Report work
Week 3	14	10/05/2024	Report submission