An Internship Report

on

Process Mining Virtual Internship

Submitted in partial fulfilment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering (AI & ML)

by

J Naga Vyshnavi **(224G1A3360)**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

(AI & ML)

Srinivasa Ramanujan Institute of Technology
(AUTONOMOUS)
Rotarypuram Village, B K Samudram Mandal, Ananthapuramu - 515 701

2024 - 2025

Srinivasa Ramanujan Institute of Technology
(AUTONOMOUS)
Rotarypuram Village, B K Samudram Mandal, Ananthapuramu - 515 701

Department of Computer Science & Engineering (AI & ML)

Certificate

This is to certify that the internship report entitled "Process Mining Virtual Internship" is the bonafide work carried out by **J NAGA VYSHNAVI** bearing Roll Number 224G1A3360 in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering (AI &ML)** for three months from April 2024 to June 2024.

Internship Coordinator

Mr. P. Veera Prakash, M. Tech.,(Ph.D),

Assistant Professor

Head of the Department

Dr. P. Chitralingappa, M.Tech., Ph.D.,

Associate Professor

Date:
Place: Ananthapuramu

EXTERNAL EXAMINER

PREFACE

All India Council for Technical Education (AICTE) has initiated various activities for promoting industrial internship at the graduate level in technical institutes and Edu-Skills is a Non-profit organization which enables Industry 4.0 ready digital workforce in India. The vision of the organization is to fill the gap between Academic and Industry by ensuring world class curriculum access to the faculties and students. Formation of the All-India Council for Technical Education (AICTE) in 1945 by the Government of India.

Purpose: With a vision to create an industry-ready workforce who will eventually become leaders in emerging technologies, Edu-Skills & AICTE launches 'Virtual Internship' program on Informational Technology, supported by Process Mining terminology. Process mining technology, businesses can improve their process intelligence to create ideal workflows and operations. As a result, more and more companies are leaning toward adopting this software to reach their maximum efficiency potential

Company's Mission Statement: The main mission of these initiatives is enhancement of the employability skills of the students passing out from Technical Institutions.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned our efforts with success. It is a pleasant aspect that I have now the opportunity to express my gratitude for all of them.

It is with immense pleasure that I would like to express my indebted gratitude to my internship coordinator **Dr. P. Veera Prakash, Assistant Professor, Department of Computer Science and Engineering** , who has supported me a lot and encouraged me in every step of the internship work. I thank him for the stimulating support, constant encouragement and constructive criticism which have made possible to bring out this internship work.

I am very much thankful to **Dr. P. Chitralingappa, Associate Professor & HOD, Computer Science and Engineering (AI & ML),** for his kind support and for providing necessary facilities to carry out the work.

I wish to convey my special thanks to **Dr. G. Balakrishna, Principal** of **Srinivasa Ramanujan Institute of Technology** for giving the required information in doing my internship. Not to forget, I thank all other faculty and non-teaching staff, and my friends who had directly or indirectly helped and supported me in completing my internship in time.

I also express our sincere thanks to the Management for providing excellent facilities and support**.**

Finally, I wish to convey my gratitude to my family who fostered all the requirements and facilities that I need.

J NAGA VYSHNAVI
(224G1A3360)

Index

LIST OF FIGURES

LIST OF ABBREVIATION

AI	Artificial Intelligence
BPM	Business Project Management
CRM	Customer Relationship Management
EMS	Execution Management System
EPM	Educational Process Mining
ERP	Enterprise Resource Planning
IT	Information Technology
KPI	Key Performance Indicator
OLAP	Online Analytical Processing
PQL	Process Query Language
P2P	Purchase-to-Pay
SCM	Supply Chain Management
SLA	Service Level Agreement

CHAPTER - 1

INTRODUCTION

Process mining is a technique designed to discover, monitor, and improve real processes (i.e., not assumed processes) by extracting readily available knowledge from the event logs of information systems. Process mining focuses on different perspectives, such as control-flow, organizational, case, and time. While much of the work around process mining focuses on the sequence of activities i.e., control-flow the other perspectives also provide valuable information for management teams.

1.1 Process Mining Cycle

Data Transformation

Visualize your data by extracting it to deliver real-time multidimensional process models. The data can be transformed and cleaned up, so it fits the expected input of the app. You can add business logic and enrich the data to enhance the possibilities for analysis.

Data Analysis

Use process apps to do initial analysis and configure the app to make it ready for business roll-out. Incorporate business procedures and rules to develop detailed reference for every business transaction across teams.

Continuous Monitoring

Deploy the app to business users for continuous monitoring of your process. Discover key business risks and enable opportunities for optimization and transformation.

1.2 Process mining

Process Mining is the combination of two disciplines: **Data Science** and **Business Process Management**. Process Mining essentially uses Data Science techniques, such as Big Data and AI, to address Process Science problems such as process improvement and automation

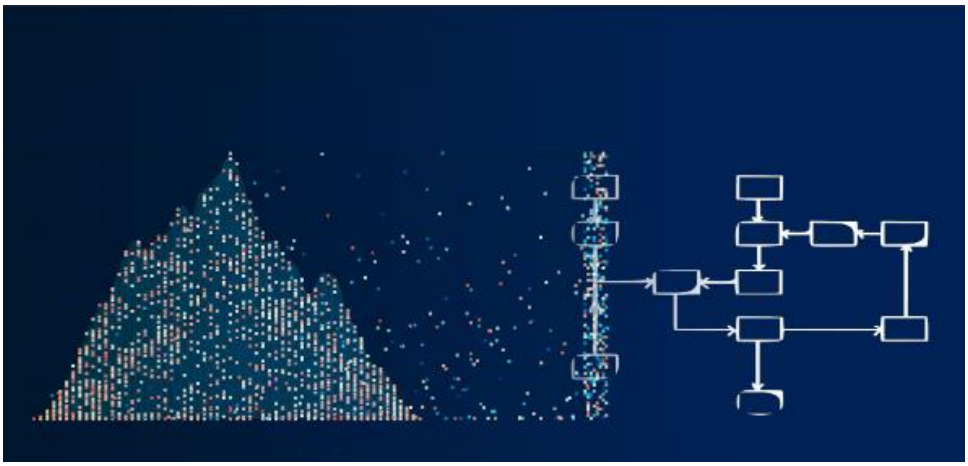

Fig. No. 1.1: Process Mining

Data and AI, to address Process Science problems such as process improvement and automation.Process Mining achieves this union by taking the **digital footprints** that are created in IT systems and using them to **reconstruct and visualize process flows**.

1.3 History of Process Mining

This specialized algorithm for identifying data trends, patterns, and inefficiencies from event logs started to take form in the late 1990s. A Dutch computer scientist and professor, Wil van der Aalst, is the face behind the academic research that went into discovering the details of how a process unfolds.

Also known as the godfather of process mining, van der Aalst and his team of researchers encountered a resolution to utilize data in assortment with analytics and visualization tools.

However, the study remained limited to the academic world for quite some time. Eventually, businesses realized that merging process mapping with data analysis could deliver the most efficient results. Around this time, process mining's economic prospects became known across industries.

Of course, modern process mining software solutions are much more advanced for competing in today's business world. They bring an amalgamation of data science, data mining, business intelligence, and analytics to deep dive into the workings of an organizational model.

As a result, the newer wave of process mining not only visualizes and optimizes

processes but also measures KPIs, benchmark and compare operations, collaborate with your existing IT infrastructure, and solve problems with data-centric solutions.

1.4 Importance of process mining

Reduce costs

Process mining technology helps uncover hidden inefficiencies, bottlenecks, and the potential for automation to reduce overall process costs.

Increase transparency

There is no shortage of data available within business operations but having the tools to use that data and create actionable insights that improve efficiency is another story. Process mining technology allows you to locate the right data and understand that data better.

Reduce multi-hops

Multi-hops create inefficiencies in processes and decrease productivity. Process mining technology empowers companies to identify and minimize these productivity slowdowns.

Chapter 2

Understanding of Process Mining

2.1 What is a process?

One easy example for that could be a pizza delivery process. It starts with placing the order by calling the pizza company or via their website. Then, the order is assigned to a pizzamaker. The pizzamaker bake the pizza, the pizza is packaged, a delivery person delivers it to the assigned address and the payment is received. The problem is this is the ideal scenario of pizza delivery process. But in practice, there are so many things that can go wrong on the way there. The pizzamaker might put the wrong ingredients, the delivery person might go to a different address, or the paymentfail. Therefore, we can say processes are the engine of every experience.

Understanding these processes and optimizing are crucial for successful businesses.

2.2 Why processes are Important?

At the same they are confronted by the regular demand on the market. For example, things happening to digital transformation, supply chain becoming more digital, automating workflows and many more!

All these challenges both on the customer and the market side make it difficult for businesses to survive in the long run. When we compare the Fortune 500 companies in 1955 and in 2015, we see that only 12% of them managed to remain on the list. Taking a closer look at the customer expectations, please take a moment and think about how you, as a customer, were disappointed by a business the last time.

2.3 What is process mining?

Process mining is a data-driven approach that uses event data recorded in information systems to analyse, visualize, and improve business processes. It leverages techniques from data mining, process modelling, and business process management to provide insights into how processes are executed within an organization. By examining event logs and transaction data, process mining aims to uncover patterns, inefficiencies, and opportunities for optimization in processes

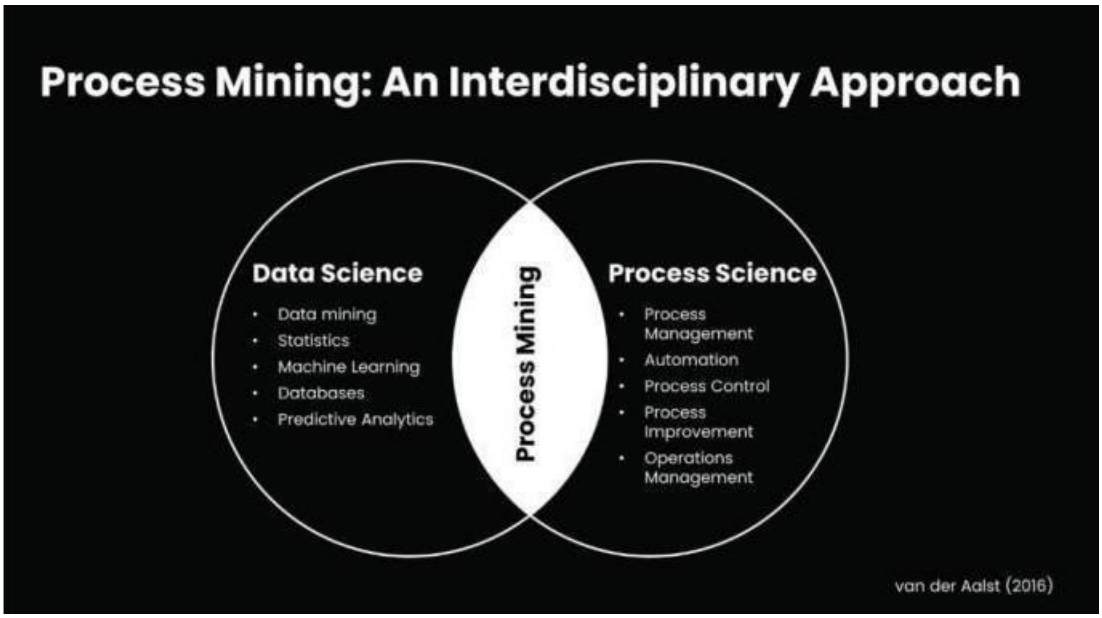

Fig. No. 2.1 Process Mining Approach

Process Mining is the combination of two disciplines: **Data Science** and **Business Process Management**. Process Mining essentially uses Data Science techniques, such as Big Data and AI, to address Process Science problems such as process improvement and automation.

Process Mining achieves this union by taking the digital footprints that are created in IT systems and using them to reconstruct and visualize process flows. From here, Process Mining technology can identify patterns and deviations.

Process mining includes:

- Automated process discovery (extracting process models from an event log)
- Conformance checking (monitoring deviations by comparing model and log)
- Social network/organizational mining
- Automated construction of simulation models
- Model extension
- Model repair
- Case prediction

- History-based recommendations

Process Mining Importance

Process mining takes a more data-driven approach to BPM, which has historically been managed more manually. BPM generally collects data more informally through workshops and interviews, and then uses software to document that workflow as a process map. Since the data that informs these process maps is more qualitative, process mining brings a more quantitative approach to a process problem, detailing the actual process through event data.

Increasing sales is not the sole avenue to generate revenue. Six sigma and lean methodologies illustrate how reducing operational costs can enhance return-on investment (ROI). Process mining aids businesses in cost reduction by quantifying inefficiencies in operational models, enabling leaders to make informed decisions regarding resource allocation. Identifying bottlenecks not only cuts costs and streamlines process improvement, but also fosters innovation, elevates quality, and boosts customer retention. Despite being a nascent discipline, process mining faces several challenges:

- **Data Quality:** The process of finding, merging, and cleaning data is often necessary for process mining. Data may be scattered across various sources, incomplete, or possess varying labels and granularity levels. Accounting for these disparities is crucial for the accuracy of the process model.

- **Concept Drift:** Occasionally, processes undergo changes during analysis, leading to concept drift.

Process mining intersects business process management (BPM) and data mining. While both deal with data, their datasets differ in scope. Process mining utilizes event log data to create process models for uncovering, comparing, or enhancing processes. On the other hand, data mining has a broader scope, encompassing various datasets to observe and predict behaviors. Its applications range from customer churn and fraud detection to market basket analysis.

Process mining adopts a data-centric approach to BPM, traditionally managed through more manual means. BPM typically gathers data informally via workshops and interviews, documenting workflows as process maps using software. As the data informing these process maps tends to be qualitative, process mining introduces a quantitative perspective to process issues, delineating the actual process through event data.

CHAPTER 3

PROCESS MINING TECHNIQUES

Process mining techniques benefit companies of any size and workflow. Process mining solutions can focus on various elements like the flow of a process, the organizational or time management with data mining and machine learning integration. There are three main classes of process mining techniques:

1. Process Discovery
2. Conformance Check
3. Analysis and Enhancement

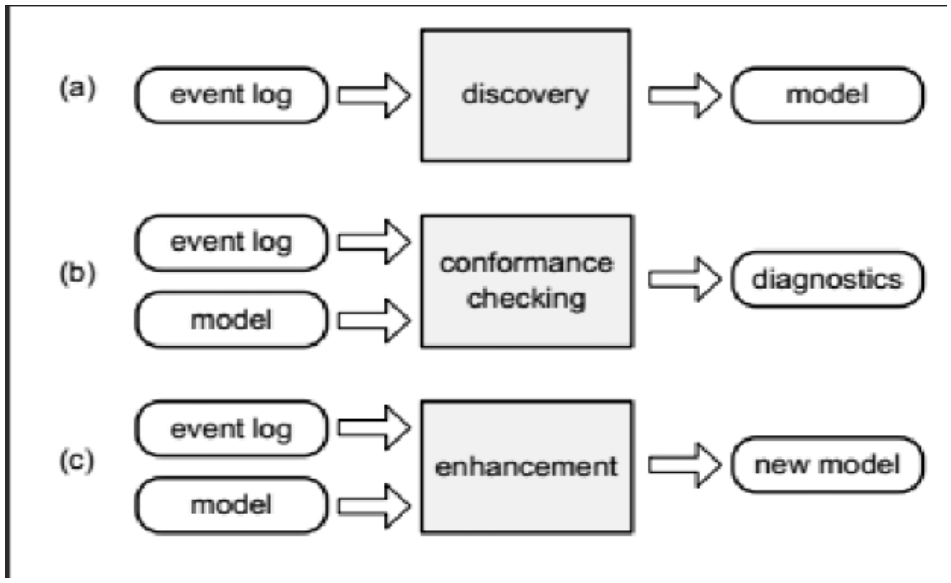

Fig. No 3.1 Process Mining Techniques

3.1 Process Discovery

Process discovery uses event log data to create a process model without outside influence. Under this classification, no previous process models would exist to inform the development of a new process model. This type of process mining is the most widely adopted. The first step in process mining. The main goal of process discovery is to transform the event log into a process model. An event log can come from any data storage system that records the activities in an organization along with the timestamps for those activities.

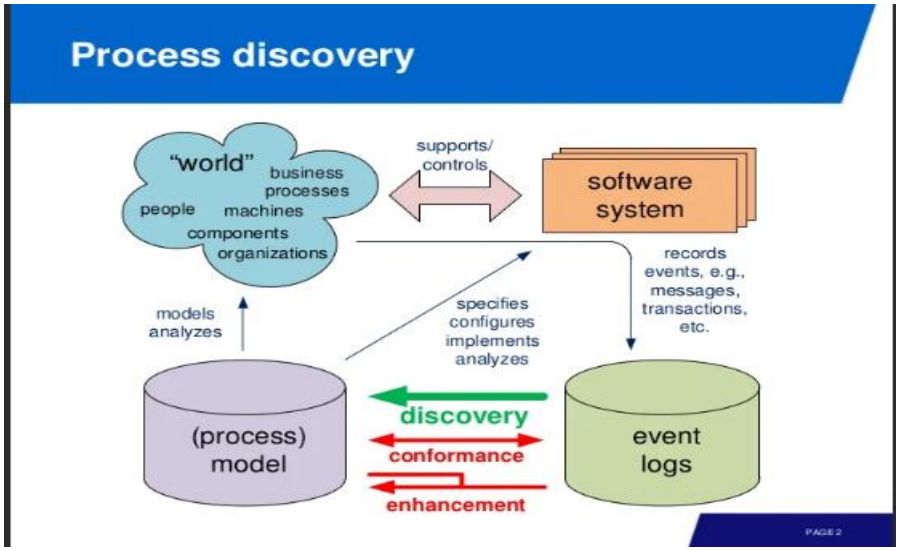

Fig. No. 3.2: Process Discovery

3.1.1 Benefits of Process Discovery

There are numerous advantages of process discovery. Some of them are outlined below.

- - Enhanced Quality and performance
- - Competitive Edge
- - Enhanced Visibility
- - Reduced Risks
- - Cost Effectiveness
- - Improved Scalability

3.2 Conformance Check

Conformance validation verifies if the intended procedural model is effectively implemented. This form of process analysis contrasts a process depiction with an established process model using its event log data, pinpointing any discrepancies from the envisioned model. Conformance validation methodologies require a process model and event log as input, yielding a collection of disparities between the conduct depicted in the process model and the conduct captured in the event log. Conformance validation assists in guaranteeing that all process deviations from the target procedure or reference model are recognized, consequently reducing the likelihood of audit complications or legal infractions.

3.3 Analysis and Enhancement

This type of process mining has also been referred to as extension, organizational mining, or performance mining. In this class of process mining, additional information is used to improve an existing process model. For example, the output of conformance checking can assist in identifying bottlenecks within a process model, allowing managers to optimize an existing process

Once you have the necessary information on the current process deficiencies from process discovery and conformance checks like bottlenecks, process loops, and unwanted deviations, you can narrow down the steps that have the potential to be optimized to the highest standard.

Therefore, the objective behind the model analysis and enhancement is to determine how to optimize the target process model to its optimum potential. The enhanced model minimizes the risk of existing bottlenecks, making the overall process more efficient with automated workflow

Process enhancement, sometimes referred to as model enhancement, is a process mining technique that's used to extend or enhance a target model or reference model using discovered information about the actual process. For example, analysimay uncover bottlenecks or unplanned process sequences that can be eliminated from the model to make it a better representation of the ideal process.

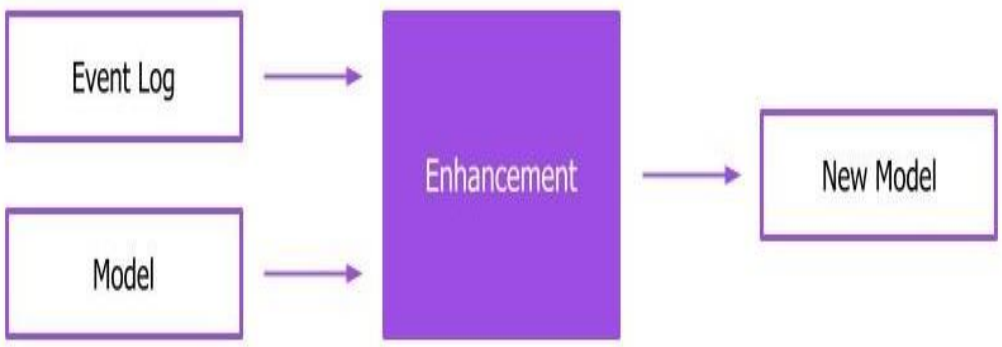

Fig. No. 3.3: Process Enhancement

Therefore, the objective behind the model analysis and enhancement is to determine how to optimize the target process model to its optimum potential. The enhanced model minimizes the risk of existing bottlenecks, making the overall process more efficient with automated workflow. There some steps in process enhancement. They are mentioned below:

1. Analyse the data.
2. Identify optimization potentials.
3. Adapt the target process models.
4. Check the continuous improvement process.

Suppose you perform a process mining analysis with the log file of a business process. Using the process discovery and conformance checking methods, you identify various process weaknesses, including bottlenecks, process loops, and unwanted process deviations. This is all valuable information. Now you know where to find which types of optimization potential.

But what do you do with this knowledge? You specifically adapt your target model, which serves as the standard guideline for process implementation. This means that you change the process in such a way that the risk of bottlenecks is reduced, or certain process sequences are no longer possible. The success or usefulness of the process changes can only be determined after the new process has been implemented for some time. Therefore, process mining is an excellent way to continuously improve your process

CHAPTER 4

APPLICATIONS

Applications of process mining involve using process mining techniques to analyze and improve processes as they occur, providing insights and interventions in real-time. Here are some examples of real-time applications of process mining:

- **Operational Monitoring and Alerts**: Process mining can be used to monitor ongoing processes in real-time and generate alerts when deviations or anomalies are detected. This allows organizations to take immediate action to address issues and maintain process efficiency.
- **Dynamic Resource Allocation**: In scenarios where resources need to be allocated dynamically, such as in manufacturing or service industries, real-time process mining can help optimize resource allocation based on the current state of the process and demand.
- **Customer Support and Service**: Real-time process mining can analyze customer support interactions and service processes as they happen. It helps identify areas where customer queries are getting delayed, allowing support teams to intervene promptly and provide timely assistance.
- **Supply Chain Visibility:** Monitoring supply chain processes in real-time using process mining can provide visibility into the movement of goods, inventory levels, and potential disruptions. This enables organizations to respond quickly to changes in demand or supply.
- **Healthcare Patient Pathway Optimization:** In healthcare settings, real-time process mining can analyze patient pathways, identify delays, and optimize the allocation of medical resources to ensure timely patient care.
- **Energy Management:** Real-time process mining can be applied to monitor energy consumption patterns in buildings or industrial processes. It helps in identifying energy wastage and suggesting real-time adjustments to optimize energy usage.

- **Fraud Prevention:** In financial transactions, real-time process mining can detect unusual patterns or behaviours that might indicate fraudulent activities. Immediate alerts can be triggered for further investigation.
- **Emergency Response Management:** During emergency situations or crisis events, real-time process mining can help organizations manage response processes effectively by identifying bottlenecks, allocating resources, and adapting to changing conditions.
- **Retail Operations:** In retail, real-time process mining can track in-store customer movements, analyze checkout processes, and optimize staff allocation based on real-time foot traffic.
- **Manufacturing Process Control:** Real-time process mining can monitor manufacturing processes, identify deviations from optimal conditions, and trigger adjustments to maintain quality and efficiency.

These examples highlight how real-time process mining can provide valuable insights and enable organizations to make informed decisions and interventions on the fly, ultimately improving operational efficiency, customer satisfaction, and resource utilization.

CHAPTER 5

MODULES

5.1 Module 1: Introduction to Process Mining

Process Mining offers a data-driven and therefore more objective and holistic approach to understanding business processes. As a result, Process Mining has come to dominate a large majority of operational excellence, automation, and digitization ambitions within industry.

Process Mining is the leading new technology when it comes to talking about algorithmic businesses - in other words, businesses that use algorithms and large amounts of real-time data to create business value. This has only become possible through the advent of information systems and administrative tools (e.g. Enterprise Resource Planning or Customer Relationship Management systems) which provide as good data source for process analytics.

Process Mining is a solution to costly and time-intense efforts to get data-driven insights into a business, as acknowledged by the industry research firm Gartner.

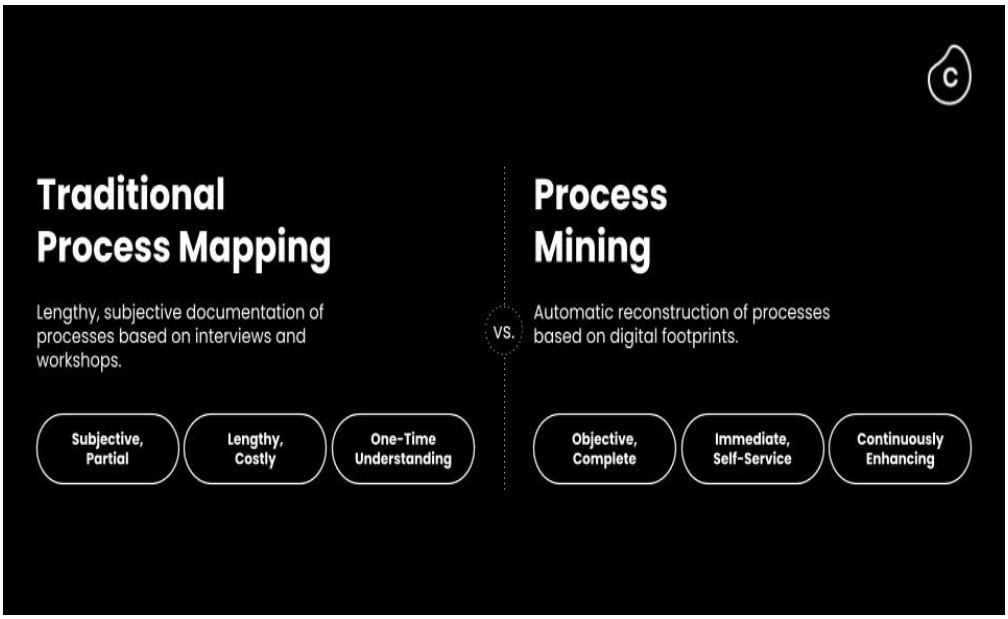

Fig. No. 5.1: Traditional Process Mapping vs. Process Mining

5.1.1 Early Stages

Process Mining originally emerged from academic research into how **event log** data retrieved from Information Systems could be used to discover, monitor, and improve real processes. This real data can facilitate several aspects of **Business Process Management** including:

- Process discovery
- Conformance checking
- Organizational mining, i.e.using data to analyze the roles and people involved in a process
- Automation
- Simulation, i.e.foreseeing and testing the outcome of a process depending on the variation of variables
- Prediction
- History-based recommendations

5.1.2 Event Logs

Event Logs are the format in which we can retrieve our digital footprints from the underlying IT systems. They are essentially the logbooks that IT systems keep recording what events take place for each Case ID and at what time.

The Event Log information can be retrieved from several types of IT systems such as Enterprise Resource Planning (ERP), Supply Chain Management (SCM) or Customer Relationship Management (CRM) systems. These systems typically generate and store Event Log information in real time. Event Log information might also be retrieved in various situations and contexts from automated payment to customer journeys.

An Event Log contains each of the three key pieces of information that our digital footprints have:

- **A Case ID:** A unique identifier such as a purchase order item, invoice number or order number
- **An Activity:** The description of what has happened - for example, the creation of a purchase order or the receipt of goods
- **Timestamp:** The date and time that the activity took place

With these data points, you can reconstruct a process flow for a particular Case ID and aggregate the information across all Case IDs. A process is very simply a series of linked actions or steps taken to achieve a particular end. For customer service, these could be the steps to resolve a ticket. For sales, it could be the steps to progress an opportunity from a lead to closure Take order management, for example. This could be the steps from a customer ordering goods, to you shipping, and then ultimately getting paid for them.

Example

Purchasing a gift online, ordering dinner via an app, or driving processes forward at work. Our everyday lives are full of activities that leave digital footprints.
Let's take a closer look at these digital footprints to understand what they consist of and how you need to structure them to derive a story out of your data.

Let me give you an example.
Imagine you were ordering dinner via an app, and you get your unique order number. All related activities are assigned to that order number.

The activities that take place are the selection of items in a menu, the payment, the preparation of the meal, the meal pick-up, and finally the delivery arriving at your doorstep.

Each of these activities leaves a digital footprint with a timestamp indicating precisely when each activity took place. This is all the information needed to create an event log so that we can start process mining.
This is quite a simple concept.

5.2 Module 2: Process Mining Fundamentals

Process mining is an analytical discipline for discovering, monitoring, and improving processes as they are Process Mining works by extracting knowledge from event logs (also called digital footprints) readily available in today's information systems, to visualize business processes and their every variation as they run. The Celonis Execution Management System (EMS) extends process mining by executing on insights automatically and orchestrating your existing technologies.

Some organizations spend their resources trying to reconstruct the process only to see pieces of the entire picture, and only at a certain point in time. Others use the digital footprints from their transactional systems to get an objective, real-time perspective on their process. Congrats, your organization is of the latter type.When interacting with the dynamic visual representation and drilldown tools such as tables and charts, one can take an exploratory approach or a confirmatory approach.

- **Exploratory Approach:** An exploratory approach is one where you simply explore the data and see what value opportunities jump out at you. You are diving into the data without specific expectations and with an open mind. Analysis tools such as the Process Explorer, the Variant Explorer, and the Conformance checker are ideal for this.
- **Confirmatory Approach:** With the confirmatory approach, you are examining the data to see if it confirms or denies a hypothesis. Using your Celonis Analysis, specifically by filtering on attributes and using drilldown tables, you can find out whether the data confirms or denies that these perceived pain points exist and have a significant impact.

Beyond uncovering inefficiencies and their root causes using Celonis Analysis, our customers choose to use Celonis tools such as Action Flows (process automation) and Celonis Apps to maximize their organization's performance capacity. In this sense, they do not stop at Process Mining and leverage all that the Celonis Execution Management System (EMS) has to offer.

- **Process:** A process is a series of linked steps taken to achieve a particular goal.
- **Activity**: An activity is a step that occurs in the process. Process activities are actions that initiate or terminate a process or take place during it. Each activity consists of one or more tasks that together are a milestone in the process.

- **Case:** A case is an “item” or “object” you follow through the process. Even for the same business process, the case differs from company to company, depending on how granular they want to get.

5.2.2 Variant Explorer

As the name implies, using the Variant Explorer, you can discover all the process variants that is all the different ways the process flows in your organization. The Variant Explorer is one of the Analysis tools to help you take an "exploratory" approach to find out how your process is performing. Watch the video below for an overview.

In the images so far and in the guided tour, the Variant Explorer was set to the Case Frequency KPI. Represented by a number, a Key Performance Indicator (KPI) allows you to quickly assess how your process is performing.

- **Case Frequency:** case frequency KPI reflects the number of unique cases associated with an activity or connection. In a single variant, naturally, the number is the same across the activities and connections.

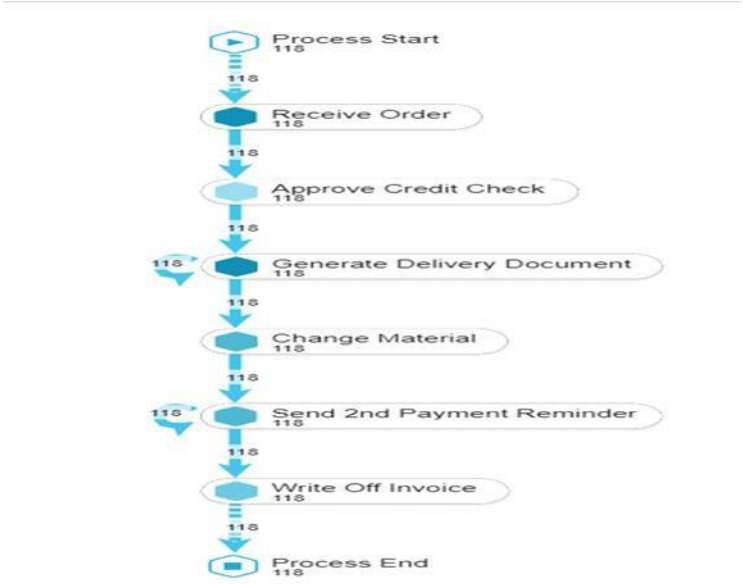

Fig. No. 5.2: Case Frequency

Case Frequency shows 118 cases associated with this variant. Therefore, the count is the same across this entire single variant.

- **Activity Freqeuncy:** Activity Frequency shows how many times each activity occurred in total (236 times) for the 118 cases in the variant.

- The activity frequency on "Generate Delivery Document" (236) is exactly double the case frequency (118); this reflects the fact that each case in this variant goes through "Generate Delivery Document" twice, as indicated by the loop. The same is true for the activity, "Send 2nd Payment Reminder."
- Above is a side-by-side comparison of the same variant with Case Frequency and Activity Frequency KPIs. Notice the difference in the count for the "Generate Delivery Document" activity.

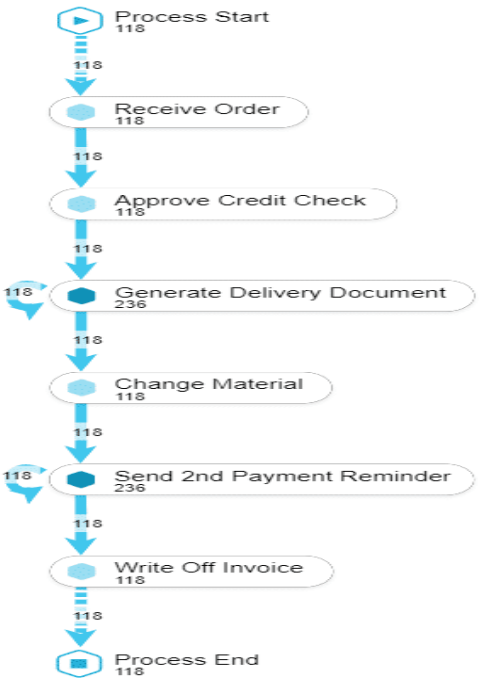

Fig. No. 5.3: Activity Frequency

5.2.3 Process Explorer

The Process Explorer is another analysis tool to use when taking an exploratory approach. It is especially useful for quickly revealing activities beyond the most common ones. It also allows you to narrow your focus on a single activity, for example an undesired activity, to see which activities cases typically come from and which activities they are going to.

In the Process Explorer, if you display the Throughput Time KPI, you are looking at the time it took all the cases in the analysis to go directly between the two displayed activities. That is unlike in the Variant Explorer where the time is reflective of the cases in the variant or variants selected.

These metrics and KPIs are customizable by the person who creates the analysis. A common custom KPI is automation rate; that is the percentage of time when the activity was completed automatically and not manually.

5.2.4 Analysis Charts

A dimension is a category of attributes; for example, the dimension "customer name" is a category for individual customer names. Other examples of dimensions, depending on the nature of the process, can include vendor name, sales organization, region, and material group.

Key Performance Indicators (KPIs) are used to calculate and add aggregated values, for example, case count, order value, invoice value, throughput time, and automation rate.

In case you are a bit fuzzy on these concepts, do not worry, it'll get clearer as we look at dimensions and KPIs in a column chart and an OLAP table.

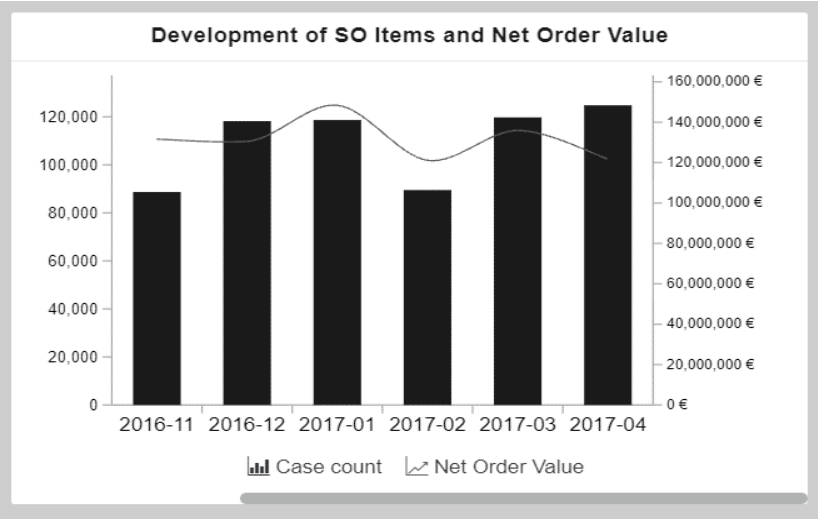

Fig. No. 5.4: Development of SO items and Net Order Value

This chart shows the development of sales order items (KPI) and the corresponding net order value (KPI) over a period of time (dimension). The x-axis displays the dimension, the creation date of sales order, grouped by months.

The two y-axes display the KPIs: The columns display the number of sales order items (case count) and the line displays the net order value.This OLAP table is currently displaying three KPIs for all the sales organizations.

The first column displays the dimension, Sales Organization. The other three columns show KPIs: number of sales orders, average cycle time, and order value.

5.3 Module 3: Rising Star Technical

In the course of digitization, an increasing number of log data is recorded in IT systems of companies worldwide. This data is precious, as it represents how business processes are running inside a company. Process Mining comprises data-driven methods to discover, enhance and monitor processes based on such data. The heart of Process Mining are the Event Logs.

5.3.1 Celonis PQL

To gain valuable process insights, it is essential for Process Mining users to formalize their process questions as executable queries. For this purpose, we present the Celonis Process Query Language (Celonis PQL), which is:

- a domain-specific language
- tailored towards a particular process data model and
- designed for business users.

It translates process-related business questions into queries and executes them on a custom-built query engine, the Celonis PQL Engine.

5.3.2 Celonis Software Architecture

As you can observe in the graphic below, Celonis PQL is an integral component of the Celonis Software Architecture. All Celonis applications use this language to query data from a data model.

- Source System
- Data Model
- Data
- Celonis PQL Engine
- Applications

Metadata is data about the data or documentation about the information which is required by the users. In data warehousing, metadata is one of the essential aspects. Metadata is used for building, maintaining, managing, and using the data warehouses. Metadata allow users access to help understand the content and find data.

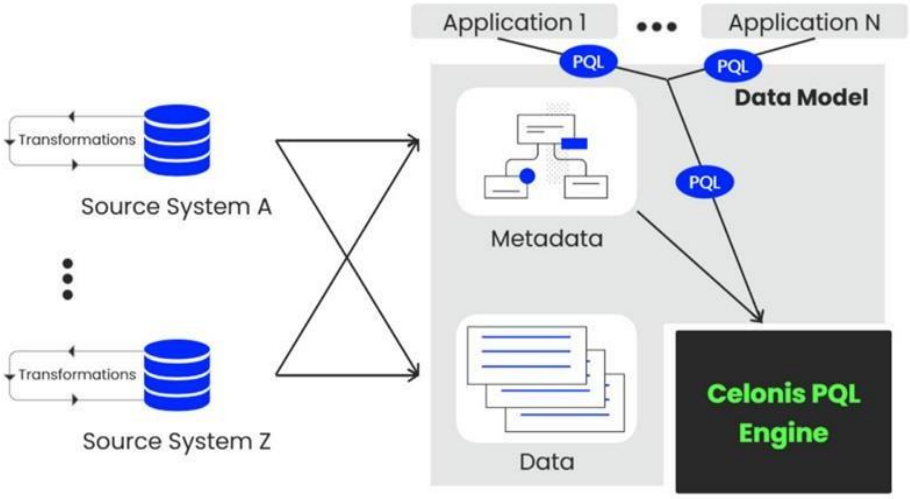

Fig. No. 5.5: Celonis Software Architecture

5.3.3 PQL Queries

PQL can be written in a lot of different applications. You can apply it in Analyses, Knowledge Models, Action Flows and so on. But when it comes to writing queries, you shouldn't be worried about visualization or design. You'd want to write a query, see its output and validate it to see if it is what you expect. And this is where the Data Explorer comes into play.

PQL is a declarative language that is based upon temporal logic. Temporal logic is an extension of traditional propositional logic with operators that refer to the behavior of systems over time. These behavioral operators, called predictes in PQL, provide PQL with a mathematically precise means for expressing properties about the relation between activities and events in process instances.

The design of the PQL language follows seven principles:

- **Compactness:** PQL queries should allow capturing intents in short, succinct programs that avoid ungrounded code redundancy.
- **Decidability:** PQL queries should be solvable by algorithms on a wide range of inputs.
- **Efficiency:** PQL queries should require reasonable and attainable amounts of computational resources.
- **Expressiveness:** PQL queries should allow describing many ideas.
- **Intuitiveness:** PQL queries should be easy-to-read and easy to comprehend.

- **Portability:** PQL queries should be independent of execution environments and data formats.
- **Usefulness:** PQL queries should allow fulfilling many practical tasks.

5.3.4 The P2P Process

P2P is the process of purchasing goods as a company. After creating a purchase order in the system containing information about the products and the vendor, the company receives the goods and pays the invoice from the vendor. As we want to analyze the process on a very granular level using Celonis PQL, the cases we are following through the process are purchase order items.

Activity table is structured, we see that for a given purchase order item number, there are different process steps/activities, such as creating the request, creating the item, receiving the goods and invoice. And every single process step has a corresponding event time. One of the earliest standard aggregation techniques you learned as a child was probably counting - for example, to express your age or how many matchbox cars you possess.

Taking the example of matchbox cars further, you might have wondered how many distinct types your car collection consists of and how many cars you owned per type on average. Since your cars could have attributes like size and price, you could have also described your collection in terms of the total monetary value your cars sum up to altogether.

You can easily map this intuitive example to what you can do with your data using the standard aggregation functions with Celonis PQL. Besides, counting, sum, distinct counting, and average, many more standard aggregations such as summary statistics (min, max, median, quartiles, standard deviation) await you.

5.3.5 Joins and Filters

The tables in a Data Model are connected via specific relationships to associate rows of one table with rows of another table. This is done using a **foreign key**. In general, these relationships can be classified as:

- One-to-many or 1:N
- One-to-one or 1:1
- Many-to-many or N:M

Depending on the number of rows of one table that can be matched with a row of another table. In Celonis Data Models only one-to-many (1:N) relationships are supported.

Every asset in Celonis has an underlying Data Model with multiple tables and **1:N relationships**. The joins between those tables are **left-outer joins**, where the N-side is on the left.

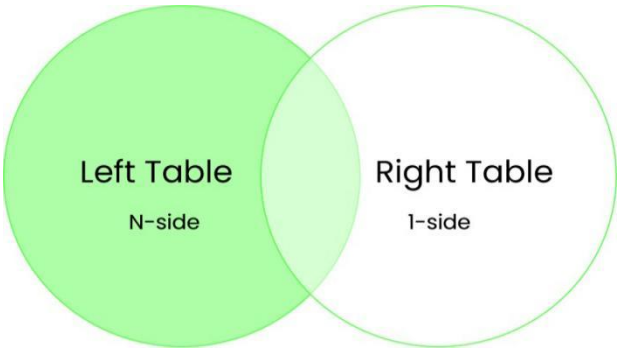

Fig. No. 4.6: Left Join

When you write a PQL-query, all tables involved are pulled to a common table first, which means that the tables are **joined implicitly**. This has consequences, for example, in KPI calculation involving several tables. The common table is always the table on the most N-side.Filters can be created on dashboards and used to modify the information displayed on all dashboards. With filters, you can create subsets of your data to have a closer look at particular parts of the process. Filters can be created from the header bar and the Filters panel.

5.3.6 Data Integration

As a data engineer or analyst working in Data Integration (formerly known as Event Collection), you're responsible for bringing in clean, real-time process data into the EMS. In other words, you build the data pipeline.Process Data is a set of connected activities with timestamps following one specific case, or object. Every activity is an "event" and your task is to collect these events and organize them in the right order. Well, that's precisely what Data Integration is for. It helps you **connect** to source systems, **extract** the relevant data, **transform** it to your needs, and **load** it into a polished Data Model. You can think of the Data Model as the fuel to all other work in your EMS. Once it's ready, your team picks it up and can get started on analyzing it and acting on it.

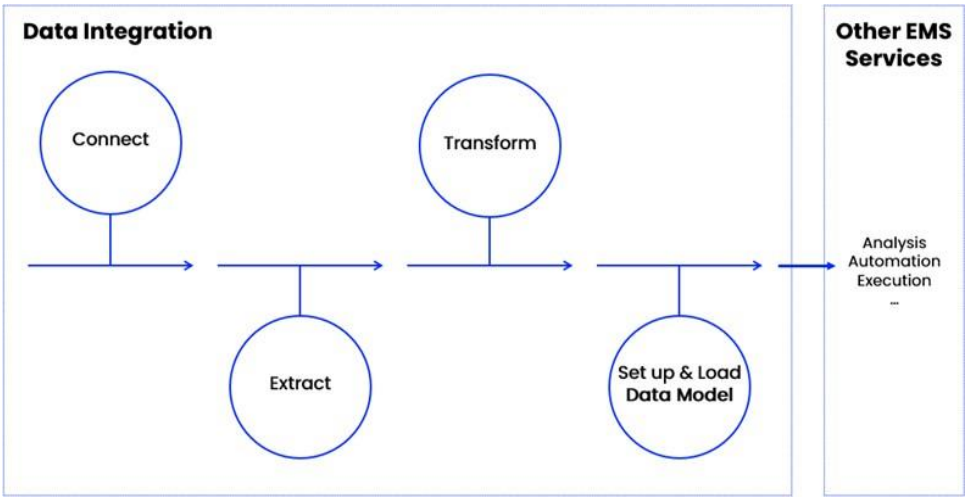

Fig. No. 4.7: Data Model Built

- **Connecting to Multiple Data Sources**:

Celonis connects to various data systems (e.g., ERP, CRM, financial systems) using different integration technologies such as APIs, direct database access, and system-specific connectors. This ensures that process data from multiple sources is pulled into the platform.

- **Extracting Data**:

  Data from the source systems is extracted in two ways:

  - **Full Extraction**: A complete extraction of entire tables or datasets.
  - **Delta Extraction**: Partial extraction, pulling only new or changed data, allowing for more efficient and incremental data updates.

- **Transform Data**: The goal here is to prepare your data for building a Process Data Model. The key task is creating the **Activity Table** (also known as the event log), which forms the foundation for visualizing the process flow.

- **Data Model**: This is a visual representation of an information system or its components, showing the relationships between data points and structures, helping communicate how the process flows.

- **Execution Management System (EMS)**: Commonly used in the financial industry, an EMS is a software platform for executing trades in financial markets. It

provides tools for routing orders, accessing market data, managing orders, and executing trades efficiently across multiple venues and asset classes.

Importance of Data Integration in Celonis Process Mining:

- **Comprehensive Process Insights**: By integrating data from different sources, Celonis enables end-to-end process visibility, allowing organizations to gain insights across different systems and departments.
- **Data-Driven Process Optimization**: Integrated data ensures that the process mining results are based on the actual behavior of systems and users, not just isolated data points.
- **Real-Time Monitoring**: With continuous data integration, organizations can monitor their processes in real-time, enabling immediate identification of inefficiencies, bottlenecks, and deviations.

CHAPTER 6

PROCESS MINING BENIFITS AND ITS USE CASES

6.1 Use Cases

The potential of process mining is not limited to any specific industry or business model. Any enterprise that follows processes can use this, technology to reach maximum efficiency.

Process mining allows financial organizations to discover the possibility of automation

- **Purchase-to-Pay (P2P) Process Optimization**

  Analyze and improve the procurement process, from creating purchase orders to paying invoices, to reduce cycle times and improve vendor relationships.
- **Order-to-Cash (O2C) Process Analysis**

  Identify inefficiencies in order fulfillment, invoicing, and payment processes to speed up cash flow and improve customer satisfaction.
- **Customer Service Process Improvement**

  Optimize customer support workflows, reducing ticket resolution times and improving overall service quality.
- **Compliance Auditing**

  Ensure that all steps in business processes adhere to regulations and internal policies, preventing compliance violations.
- **IT Service Management (ITSM)**

  Analyze IT service processes, such as ticket handling and incident management, to improve service delivery and reduce resolution times.
- **Supply Chain Optimization**

  Enhance supply chain operations by identifying inefficiencies in logistics, warehousing, and delivery processes to reduce lead times and costs.
- **Healthcare Process Optimization**

  Analyze patient journeys and hospital administrative processes to improve patient outcomes and reduce wait times.
- **Financial Processes (Accounting and Auditing)**

  Optimize accounting and financial processes, such as invoicing and account reconciliation, to improve efficiency and accuracy.

- **Education:** Educational process mining (EPM) allows administrators to analyze and visualize students' learning behavior by applying specialized algorithms. The student activity logs provide insights into tracking and monitoring their academic performance.

6.2 Benefits

Process mining technology, businesses can improve their process intelligence to create ideal workflows and operations. As a result, more and more companies are leaning toward adopting this software to reach their maximum efficiency potential because process mining:

- **End-to-End Process Transparency**

  Process mining provides a clear view of how processes are executed from start to finish, making deviations, bottlenecks, and inefficiencies visible.
- **Data-Driven Decision Making**

  By using real event logs, process mining provides objective, fact-based insights to inform business decisions.
- **Improved Process Efficiency**

  Identifies inefficiencies such as unnecessary steps or bottlenecks, allowing businesses to optimize workflows and reduce delays.
- **Root Cause Analysis**

  Helps pinpoint the exact causes of process inefficiencies or deviations, enabling targeted interventions.
- **Continuous Process Improvement**

  Enables ongoing monitoring and optimization of processes, allowing businesses to iteratively improve their performance.
- **Compliance and Auditability**

  Ensures that business processes follow regulatory and internal guidelines, helping to prevent compliance violations.
- **Better Customer Experience**

  Optimizing processes improves customer satisfaction by reducing response times, errors, and delays in service delivery.

CHAPTER 7

LEARNING OUTCOMES

After completion of this training, we should be able to:

- Understand what Process Mining is and the basics of how it works.
- To extract insights from event logs, identify bottlenecks, inefficiencies, and opportunities for optimization.
- To extract and create visual representation of processes to aid decision making and process improvement efforts.
- Attain skills in using process mining tools and interpreting the results to enhance organizational efficiency and effectiveness.
- Summarize what an Event Log is and why we need it for Process Mining.
- Identify business use cases for Process Mining.
- Understanding how to discover, analyze and improve business process using data driven techniques.

CONCLUSION

Process mining is a powerful methodology that offers organizations valuable insights into their operational processes, enabling them to enhance efficiency, compliance, and overall performance. It was a valuable experience. It helped to identify, where improvements could be made to make things run smoother and more efficiently. This internship taught me practical skills, like working with data and collaborating with different experts. Overall, it was a great opportunity to learn and contribute to making processes better

INTERNSHIP CERTIFICATES

REFERENCES

[1] Celonis. "Process Mining Software | Execution Management System." *Celonis*, 2023, https://www.celonis.com

[2] Explore the Best Practices and Tools September 16, 2022 by Washija Kazim "https://www.g2.com/articles/process-mining"