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 (Data Science)

by

S Sarfaraaz Ahmed (214G1A3297)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY
(AUTONOMOUS)

(Affiliated to JNTUA, accredited by NAAC with 'A' Grade, Approved by AICTE, New Delhi & Accredited by NBA (EEE, ECE & CSE))
Rotarypuram village, B K Samudram Mandal, Ananthapuramu-515701.

2023 - 2024

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY
(AUTONOMOUS)

(Affiliated to JNTUA, accredited by NAAC with 'A' Grade, Approved by AICTE, New Delhi & Accredited by NBA (EEE, ECE & CSE))
Rotarypuram village, B K Samudram Mandal, Ananthapuramu-515701.

Department of Computer Science & Engineering (Data Science)

Certificate

This is to certify that the internship report entitled "Process Mining" is the bonafide work carried out by **S Sarfaraaz Ahmed** bearing Roll Number 214G1A3297 in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering (Data Science)** for four months from June 2022 to September 2022.

Internship Coordinator

Mr. P. Veera Prakash, M. Tech., (Ph.D.),
Assistant Professor & HOD of CSE

Head of the Department

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

Date:

Place: Ananthapuramu

EXTERNAL EXAMINER

P

PREFACE

All India Council for Technical Education (AICTE) has initiated various activities for promoting industrial internship at the graduate level in technical institutes and Eduskills 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 Academicand 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, EduSkills & AICTE launches 'Virtual Internship' program on Celonis Process Mining. This field is one of the most in-demand, and this internship will serve as a primer.

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 **Mr. P. Veera Prakash, Assistant Professor & HOD, 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 (Data Science),** 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.

S Sarfaraaz Ahmed

(214G1A3297

INDEX

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
BPM	Business Process Management
CSV	Comma Separated Value
EMS	Execution Management System
ERP	Enterprise Resource Planning
IT	Information Technology
KPI	Key Performance indicator
PQL	Process Query language
SCM	Supply chain management
XES	Extensible Event Stream

CHAPTER - 1

INTRODUCTION

Process mining is a revolutionary approach that unveils the hidden dynamics of business processes through data analysis. In a world where every digital interaction generates a trail of data, process mining harnesses this information to create a vivid map of how processes truly unfold. It transcends traditional process documentation, offering a real-time, data-driven perspective on how work gets done.

- This manual modelling could take months, cost $100,000 in consultancy fees, or employee wages and was subject to human error, like any other manual activity. And the data it was based on very often coming from assumptions human observations was incomplete at best, and simply inaccurate at worst. Now we know something much better exists it's called process mining

- At its core, process mining is about discovery, transparency, and optimization. It transforms raw event data—capturing each task, decision, and interaction into actionable insights. These insights expose inefficiencies, highlight bottlenecks, and reveal the paths that deviations take. Process mining doesn't just rely on assumptions or theoretical models; it's grounded in the reality of Organizations and operations

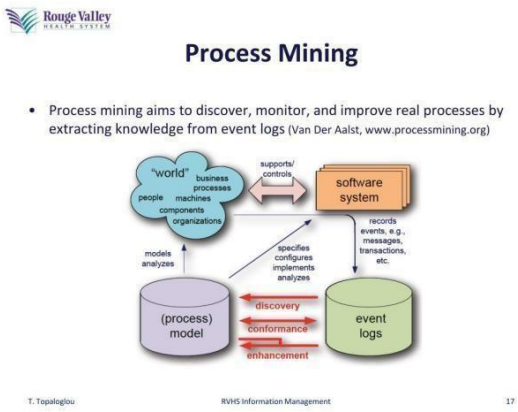

Fig. No. 1.1: Process Mining

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 paymentis 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 goto a different address, or the payment fail. 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 in1955 and in 2015, we see that only 12% of them managed to remain on the list. Takinga closer look at the customer expectations, please take a momentand 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 analyze, visualize, and improve business processes. It leverages techniques from data mining, process modeling, 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.

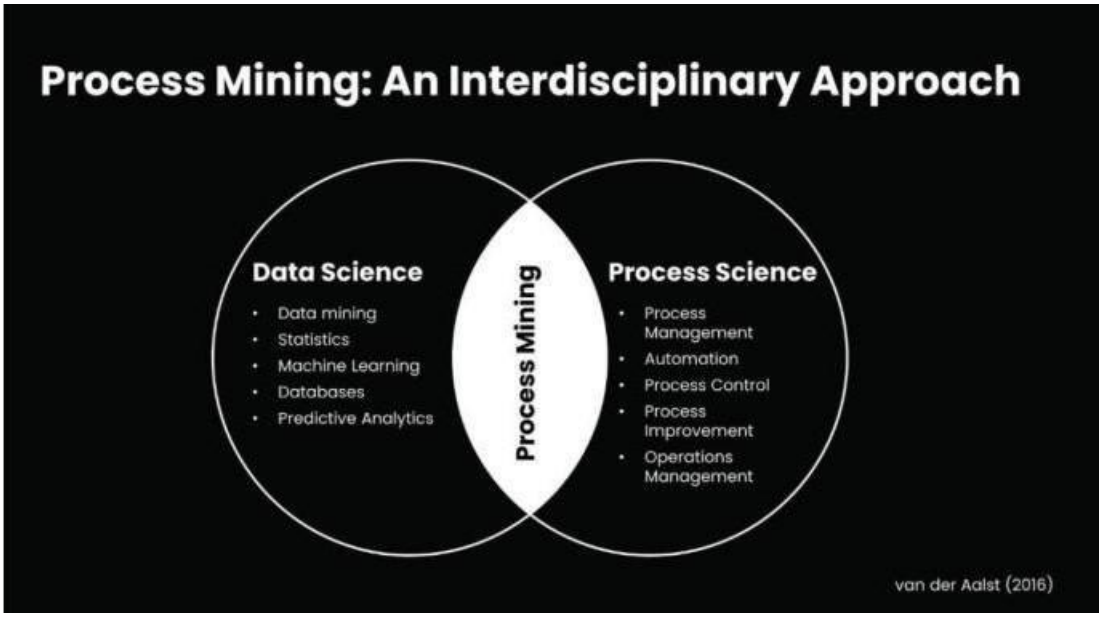

Fig. No. 2.1: Process Mining Approach

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

2.4 Process Mining Techniques:

Process mining encompasses several techniques that help organizations analyze event data and gain insights into their business processes. These techniques enable organizations to discover, understand, and improve their processes. Here are some prominent process mining techniques.

1. **Process Discovery**: This technique involves extracting process models from event logs. It aims to automatically generate a visual representation of the process flow basedon the recorded events. Common process discovery algorithms include Alpha-algorithm, Heuristic Miner, Inductive Miner, and others.

2. **Conformance Checking:** Conformance checking compares the actual process execution with the expected process model. It identifies deviations, discrepancies, and non-compliance issues between the modeled process and the real execution. This helpsorganizations assess how closely their processes adhere to the intended design.

3. **Performance Analysis:** Performance analysis focuses on extracting quantitative performance metrics from event logs to assess process efficiency, cycle times, waiting times, and resource utilization. Throughput analysis, bottlenecks identification, and efficiency metrics are part of this technique.

4. **Variant Analysis:** Variant analysis aims to understand the different paths and variations that processes can take based on event data. It helps organizations identify both the most common and exceptional paths that processes follow, providing insightsinto process flexibility and adaptability.

5. **Predictive Analysis:** Process Mining involves using historical event data to forecast for future process behaviour. By applying Predictive models to the event logs organizations can make informed decisions and anticipate the potential process outcomes.

2.5 Understanding of Event data:

What is Event data?

Event data in process mining refers to the collection of detailed information about the various events, actions, and interactions that occur within a business process. This data includes timestamps, identifiers, and attributes associated with each event, and it serves as the basis for conducting process mining analyses. Event data provides a chronological record of the steps taken during the execution of a process, enabling organizations to gain insights into process behavior, identify patterns, and make informed decisions for process improvement.

key components of event data in process mining include:

1. **Events:** Events are individual occurrences or actions that take place within a process.These actions can be carried out by humans, systems, or machines. Each event represents a specific activity or interaction.
2. **Timestamps:** Timestamps indicate when each event occurred. They help establish the temporal order of events and allow for the reconstruction of the process timeline. Timestamps are essential for analyzing the duration between events and identifying potential bottlenecks or delays.
3. **Event Attributes:** Event attributes provide additional context and information about each event. These attributes can include the type of activity, the resource responsible for the event, the outcome or result of the event, and any other relevant metadata.
4. **Event Logs:** Event logs are structured collections of recorded events. They are oftenstored in formats such as CSV, XES (Extensible Event Stream), or database tables. Event logs serve as the input data for process mining analyses.
5. **Process Analysis:** Event data is used for various process analysis tasks, including conformance checking, performance analysis, and compliance monitoring.Through analysis, organizations can identify deviations from expected behavior and opportunities for improvement.

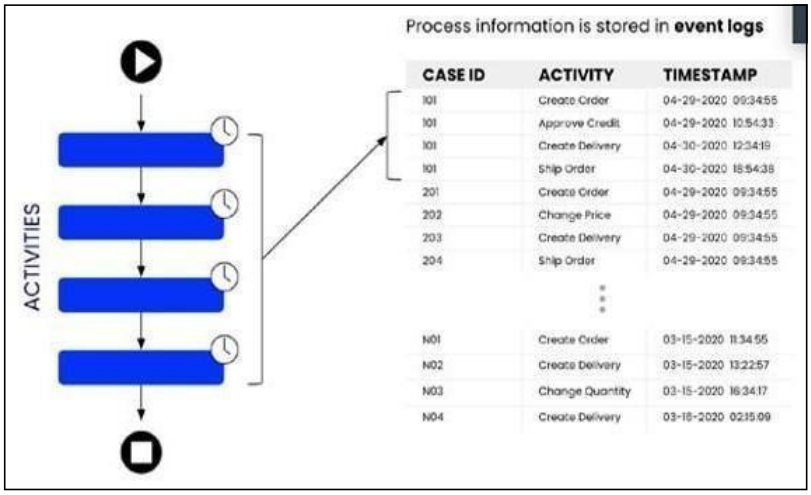

Fig. No. 2.2: Event Logs

Event data is processed and analyzed using various process mining techniques to achieve different goals, such as:

1. **Process Discovery:** Automatically generating process models based on theevent data to visualize the actual process flow.
2. **Conformance Checking:** Comparing the event data against a predefined process model to identify deviations and non-compliance.
3. **Performance Analysis:** Analyzing event timestamps and attributes to evaluateprocess efficiency and identify areas for optimization.

Overall, event data is the foundation of process mining and plays a critical role in understanding, analyzing, and improving business processes by providing data-driven insights.

CHAPTER - 3

FUNDAMENTALS OF PROCESS MINING

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!

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 ExecutionManagement 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

3.1 Variant Explorer:

Variant Explorer is a Celonis EMS Analysis tool that helps you explore how a specific process flows through your organization.

If we think about a process as a road trip, each process variant would be a potential route. Each activity within a process would be a waypoint along a route, and the connections between activities are like the roads that connect the stops. And, each trip a person makes along a particular route would be a case.

Using Variant Explorer, you can see the individual activities within each process variant and the frequency of each variant (based on number of cases). You can also compare variants to each other and see metrics for individual variants, such as Activity Frequency and Throughput Time.

In short, Variant Explorer gives you a quick way to see whether most process cases follow an acceptable flow of activities or not and helps you develop your first analysis questions.

1. 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.

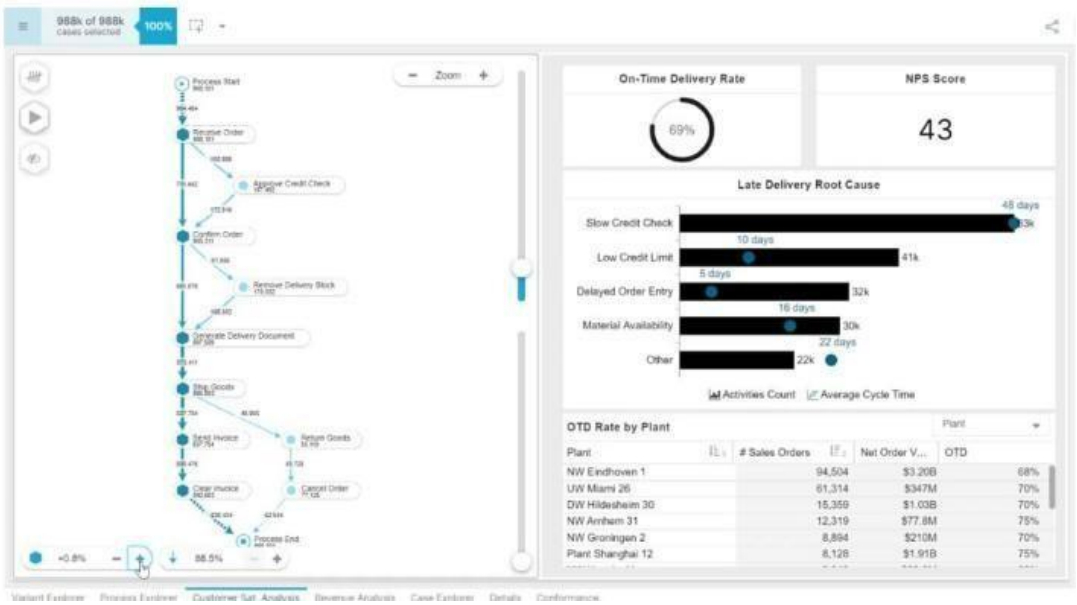

Fig. No. 3.1: Case Frequency

2. 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."

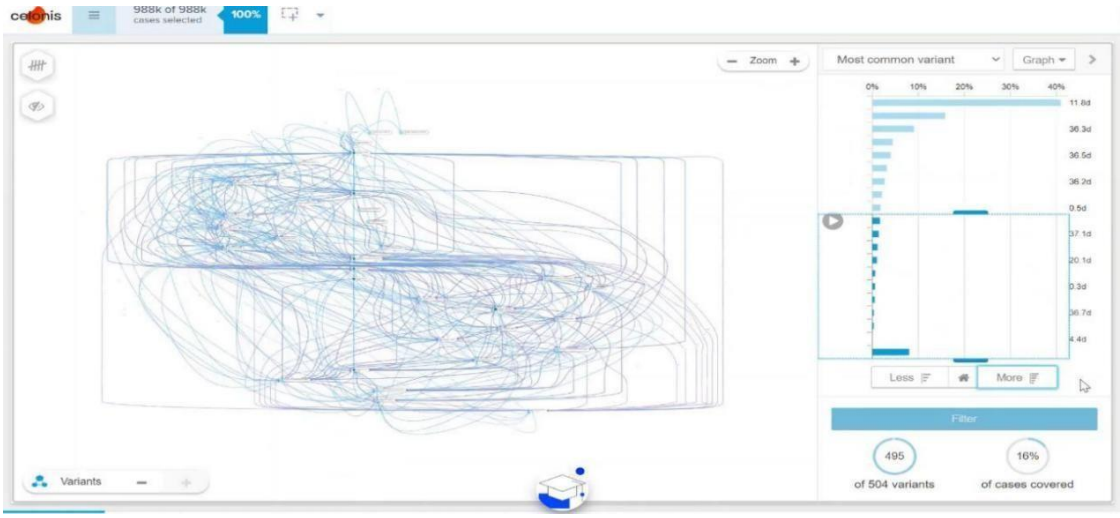

Fig. No. 3.2: Graphical Division

3.2 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.

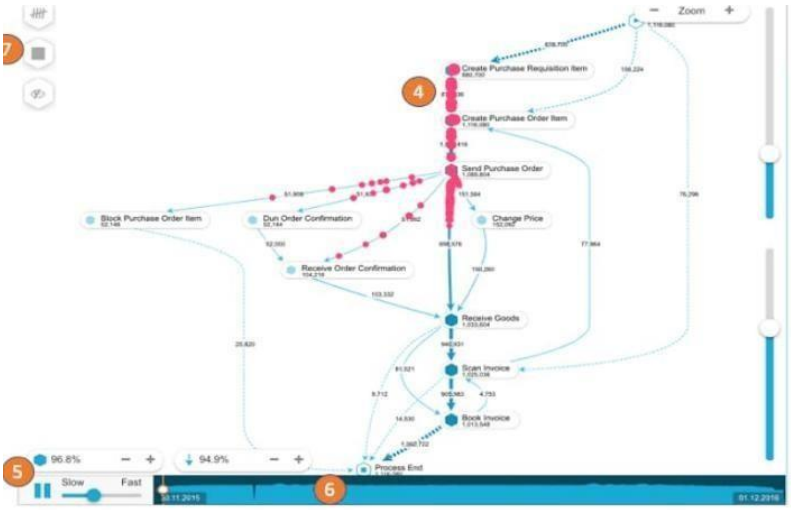

Fig. No. 3.3: Process Explorer

3.3 Analysis Charts :

Charts:

You can visually display your data using area, bar, line, lollipop, scatter plot, and stacked charts in your studio view. Each chart type can then be customized further, giving you control over their look and feel.

Types of charts

- Area
- Bar
- Line
- Lollipop
- Scatterplot
- KPI

1.Area

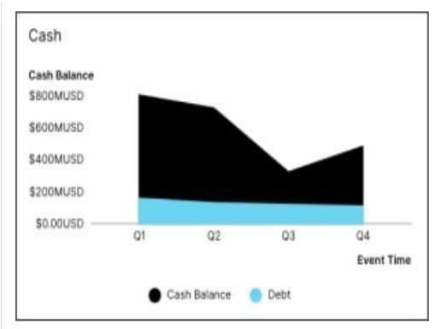

Fig. No. 3.4: Area

2.Bar

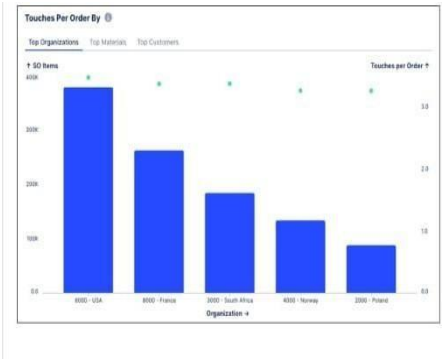

Fig. No. 3.5: Bar

3.Line

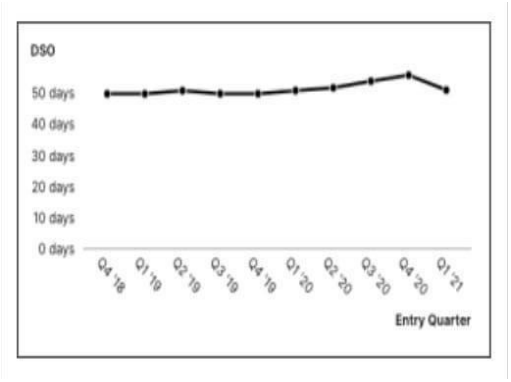

Fig. No. 3.6: Line

4.Lollipop

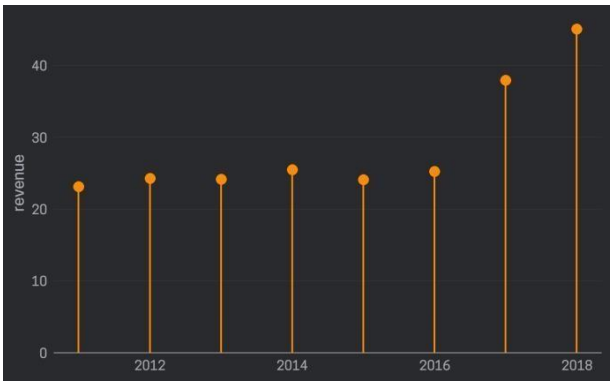

Fig. No. 3.7: Lollipop

CHAPTER - 4

MODULES

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 Mininghas come to dominate a large majority of operational excellence, automation, and digitalization 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 becomepossible 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.

Module-2: Process Mining Fundamentals

The key Fundamentals of process mining refer to the core concepts and principles that underlie the analysis and improvement of business processes using process mining techniques. It involves extracting insights from event data to understand how processes function, identifying bottlenecks, inefficiencies, and opportunities for optimization. Key components include data extraction, process discovery, conformance checking, and process enhancement**.** Fundamentals of process mining refer to the core concepts and principles that underlie the analysis and improvement of business processes using process mining techniques.

It involves extracting insights from event data to understand how processes function, identifying bottlenecks, inefficiencies, and opportunities for optimization. Key components include data extraction, process discovery, conformance checking, and process enhancement. The process mining fundamentals are event log, process discovery, conformance checking, performance analysis, process enhancement,Root cause analysis, variants, and patterns etc.

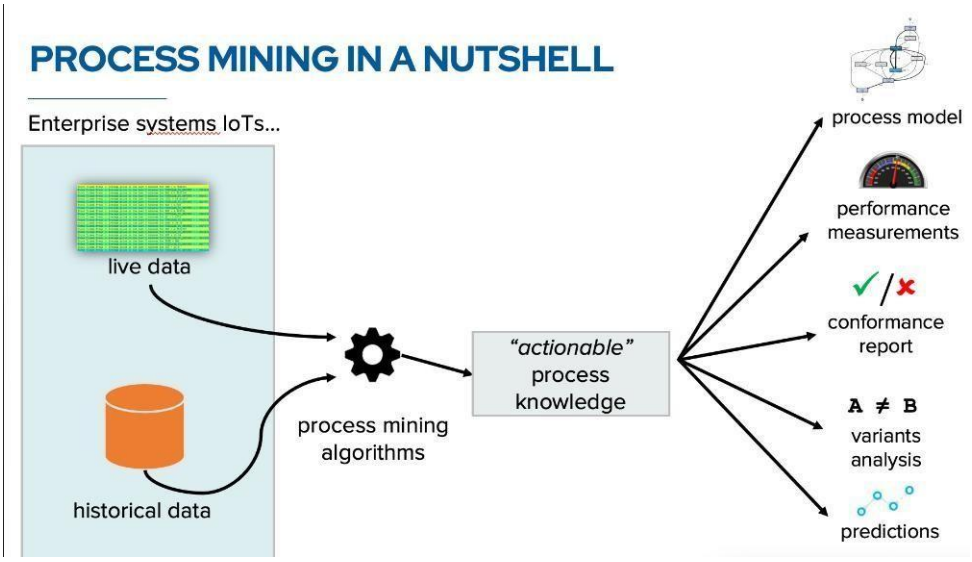

Fig. No. 4.1: Fundamentals of Process Mining

Variant Explorer:

The Variant Explorer is one of the Analysis tools to help you take an "exploratory"approach to find out how your process is performing.

Fig. No. 4.2: Variant Explorer

Module 3: Rising Star Technical

3.1 PQL Queries

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.

3.1.1 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.

3.1.2 Celonis Software Architecture

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.

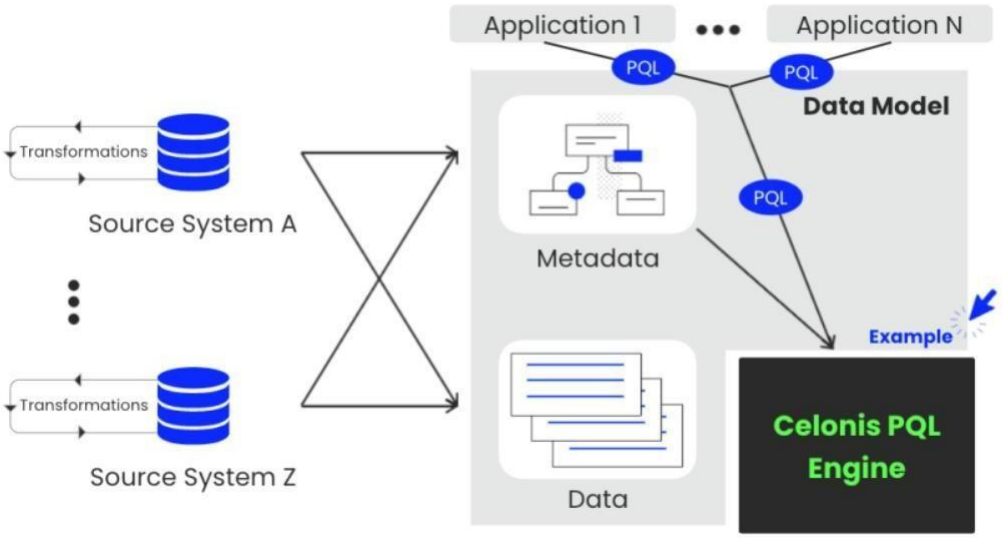

Fig. No. 4.3: Celonis PQL Engine

3.1.3 PQL Queries

PQL enables the user to translate process-related business questions into queries, which are then executed by a custom-built query engine. PQL covers a broad set of operators, ranging from process-specific functions to aggregations and mathematical operators. Its syntax is inspired by SQL, but specialized for process-related queries.

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 five principles:

- **Compactness:** PQL queries should allow capturing intents in short, succinct programs that avoid ungrounded code redundancy.
- **Efficiency:** PQL queries should require reasonable and attainable amounts of computational resources.
- **Expressiveness:** PQL queries should allow describing many ideas.
- **Portability:** PQL queries should be independent of execution environments and data formats.
- **Usefulness:** PQL queries should allow fulfilling many practical tasks.

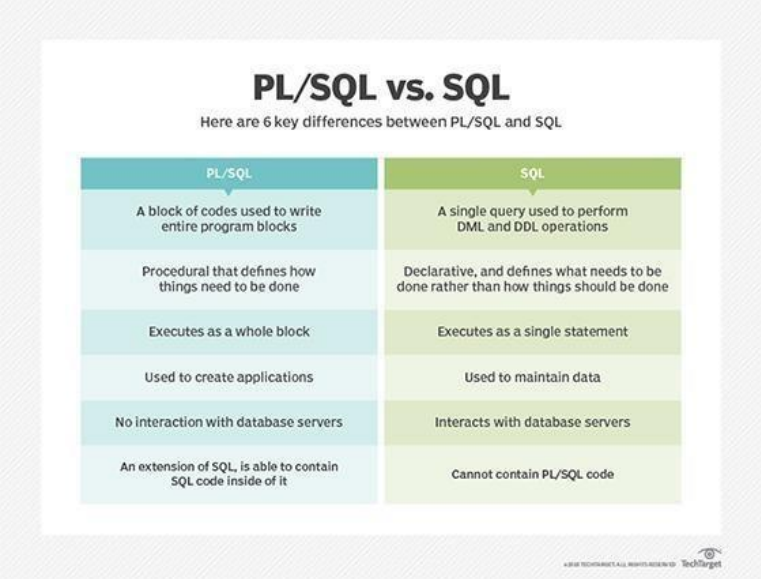

Fig. No. 4.4: PQL VS SQL

3.2 Get data into EMS

3.2.1 Data Integration:

Data integration is the process of bringing data from disparate sources together to provide users with a unified view. The premise of data integration is to make data more freely available and easier to consume and process by systems and users. Data integration done right can reduce IT costs, free-up resources, improve data quality, and foster innovation all without sweeping changes to existing applications or data structures. And though IT organizations have always had to integrate, the payoff for doing so has potentially never been as great as it is right now.

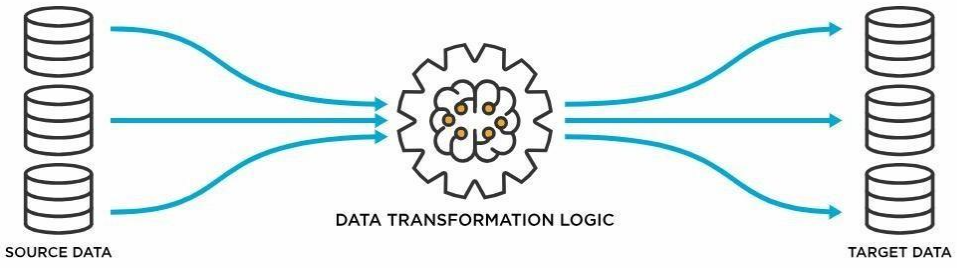

Fig. No. 4.5: Data Integration

3.2.2 Data Extraction:

Data extraction is the process of obtaining raw data from a source and replicating that data somewhere else. The raw data can come from various sources, such as a database, Excel spreadsheet, an SaaS platform, web scraping, or others. It can then be replicated to a destination, such as a data warehouse, designed to support online analytical processing (OLAP). This can include unstructured data, disparate types of data, or simply data that is poorly organized. Once the data has been consolidated, processed, and refined, it can be stored in a central location on-site, in cloud storage, or a hybrid of both to await transformation or further processing.

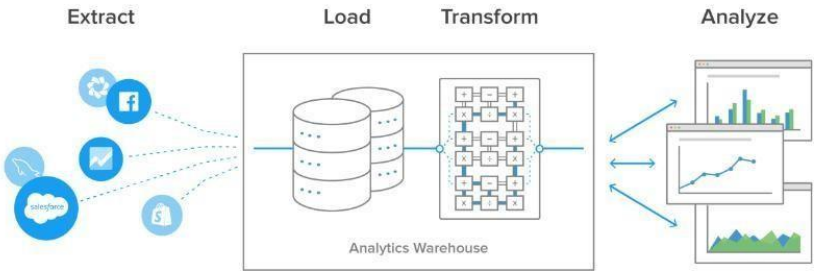

Fig. No. 4.6:Data Extraction

3.2.3 Data Transformation:

Data transformation is used when data needs to be converted to match that of the destination system. This can occur at two places of the data pipeline. First, organizations with on-site data storage use an extract, transform, load, with the data transformation taking place during the middle 'transform' step.

Organizations today mostly use cloud-based data warehouses because they can scale their computing and storage resources in seconds. Cloud based organizations, with this huge scalability available, can skip the ETL process. Instead, they use a transformation process that converts the data as the raw data is uploaded, a process called extract, load, and transform. The process of data transformation can be handled manually, automated or a combination of both.

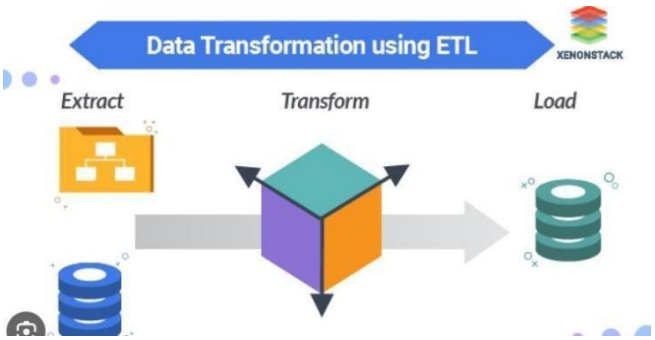

Fig. No. 4.7:Data Transformation

3.2.4 Data Loading:

Data loading defines the LOAD component of the ETL process. ETL stands for Extraction, Transformation, and Load. Extraction deals with the retrieval and combining of data from multiple sources. Transformation deals with cleaning and formatting of the Extracted Data. Data Loading deals with data getting loaded into a storage system, such as a cloud data warehouse.

ETL aids in the data integration process that standardizes diverse data types to make them available for querying, manipulation, or reporting for many different individuals and teams. Because today's organizations are increasingly dependent upon their own data to make smarter, faster business decisions, ETL needs to be scalable and streamlined to provide the most benefit.

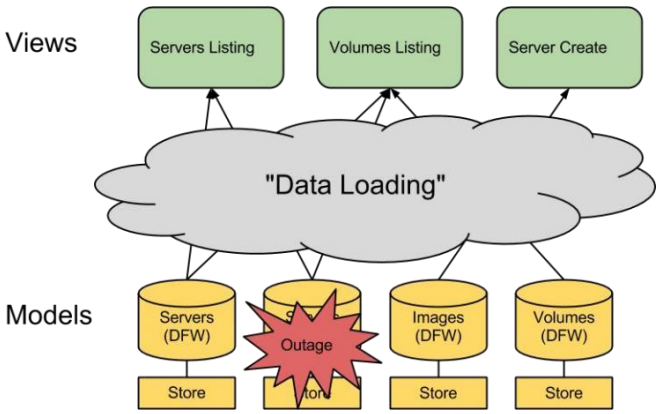

Fig. No. 4.8:Data Loading

CHAPTER - 5

REAL TIME APPLICATIONS

Process mining has a wide range of real-world applications across various industries. Here are some notable examples**:**

1. **Supply Chain Management (SCP):** Process mining helps optimize supply chain processes by analyzing the flow of goods, materials, and information. It identifies bottlenecks, inefficiencies, and areas for improvement, leading to enhanced inventory management,reduced lead times, and improved demand forecasting.

2. **Manufacturing:** In manufacturing, process mining can improve production efficiency by analyzing workflows, identifying production bottlenecks, and streamlining production processes. It also aids in quality control and compliance with manufacturing standards.

3. **Healthcare:** Process mining is used to improve patient care and hospital operations.It assists in optimizing patient flow, reducing wait times, and enhancing resource allocation in healthcare facilities. Process mining can also be employed to analyze treatment pathways and diagnose potential process-related issues.

4. **Financial Services:** In the financial industry, process mining can be used to optimizeloan approval processes, fraud detection, and compliance monitoring. It helps identify irregular patterns and deviations from expected process behavior.

5. Digital Transformation: Process mining is frequently used in larger-scale digital transformation initiatives because it can give you the precise insights needed for process improvement, allowing systems to run more quickly, smoothly, and efficiently, as well as objective data-driven insights into the causes of delays and inefficiencies within business processes.

6. Information Technology: IT processes can be analyzed to improve system deployment, software development, and help desk operations. Process mining aids in identifying opportunities for automation, reducing downtime, and optimizing IT servicemanagement.

7. Human Resources: Process mining can optimize HR processes such as recruitment, on boarding, and performance management. It helps streamline hiring procedures and improve employee satisfaction by identifying areas for efficiency gains.

8. Public Sector: Process mining finds applications in government agencies to optimize administrative processes, increase transparency, and enhance service deliveryto citizens.

9. Pharmaceuticals: In pharmaceuticals, process mining can be employed to improve drug development processes, regulatory compliance, and supply chain management

10. Telecommunications: Process mining helps analyze customer interactions, billing processes, and network operations. It optimizes service delivery and enhances customersatisfaction.

11. Process Optimization: Process mining helps organizations identify inefficiencies, bottlenecks, and deviations from the desired process flow. By analyzing digital footprints, organizations can optimize processes to reduce cycle times, streamline workflows, and allocate resources more effectively.

CHAPTER 6

LEARING OUTCOMES

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

- Understand about what Process Mining is about.
- Interpret process visualizations and leverage analyses to identify process inefficiencies.
- Put your Knowledge about the theoretical foundations of Process Mining into practice.
- Identify business use cases for process mining.
- You will learn how to load data, extract data and transform data into the EMS.
- You will gain skills in using process mining tools and interpreting the resultsto enhance organizational efficiency and effectiveness.
- Understanding process behaviour and it's applications in day life.
- Understanding and getting knowledge on the PQL to perform the technical actions in the process mining
- You will get know about the difference between the PQL and SQL.
- Understanding how process mining helps to uncover inefficiencies and methods for insufficient process.
- We will get to know about the variant explorer, process explorer and charts for the celonis process mining fundamentals.

CONCLUSION

Process mining is a tool that businesses may rely on to conduct audits that are more accurate and efficient while avoiding conjecture and subjective conclusions. This results in less time being lost on doubt and subsequent retesting.

Process mining's ability to visualize process flows, detect bottlenecks, and pinpoint deviations from the ideal path enables organizations to make informed decisions aimed at enhancing efficiency, reducing costs, and improving overall performance. The real-time applications of process mining are particularly noteworthy, as they empower businesses to respond promptly to changing circumstances, address issues as they arise, and ensure that processes operate at their optimal levels.

Process mining Internship 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.

CERTIFICATE

REFERENCES:

1. https://internship.aicte-india.org/

2. Process Mining | Celonis | The Leader in Process Mining

3. https://www.analyticssteps.com/blogs/types-and-applications-process-mining