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

by

Urichintala Ezaaz Basha **214G1A0526**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY
(AUTONOMOUS)

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

Certificate

This is to certify that the internship report entitled "Process Mining Virtual Internship" is the bonafide work carried out by **Urichintala Ezaaz Basha** bearing Roll Number 214G1A0526 in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** for Three months from March 2023 to May 2023.

Internship Coordinator

Mr. P. Veera Prakash,
MTech,(Ph.D.),MIEI,MCSI
Assistant Professor

Head of the Department

Mr. P. Veera Prakash,
MTech,(Ph.D.),MIEI,MCSI
Assistant 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 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 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, 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 **Mr. P. Veera Prakash, Assistant Professor, Department of Computer Science and Engineering**, 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.

Urichintala Ezaaz Basha

214G1A0526

INDEX

CHAPTER 1

INTRODUCTION

In the world and within organizations, a multitude of processes exist. These processes encompass various aspects such as purchasing and order management. Organizations are confronted with intricate, worldwide, and occasionally flawed processes every day. Conversely, ***Frictionless*** processes guarantee:

- Discovering the appropriate groceries at the supermarket,
- Ensuring timely arrival of airplanes,
- Minimizing waiting periods for patients in hospitals.

1.1 Introduction to Process Mining

Process Mining is a valuable tool that bridges the gap between business process management and data mining. This innovative technique offers numerous benefits, including enhanced speed, accuracy, and auditability. By analyzing event logs, Process Mining enables organizations to gain insights that drive improved efficiency, effectiveness, and compliance in their business processes.

Utilizing modern Process Mining technology allows for the quick and reliable extraction of information from event and transaction logs. These extracted data are then transformed into visually engaging real-time process models that accurately represent current processes.

With its ability to uncover hidden patterns and bottlenecks within business processes, Process Mining provides organizations with a competitive edge by enabling them to identify areas for improvement. By leveraging these insights, businesses can achieve significant cost savings while also reducing time-to-market.

Overall, Process Mining is a powerful approach that combines aspects of process management with data analysis to drive organizational success through optimized workflows.

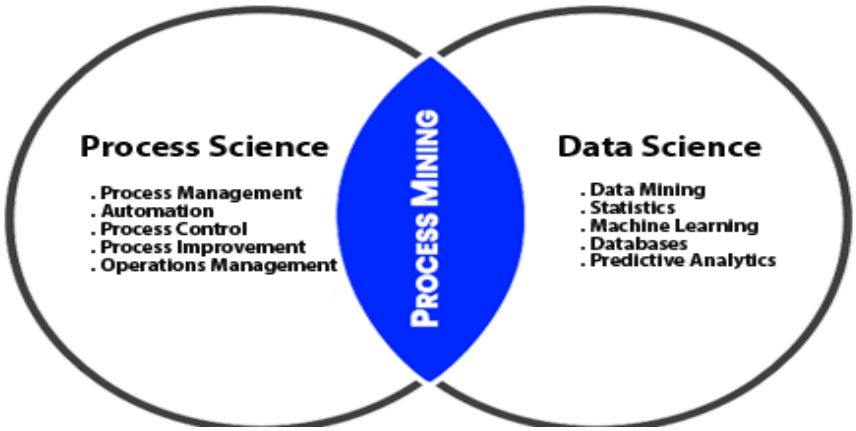

Fig:1.1 Process Mining Interaction between Process and Data Science

The real-world examples of process mining include procurement, order management, compliance, intelligent automation, digital transformation, KPI reporting, accounts payable, accounts receivable, auditing, IT development, service management, logistics, and many more.

Understanding Process Mining:

Process mining is a vibrant field that focuses on extracting valuable insights from event logs to discover, analyze, and optimize real-world organizational processes. This discipline effectively combines data science with process management, allowing us to gain a comprehensive understanding of how processes are carried out and identify areas where enhancements can be made. By uncovering the intricacies of process execution, process mining opens up opportunities for improvement within an organization.

Internship Goals:

During my internship, I have set several goals that I aim to achieve. These goals include becoming proficient in using leading process mining tools such as Disco, ProM, and Celonis. By familiarizing myself with these tools and learning various techniques like process discovery, conformance checking, and data visualization, I will be able to apply them effectively in real-world scenarios. Another goal of mine is to gain hands-on experience in working with raw event data. This involves preparing and transforming the data into a structured format suitable for analysis. I will focus on tasks such as data cleaning, preprocessing, and organizing event logs to ensure accurate insights. Process discovery is also an area that I want to excel in during this internship. It entails uncovering underlying process models from event logs. Once these models are identified, my aim is to analyze them thoroughly in order to identify key process elements, decision points, and variations. Conformance checking plays a vital role in improving processes. Through this analysis technique, I plan on comparing the discovered process models with the actual execution of those processes. By identifying discrepancies between the two and pinpointing areas for improvement within the processes themselves can lead ultimately lead us towards enhanced efficiency. Overall, I am eager about this opportunity provided by my internship. I see it as a chance for personal growth while gaining valuable skills that will benefit me throughout my career journey I understand the significance of effectively communicating complex data insights. To enhance my ability to convey information clearly, I focus on developing my skills in data visualization. Through creating informative dashboards and visual representations, I aim to effectively communicate my findings to different stakeholders. This will lead to improved efficiency and effectiveness in conveying the message.

CHAPTER 2

FOUNDATION OF PROCESS MINING

Process mining encompasses a variety of techniques utilized to gain knowledge and extract insights from processes by analyzing event data generated during their execution. This training program offers a combination of theoretical and practical foundations in the field of process mining. It involves reading the data, transforming it into an event log, and creating visual representations that provide comprehensive analytics for the entire process. An event log consists of every step performed in the process (referred to as activities), along with the corresponding timestamps and case IDs. Utilizing this event log, algorithms generate a process model that accurately reflects the timing of each step and captures all variations within the process.

In simpler terms, a process can be defined as a series of interconnected steps undertaken to achieve a specific objective. Each individual item or object that undergoes this series of steps is referred to as a case. Activities refer to events that occur throughout this process.

Through thorough analysis using advanced techniques, such as converting raw data into meaningful visualizations, organizations can gain valuable insights into their processes' efficiency and identify areas for improvement. Process mining enables businesses to optimize their operations by identifying bottlenecks or inefficiencies in real-time while providing an accurate representation of how processes unfold in practice.

The training program called "Review and Interpret Analyses" is specifically designed for individuals working as data and business analysts, process experts, and process improvement specialists. It is important to note that this program primarily focuses on enhancing product knowledge rather than developing business acumen. If you are interested in complementing your existing experience in strategically identifying and prioritizing process inefficiencies, as well as planning and implementing improvement measures, we suggest considering the "Deliver Business Value with Celonis" training track after completing this one.

Now let's give you a glimpse of what you can expect from the Review and Interpret Analyses training track.

Self-paced Reading and video Demos

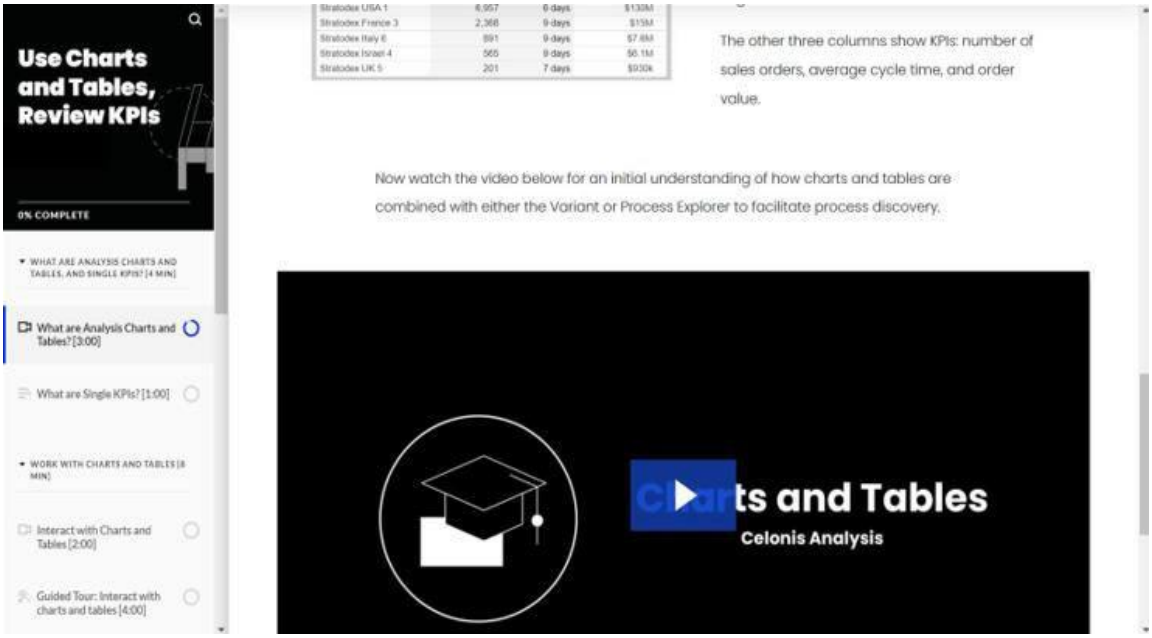

Fig 2.1 Module of Use of Charts and Tables KPIs

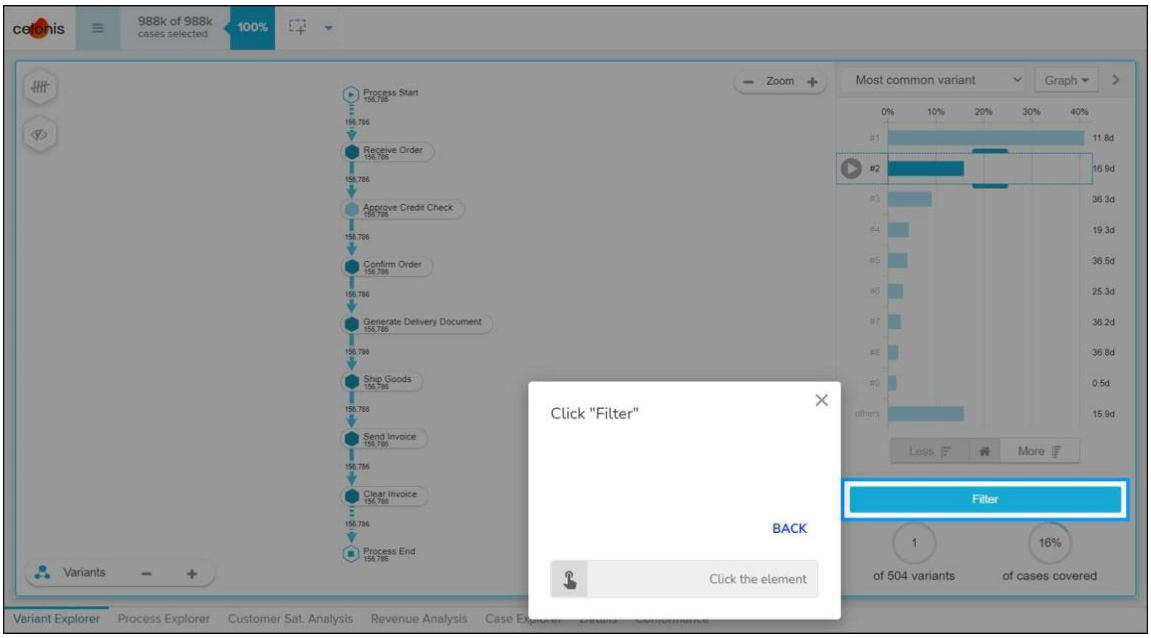

Fig 2.2: Variant Explorer

Here are a few concepts to familiarize yourself with:

Guided Learning Tours

During the courses, there will be opportunities for you to participate in guided tours within a public Analysis demo. These tours have been specifically designed for this training and do not require any login information. Think of it as having a mentor who will guide you through Celonis Analysis.

Hands-on Exercises

Towards the end of most courses, you will also be given exercises that involve answering questions and interacting with the Analysis demo firsthand. These exercises encourage reflection on what you have learned and provide an opportunity to apply your knowledge directly within the demo.

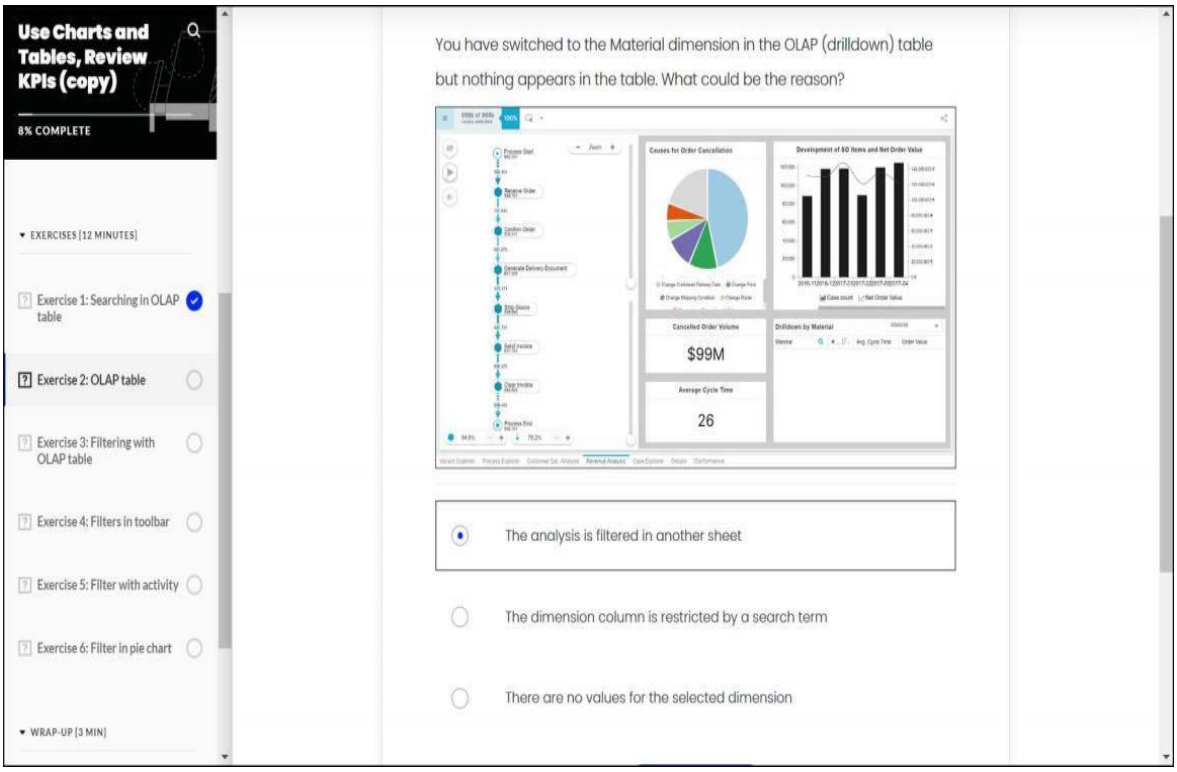

Fig 2.3: Exercise of Process Mining

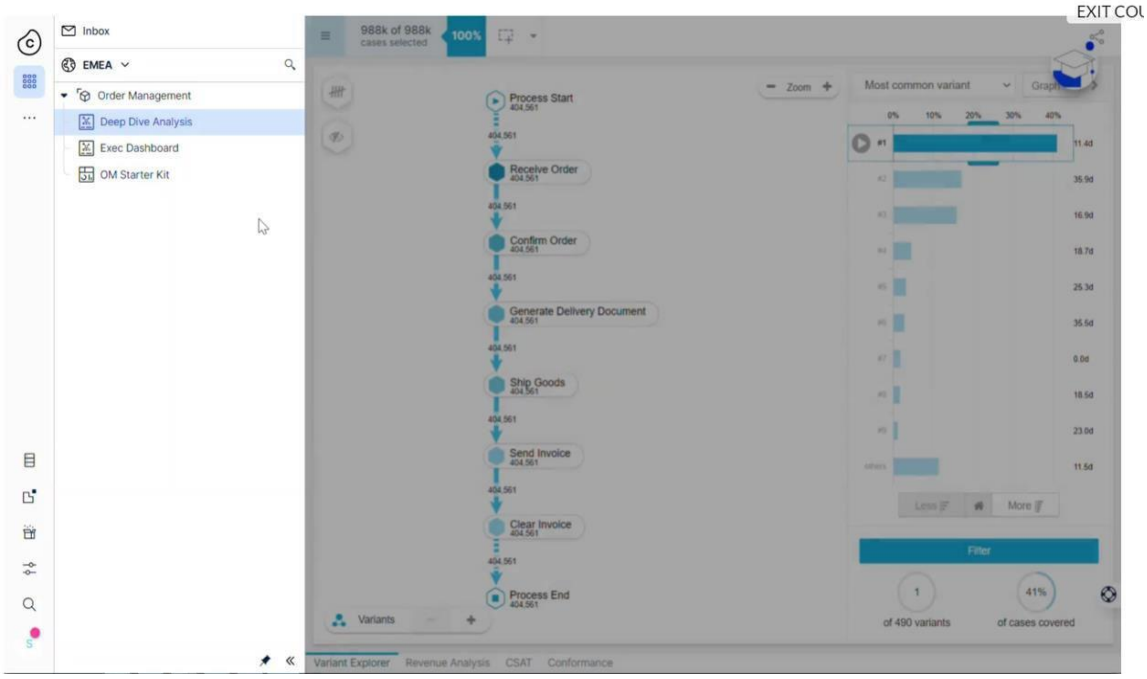

Fig 2.4: Example of Variant Explorer (DDA)

Process-Specific Examples

By the conclusion of the training program, you will have become familiar with the most commonly utilized components of Celonis Analysis. While Celonis Analysis tools operate in a uniform manner across processes, each organization's unique process has its own intricacies. As a result, it will be necessary for you to exercise critical thinking skills in order to establish connections between this training and your specific process. However, we will provide assistance along the way by encouraging you to reflect on your ideas and take notes that can be referenced during your work.

2.2 Navigating to an Analysis

Within the concept of navigating to an analysis, there are three key words: space, package, and analysis. These terms are organized hierarchically and serve as visual representations of data sets. By examining these visuals and their accompanying graphs, we can effectively analyze the data at hand.

The two screenshots provided depict demos from public analyses. Depending on which training track you are undertaking - *Reviewing and Interpreting Analyses* or *Monitoring KPIs* in Analysis Dashboards - you will interact with one of these two demos accordingly.

2.3 Use the Variant Explorer

The provided screenshots display two different public Analysis demos. Depending on the training track you are undertaking, whether it is Review and Interpret Analyses or Monitor KPIs in Analysis Dashboards, you will have the opportunity to interact with one of these demos.

Variant Explorer is an Analysis tool within Celonis EMS that enables users to explore how a particular process flows within their organization.

Analogously, if we envision a process as a road trip, each variant of that process represents a potential route. Each activity within the process can be likened to waypoints along that route. The connections between activities resemble roads connecting various stops on the journey. Furthermore, individual trips made along specific routes are equivalent to cases.

In summary, Variant Explorer provides users with an efficient means of assessing whether most cases in a given process adhere to an acceptable flow of activities or not. Additionally, this tool aids in formulating initial analysis questions for further investigation and evaluation purposes.

2.4 Use the Process Explorer

Discover the power of Process Explorer, a Celonis EMS Analysis tool designed to delve into the intricate connections within process activities. Instead of focusing on specific process variants, this tool highlights the most prevalent activities and their corresponding links.

To illustrate this concept further, let's revisit our road trip analogy. Rather than displaying various routes (variants) people took on a particular journey (case), Process Explorer reveals the waypoints (activities) and roads (connections) that are most frequently traversed along the way.

Process Explorer proves to be an invaluable tool when it comes to identifying uncommon activities that may otherwise go unnoticed with Variant Explore, as these unique activities might not show up in the more common variants.

2.6 Selection Views

Selection Views provide a wider range of options for filtering cases compared to the filtering capabilities offered by analysis sheets components.

Selection Views Button

By clicking on the Selection Views button in the analysis toolbar, you can access any of the six available Selection Views from anywhere within the analysis.

Subsequently, you have the opportunity to choose one of six selection types based on your specific search requirements.

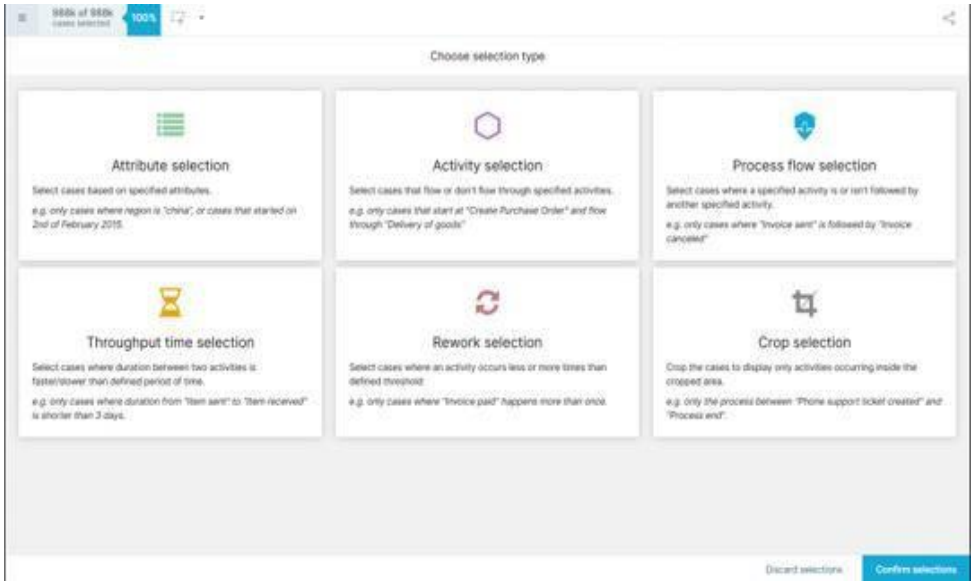

Fig. 2.5 Selection types

There are Six selection types they are:

- Choosing attributes
- Selecting activities
- Deciding on process flow
- Determining throughput time
- Selecting rework options
- Picking crops.

CHAPTER 3

ENHANCE YOUR ANALYSIS BUILDING SKILLS AND LEARN TO GET INTO CELONIS.

3.1 Writing PQL Queries:

The origins and objectives of PQL:
PQL, or Process Query Language, was developed with five key principles in mind:

1. Ease of use
2. Adaptability
3. Focused on event logs.
4. Business-oriented
5. Interactive frontend

1. Ease of Use (Simplicity)

- From business user to data engineer, it is simple to use.
- It converts intricate process inquiries into data queries.
- Process Mining is made accessible to every Celonis user.

2. Adaptability (flexibility)

- A collection of general functions and operators.
- Brings together a vast array of permutations.
- Constructs inquiries without constraining the method used.

3. Focused on Event logs

- It offers features for process mining.
- Link cases to an event.
- Specialized functions for processes are focused on the provided event log.

4. Business Oriented

- This solution combines the capabilities of process mining and business intelligence.
- It enhances event data by adding extra business information.
- Offers a wide range of SQL functions to choose from.

5. Interactive Fronted

- The presence of a graphical user interface (GUI) contributes to increased acceptance, usage, and adaptation by users.
- This leads to a higher level of acceptance, usage, and adoption by the users due to the support of a graphical user interface (GUI).

Celonis PQL Engine

The Celonis PQL Engine is a critical part of the Celonis Software Architecture, as depicted in the image provided. This language is utilized across all Celonis applications to retrieve data from a data model.

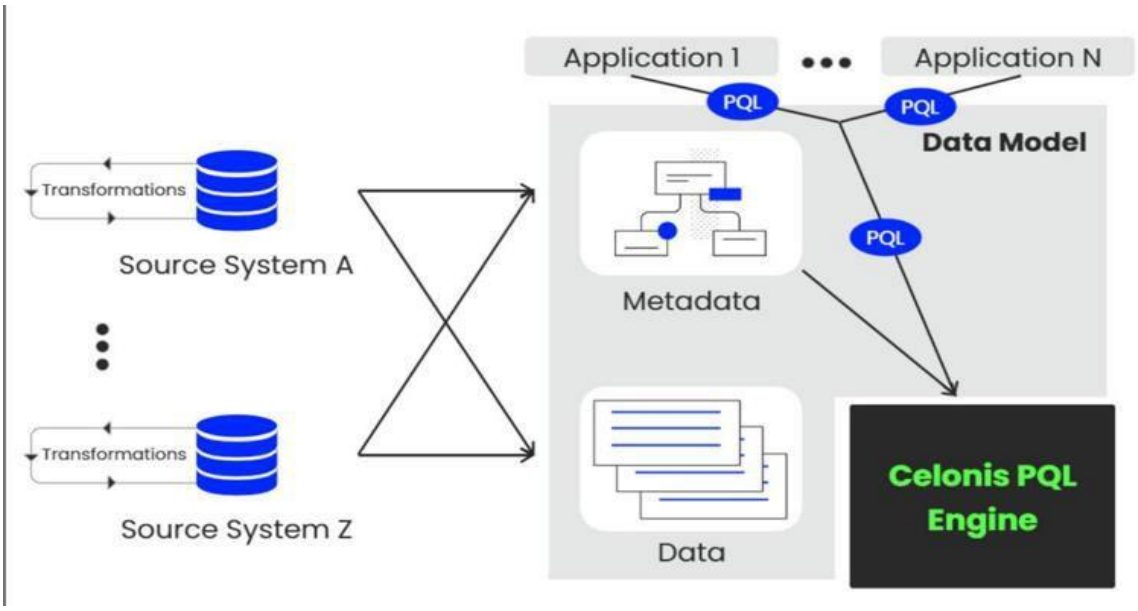

Fig 3.1: PQL Engine

SQL vs PQL

Both have distinct characteristics; PQL is not capable of executing data manipulation and data definition language tasks due to its design based on specific requirements. PQL only supports aggregate functions.

3.2 Importing Data into EMS

In this section, we will explore two types: setting up a data pipeline and refining your data pipeline. Under setting up a data pipeline, there are subcategories which include:

- Basics of Data Integration
- Connecting to Systems
- Extracting Data
- Transforming Data
- Loading a Data Model

Under refining your data pipeline, the categories are as follows:

- Schedule Data Jobs
- Monitoring and validating your Data Pipeline
- Handling Multiple Processes and Systems
- Enhancing Your EMS SQL Transformations
- Connecting Custom Processes
- Quality Assurance for Your Data Pipeline

Basics of Data Integration - What is it used for?

As a data engineer or analyst working in the field of data integration (previously referred to as event collection), you have the responsibility of bringing clean real-time process data into the EMS platform by building an efficient and effective data pipeline.

2. Establishing Connections to Systems

The initial step in pulling process data into the Celonis EMS is to establish connections with source systems. The EMS employs a wide range of technologies, including message queues, Restful APIs, Soap APIs, direct database access, and system-specific solutions for this purpose.

Data Integration involves setting up connections and configuring your data pipeline. There are several methods available for bringing data into the EMS:

- Process Connectors
- Extractors (Data Connections)
- Extractor Builder
- File Uploads
- Data Push API
- Celoxtractor

3. Extracting Data

Regardless of which system you are working with when extracting data, it is important to have a clear understanding of the business process in order to determine the specific tables required.

Why don't we simply extract entire databases and simplify our lives? There are several reasons why this approach is not feasible:

- It would be time-consuming.
- It could put strain on source systems.
- It would consume unnecessary cloud storage space.

Moreover, it would also be expensive!

4. Rearrange Data

The process table embodies your workflow and consistently includes at least these three columns that represent your process:

The unique identifier or case key,

The steps or activities carried out for different case keys, and the time stamps or event times for each activity.

In the Purchase-to-Pay procedure, the Purchase Order Item Number serves as our central reference point.

Each Purchase Order Item progresses through a series of activities such as request creation, item creation, goods reception, and invoice payment. Furthermore, each activity is associated with a corresponding event time.

In essence, every Purchase Order Item possesses an individual case key that undergoes diverse activities at different points in time. Together, these three columns form the foundation of your process flow.

5. Load a Data Model

Merely including the Activity table on its own within a Data Model does not suffice. To delve deeper into case details, we require both the Case table and other master data tables.

As you are aware in Celonis software package, the Case table comprises one row dedicated to each individual case analyzed by this application; hence this particular table presents information pertaining to every "process path" (i.e., pathway following a specific case).

The "EKPO" table contains data where each case corresponds to one line. Therefore, it is necessary to choose this table as the Case table.

By selecting a Case table, you can utilize pre-defined KPIs in Celonis analysis, such as a case count. The case count will now specifically pertain to the "EKPO" table and will always calculate the number of entries in this table based on the applied filters. The Studio screenshot provides a visual representation of what is being measured by this KPI.

Fig 3.2: KPIs

CHAPTER 4

APPLICATIONS

Celonis

Celonis, a prominent process mining platform, aids organizations in visualizing and analyzing their processes through the utilization of data from diverse systems. It provides valuable features such as process discovery, conformance checking, and performance analysis.

Features

Process Discovery is one of Celonis' key features as it automatically generates process models by carefully analyzing event logs. This allows users to gain a clear visualization of how processes are executed in reality.

Conformance Checking is another important capability offered by Celonis. It compares the actual execution of processes with expected models to identify any deviations or compliance issues that may arise.

Performance Analysis is yet another valuable feature provided by Celonis. It measures the efficiency of processes, identifies bottlenecks within those processes, and offers insights for optimization strategies.

Benefits

The benefits that Celonis brings include streamlined operations within organizations as well as a reduction in inefficiencies. Additionally, it enhances compliance efforts and empowers decision-making through data-driven insights.

Disco by Fluxicon

Another notable platform in this field is Disco by Fluxicon. Recognized for its user-friendly interface, Disco offers comprehensive process mining capabilities to its users.

Key Features:

Intuitive User Interface: Created with user-friendliness and accessibility in mind, making it suitable for a diverse range of users.

Visual Representation of Processes: Generates process maps and diagrams that effectively illustrate the actual flow of processes.

Identifying Root Causes: Aids in pinpointing the underlying reasons behind process deviations and bottlenecks

.

Advantages

Disco offers a straightforward yet robust set of tools for uncovering valuable insights from event data and driving improvements in processes.

ProM

ProM is a toolset for process mining that is widely used in both research and industry, available as an open-source solution.

Features:

Modular Architecture with Plugins: Offers an extensive collection of plugins to cater to various process mining tasks, allowing users to select specific techniques.

Customizability: Enables users to customize and experiment with different process mining algorithms.

Process Gold

Process Gold is a company that provides process mining solutions with a primary focus on comprehending and enhancing processes. Their offerings include advanced analytics, which combines process mining with sophisticated data analysis techniques to provide deeper insights and predictive capabilities. Additionally, ProcessGold automates the creation of visual process models from event data, making it easier for organizations to gain an understanding of their operations.

Minit

Minit, on the other hand, specializes in simplifying the process discovery and analysis for businesses. Their user-friendly interface caters specifically to non-technical users such as process analysts and business professionals. Minit also offers performance metrics that enable organizations to measure their process efficiency using key indicators.

Both Process Gold and Minit have similar goals of empowering organizations by making process mining more accessible. By utilizing these tools, businesses can identify areas for improvement within their processes and predict future outcomes. Ultimately, this leads to enhanced efficiency gains and overall optimization across various industries.

In summary, Process Gold focuses on offering advanced analytics alongside its automatic generation of visual models from event data while Minit prioritizes simplicity through its user-friendly interface catered towards non-technical users along with providing performance metrics for measuring process efficiency.

QPR Process Analyzer

QPR Process Analyzer is a tool that combines process mining and advanced analytics capabilities. It allows organizations to obtain comprehensive insights and predictions by integrating process data with advanced analytics. One of its key features is real-time monitoring, which enables the identification of deviations and opportunities for improvement as processes unfold. By using QPR ProcessAnalyzer, organizations can gain a holistic view of their processes, enabling them to make data-driven decisions and enhance operational performance.

When choosing among various process mining applications, it's important to consider factors such as your organization's specific goals, technical requirements, and the complexity of the processes you want to analyze and optimize. Each application offers its own unique set of features and benefits tailored to different needs. Therefore, it's crucial to assess these factors before making a decision on which application best suits your requirements.

CHAPTER 5

LEARNING OUTCOMES

Upon finishing this Training Track, you will have the capability to:

- Analyze process visualizations and utilize analyses to pinpoint inefficiencies in the process.
- Formulate your process by categorizing activities and cases.
- Store an analysis selection for future use and share it with your team; export visualizations and data related to the process.
- Execute fundamental tasks required for constructing Celonis analyses.
- Gain familiarity with Analysis Settings and Permissions.
- Publish analyses following version control best practices.
- Apply your understanding of the theoretical principles behind Process Mining in practical situations.

CONCLUSION

In summary, process mining is an effective and adaptable technology that provides valuable insights into how organizational processes function. By analyzing the event data generated during process execution, process mining reveals hidden patterns, identifies inefficiencies, and offers actionable recommendations for optimizing processes. This technology has the potential to drive improvements in various industries such as manufacturing, healthcare, finance, logistics, customer service, and more.

The visualization of process flows, detection of bottlenecks, and identification of deviations from the ideal path are some of the key capabilities that make process mining a powerful tool for organizations. These capabilities enable informed decision-making focused on enhancing efficiency while reducing costs and improving overall performance. The real-time applications of process mining are particularly noteworthy as they empower businesses to respond swiftly to changing circumstances by promptly addressing issues that arise and ensuring optimal operation of their processes.

As technology progresses, the techniques used in process mining are expected to become more advanced and seamlessly integrated with other data-driven methods. This integration will enhance their capacity to drive process excellence even further. However, in order to successfully implement process mining, it is crucial to have a thorough understanding of both the technology itself and the underlying business processes. Organizations that embrace this practice have the opportunity to gain a competitive advantage by utilizing data-driven insights to continuously improve their operations and achieve greater levels of efficiency and effectiveness.

INTERNSHIP CERTIFICATE

REFERENCES

- ✓ [1]. Ref website:https://academy.celonis.com/learning-paths/execution-management.
- ✓ [2] Execution management link : https://tinyurl.com/2pvuuzmd.
- ✓ [3] Login page : https://academy-login.celonis.com/s/login/.