An Internship Report

on

PROCESS MINING VIRTUAL INTERNSHIP

Submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering (DATA SCIENCE)

by

MOUNIKA P (224G1A3257)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)

SRINIVASA RAMANUJAN INSTITUTE OFTECHNOLOGY (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.

2024-2025

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 Virtual internship" is the bonafide work carried out by **MOUNIKA P** bearing Roll Number 224G1A3257 in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering (Data Science)** for 10 weeks from May – July 2024.

Internship Coordinator

Mr.P.Veera Prakash M. Tech. (Ph.D).
Associate Professor

Head of the Department

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

Date:

Place:Anantapuramu

EXTERNAL EXAMINER

PREFACE

I completed a Virtual Internship in Process mining that took place from April to June 2024.

Celonis institutions refer to educational and research bodies that either partner with Celonis or offer programs and certifications in process mining, a technology that Celonis leads. These institutions provide training, certification, or research opportunities in business process optimization using Celonis' process mining tools Celonis Process Mining is a technology that helps businesses analyze and improve their operational processes by providing real-time, data-driven insights. It automatically extracts data from IT systems (such as ERP, CRM, or other business management tools) and visualizes how processes are actually being executed. This allows organizations to identify inefficiencies, bottlenecks, and areas for improvement.

This internship was part of my third-year B.Tech program at **Srinivasa Ramanujan Institute of Technology** in Anantapur. This program follows the **AICTE** (All India Council for Technical Education) model curriculum, which is designed by leading academicians in India to produce graduates who are ready for the job market and have the skills that industries require. Through this internship, I was able to gain a wealth of practical skills and insights into the fast-growing field of business process optimization. Interns gain hands-on experience using Celonis software to extract and analyze data, visualize business process flows, and identify inefficiencies. This opportunity helps develop strong data analytics and business intelligence skills, allowing interns to work with large datasets, create insightful dashboards, and provide data-driven recommendations for process improvements. In addition, the internship exposes interns to cutting-edge technologies like the Celonis Execution Management System (EMS), which integrates process mining with machine learning and real-time process monitoring. The experience has helped me prepare for a future career in cybersecurity, and thanks to Eduskills, I was also exposed to several job opportunities through their placement drives, which could lead to employment in different companies.

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, Associate Professor, Department of Computer Science & Engineering** , who has supported me a lot and encouraged me in every step of the internship work. I thank him/her 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 Department of Computer Science & Engineering (AI & ML and Data Science),** for his kind support and for providing necessary facilities to carry out the internship.

I wish to convey my special thanks to **Dr. G. Bala Krishna, 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.

MOUNIKA P

224G1A3257

INDEX

Contents

LIST OF FIGURES

Chapter 1

Introduction

Our world and the organizations in it are full of processes. From purchasing to order management, organizations deal with complex, global and sometimes faulty processes on a daily basis. Frictionless processes, on the other hand, ensure:

- that you can find the right groceries at the grocery store,
- that planes land on time,
- that patient waiting times at hospitals are kept to a minimum.

By contrast, 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 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 become possible through the advent of information systems and administrative tools (e.g. Enterprise Resource Planning or Customer Relationship Management systems) which provide a 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.

1.1 The Process Movement

The processes is everywhere in daily life and that they're crucial for frictionless operations. Business Process Management has traditionally examined these processes by talking to the people involved. The problem is that, by gathering information from people, you also gather their (false) assumptions and subjective or fractured observations. Traditional approaches fail to understand the real-life complexity of processes and also struggle to provide complete insights and visibility given the vast amounts of data that are now available.

1.2 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

1.3 Process Mining Enters the Business World

Since then, these ideas and the academic concepts behind Process Mining have bridged the gap to enter the business world. A variety of software vendors have ventured into the market and even expanded its capability from analytics to business execution. They achieve this through stronger operational links to automation frameworks and IT source systems which allow daily users to receive prompts and take direct action to improve processes. One very recent example of this is the creation of the new Execution Management Software category by software vendor Celonis.

1.4 IEEE Task Force

One key milestone for the Process Mining technology category was the formation of the Institute for Electrical and Electronic Engineers (IEEE) Task Force on Process Mining. The IEEE Task Force brings together both vendors and researchers interested in the field and actively works to define and drive the field further. One of their most important achievements is the publishing of their Process Mining Manifesto.

Chapter 2

Process Mining

Process mining applies data science to discover, validate and improve workflows. By combining data mining and process analytics, organizations can mine log data from their information systems to understand the performance of their processes, revealing bottlenecks and other areas of improvement. Process mining leverages a data-driven approach to process optimization, allowing managers to remain objective in their decision-making around resource allocation for existing processes.

Information systems, such as Enterprise Resource Planning (ERP) or Customer Relationship Management (CRM) tools, provide an audit trail of processes with their respective log data. Process mining utilizes this data from IT systems to create a process model, or process graph. From here, the end-to-end process is examined, and the details of it and any variations are outlined. Specialized algorithms can also provide insight into the root causes of deviations from the norm. These algorithms and visualizations enable management to see if their processes are functioning as intended, and if they aren't, they arm them with the information to justify and allocate the necessary resources to optimize them. They can also uncover opportunities to incorporate robotic process automation into processes, expediting any automation initiatives for a company.

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. Organizational perspectives can surface the various resources within a process, such as individual job roles or departments, and the time perspective can demonstrate bottlenecks by measuring the processing time of different events within a process.

In 2011, the Institute of Electrical and Electronics Engineers (IEEE) published the Process Mining Manifesto (link resides outside IBM) in an effort to advance the adoption of process mining to redesign business operations. While proponents of process mining, like the IEEE, promote its adoption, Gartner notes that market factors will also play a role in its acceleration. Digital transformation efforts will prompt more investigation around processes, subsequently increasing the adoption rate of new technologies, such as artificial intelligence, task automation, and hyper automation. The pace of these organizational changes will also require businesses to apply operational resilience to adapt as well. As a result, enterprises will increasingly lean on process mining tools to achieve their business outcomes.

2.1 Types of process mining

Wil van der Aalst, a Dutch computer scientist and professor, is credited with much of the academic research around process mining. Both his research and the above-mentioned manifesto describe three types of process mining, which are discovery, conformance, and enhancement.

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.

Conformance: Conformance checking confirms if the intended process model is reflected in practice. This type of process mining compares a process description to an existing process model based on its event log data, identifying any deviations from the intended model.

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.

2.2 Why is process mining important?

Increasing sales isn’t the only way to generate revenue. Six sigma and lean methodologies also demonstrate how the reduction of operational costs can also increase your return-on- investment (ROI). Process mining helps businesses reduce these costs by quantifying the inefficiencies in their operational models, allowing leaders to make objective decisions about resource allocation. The discovery of these bottlenecks can not only reduce costs and expedite process improvement, but it can also drive more innovation, quality, and better customer retention. However, since process mining is still a relatively new discipline, it still has some hurdles to overcome. Some of those challenges include:

Data Quality: Finding, merging and cleaning data is usually required to enable process mining. Data might be distributed over various data sources. It can also be incomplete or contain different labels or levels of granularity. Accounting for these differences will be important to the information that a process model yields.

Concept drift: Sometimes processes change as they are being analyzed, resulting in concept drift.

2.3 Process mining use cases

Process mining techniques have been used to improve process flows across a wide variety of industries. Since process maps highlight the key performance indicators (KPIs) which impact performance, they have spurred businesses to reexamine their operational inefficiencies. Some use cases include:

Education: Process mining can help identify effective course curriculums by monitoring and evaluating student performance and behaviors, such as how much time a student spends viewing class materials.

Finance: Financial institutions have used process mining software to improve inter-organizational processes, audit accounts, increase income, and broaden its customer base.

Public works: Process mining has been used to streamline the invoice process for public works projects, which involve various stakeholders, such as construction companies, cleaning businesses, and environmental bureaus.

Software Development: Since engineering processes are typically disorganized, process mining can help to identify a clearly documented process. It can also help IT administrators monitor the process, allowing them to verify that the system is running as expected.

Healthcare: Process mining provides recommendations for reducing the treatment processing time of patients.

E-commerce: It can provide insight into buyer behaviors and provide accurate recommendations to increase sales.

Manufacturing: Process mining can help to assign the appropriate resources depending on case—i.e. product—attributes, allowing managers to transform their business operations.

They can gain insight into production times and reallocate resources, such as storage space, machines, or workers, accordingly.

2.4 History

The term "Process mining" was first coined in a research proposal written by the Dutch computer scientist Wil van der Aalst. Thus began a new field of research that emerged under the umbrella of techniques related to data science and process science at the Eindhoven University in 1999.

Process mining is a relatively recent field that emerged at the intersection of data mining, business process management, and information systems. It involves the analysis of event logs from various systems to gain insights into business processes. The history of process mining can be summarized as follows

1.Emergence of Process Mining (Late 1990s - Early 2000s): The roots of process mining can be traced back to the late 1990s when researchers like Wil van der Aalst began exploring ways to analyze event logs generated by information systems. These logs contained valuable data about how business processes were executed, but extracting meaningful insights from them was a challenge.

2.Early Research and Algorithm Development (Early 2000s - Mid 2010s): In the early 2000s, researchers started developing algorithms and methodologies to analyze event logs and extract process-related information. The "workflow nets" proposed by van der Aalst were among the early formalisms used to model and analyze processes. The Alpha algorithm (2004) and Heuristics Miner (2007) were significant developments during this time, enabling automated process discovery from event logs. As organizations recognized the value of data-driven insights into their processes, process mining started to be adopted in various industries. Commercial process mining tools emerged, offering user-friendly interfaces and advanced analysis capabilities.

3.Industry Adoption and Commercial Tools (Mid 2000s - Mid 2010s): As the potential value of process mining became evident, industries such as finance, healthcare, manufacturing, and telecommunications started adopting process mining techniques.

Organizations recognized that analyzing event data could lead to better process optimization and compliance. Commercial process mining tools like ProM (open-source) and Celonis (commercial) emerged to support these efforts.

4.Expansion of Process Mining Capabilities (Late 2010s - Early 2020s): Process mining continued to evolve, incorporating more advanced capabilities. Conformance checking, which compares actual process execution to a model to identify deviations, gained prominence. Predictive analytics started being applied to forecast process behavior, while prescriptive analytics provided recommendations for process improvement. The foundations of process mining were laid in the early 2000s, with researchers like Wil van der Aalst and others exploring ways to extract process-related information from event logs generated by information systems.

5.Integration with Advanced Technologies (Late 2010s - Present): With the rise of artificial intelligence and machine learning, process mining began to integrate these technologies. This enabled more sophisticated analysis, anomaly detection, and even automated decision-making within processes. The combination of process mining and robotic process automation (RPA) allowed for end-to-end process optimization. The integration of process mining with other technologies like machine learning and artificial intelligence became more common, enabling more sophisticated analysis and automation. rocess mining continued to evolve, incorporating more advanced techniques such as conformance checking, predictive analytics, and prescriptive analytics. The integration of process mining with other technologies like machine learning and artificial intelligence became more common, enabling more sophisticated analysis and automation.

6.Real-time Process Monitoring (Present): Recent advancements have focused on real- time process monitoring. This involves analyzing event data as it's generated, allowing organizations to respond to process deviations and anomalies in near real-time.

7. Future Directions: The future of process mining may involve deeper integration with emerging technologies like blockchain, IoT (Internet of Things), and more advanced AI algorithms. This could provide even greater insights into complex, interconnected processes.

Chapter 3

Fundamentals of Process Mining

3.1 Celonis Analysis

Most businesses face numerous improvement opportunities, also called value opportunities inefficiencies in their processes that prevent them from realizing their full potential. Of course, they can address these value opportunities by improving how the process runs, but it turns out they're usually aware of some, not all of the value opportunities. In addition, they may have incorrect assumptions about the cause of certain inefficiencies. 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.

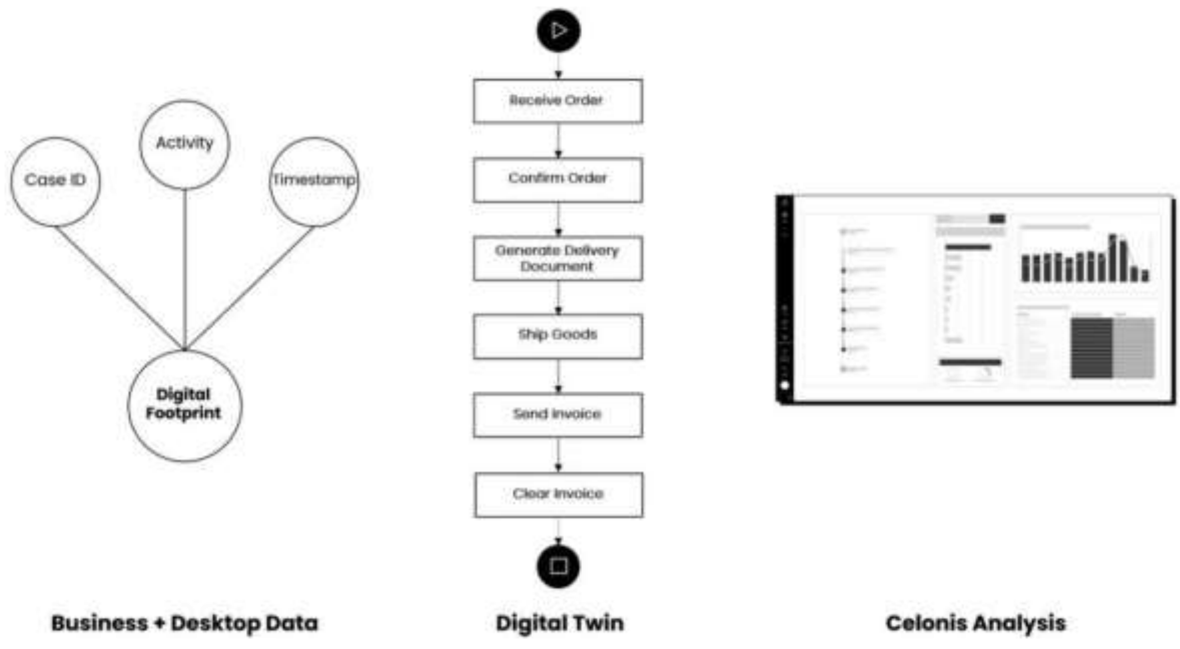

Fig 3.1 Celonis Analysis

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.

3.1.1 Elementary Apporach

An exploratory approach is one where you simply explore the data and see what value opprtunities 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.

3.1.2 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 drill down table, 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 don't stop at Process Mining and leverage all that the Celonis Execution Management System (EMS) has to offer.

3.1.3 Process, Activity, and Case

A process is a series of linked steps taken in order to achieve a particular goal.

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.

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.

3.1.4 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.

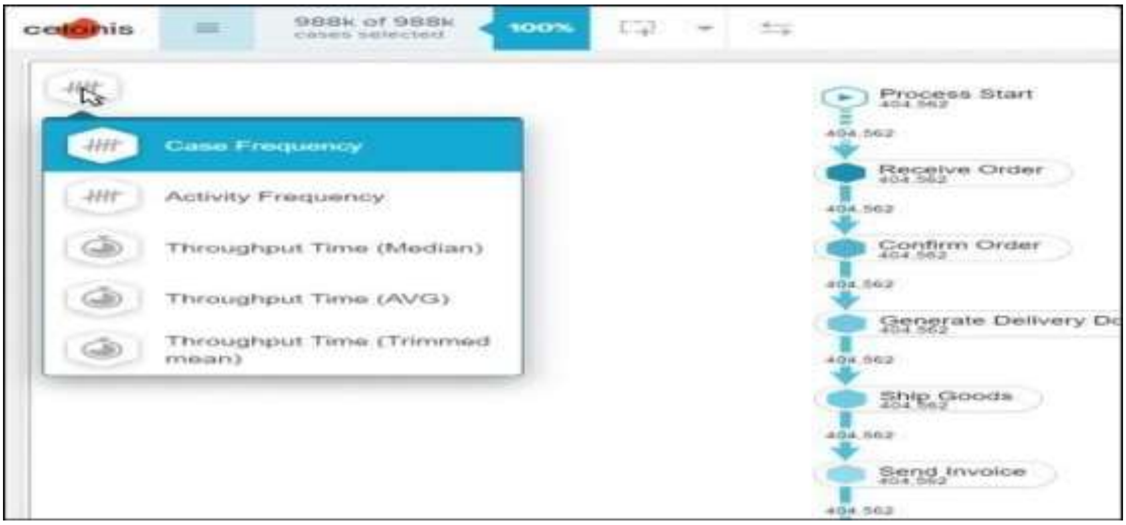

Fig 3.2 Case Frequency

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. The 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.

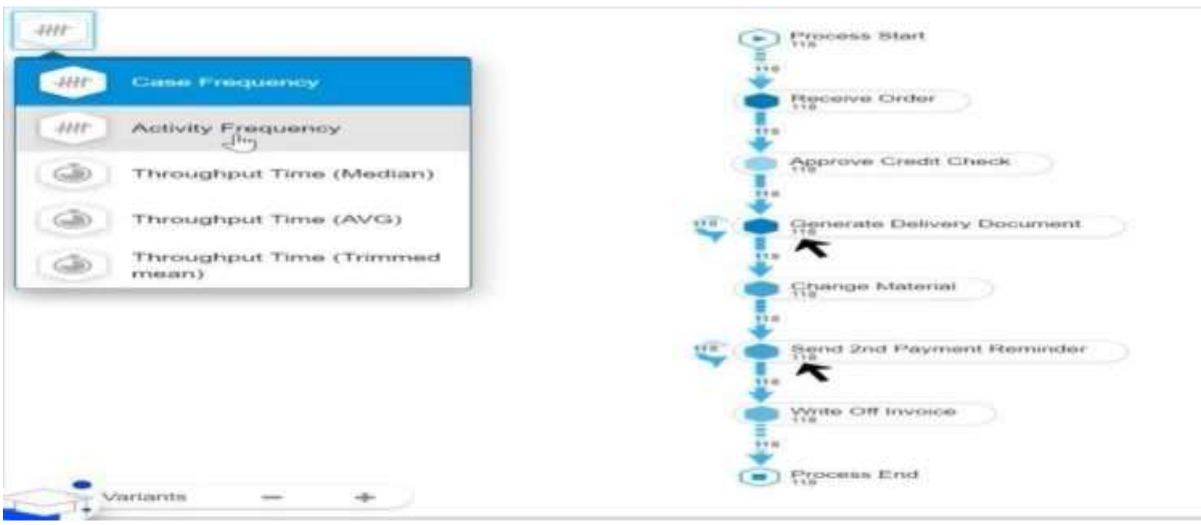

Fig 3.3 Activity Frequency

Case Frequency shows 118 cases associated with this variant. Therefore, the count is the same across this entire single variant. 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. "The Throughput Time KPI is especially powerful to review when undesired activities take place, such as: the Throughput Time KPI is especially powerful to review when undesired activities take place, such as:

- **Order Management**
  a. change confirmed delivery date
  b. change shipping conditions
  c. change route
  d. set delivery block
  e. write off invoice
  f. return goods
  g. cancel order
- **Account Receivable**
  a.change price
- **Procurement**
  a.change price
- **Account Payable**
  a.change baseline date

3.1.5 Selection Views

Selection Views offer a more comprehensive set of options to filter on cases as compared to what you can do in the components in analysis sheets. It's worth noting that the person creating the Analysis determines whether or not to enable Selection Views for users. Selection Views offer a more comprehensive set of options to filter on cases as compared to filtering you can do using the components in analysis sheets. You can access the six Selection Views from anywhere in the analysis by clicking on the Selection Views button located in the analysis toolbar.

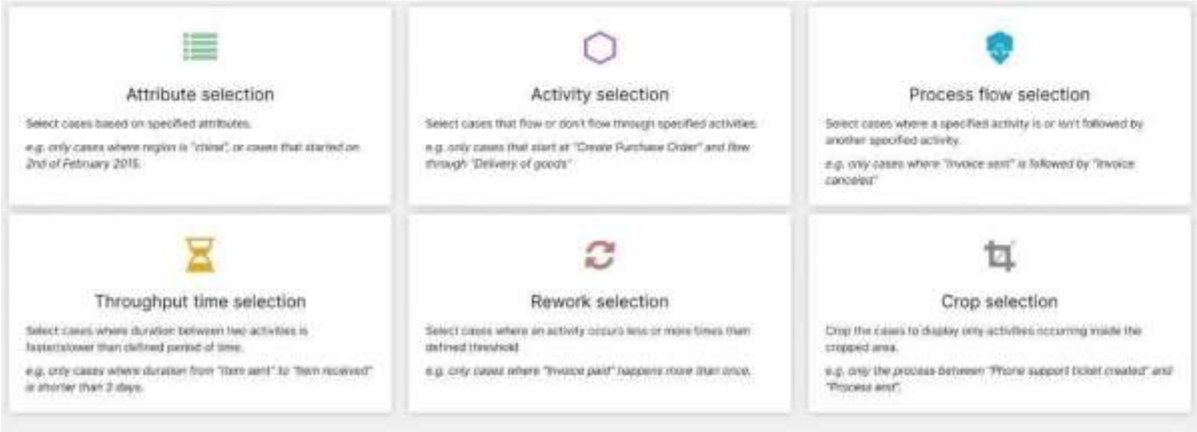

Fig 3.4 Selection Views

Attribute selection: Attribute selection since it gives you access to all the dimensions and their attributes in the data model.

Activity Selection: Activity Selection to select cases that flow through or don’t flow through one or more specific activities and even further specify whether they start or end with a particular activity. Activity selection empowers you to quickly filter on cases without having to look for the activities in the Variant or Process Explorer. And a certain combination of activity selections can only be done in Activity selection.

Process Flow Selection: Process Flow Selection is useful when you need to filter on cases where a specific activity is or isn’t followed by another specific activity.

Throughput time Selection: Throughput time refers to the total amount of time that it takes to run a particular process in its entirety from start to finish. For example, a manufacturer can measure how long it takes to produce a product, from initial customer order to sourcing raw materials to manufacturing to sale.

Rework Selection: Rework Selection want to look at cases where an activity occurred more or fewer than a specific number of times or between two specific thresholds; for example, more than 2 times, under 4 times, or between 2 and 5 times.In the chart below, you see the distribution of cases in two columns with regard to a specific activity ("Change Price").

Crop Selection: Crop Selection when you want to focus your analysis on all activities and connections between two specific activities. This is especially useful if you are working in a department or team that is only responsible for a certain part of the process.

Fig 3.5 Crop Selection

As shown by the number zero (0) under the column, most cases in the analysis did not go through the specified activity. We're looking for rework, specifically for cases where the "Change Price" activity occurred more than once. So the column on the right does include those cases that include rework but it also includes cases that have gone through the activity just once (note 1 - 3 under the column).

3.2 Build Analyses

As an analysis builder, your work begins once the Data Engineer has brought in the data. It's common for analysis builders to become very familiar with process data tables over time such that when storyboarding the analysis, they already know whether the tables are missing needed data. But it's also possible that you only realize more data is needed in the process of creating analysis components. In either case, you'll want to collaborate with the Data Engineer to bring in additional data or define specific KPIs. Once you've done the preparation work, including gathering user requirements and reflecting on the best visualizations, you can create the analysis asset and start to build the first draft of the analysis.

3.2.1 Package and Asset Keys

The key of a Package or Asset allows you to uniquely identify it throughout your Celonis team. You can have multiple Packages with the same name but their keys need to be different. The same goes for Asset names and keys. As an example, if you're creating a Package with a name that already exists, you'll get a prompt to modify the key (but not the name).Also, it's not possible to change the key once you've created the Package or Asset. The key won't change even if you change the name of the Package or Asset.

We can't talk about working in Studio without first understanding these elements: Space, Package, and Asset. To get started, we'll create a Package called "Build Analyses (Training)" and an Analysis Asset called "P2P_Analysis." Although not visible in the image, we'll link the analysis to a "Purchase-to-Pay" data model we've prepared for training purposes. Access Studio in your training environment and do the following.

Create a Space: Name it whatever you'd like

Create a Package: Name it "Build Analysis (Training)"

Click "Choose data model" and from the list select "Purchase_to_PayTaining_EN"

Create an Analysis Asset: Name it "P2P_Analysis"

You can have multiple analyses in a Package and link each analysis to a different Data Model or Knowledge Model. Studio gives you the flexibility to organize analyses and other assets according to your needs. Besides Packages, you have another layer available to you for organizing your assets: "Spaces." You can create and name multiple Spaces to house packages in. Here, we have renamed the "Default" Space to "Automation Rate" and added other Spaces.

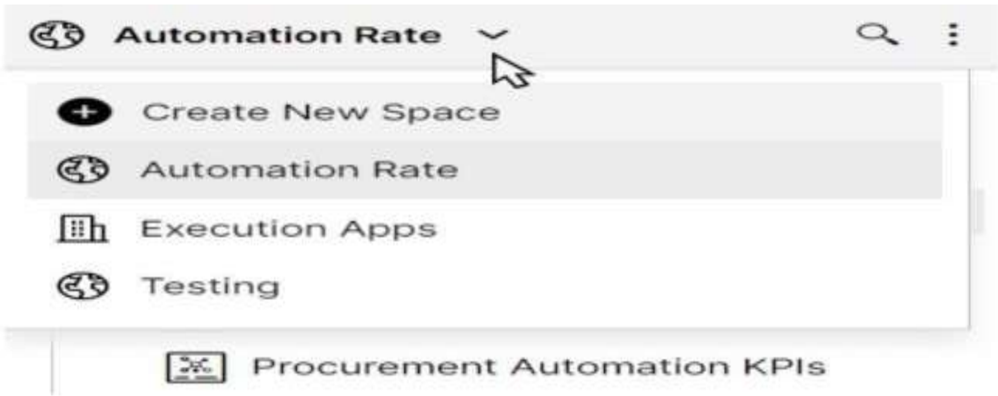

Fig 3.6 Analysis Asset

3.2.2 Configure dimensions and KPIs

Anytime you add a table or chart to the analysis, you'll need to select the dimension(s) and KPI(s) to display. You might remember these from the training track, Review and Interpret Analyses. Dimensions are used to determine which attributes from the analysis are displayed; for example, vendor name, customer name, date, region, or document type. Dimensions represent the columns of the data tables (for example, Dimension = Customer name, Attribute = Acme Bread Company). KPIs, or Key Performance Indicators, are used to calculate and add aggregated values; for example, case count, order value, invoice value, throughput time, or automation rate. KPIs are functions that consolidate a set of values belonging to a single occurrence inside a dimension into one single value. To add dimensions and KPIs to an analysis component, you'll need to work with the data tables in the analysis. In the SAP Purchase-toPay (P2P) data model we work with for this training, we have four tables regardless of whether we're selecting dimensions or KPIs. It's important that you have a conceptual understanding of the following four data tables.

- Purchasing Document Header (EKKO)
- Purchasing Document Item (EKPO)
- Vendor Master (LFA1)
- Activity Table—also known as the Event Log

Configure an OLAP Table: We'll add a blank sheet and add an OLAP table to it, then we'll configure a dimension (vendor name) and a KPI (case count).

Dynamic Drilldown Table: The "Use Tables and Charts course," that it's possible to have a dynamic drill-down table, where the user changes the dimension using a dropdown menu. You can find out how to configure this component later in this training track, specifically the course, "Create Dynamic Analyses(opens in a new tab)."

Fig 3.7 Drilldown Sales

configure a column chart with the creation date of purchase orders as a dimension and case count as the KPI. You'll get to know the date functions available.

Configure a Pie Chart: The principle of creating table and chart components in Celonis is always the same. Although, you'll need to select only one KPI to display in a Pie Chart. As you can see in the image, each piece of the pie (each row) represents one attribute from the dimension, Material Group. In this case, it would be really difficult to select beyond the first handful of rows. We need to limit the pieces of pie (rows) displayed. Fortunately, we can limit the rows to something manageable. We updated the field, "Maximum elements shown" from 100 to 5. This way, the chart displays only 5 rows (pieces). You would switch the dropdown from "Limit" to "First k rows." The difference with the first approach is that besides the top 5 rows, all other rows are summarized under a sixth piece called "others." When the user selects the "Others" row, the chart updates to display the next 5 items along with a new "Others" row and so on.

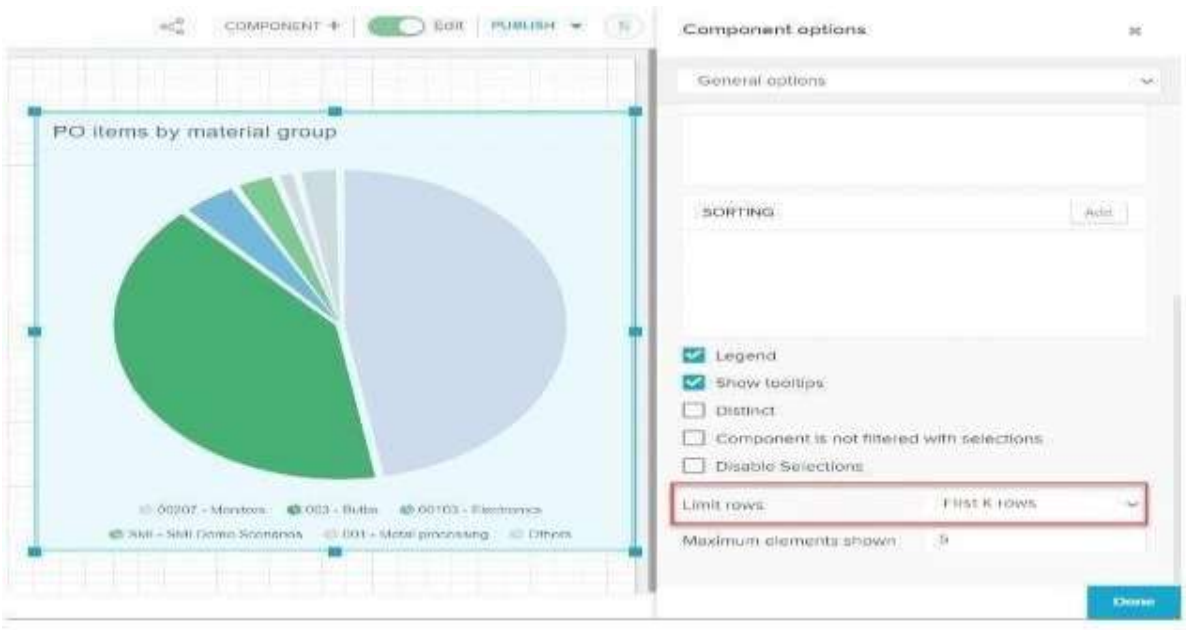

Fig 3.8 Pie Chart

Configure Single KPI Components: Celonis Analysis includes four types of single KPI components. The most common use cases for the single KPI component include the case count and net value. For both, you would use the Number KPI. Aside from the Number, you might choose other Single KPI components such as Gauge, Fill, and Radial, depending on what you need to display. The Radial, for example, would be appropriate for displaying percentages. After adding the Number component onto the sheet, we need to define the

formula for the KPI. We do so by clicking the formula button or f(x). By default, when the user selects cases (filters on cases) in the analysis, all the components in the analysis get filtered too, including Single KPI components. To exclude a component from the filter, you can activate the "Component is not filtered with selection" checkbox. This way, the component will be static. In that case, note the fact that the KPI is static in the title. In the course “Configure Tables and Charts,” you became familiar with the three simplified data tables we use in this training when configuring components that require dimensions and KPIs. These tables are:

- The Purchasing Document Header Table (EKKO).
- The Purchasing Document Item Table (EKPO).
- The General Vendor Master Table (LFA1).

To create advanced KPIs and formulas such as automation rate, you'll use another table: the Activity Table (also called the Event Log). The Data Engineer can add more columns to the

Activity Table with activity-specific information, for example, add a “User Type” column which indicates whether an activity has been conducted by an automated or a manual user. In an SAP system, automated users are marked with a “B” while manual users are marked with an “A”. You can use the “User Type” to calculate automation rates. Depending on the nature of information you need to display and visualize in the analysis, you might need to work with the Data Engineer to bring in additional columns to the Activity Table.

Visual Editor: Use Standard Process KPIs and the Visual Editor," you used the Visual Editor to customize some of the Standard Process KPI you selected; you started with a KPI then customized it in the Visual Editor. Here, after selecting Add KPI, we'll start in the Visual Editor, as such, you'll see 8 different categories are available. Once you select a category, the Visual Editor displays a formula creation screen with a pre-defined template, according to your category selection.

Ratio vs. Process Flow Selection Formula Builder: Looking at the formula builder for Process flow selection, we see the configuration looks similar to the one we already got to know when configuring the Ratio formula for maverick buying. But notice the "Process flow selection" formula counts the absolute number of items ("Count of items where") versus the ratio ("Ratio of cases where").Keep in mind that both metrics are very important to analyze process deviations. However, if your company only ordered 5 items from that vendor, then the absolute number would be 4 and would probably not be worth further investigation.

Chapter 4

Raising 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. Those Event Logs are a collection of process events that can be described by the following attributes:

Case Attribute: The case attribute indicates which process instance the event belongs to. A process instance is called a case, usually consisting of multiple events.Let's consider an example: Imagine you are running a restaurant with food delivery. Each order has a specific number, the order number. This number is the unique case ID, and all related activities are assigned to this ID.

Activity Attribute: The activity attribute describes the action that is captured by the event.In our food delivery example, these are all the steps an order has passed through, from receiving the order, to cooking the meal, delivery and payment.

Each activity leaves a digital footprint with a timestamp, indicating precisely when each event took place. With the help of timestamps, we know precisely in which chronological order the different activities have runoff. 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.

4.1 The Celonis PQL Engine

Celonis PQL is an integral component of the Celonis Software Architecture. All Celonis applications use this language to query data from a data model. Even though Celonis PQL is inspired by SQL, there are major differences between the two query languages. On a high level, Celonis PQL varies along four key dimensions:

Language Scope: Celonis PQL does not support all operators that are available in SQL.This is because customer requirements drive the development of the language, and only operators needed for the target use cases are implemented.

Data Manipulation Language: Celonis PQL is not supported by a data manipulation language (DML).As all updates in the Process Mining scenario should come from the source systems, there is no need to manipulate and update the data through the query language directly. **Data Definition Language :** Celonis PQL does not provide any data definition language (DDL).As the data model is created by a visual data model editor and stored internally, there has not been any need for creating and modifying database objects.

Integrate: In contrast to SQL, Celonis PQL is domain-specific and offers a wide range of Process Mining operators not available in SQL. Consequently, Celonis PQL seamlessly integrates the data with the process perspective.

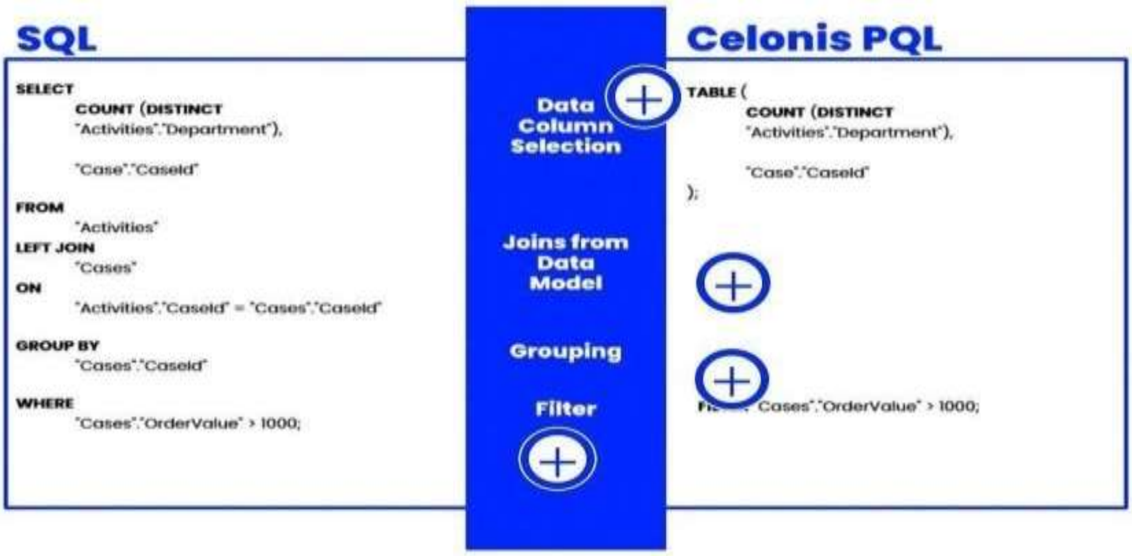

Fig 4.1 SQL vs PQL

4.2 Process Querying Framework

The Process Querying Framework is an abstract system consisting of a set of generic components to define a process querying method. Celonis PQL covers many of these components. This section describes the integration of Celonis PQL into the PQF. Celonis PQL instantiates the main components of the PQF. The first part of the framework (Model, Record, and Correlate) retrieves or creates the behavioral models and formalizes the business questions into process queries. The event logs are recorded by information systems like ERP or CRM systems, and extracted from these source systems into the Celonis IBC plat form.The process models are either manually modeled (e.g., in Celonis IBC or in an external tool) or discovered by process mining techniques, like the Inductive Miner. The correlation models are created by the conformance checking operator , relating activities of the event log to tasks in the process model. However, the different kinds of model repositories overlap due to their related storage, as relational data within the same data model. For example, the result column of the conformance checking operator (correlation model) is added to the activity table(event log).The query intent of Celonis PQL is limited to create and read. While all supported kinds of behavioral models can be read, process models and correlation models can also be created by Celonis PQL queries (e.g., by process discovery and conformance checking). The update and delete query intents are not included especially for the event logs as they should always stem from the source systems.

Therefore, event log updates can be achieved by delta loads that regularly extract the latest data from the source systems. The process querying instruction is usually defined by an analyst through a user interface. For example, the user defines the columns to be shown in a table, which can be considered as the query conditions. The selections from the user interface are then formalized into a Celonis PQL query. The Prepare part of the framework focuses on increasing the efficiency of the query processing. The Celonis PQL Engine—that processes the queries—maintains a cache for query results, refer to Sect. 6. After the application starts, it warms up the cache with the most relevant queries derived from the Process Querying Statistics to provide fast response times. The Indexing component does not only include classical index structures but also all kind of data structures for an efficient retrieval of data records. It is covered by the dictionary encoding of columns. The Execute part of the framework combines an event log with an optional process model and a Celonis PQL query

into a query result, which can be either a process model, KPIs, filtered and processed event log data, or conformance information. The concrete input and output of the query depend on the selected query intent and the query conditions. The Filtering component reduces the input data of the query. This can either be achieved by the REMAP_VALUES operator and the filter column of the SOURCE and TARGET operators.

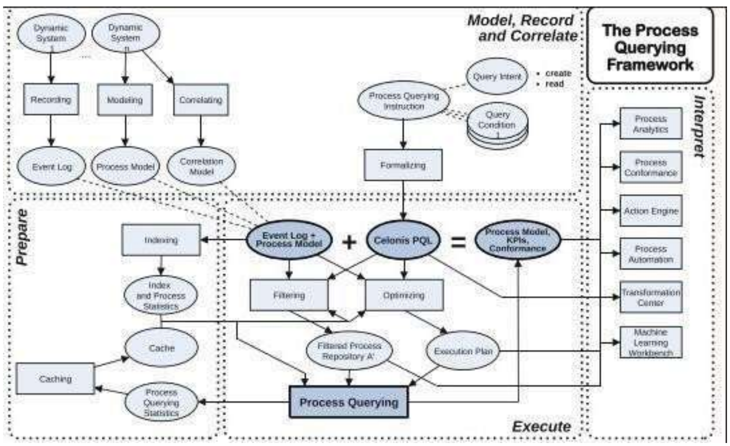

Fig 4.2 Framework

component uses basic database technology to rewrite the query and create the Execution Plan, which describes a directed graph of operator nodes. The Process Querying component then executes the execution plan on the filtered data. It also retrieves data from the cache to avoid re-computation of either the full query or certain parts of it, which are shared with previous queries. The Interpret part of the framework communicates the query results to the user and improves the user's comprehension of them. The applications in the Celonis IBC platform incorporate Celonis PQL and make the results accessible to the user. The Process Analytics presents the query results as process graphs, charts, and tables. Beyond pure visualization, it is highly interactive with dynamic filtering to drilldown the processes to specific cases of interest. This interactivity offered by all GUI components is achieved through the dynamic creation of Celonis PQL queries. Process Conformance shows the deviations between process model and event log in a comprehensive view, including a comparison of KPIs between conforming and non-conforming cases. In contrast to this, the focus of the Action Engine is not to present query results, but to trigger user actions, for instance, by informing about deliveries that are expected to be late. The Action Engine canal so trigger automated workflows that are executed by the Process Automation component. Within this component, the workflows can query data from the event log using Celonis

PQL. Transformation Center supports process monitoring. It historicizes the query results to show how the processes evolved over time. Finally, Machine Learning Workbench provides a platform for user-defined machine learning analyses over event logs and retrieves the event data using Celonis PQL queries.

4.3 Language Overview

Across all classes of operators, Celonis PQL follows four language features:
First, operators usually create and return a single column that is either added to an existing table (e.g., the case or activity table) or to a new, temporary result table. But note, there are also a few operators that create and return one or more tables with multiple columns (e.g., for computing a process graph). Second, the supported data types comprise STRING, INT, FLOAT, and DATE. Third, Boolean values are not directly supported, but can be represented as integers. Fourth, each data type can hold NULL values. Celonis PQL operates as follows- In aggregations: treats NULL values as non-existing and ignores them. In row-wise operations: returns NULL if one of its inputs is NULL.

PQL is the query language to formulate your process questions and calculate KPIs. This is why you can apply it in a multitude of applications in the EMS. t offers not only the possibility to use a visual editor and switch between visual editor and Code editor, but also visualizations to validate your queries and to make sure that this is exactly the query you need. he central place for all your records, KPIs, filters, etc. is the Knowledge Model. This is where you can add new PQL queries and save them for future usage across the different assets. Whenever you configure Views, you use the PQL queries you defined in your Knowledge Model. Although your analysis should ideally be connected to your Knowledge Model, you can also write individual PQL queries inside analyses. But note that those can't be reused in the other analyses or views afterwards. When building Action Flows, sometimes you have to filter on specific subsets in your data or build a logic based on values in your data model. When working with ML Workbench and the by celonis package, you can interact with Celonis objects as native objects, e.g. copy an analysis, pull & push data, reload data models, etc. Welcome to another group of operators within the Celonis PQL library: Data Flow. Data Flow operators cover a variety of functions to return or change particular values based on conditions defined by the user.

4.4 Predicate Functions and Operators

Predicate functions and operators check if the input argument satisfies a condition.

Between Operator: PQL is an operator based language where an operator can be seen as a logical building block of the query. Thus, PQL is the connection of several operators. Since Odysseus differentiates between logical operators and their physical operators, which are the implementing counterpart, PQL is based upon logical operators.

Comparison Operator: Comparison Operators are important and just as easy as the BETWEEN operator when working in PQL, etc…

- Use both a left and right side argument to satisfy the requirements of binary operators. A DATE column can be compared to another DATE column or a DATE constant.
- A FLOAT column can be compared to a column or a constant of either type FLOAT or INT. The same applies to INT.
- A STRING column can be compared to another STRING column or a STRING constant based on lexicographic ordering.

4.5 Joining & Aggregating Data

If you have worked with PQL before, you might have encountered the following: You used the same formula in different components, but got different results. The activity table contains the data about the case identification, activity and timestamp for every process you want to analyze in Celonis. In addition, you always have a case table with one row for each of your cases and an arbitrary number of tables with additional information about for example your vendors, material, customers, routes, etc. This data is gathered from one or many source systems. The set of tables containing all the data is called the Data Model. In Celonis, every asset you build is based on an underlying Data Model. To obtain the desired information and interpret results or KPIs, it's crucial that you know the structure of the Data Model you're working with. 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. The database server provides support for using a subset of the ANSI join syntax.

This syntax that includes the following keywords:
ON keyword to specify the join condition and any optional join filters LEFT OUTER JOIN keywords to specify which table is the dominant table .In an ANSI outer join, the database server takes the following actions to process the filters: Applies the join condition in the ON clause to determine which rows of the subordinate table (also referred to as inner table) to join to the outer table Applies optional join filters in the ON clause before and during the join If you specify a join filter on a base inner table in the ON clause, the database server can apply it prior to the join, during the scan of the data from the inner table. Filters on a base subordinate table in the ON clause can provide the following additional performance benefits: Fewer rows to scan from the inner table prior to the join Use of index to retrieve rows from the inner table prior to the join .Fewer rows to join Fewer rows to evaluate for filters in the WHERE clause For information about what occurs when you specify a join filter on an outer table in the ON clause. Applies filters in the WHERE clause after the join. Filters in the WHERE clause can reduce the number of rows that the database server needs to scan and reduce the number of rows returned to the user.

The term post-join filters refers to these WHERE clause filters. When distributed queries that use ANSI-compliant LEFT OUTER syntax for specifying joined tables and nested loop joins are executed, the query is sent to each participating database server for operations on local tables of those servers. For example, the demonstration database has the customer table and the cust_calls table, which tracks customer calls to the service department. Suppose a certain call code had many occurrences in the past, and you want to see if calls of this kind have decreased. To see if customers no longer have this call code, use an outer join to list all customers.

4.6 PU-Functions

PU_SUM: Calculates the sum of the specified source column for each element in the given target table. Like the regular SUM operator, the column can either be an INT or FLOAT column. The data type of the result is the same as the input column data type.

PU_AVG: Calculates the average of the specified source column for each element in the given target table. Like the regular AVG operator, the column can either be an INT or FLOAT column. The data type of the result is always a FLOAT.

PU_COUNT, PU_COUNT_DISTINCT: Calculates the (distinct) number of elements in the specified source column for each element in the given target table.PU_COUNT and PU_COUNT_DISTINCT can be applied to any data type. The data type of the result is always an INT.

PU_MAX, PU_MIN: Calculates the maximum/minimum of the specified source column for each element in the given target table. Like the regular MAX/MIN operator, PU_MAX/PU_MIN can be applied to any data type. The data type of the result is the same as the input column data type.

PU_MEDIAN: Calculates the median of the specified source column for each element in the given target table. Like the regular MEDIAN operator, the column can either be an INT, FLOAT, or DATE column. The data type of the result is the same as the input column data type.

PU_QUANTILE: Calculates the quantile of the specified source column for each element in the given target table. Like the regular QUANTILE operator, the column can either be an INT, FLOAT, or DATE column. The data type of the result is the same as the input column data type. The given quantile has to be a float number between 0 (same as PU_MIN) and 1.0 (same as PU_MAX).

PU_FIRST, PU_LAST: Returns the first/last element of the specified source column for each element in the given target table. An "ORDER BY" expression can be set as last parameter to define the order that should be used to determine the first/last element. PU_FIRST, PU_LAST can be applied to any data type. The data type of the result is the same as the input column data type.

PU_STRING_AGG: Returns the concatenation of strings from the specified source column for each element in the given target table. The delimiter will be always inserted between the concatenation of the strings. Multiple order by expressions can be used in order to determine the order of the concatenation.

Chapter 5 Modules

Module 1: Techniques of Process Mining

All company's operations are centered on its business processes. Process mining techniques help you gain an understanding of your processes and the workflows that lead to the successful operation of your business. This lets you know how well these processes are working, what problems may exist within them, and how you can improve them to make your business run more smoothly. The technique an organization employs for process mining depends on the stage at which its process models are stored. The following are the most widely used process mining techniques:

Automated Process Discovery

Automated process discovery is a subset of process mining that defines the data-driven visualization of a process. It provides an intuitive, visual, and interactive method for exploring each and every step of processes in order to identify bottlenecks. It uses both machine learning and artificial intelligence to keep track of all the possible ways in which a process might be carried out and to suggest ways to automate them. As a result, process discovery aids in creating workflows and deploying automated processes quickly and efficiently.

Module 2: Tools

Process mining tools are largely used by organizations that want to optimize their processes, conform existing processes to certain specifications, create harmony between distinct processes, or get future predictions regarding their processes. Process mining tools can help you collect data about your processes and then use that data to find ways to improve them in real time, which will ultimately lead to better outcomes across the board. The following is a list of the most widely used process mining tools.

ABBYY Timeline

ABBYY Timeline is a cloud-based, AI-driven process mining platform that enables businesses to develop a visual model of their processes, analyze them in real-time to detect bottlenecks, and predicting possible results to help with technology investment decisions. It is aimed to benefit healthcare providers, insurance corporations, banks, and other similar entities in optimizing their internal operations. A transparent pricing procedure, analytics, data integration, and the ability to scale quickly are just a few of the features featured in this intelligent solution.

ARIS Cloud Process Mining

ARIS Cloud is a process-centric management solution that offers process mining solutions. It is accessible and versatile, allowing for additional resources to be deployed as needed. ARIS Cloud includes features such as process versioning, release cycle management, content merging, social collaboration, document management, and customer journey mapping in addition to process design, modeling conventions, method filters, content languages, and process versioning. There are multiple levels of service packages available, from free trials to high-end cloud services to enterprise clouds.

Module 3:Process KPIs

Fig 5.1 Process KPI

Process effectiveness is the delivery of a qualified service or product in a way that it satisfies the customers. Some examples for process effectiveness KPIs include:

1.Quality: The output (service or product) meets the client standards, internal QA and budget.

2.Error rate: The number of units, products or service that failed during the entire process cycle

3.Customer Satisfaction: How well the process meets the customer’s expectations.

4.Conversion rate: The number of prospects that company interacts to become customers

Chapter 6

Pros and Cons of Process Mining

6.1 Pros of Process Mining

Improved efficiency: By analyzing data from information systems, process mining can identify areas of inefficiency and help streamline processes.

Increased visibility: Process mining provides a detailed view of business processes, making it easier to identify bottlenecks and areas for improvement.

Better compliance: Process mining can help organizations ensure they are following established procedures and regulations.

Data-driven decision making: Process mining provides valuable data that can be used to inform decision-making and drive business improvement.

Minimize manual effort: Process mining is automated, reducing the need for manual effort and freeing up resources for other tasks.

6.2 Cons of Process Mining

Complexity: Process mining can be a complex process that requires specialized skills and expertise.

Data quality: The quality of process data can have a significant impact on the results of process mining. Poor quality data can lead to incorrect conclusions and ineffective process improvement initiatives.

Technical expertise: Process mining requires technical expertise to set up and configure. Organizations may need to invest in training or hire specialized personnel to get the most out of the technology.

Time and cost: The process of collecting, cleaning, and analyzing process data can be time-consuming and costly, especially for large organizations with complex processes.

Resistance to change: Some organizations may be resistant to change, especially if the results of process mining highlight areas for improvement.

Chapter 7
Learning Outcomes

After you complete this training, you should be able to:

- Understand what process mining is and the basics of how it works.
- Understanding how process mining helps in Business world.
- Summarize what an event log is and why we need it for processing.
- Identify business use cases for process mining.
- Learn how to find training courses to get started.
- Understanding how to discover, analyses, and improve business process using data driven techniques.
- You will learn to extract insights from event logs, identify bottlenecks, inefficiencies, and opportunities for optimization.
- And also, you will learn to extract to create visual representations of processes to aid decision making and process improvement efforts.
- You will gain skills in using process mining tools and interpreting the results to enhance organizational efficiency and effectiveness.
- After completing this course you will learn about process query language.
- Understanding process behavior and it’s applications in day life.

Conclusion

In conclusion, process mining is a powerful tool that enables organizations to optimize their processes by analyzing event data. It enhances efficiency, identifies bottlenecks, and supports data-driven improvements. However, successful implementation requires careful attention to data quality, privacy considerations, and alignment with business objectives to ensure meaningful and actionable insights.

INTERNSHIP CERTIFICATE

References

- Process Mining | Celonis | The Leader in Process Mining
- https://internship.aicte- india.org/
- https://www.altexsoft.com/blog/process-mining/
- https://www.qentelli.com/thought-leadership/insights/process-mining-techniques- tools-you-should-adopt
- https://research.aimultiple.com/process-kpis/