

An Internship Report

on

Process Mining Virtual Internship

Submitted in partial fulfilment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering (AI&ML)

by

S.ALIYA
(224G1A3303)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

(ARTIFICIAL INTELLEGEANCE & MACHINE LEARNING)

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.

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 (AI&ML)



Certificate

This is to certify that the internship report entitled “**Process Mining Virtual Internship**” is the bonafide work carried out by **S.ALIYA** bearing Roll Number **224G1A3303** in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering (AI&ML)** for 10 weeks from April – June 2024

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

PREFACE

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

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

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

Business Activities:

- KPI Tracking
- Process Discovery
- Process Optimization
- Predictive Analysis
- Task Prioritization
- Data Integration

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 (AI&ML)**, for his kind support and for providing necessary facilities to carry out the work.

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

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

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

S.ALIYA

(224G1A3303)

INDEX

Contents		Page No.
List of Figures		vi
List of Abbreviation		vii
Chapter 1	Introduction	1
	1.1 Process Mining Cycle	
	1.2 Process Mining	
	1.3 Process Mining Importance	
Chapter 2	Technology	3
	2.1 Process Discovery	
	2.1.1 Benefits of Process Discovery	
	2.2 Conformance Check	
	2.3 Analysis and Enhancement	
Chapter 3	Applications of Process Mining – Operational Monitoring and Alerts, Dynamic Resource Allocation, Customer Support and Service, Supply Chain Visibility, Energy Management	6
Chapter 4	Modules Explanation	8
Chapter 5	Real time example of Process Mining – Education, Sales, Supply Chain Management, Finance, IT & Software Benefits of Process Mining – Identifies process Bottlenecks, Increases Transparency, Reduces expenses	20
Chapter 6	Learning outcomes of Process Mining	22
Conclusion		23
Internship certificates		24
References		26

LIST OF FIGURES

Fig.No.	Description	Page No.
1.1	Process Mining	2
2.1	Process Mining Techniques	3
2.2	Process Discovery	4
2.3	Analysis and Enhancement	5
4.1	Traditional Process Mapping vs. Process Mining	8
4.2	Case Frequency	12
4.3	Activity Frequency	13
4.4	Development of SO items and Net Order Value	14
4.5	Celonis Software Architecture	16
4.6	Left Join	18
4.7	Data Model Built	19

LIST OF ABBREVIATION

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

CHAPTER 1

Introduction

Process mining technology transforms event logs into a process map capable of providing insights to better monitor and improve processes through identifying variants, rework patterns, and bottlenecks based on true data. Process mining is a technique that involves extracting insights and knowledge from event logs generated by various IT systems. It provides a visual representation of real processes, allowing organizations to analyze, optimize, and improve their operations. Through process mining, you can identify bottlenecks, deviations, and inefficiencies in workflows, ultimately leading to more informed decision-making enhanced process efficiency. 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. Process Mining achieves this union by taking the digital footprints that are created IIT systems and using them to reconstruct and visualize process flows.

- Process Mining virtual internship is under the AICTE Organization.
- All India Council for Technical Education (AICTE) is a national level council for technical education, under the Department of Higher Education.
- AICTE is associated with Eduskills internship, it has corporate partner Celonis Process Mining academy through this platform they provided Process Mining internship.

Business process management (BPM), data mining, and process mining are all related concepts. Despite the fact that both process mining and data mining utilize data, each dataset's size varies. To create process models that may be used to find, contrast, or improve a certain process, process mining especially leverages event log data. Data mining encompasses a range of different data sources and has a much larger application. Applications include customer churn analysis, fraud detection, and market basket analysis, to mention a few. It is used to monitor and forecast behaviors. In contrast to BPM, which has traditionally been managed more manually, process mining is more data-driven. BPM often employs workshops and interviews to gather data in a more informal manner before using software to map out the workflow. Process mining applies a more quantitative approach to a

process problem by describing the actual process through event data since the data that supports these process maps is more qualitative.

1.1 Early Stages

Process Mining originally emerged from academic research into how event log data retrieved from Information Systems could be used to discover, monitor and improve real processes.

This real data can facilitate several aspects of Business Process Management including:

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

1.2 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.

1.3 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
Department of Computer Science & Engineering (AI&ML)

CHAPTER 2

Process Mining

Process Mining is family of techniques relating the fields of data science and process mining management to support the analysis of operational processes based on event logs. The goal of process mining is to turn event data into insights and actions. Process mining is an integral part of data science, fueled by the availability of event data and the desire to improve and the in processes.^[1] Process mining techniques use event data to show what people, machines, and organizations are really doing. Process mining provides novel insights that can be used to identify the execution paths taken by operational processes and address their performance and compliance problems. Process mining starts from event data. Input for process mining is an event log. An event log views a process from a particular angle. Each event in the log should contain a unique identifier for a particular process instance (called case id), an activity (description of the event that is occurring), and a timestamp. There may be additional event attributes referring to resources, costs, etc., but these are optional. With some effort, such data can be extracted from any information system supporting operational processes. Process mining uses these event data to answer a variety of process-related questions.

There are three main classes of process mining techniques: *process discovery*, *conformance checking*, and *process enhancement*. In the past terms like *Workflow Mining* and *Automated Business Process Discovery* (ABPD) were used.

2.1 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. In the early days, process mining techniques were often convoluted with the techniques used for workflow management. In the year 2000, the very first practically applicable algorithm for process discovery, "Alpha miner" was developed. The very next year, in 2001, a much similar algorithm based on heuristics called "**Heuristic miner**" was

introduced in the research papers. Further along the link more powerful algorithms such as inductive miner were developed for process discovery. As the field of process mining began to evolve, conformance checking became an integral part of it. The year 2004 earmarked the development of "Token-based replay" for conformance checking purposes. Apart from the mainstream techniques of process discovery and conformance checking, process mining branched out into multiple areas leading to the discovery and development of "Performance analysis", "Decision mining" and "Organizational mining" in the year 2005 and 2006 respectively. In the year 2007, the first-ever commercial process mining company "Futura Pi" was established. The "IEEE task force on PM", a governing body was formed in the year 2009 that began to overlook the norms and standards related to process mining. Further techniques were developed for conformance checking which led to the publishing of "Alignment-based conformance checking" in the year 2010. In 2011, the first-ever Process mining book was published. Further along in 2014, a MOOC course was offered by Coursera on Process mining. By the year 2018, nearly 30+ commercially available process mining tools were in the picture. The year 2019 earmarked the first Process mining conference. Today we have over 35 vendors offering tools and techniques for process discovery and conformance checking.

2.2 Categories

There are three categories of process mining techniques.

- **Process Discovery:** The first step in process mining. The main goal of process discovery is to transform the event log into a process model. An event log can come from any data storage system that records the activities in an organisation along with the timestamps for those activities. Such an event log is required to contain a case id (a unique identifier to recognise the case to which activity belongs), activity description (a textual description of the activity executed), and timestamp of the activity execution. The result of process discovery is generally a process model which is representative of the event log. Such a process model can be discovered, for example, using techniques such as alpha algorithm (a didactically driven approach), heuristic miner, or inductive miner. Many established techniques exist for automatically constructing process models (for

example, Petri nets, BPMN diagrams, activity diagrams, State diagrams, and EPCs) based on an event log. Recently, process mining research has started targeting other perspectives (e.g., data, resources, time, etc.). One example is the technique described in (Aalst, Reijers, & Song, 2005),^[18] which can be used to construct a social network. Nowadays, techniques such as "streaming process mining" are being developed to work with continuous online data that has to be processed on the spot.

- **Conformance checking:** Helps in comparing an event log with an existing process model to analyse the discrepancies between them. Such a process model can be constructed manually or with the help of a discovery algorithm. For example, a process model may indicate that purchase orders of more than 1 million euros require two checks. Another example is the checking of the so-called "four-eyes" principle. Conformance checking may be used to detect deviations (compliance checking), or evaluate the discovery algorithms, or enrich an existing process model. An example is the extension of a process model with performance data, i.e., some *a priori* process model is used to project the potential bottlenecks. Another example is the *decision miner* described in (Rozinat & Aalst, 2006b),^[19] which takes an *a priori* process model and analyses every choice in the process model. The event log is consulted for each option to see which information is typically available the moment the choice is made. Conformance checking has various techniques such as "token-based replay", "streaming conformance checking" that are used depending on the system needs. Then classical data mining techniques are used to see which data elements influence the choice. As a result, a decision tree is generated for each choice in the process.
- **Performance Analysis:** Used when there is an *a priori* model. The model is extended with additional performance information such as processing times, cycle times, waiting times, costs, etc., so that the goal is *not* to check conformance, but rather to improve the performance of the existing model with respect to certain process performance measures. An example is the extension of a process model with performance data, i.e., some prior process model dynamically annotated with performance data. It is also possible to extend process models with additional information such as decision rules and national information (e.g., roles).

2.3 Process Mining Cycle

Data Transformation

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

Data Analysis

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

Continuous Monitoring

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

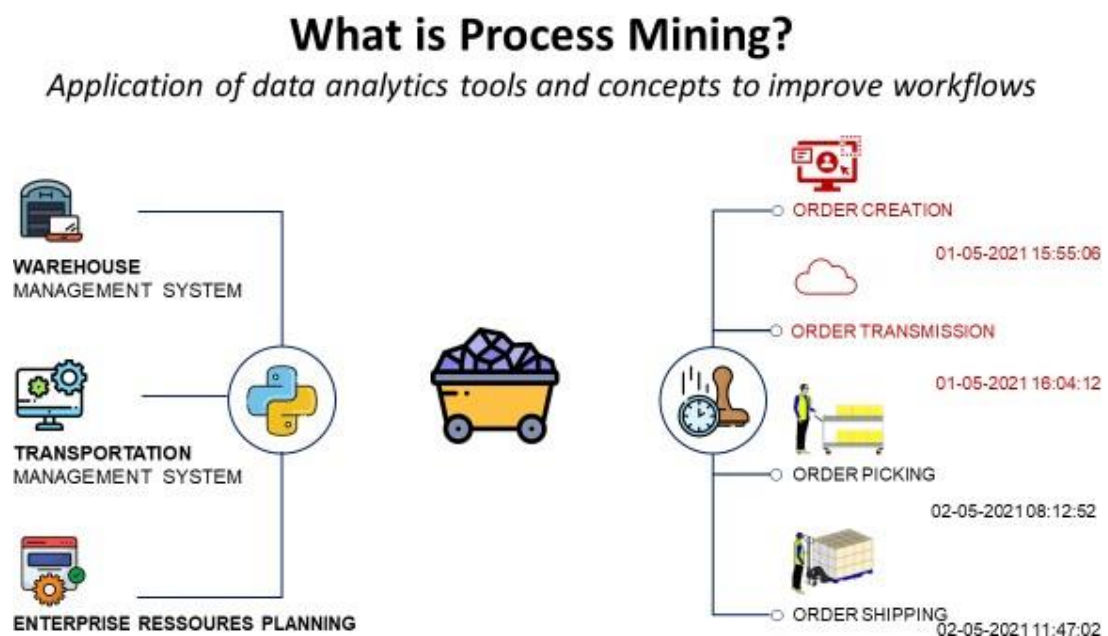


Fig 2.1:Process Mining

Process mining is a type of data analytics that focuses on the discovery, monitoring, and improvement of business processes. 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 and a 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. It involves analyzing data from various sources, such as process logs, to understand how a process is actually being executed, identify bottlenecks and inefficiencies, and suggest ways to improve the process.



Fig 2.2: Example of Supply Chain Information Systems for Process Mining

Why is process mining important?

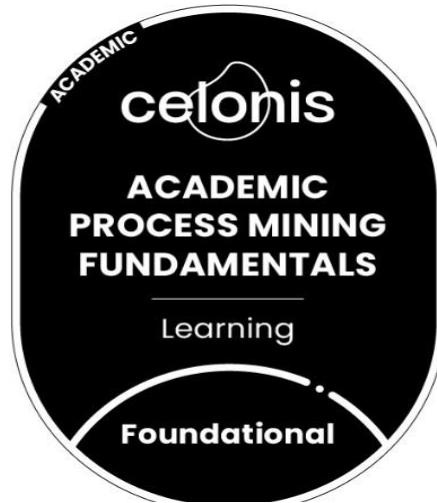
Increasing sales isn't the only way to generate revenue. Six sigma and lean methodologies also demonstrate how the reduction of operational costs can also increase your return-on-investment (ROI). Process mining helps businesses reduce these costs by quantifying the inefficiencies in their operational 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 that.

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.

CHAPTER 3

Fundamentals of Process Mining



Academic Process Mining Fundamentals

The badge earner is able to describe and apply the science of Process Mining inside of Celonis Analyses. They can use the Celonis software to import and analyze process data and identify friction points.

Process Mining Use cases

1. Increase process efficiency by discovering bottlenecks

Process mining can be applied to different financial processes that generate event logs such as account payables (AP), receivables (AR), and procurement to visualize process execution and identify bottlenecks.

Process Mining can provide strategies and recommendations to eliminate process inefficiencies. For example, in claims settlement, mining can identify root causes by measuring the average amount of time to settle a claim.

2. Discover automation opportunities

Leveraging process mining provides insights about processes that can benefit from [automation](#). For example, process mining allows financial organizations to discover the possibility of automation in transactions such as:

- purchase-to-pay processes that take longer due to mistakes and manual interventions can be enhanced by implementing automation solutions such as RPA.
- Invoice processes where automation can enable quicker and less costly invoicing by automating repetitive tasks such as data extraction from PDFs.

You can read our in-depth article to learn more on how to implement RPA in finance.

3. Compliance checking

Process mining enables financial institutions to monitor and log process improvement overtime to ensure that their processes are audit-ready. Process mining also provides conformance check and root cause analysis, allowing financial institutions to compare their processes against rules and regulations and analyze the reasons behind the deviations.

To learn more on specific applications of process mining in compliance and audit:

- 6 Ways Process Mining Enhances Compliance
- 7 Process Mining Use Cases in Auditing

4. Maverick buying in purchase-to-pay

Banks can leverage process mining in purchase-to-pay, which refers to a company's entire purchasing process. For example, Process mining can help reduce maverick buying (i.e. purchases of employees without informing the procurement department). Process mining enables the user to check the necessary steps in a P2P process, including:

- Generating a receipt after a purchase order (PO)
- Matching the PO to a contract: There should not be a PO without a contract (especially if the amount of orders is large in quantity and regular)

5. Root causes for delays and incorrect invoices

Process mining enables users to discover the root causes of delays within the departments. Banks can uncover the reasons for the late fees (e.g. credit card debt payment, loan payments)

debts on time. Also, process mining allows banks to identify the root causes for mistakes and duplicate payments that generate extra workload.

6. Customer satisfaction

Some process mining techniques (e.g. process discovery) can constantly monitor processes in real-time. Such monitoring enables companies to optimize their processes which increase customer satisfaction.

For example, process mining can help financial service providers to understand customer behavior by discovering customer engagement insights (e.g. length of call, pattern in issues). Banks that understand customer perspectives can change their operations to improve customer experiences. Process mining allowed many retail banks to discover customers' difficulties during the bank account confirmation process and provide fast confirmation for opening a bank account.

7. Risk mitigation

Process mining helps financial service providers to avoid potential risks when they innovate or modify processes in their systems by constantly monitoring processes and providing data-driven insights and optimization opportunities.

3.1 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.

3.1.1 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.

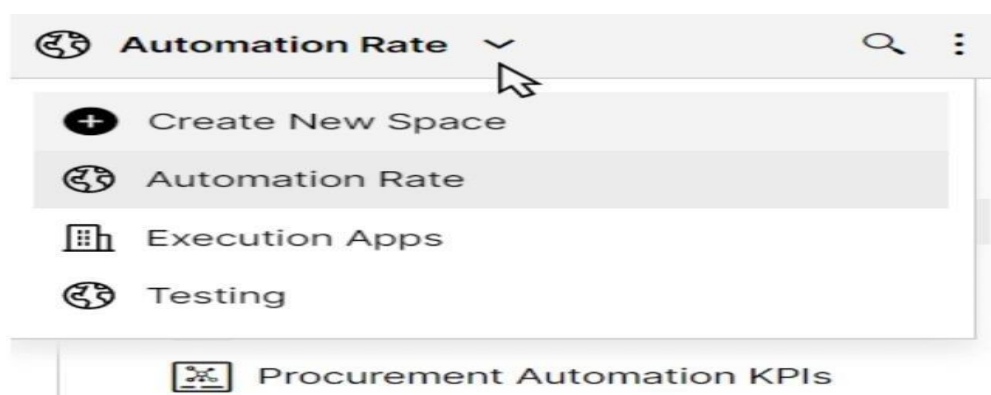


Fig 3.1 Analysis Assest

3.1.2 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-to-Pay (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)."



Sales Organization	# Sales Orders
Vertimode Germany 2	791
Vertimode Netherlands 1	120
Stratodex Germany 2	54
Stratodex China 8	10
Stratodex USA 1	6
Stratodex France 3	2
Stratodex Italy 6	
Stratodex Israel 4	
Stratodex UK 5	

Dropdown menu options: Customer, Material, Sales Organization, Company Code, Plant, Type.

Fig 3.2 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."

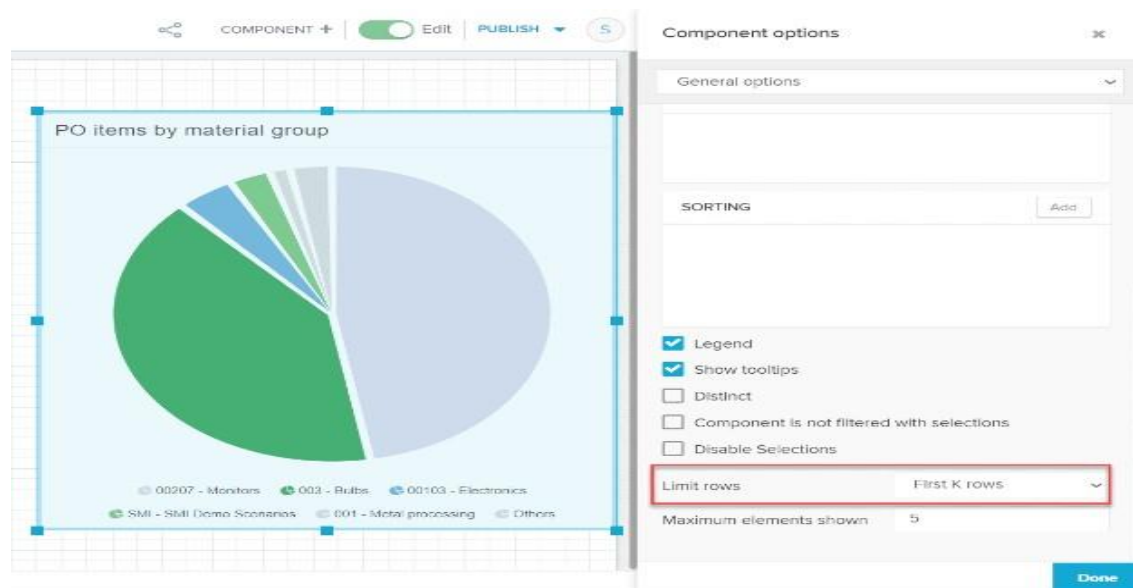


Fig 3.3 Pie Chart

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 select. Once a category is selected, the Visual Editor displays a formula creation screen with a predefined template, according to your category selection. As you click on the template, it offers either functionalities or table columns to build up your formula.

If you have already selected a column or a table, the menu will offer the following possibilities:

- **Change:** Choose another table or column.
- **Add:** With the functions listed in the Add section, you can apply operations to your column. This usually leads to the addition of brackets, to specify.

CHAPTER 4

PQL

Process Query Language

Description

The Process Query Language (PQL) is a domain-specific language tailored towards a special process data model and designed for business users. PQL enables the user to translate process-related business questions into queries, which are then executed by a custom-built query and engine. PQL covers a broad set of operators, ranging from process-specific functions to the aggregations and mathematical operators. Its syntax is inspired by SQL, but specialized for process-related queries. Each case represents a process instance of one purchase order. In the first case, the order item is created in the system, an approval for purchasing is requested and approved. After the approval, the order is sent to the vendor. Two days later, the ordered goods are received, the invoice is registered and eventually paid. In the second case, an order item is created, but the approval to actually order it is rejected. Besides the three required attributes of an event log mentioned above, the example also includes attribute Department, which specifies the executing department for each event, as well as attribute Item containing the description of the corresponding order item.

Architecture Overview

Process mining techniques that are applied on an event log to understand and improve the corresponding process can be assigned to three groups [15]: discovery, conformance and enhancement. Discovery uses the event log as input and generates a business process model as output. Conformance takes the event log and an a priori process model to detect discrepancies between the log data and the a priori model. Enhancement takes the event log and an a priori model to improve the model with the insights generated from the event log.

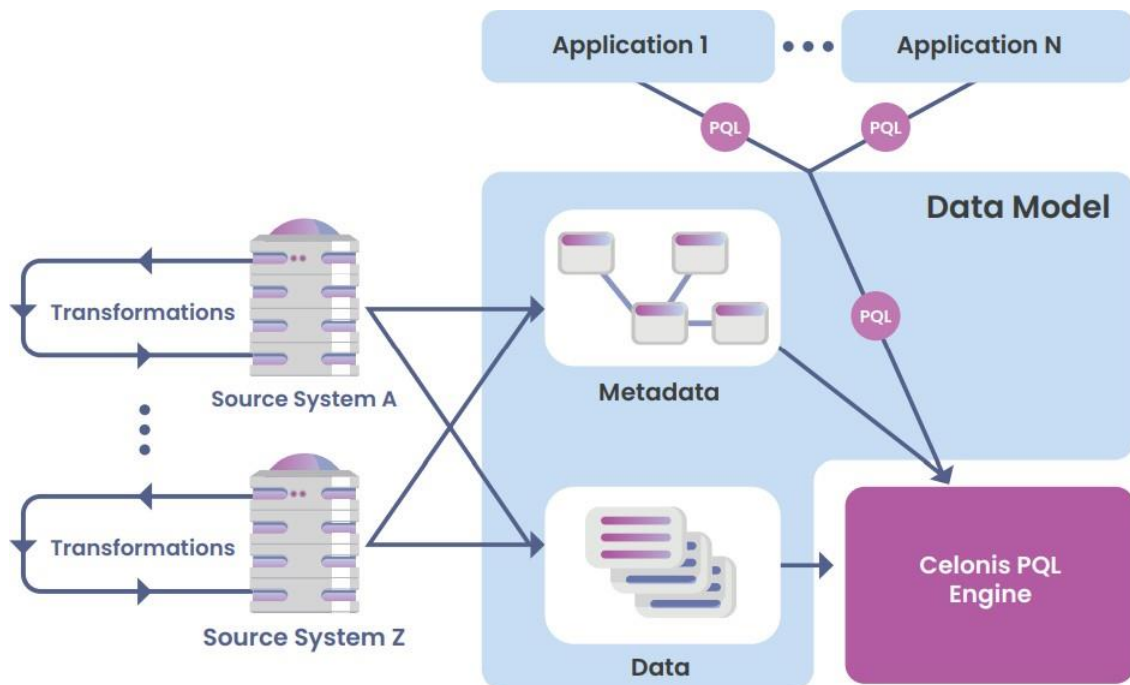


Fig 4.1: Celonis architecture overview

Celonis PQL is an integral component of the Celonis software architecture, which is shown in Figure 2. All Celonis applications use this language to query data from a data model. The data model contains metadata like schema information and the foreign key relationships between the tables, as well as the actual data from the source systems. Celonis PQL queries are evaluated by the Celonis PQL Engine. A data model combines all tables from the source system (or multiple source systems) which contain the data about a process that a user wants to analyze. In the data model, the foreign key relationships between the source tables can be defined. This is performed here because specifying joins is not part of the query language itself. The tables are arranged in a snowflake schema, which is common for data warehouses, and the schema is centered around explicit case and activity tables. Other data tables provide additional context. Figure 3 shows an example data model. It contains the event log of Figure 1 in the Activities table, including the Department column. It is linked to the Cases table, containing information about each order item. The Item attribute from the example event log of Figure 1 is contained in the Cases table, as the data model should contain normalized table schemas. Both order items (i.e. both cases) belong to the same purchase order which is sent

one vendor. Details about the purchase orders and the vendors are available in the Orders and Vendors tables of the data mode.

The Celonis Process Query Language

The intention of Celonis PQL is to provide a query language for performing process mining tasks on large amounts of event data. As described in Section 2.2, it is based on a relational data model. The event and business data as well as all results (including the mined process models) are represented as relational data. Currently, the supported data types comprise STRING, INT, FLOAT, and DATE. Boolean values are not directly supported, but can be represented as integers. Each data type can hold NULL values. In general, Celonis PQL treats NULL values as non-existing and ignores them in aggregations. Also, row-wise operations like adding the values of two columns will return NULL if one of its inputs is NULL.

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. Only a few operators (e.g. for computing a process graph) create and return one or more tables with multiple columns. However, these operators are only used internally by GUI components and are not exposed to the end-user. Currently, Celonis PQL provides more than 150 different operators to process the event data. Due to space limitations, we cannot sketch the full language. However, we can offer a brief overview of the major language features before we present selected examples to showcase the expressiveness of the language. Comprehensive documentation of the Celonis PQL operators can be accessed via the free process mining platform.

Language Overview

Even though Celonis PQL is inspired by SQL, there are major differences between the two query languages. Figure 6 shows these differences by comparing how to query the cases and the number of involved departments for all orders with a value of more than 1000 euros in both languages. Furthermore, it also illustrates the key concepts of Celonis PQL. Similar to SQL, Celonis PQL enables the user to specify the data columns to retrieve from the data model. This can either be an aggregation, which we call a KPI, or an unaggregated column, which we call a dimension. While the data columns are part of the SELECT statement in SQL, Celonis PQL requires them to be wrapped in the TABLE operator, which combines the

specified columns into a common table. In contrast to SQL, Celonis PQL does not require the user to define how to join the different tables within the query

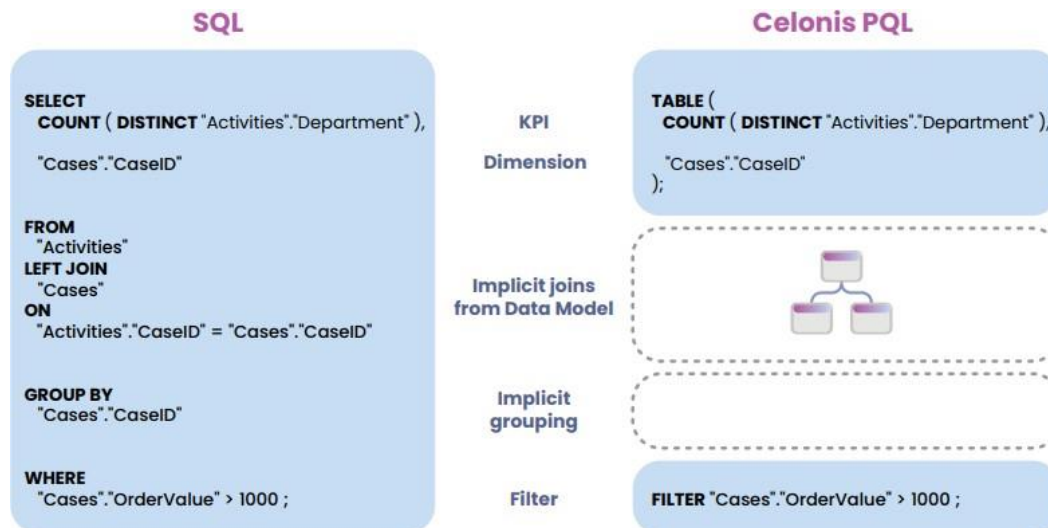


Fig 4.2 SQL vs PQL

Similar to SQL, Celonis PQL enables the user to specify the data columns to retrieve from the data model. This can either be an aggregation, which we call a KPI, or an unaggregated column, which we call a dimension. While the data columns are part of the SELECT statement in SQL, Celonis PQL requires them to be wrapped in the TABLE operator, which combines the specified columns into a common table. In contrast to SQL, Celonis PQL does not require the user to define how to join the different tables within the query. Instead, it implicitly joins the tables according to their foreign key relationships which have to be defined only once in the data model. Also, the grouping clause is not needed in SELECT COUNT (DISTINCT "Activities". "Department"), "Cases". "CaseID" FROM "Activities" LEFT JOIN "Cases" ON "Activities". "CaseID" = "Cases". "CaseID" GROUP BY "Cases". "CaseID" WHERE "Cases". "Order Value" > 1000 ; TABLE (COUNT (DISTINCT "Activities". "Department"), "Cases". "CaseID"); FILTER "Cases". "Order Value" > 1000 ;

SQL Celonis PQL KPI Implicit joins from Data Model Implicit grouping Filter Dimension

14

Celonis PQL: A Query Language for Process Mining Celonis PQL as each selected column which is not aggregated (i.e. a dimension) is implicitly used as a grouper. According to the design goals, implicit joins and groupings significantly reduce the size and complexity of the queries and make it much simpler to formulate them. Both languages offer the possibility to filter rows. While SQL requires the user to formulate the filter condition in the WHERE

clause of the query, Celonis PQL offers the FILTER statements which are separated from the TABLE statements but executed together. Splitting the data selection and the filters into different statements enables the user to define multiple filter statements in different locations inside an application, which then can be combined into the table statement to query the data. Beyond this simple structure, Celonis PQL provides a wide range of different operators which can be combined to answer complex business questions. The following list gives an overview of the most important classes of operators.

Celonis PQL and the Process Querying Framework

The Process Querying Framework (PQF) [13] 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. illustrates how 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 platform. 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 [8]. The correlation models are created by the conformance checking operator (see Section 4.4), 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 which 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.

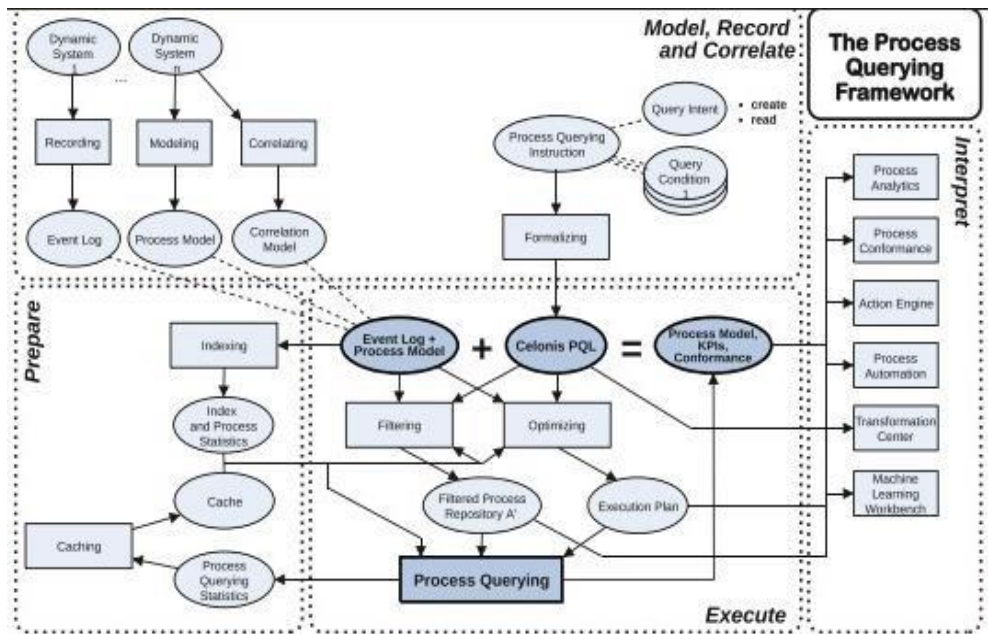


Fig 4.3 Process Querying Framework

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 Section 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. According to [13], 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, as discussed in Section 6. 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, as described in Section 4.2, or by the general filter statement shown in the example in Figure 6. The Optimizing 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 that lead to the successful operation of your business.

CHAPTER 5

Modules

Module 1: Techniques of Process Mining

Every 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.

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.

Apromore Advanced Process Automation Solution

Apromore is an open-source, collaborative platform for business process analytics that supports the entire spectrum of process mining capabilities. The goals of this platform are to enhance the customer experience by exposing and visualizing real-time business processes, as well as to monitor process compliance. Open-source solutions are typically robust because they are designed to meet the needs of the community. Businesses that require more in-depth assistance may be required to implement a commercial solution.

bupaR

bupaR is an open-source suite of R-packages for managing and analyzing business process data. bupaR offers support for various stages of process analysis, including importing and pre-processing event data, calculating, descriptive statistics, process visualization, and conformance checking. This platform is appropriate for businesses that prefer to handle their own process solutions management. The entry barrier may be higher than other desktop solutions. Often, bupaR is used as a foundation for a larger system.

Celonis

Celonis is a powerful and capable process mining suite that collects and analyzes IT data in order to generate actionable insights. It is used to identify and fix operational flaws, making the overall operation more effective. Visual reporting is used by Celonis to help find problems in existing processes. It creates a process flowchart of the company's processes by tracing any IT-supported activities. After that, they'll create models for the best solutions and

the various variants that are currently being used. You can view the complete company's operations in real time, including all active processes.

Datapolis Process Intelligence

Datapolis is a powerful workflow toolkit that unifies process management with data mining. Graphical designs, quick change management, business information, reporting, and integration to Office 365 and SharePoint are just a few of the features offered by Datapolis. This is an enterprise-level tool that offers complete business analytics and monitoring features in addition to process mining.

Fujitsu Process Analytics Software

Fujitsu provides a comprehensive business process management solution that features process management, process analytics, and a business process discovery service. Automated process discovery services give a real-time, end-to-end view of processes, showing problems like bottlenecks and delays. Activity monitoring lets users control the flow of processes in real time and see, analyze, and improve their own processes.

Icaro Tech Ever Flow

Icaro Tech Ever Flow uses log entries to provide an in-depth look at business processes, with simple visualizations. Using analytics, an organization can find potential inefficiencies and bottlenecks, and advanced analysis can lead to the discovery of new processes that can be implemented. Icaro Tech Ever Flow can be utilized for field services, operational processes, value-added processes, asset management, and IT service management.

Kofax Insight

Kofax analyzes how your processes affect customer satisfaction and profit. This solution incorporates a user analysis interface, process visibility, analytics. As a unified platform, Kofax Insight is powerful for business analytics. The analytics server, metrics server, and dashboard server are combined in a single system. It offers a drag-and-drop interface that requires no scripting and is simple to use. New users are guided through the analysis step-by-step, and the environment is customizable.

1. LANA Process Mining

LANA Process Mining, a suite of automation and process management tools, is intended to assist organizations in taking advantage of digital transformation. LANA Process Mining features include automatic root-cause analysis and KPI monitoring to name a few.

2. Minit

Minit creates maps of processes that are both visual and interactive so that companies can see how their processes work in real time and innovate with improvements. Minit is all about how people, machines, and software interact with each other. It can be used to enhance the overall structure of a company's technology infrastructure. This is the solution for you, if you want something that has been thoroughly vetted and well-documented.

3. PAF Now Process Mining

The PAF Now platform emphasizes simplicity and ease of use. It's not a free fix, but it's not overly expensive either. Microsoft Business Intelligence provides additional functionality and support by integrating it. PAF Now is a three-stage process that begins with discovery (analyzes your business processes), continues with conformance (reveals insights from actual event logs), and ends with enhancement. Power BI and PAF Now can work together to improve data mining capabilities.

4. ProM

ProM is a free, cross-platform desktop application designed specifically for process mining. The platform's adaptability and scalability are enhanced by the availability of plug-ins, and creators are encouraged to contribute new solutions. Both ProM and any solutions developed on top of it are freely available under the GPL license. ProM provides real-time monitoring and analysis of business processes and includes a number of free, interactive online training modules. If you're new to process mining, you'll benefit greatly from the community and the free online courses. ProM Lite, ProM, ProM 6, Rapid ProM, and ProM 5.2 are just some of the different iterations available. A business may choose to use a stripped-down version or an older version if that's what's best for its needs. Due to its open-source nature, ProM can be adapted to meet the specific requirements of any business.

5. QPR Process Analyzer

QPR Process Analyzer is designed to pinpoint problems in the logistics process, with a special emphasis on supply chain management through the use of logging and time stamps. QPR Process Analyzer allows users to monitor the entire logistics and handling process, from procurement to transportation to delivery.

6. Worksoft Analyze

Worksoft Analyze is an all-inclusive process discovery solution that can be used to record and verify the quality of newly developed procedures. The convenience of this cloud-based service is its portability. In addition to automating the discovery of processes, we can also generate documentation for them without human intervention and streamline testing procedures. In order to boost productivity and lower risk, businesses are increasingly turning to automated process analysis.

Moreover, businesses can save time previously spent on documentation and redirect that effort toward improving the quality of their processes. Worksoft Analyze is great for investigating previously unexplored processes or bringing businesses up to date on regulatory requirements.

7. SAP Signavio Process Transformation Suite

SAP Signavio is a portfolio of products designed to streamline operational procedures and facilitate organizational transformation. This web-based service requires little setup, and it can be tested out for free before committing to a full subscription. Signavio Process Intelligence's web-based platform makes it a valuable tool because it enables users to access their data from any location, making it ideal for modeling, analysis, and improvement. With the help of a collaborative solution like Signavio, businesses and their employees can easily work together to boost process efficiency while also analyzing the full cost of operations the cache to avoid re-computation of either the full query or certain parts of it queries.

Module 3: Benefits of Process Mining

Make Data-Driven Decisions

Unfortunately, most organizations don't have everything they need to make good data-driven decisions. Process mining provides all these capabilities by using your raw data or event logs to extract the true process sequence and close data gaps. As a result, this tool unlocks accurate, unbiased insights into everything that occurs in your end-to-end processes so you can make more informed decisions.

Save Money and Time Through Targeted Automation

Process mining helps you zero in on the tasks within your processes that will benefit most from automation, allowing you to spend your time and money as efficiently as possible. And when you automate these processes, you save time through better productivity, save money by removing unnecessary steps and eliminating human error, and improve your employees' work lives.

Gain Transparency and Justification for Your Digital Transformation Goals

You understand the value and urgency of digital transformation, but your decision-makers may not. That's why process mining is such a powerful ally in your quest to get executive buy-in and employee approval. Executives are constantly being pulled in every direction, so it can be tough to help them understand why your initiative is worth their time. Additionally, employees are often content with "the old way of doing things," and the prospect of change can be scary. This technology uses data to help executives and employees visualize how current systems and processes are chipping away at valuable money, time, work-life quality, and company reputation. It also showcases the immense value digital transformation will bring.

Gain a Comprehensive Understanding of Your Processes As-Is

To make a wise technology investment, you need to have a clear vision of your goals and a solid grasp of your current state. Process mining can help you understand the ins and outs of your current process mining helps you gain this understanding far faster than any employee could and without human bias or error.

Improve Process Performance Management

Process mining tools can automate the data collection of process and employee performance

enables you to continuously monitor KPIs like Time to Resolution, First Time Right, SLA, and others. Keeping close tabs on this data means that you can improve processes as you go. Plus, process mining can access historical process data to produce insights on the parameters that impact KPI measures, detect ideal processes, and improve the deviants.

Boost Customer Satisfaction

When you use process mining to optimize customer processes, you can create more seamless, intuitive, and innovative experiences that increase customer loyalty and expand your customer base.

Process mining tools also allow you to constantly monitor your processes, which enables you to react to problems faster. If an issue arises, the technology can determine its cause quickly so your team can address it. In addition, by putting out fires faster, you can prevent them from escalating, which improves the customer experience.

Identify and Resolve Process Bottlenecks

One of the biggest benefits of process mining is its ability to eliminate even the most stubborn, hidden bottlenecks. Many organizations have bottlenecks in their processes that cause significant slowdowns, but without process mining tools, these obstacles are invisible.

Reduce Risk and Increase Compliance

Process mining helps auditors analyze data faster and enables you to predict where compliance issues and risk factors are likely to exist. Plus, when you know where issues are likely to occur, it's easier to fill in the gaps and keep watch to prevent expensive fines or damaging attacks. Process mining also makes it quick and easy to build dashboards and alarms that identify real-time compliance issues.

Standardize Processes and Streamline Business Outcomes

Processes are typically more complex than a company can predefine. When you bring process mining in, it can detect non-standard procedures down to minute detail. System logs that document your actual processes act as sources to record the differences between actual and idealized processes.

1. Continuous Improvement

There is no limit to how frequently you can apply process mining tools to your processes. It's not just a one-off benefit. Whether you're aiming to determine your optimizing a process, performing a pulse check post-implementation, process mining can add real value.

CHAPTER 6

Applications

Process mining finds applications across various industries and business sectors. Here are some key applications:

- 1. Process Discovery:** Process mining can automatically reconstruct and visualize actual business processes based on event logs. This helps organizations gain a clear understanding of how processes are executed, revealing variations and deviations from the intended flow.
- 2. Process Conformance Checking:** By comparing actual process executions with predefined process models or guidelines, process mining identifies deviations, non-compliance, and bottlenecks. This is crucial for ensuring adherence to regulatory requirements and best practices.
- 3. Performance Analysis:** Process mining enables the analysis of process performance metrics, such as cycle times, waiting times, and throughput. This helps organizations identify inefficiencies, optimize resource allocation, and enhance overall operational efficiency.
- 4. Root Cause Analysis:** Process mining allows for tracing back process issues to their root causes. By analyzing event logs, organizations can pinpoint the sources of bottlenecks, delays, and errors, facilitating targeted process improvement efforts.
- 5. Resource Allocation:** Understanding how resources are utilized within a process is essential. Process mining provides insights into how people, machines, and other resources are involved, helping organizations optimize resource allocation and utilization.
- 6. Compliance Monitoring:** Industries with strict regulatory requirements, such as healthcare and finance, benefit from process mining's ability to monitor and ensure compliance with regulations and standards.
- 7. Customer Journey Analysis:** In customer-centric industries, process mining can map out the end-to-end customer journey, highlighting touchpoints, interactions, and potential pain points. This assists in delivering an improved customer experience.
- 8. Supply Chain Management:** Process mining can enhance supply chain visibility by tracking the movement of goods, identifying delays, and optimizing inventory management processes.

9. Audit and Fraud Detection: By analyzing event logs, process mining can help auditors detect fraudulent activities or irregularities in processes, contributing to improved financial integrity.

10. Continuous Process Improvement: Process mining supports a data-driven approach to process improvement. Organizations can iteratively analyze data, identify areas for enhancement, and monitor the impact of changes over time.

11. IT System Analysis: Process mining can be applied to analyze how IT systems are used within processes, helping organizations identify inefficiencies, system bottlenecks, and opportunities for automation.

12. Predictive Analytics: In some cases, process mining can be used to predict future process behavior based on historical data, aiding in proactive decision-making and planning.

13. Change Management: When organizations undergo process changes, process mining can help assess the impact of those changes on process performance and ensure a smoother transition.

14. Service Level Agreement (SLA) Monitoring: Process mining can track SLA compliance by analyzing the time taken to complete specific steps within a process.

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 behaviour and it's applications in day life.

Conclusion

Process mining is a valuable technique that helps organizations gain insights into their processes by analysing event data. In conclusion, it offers the potential to enhance process efficiency, identify bottlenecks, and make data-driven improvements. However, successful implementation requires careful consideration of data quality, privacy concerns, and alignment with business goals.

Internship Certificate



N·E·A·T
प्रौद्योगिकी के लिए राष्ट्रीय शैक्षणिक सहयोग
National Educational Alliance for Technology



अखिल भारतीय तकनीकी शिक्षा परिषद्
All India Council for Technical Education



EduSkills®
Nation Building Through Skills



Certificate of Virtual Internship

This is to certify that

Shaik Aliya

Srinivasa Ramanujan Institute of Technology

has successfully completed 10 weeks

Process Mining Virtual Internship

During April - June 2024

Supported By **celonis**



Angela-Sophia Gebert
Global Head of Academic Alliance
Celonis



Shri Buddha Chandrasekhar
Chief Coordinating Officer (CCO)
NEAT Cell, AICTE



Dr. Satya Ranjan Biswal
Chief Technology Officer (CTO)
EduSkills



Certificate ID :0ca2144a7e65f1ecf1e7391d658d6c12
Student ID :STU641C3F439d323167952803



GRADE- O (Outstanding):90-100 | E (Excellent):80-89 | A (Very Good):70-79 | B (Good): 60-69 | C (Fair): 50-59 | D (Average): 40-49 | P (Pass): 30-39 | F (Fail): Below 30

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>
- <https://research.aimultiple.com/process-kpis/>