

CHAPTER - 1

INTRODUCTION

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

1.1 Process Mining Cycle

Data Transformation

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

Data Analysis

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

Continuous Monitoring

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

1.2 Process mining

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

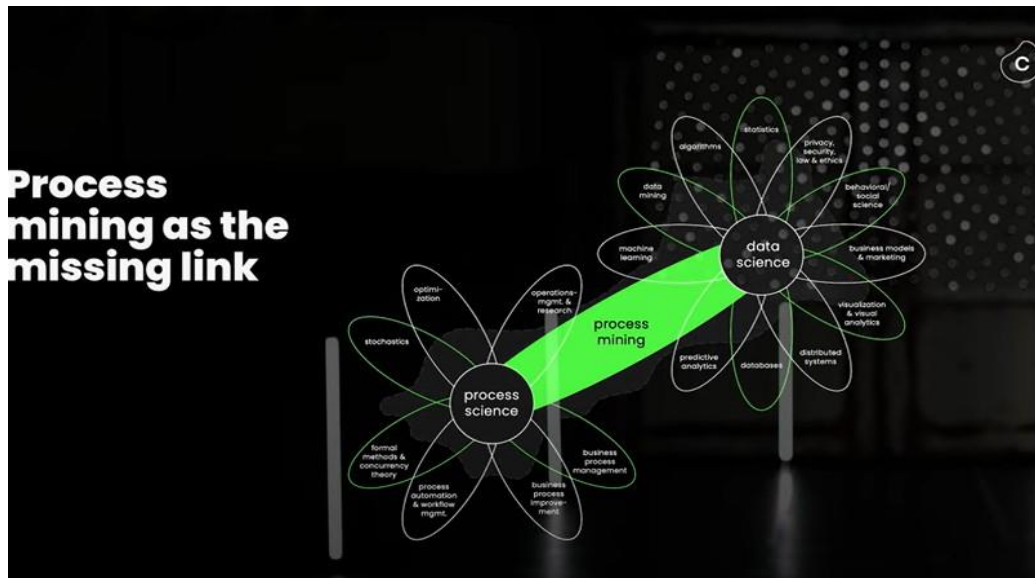


Fig. No. 1.1: Process Mining

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

1.3 Importance of process mining

Reduce costs

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

Increase transparency

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

Reduce multi-hops

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

CHAPTER - 2

PROCESS MINING TECHNIQUES

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

There are three main classes of process mining techniques:

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

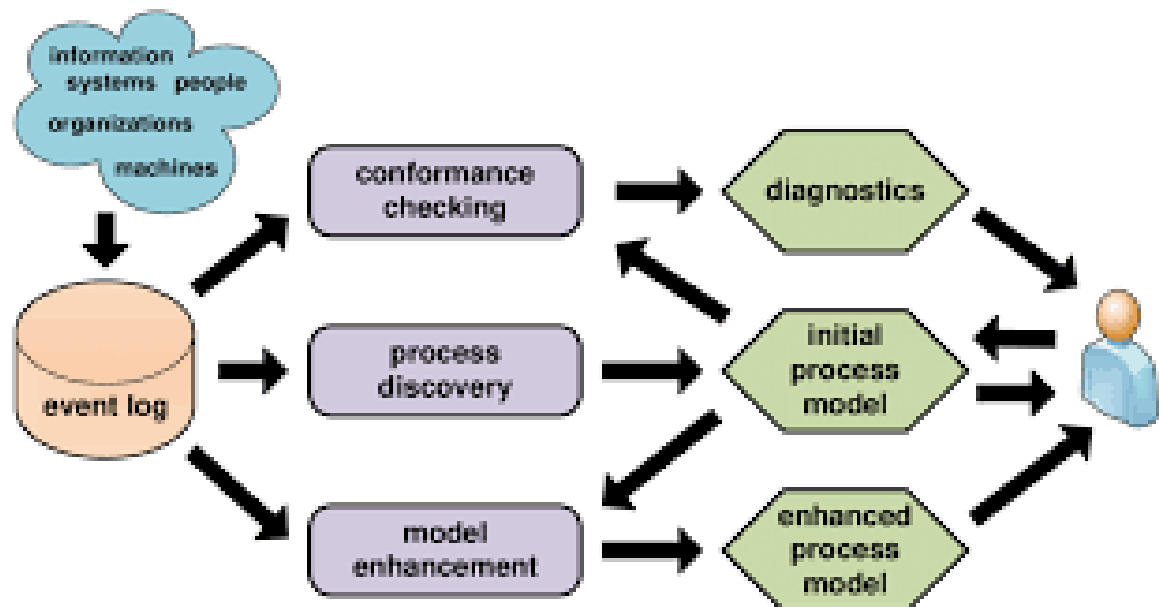


Fig. No 2.1 Process Mining Techniques

2.1 Process Discovery

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

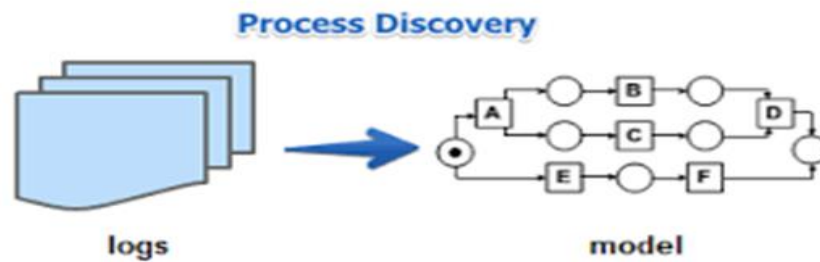


Fig. No. 2.2: ProcessDiscovery

2.1.1 Benefits of Process Discovery

There are so many benefits of process discovery. Some of them are given below.

- Improved Quality and performance
- Competitive Advantages
- Visibility
- Less Risks
- Cost Efficiency
- Improved Scalability

2.2 Conformance Check

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. Conformance checking techniques take as input a process model and event log and return a set of differences between the behavior captured in the process model and the behavior captured in the event log. Conformance checking helps ensure that all process deviations from the target process or reference model are identified, minimizing risk of audit problems or legal violations.

2.3 Analysis and Enhancement

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

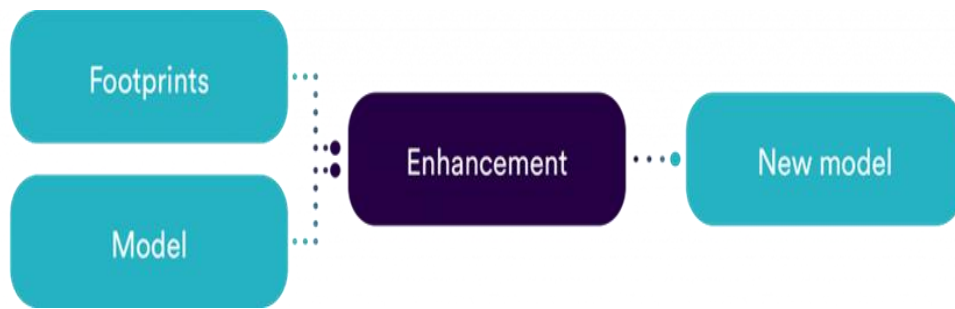


Fig. No. 2.3: Analysis and Enhancement

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

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

CHAPTER - 3

APPLICATIONS

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

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

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

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

CHAPTER - 4

MODULES

4.1 Module 1: Introduction to Process Mining

Process Mining offers a data-driven and therefore more objective and holistic approach to understanding business processes. As a result, Process Mining has come to dominate a large majority of operational excellence, automation, and 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 as good data source for process analytics.

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

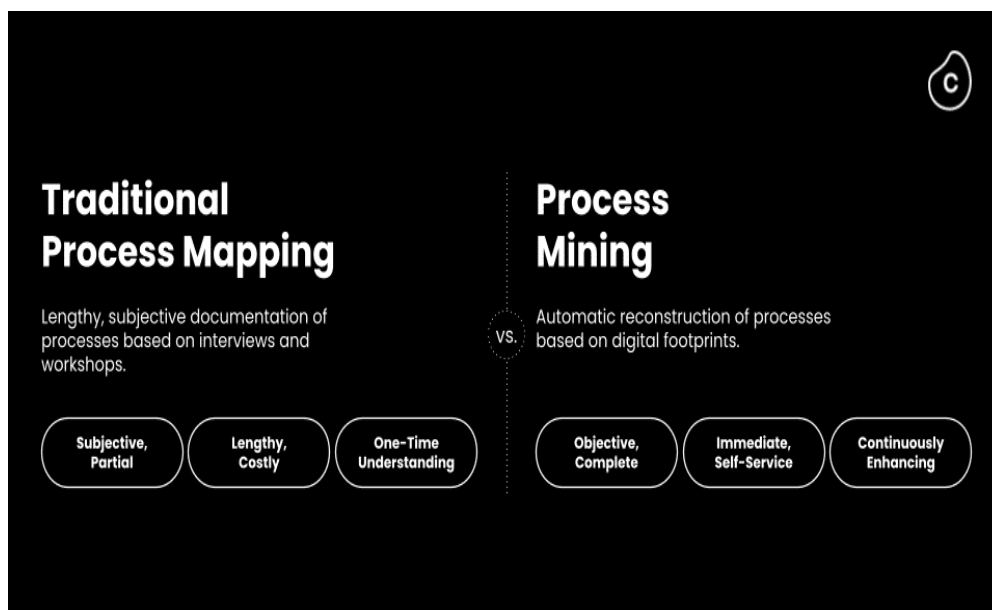


Fig. No. 4.1: Traditional Process Mapping vs. Process Mining

4.1.1 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

4.1.2 Event Logs

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

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

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

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

- **Timestamp:** The date and time that the activity took place

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

4.2 Module 2: Process Mining Fundamentals

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

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

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

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

- **Process:** A process is a series of linked steps taken to achieve a particular goal.
- **Activity:** An activity is a step that occurs in the process. Process activities are actions that initiate or terminate a process or take place during it. Each activity consists of one or more tasks that together are a milestone in the process.
- **Case:** A case is an “item” or “object” you follow through the process. Even for the same business process, the case differs from company to company, depending on how granular they want to get.

4.2.1 Variant Explorer

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

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

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



Fig. No. 4.2: Case Frequency

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

- **Activity Frequency:** 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.
- Above is a side-by-side comparison of the same variant with Case Frequency and Activity Frequency KPIs. Notice the difference in the count for the "Generate Delivery Document" activity.



Fig. No. 4.3: Activity Frequency

4.2.2 Process Explorer

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

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

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

4.2.3 Analysis Charts

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

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

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

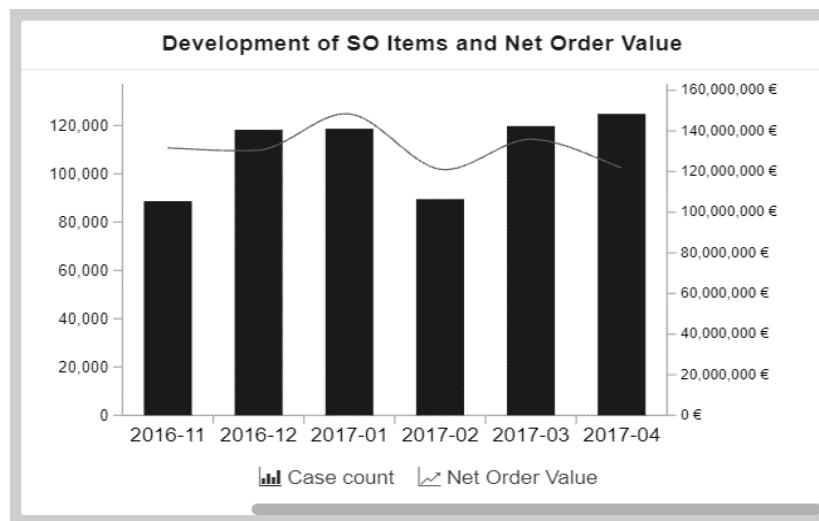


Fig. No. 4.4: Development of SO items and Net Order Value

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

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

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

4.3 Module 3: Rising Star Technical

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

4.3.1 Celonis PQL

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

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

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

4.3.2 Celonis Software Architecture

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

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

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

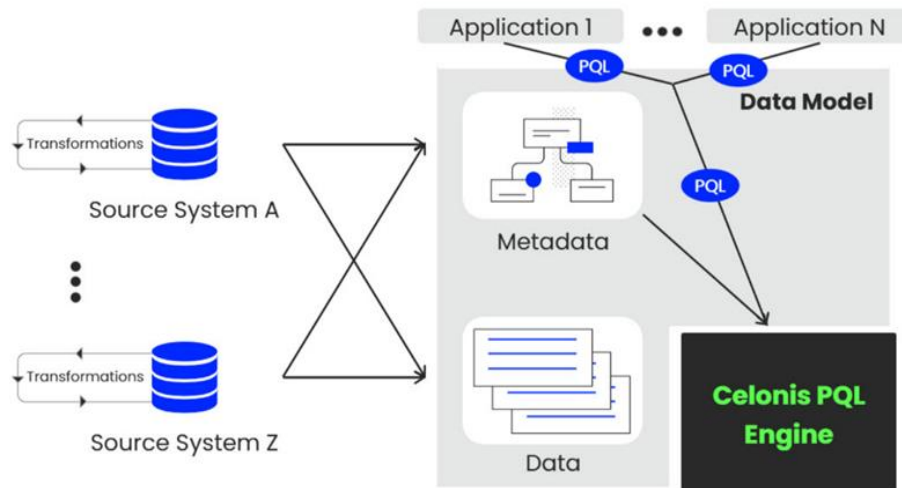


Fig. No. 4.5: Celonis Software Architecture

4.3.3 PQL Queries

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

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

The design of the PQL language follows seven principles:

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

- **Expressiveness:** PQL queries should allow describing many ideas.
- **Intuitiveness:** PQL queries should be easy-to-read and easy to comprehend.
- **Portability:** PQL queries should be independent of execution environments and data formats.
- **Usefulness:** PQL queries should allow fulfilling many practical tasks.

4.3.4 The P2P Process

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

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

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

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

4.3.5 Joins and Filters

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

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

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

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

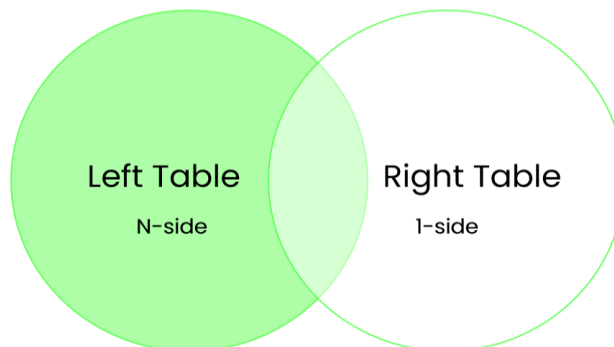


Fig. No. 4.6: Left Join

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

4.3.6 Data Integration

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

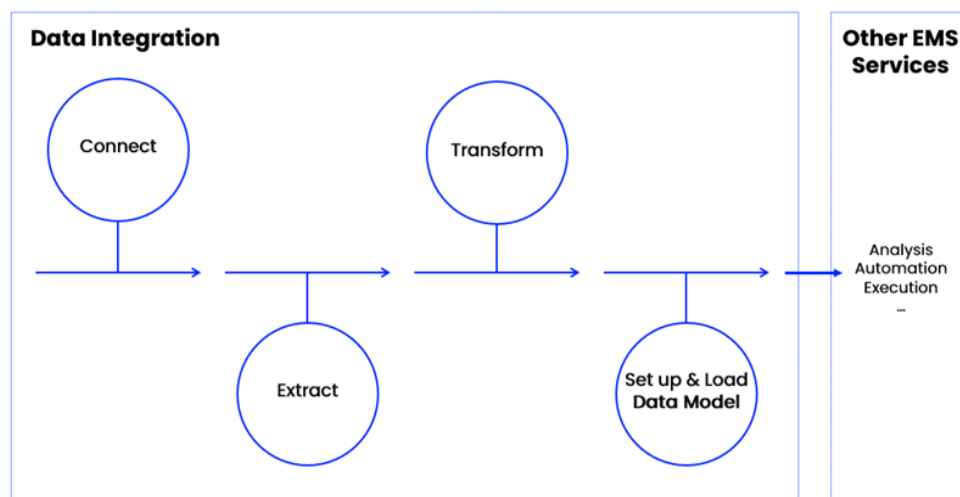


Fig. No. 4.7: Data Model Build

- **Process Connect:** Connecting to source systems is your very first step to pull process data into the Celonis EMS. The EMS utilizes a broad set of technologies like message queues, Restful APIs, Soap APIs, direct database access, or system-specific solutions to connect.
- **Extract Data:** A full extraction is when you load entire source system tables into Celonis. It draws a fresh snapshot of source system tables and completely overwrites data you previously extracted. A delta extraction is when you do a partial load of the new or updated data from your source system. Note that delta extractions with Data Jobs rely on filters rather than change logs.

- **Transform data:**The goal is to transform your data so it can become a Process Data Model. Your most important task is to create the Activity table, also called the event log. This table is the basis for the Data Model you build to visualize your process flow.
- **Data Model:**It is the process of creating a visual representation of either a whole information system or parts of it to communicate connections between data points and structures.
- **Execution Management System:** An Execution Management System is typically associated with the financial industry, particularly in the context of trading. It is a specialized software platform that facilitates the execution of trades in financial markets. EMS platforms provide traders with tools to route orders, access market data, manage orders, and execute trades efficiently across multiple venues and asset classes.

CHAPTER - 5

PROCESS MINING BENIFITS AND ITS USE CASES

5.1 Use Cases

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

Process mining allows financial organizations to discover the possibility of automation

- **Supply chain management:** Process mining software analyzes logistical functions to pinpoint any weak links in a supply chain. Adopting an optimized model makes the supply chain more resilient to unexpected disorders.
- **Finance:** Automation and process enhancement significantly benefits the finance sector. Process mining can optimize audit controls, find root causes for incorrect invoices, mitigate risk, and enhance loan processes.
- **Sales:** Since complex sales processes depend on external factors, process mining allows business leaders to envision discrepancies and resolve them to meet fixed targets. Companies can increase conversion rates, enhance strategies, and improve sales performance across the board.
- **IT & software:** IT professionals benefit from sorting out disorganized engineering processes by gaining clarity and managing the complexity of ERP migrations and implementation. They can also monitor systems in real-time to ensure everything is running smoothly.
- **Customer experience:** With process mining technology, businesses can identify which customer processes are taking longer to resolve and pinpoint their root causes while providing solutions to fix the delay.
- **E-commerce:** Organization leaders in the e-commerce sector can boost their conversion rates by getting exclusive insights into buyer behavior, market trends, and growing customer base.
- **Healthcare:** The healthcare industry has a surplus of data, from health records to appointment booking procedures. Professionals can reconstruct this data digitally with process mining software for seamless integration.

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

5.2 Benefits

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

- **Identifies process bottlenecks:** Process mining tools empower businesses to identify process bottlenecks by discovering the causes of existing discrepancies.
- **Enhances process intelligence:** With the help of process mining software, organization stakeholders better understand processes, evaluate suggested fixes, and make smarter unified decisions based on data-backed insights.
- **Improves business process management:** Process mining is an integral part of BPM. Process mining work means owners can view processes in visual flows and automate tasks for higher efficiency.
- **Increases transparency:** Process mining serves as a guide to internal processes, allowing full disclosure about how functions carry out in a business workflow. This increased transparency makes it possible for decision-makers.
- **Reduces expenses:** Organizations can reduce operating costs drastically by identifying opportunities to automate tasks and fixing problem areas beforehand. It lets businesses analyse event log data and minimizes wasted time and resources.
- **Offers superior customer service:** Users can track process performance in real time by leveraging data procurement to identify bottlenecks and reach solutions much faster. As a result, businesses can improve their decision-making skills and deliver excellent support to their clients.

CHAPTER - 6

LEARNING OUTCOMES

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

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

CONCLUSION

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

INTERNSHIP CERTIFICATES



Certificate of Virtual Internship

This is to certify that

Ramya Sree

Srinivasa Ramanujan Institute of Technology

has successfully completed 10 weeks

Process Mining Virtual Internship

during May - July 2023

Supported By **celonis**



Jerome Geyer-Klingeberg
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 : 7fff970f65d11139a50b378c88be5a1b

Student ID : STU62306498ab3281647338648

REFERENCES

- [1] Sign up for the Celonis Newsletter and stay up to date by celonis group“<https://academy.celonis.com>”

- [2] Explore the Best Practices and ToolsSeptember 16, 2022by Washija Kazim
“<https://www.g2.com/articles/process-mining>”