

An Internship Report

on

Process Mining Virtual Internship

Submitted in partial fulfilment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering (Data Science)

by

Gurram Mekhala Sumanth

(214G1A32A4)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(DATA SCIENCE)**

**SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY
(AUTONOMOUS)**

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

2023 - 2024

**SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY
(AUTONOMOUS)**

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

Department of Computer Science & Engineering (Data Science)



Certificate

This is to certify that the internship report entitled “**Process Mining**” is the bonafide work carried out by **GURRAM MEKHALA SUMANTH** bearing Roll Number **214G1A32A4** in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering (Data Science)** for four months from June 2022 to September 2022.

Internship Coordinator

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

Head of the Department

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

Date:

Place: Ananthapuramu

EXTERNAL EXAMINER

PREFACE

The **All India Council for Technical Education (AICTE)** is a statutory body, and a national-level council for technical education, under the Department of Higher Education. Established in November 1945 first as an advisory body and later on in 1987 given statutory status by an Act of Parliament, AICTE is responsible for proper planning and coordinated development of the technical education and management education system in India.

In its 25 April 2013 judgment, the Supreme Court said "as per provisions of the AICTE Act and University Grants Commission (UGC) Act, the council has no authority which empowers it to issue or enforce any sanctions on colleges affiliated with the universities as its role is to provide only guidance and recommendations." Subsequently, AICTE was getting approval from the Supreme court to regulate technical colleges on a year to year basis till January 2016, when AICTE got blanket approval for publishing the Approval Process Handbook and approve technical colleges including management for the session 2016-17 and in all future sessions.

Company's Mission Statement: To provide for establishment of an All India council for Technical Education with a view to the proper planning and co-ordinated development of the technical education system throughout the country, the promotion of qualitative improvement of such education in relation to planned quantitative growth and the regulation and proper maintenance of norms and standards in the technical education system and for matters connected therewith.

Business Activities: The business activities of AICTE include providing eligible students all over India with useful internships and training programs which lead to the growth of the student and improvement of their knowledge over the provided domain. AICTE has many bureaus to keep these business activities organised.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned our efforts with success. It is a pleasant aspect that I have now the opportunity to express my gratitude for all of them.

It is with immense pleasure that I would like to express my indebted gratitude to my internship coordinator **Mr. P. Veera Prakash, Assistant Professor & HOD, Department of Computer Science and Engineering**, who has supported me a lot and encouraged me in every step of the internship work. I thank him for the stimulating support, constant encouragement and constructive criticism which have made possible to bring out this internship work.

I am very much thankful to **Dr. P. Chitralingappa, Associate Professor & HOD, Computer Science and Engineering (Data Science)**, for his kind support and for providing necessary facilities to carry out the work.

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

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

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

Gurram Mekhala Sumanth
(214G1A32A4)

INDEX

Contents		Page No.
List of Figures		vii
List of Abbreviations		ix
Chapter 1	Introduction	1
Chapter 2	About Celonis	2
	2.1 What is Celonis?	2
	2.2 History of Celonis	2
Chapter 3	Fundamentals of Process Mining	3
	3.1 What does Process Mining do?	3
	3.2 What are processes?	4
	3.3 Why are processes important?	4
Chapter 4	Process Mining Techniques and Tools	5
	4.1 What are Process Mining techniques?	5
	4.2 How are Process Mining techniques grouped?	5
	4.3 Tools used for Process Mining	6
Chapter 5	Stages of Process Mining	8
	5.1 Main stages of Process Mining	8
Chapter 6	Process Visualization in Process Mining	10
	6.1 What is Process Visualization?	10
	6.2 How to visualize processes?	10
	6.3 Process Visualization techniques	10
Chapter 7	Modules	12
	7.1 Module -1: Introduction to Process Mining	12
	7.1.1 Five Common Use Cases of Process Mining	12
	7.2 Module – 2: Process Mining Fundamentals	13
	7.2.2 Variant Explorer	13
	7.2.3 Process Explorer	14
	7.2.4 Case Explorer	15
	7.2.5 Charts and Tables, Review KPIs	15
	7.2.6 Conformance Checker	17
	7.3 Module – 3: Rising Star Technical	18
	7.3.1 Celonis PQL Engine	18
	7.3.2 SQL vs Celonis PQL	18
	7.3.3 Applications in the EMS	19

	7.3.4 Get Data into the EMS	20
	7.3.5 Troubleshooting	25
	7.3.6 Permissions	25
Chapter 8	Applications	26
Chapter 9	Learning Outcomes	29
Conclusion		30
Internship Certificate		31
References		32

LIST OF FIGURES

Fig. No.	Description	Page No.
2.1	Celonis Logo	2
3.1	Process Mining in a Nutshell	3
3.2	Illustration of a Process	4
4.1	Conformance Analysis	5
4.2	Celonis	7
4.3	QPR	7
4.4	UiPath	7
4.5	Fluxicon	7
4.6	ProM Toolkit	7
4.7	Apromore	7
5.1	Process Mining Stages	8
5.2	A Sample Event Log of Order Handling	9
5.3	Process Visualization and Analytics	9
6.1	Flowchart Mapping Example	11
6.2	Process Mapping Example	11
6.3	Journey Mapping Example	11
7.1	Traditional Process Mapping vs Process Mining	12
7.2	Variant Explorer	13
7.3	Process Explorer	14
7.4	Case Explorer	15
7.5	Example of a Chart	16
7.6	Example of a Table	16
7.7	Conformance Overview	17
7.8	Celonis PQL Engine	18
7.9	SQL vs PQL	18

7.10	Data Integration	21
7.11	Data Pool - 1	22
7.12	Data Pool - 2	22
7.13	Data Flow Diagram	24

LIST OF ABBREVIATIONS

EMS	Execution Management System
IEEE	Institute of Electrical and Electronics Engineering
IT	Information and Technology
KPI	Key Performance Index
PQL	Process Query Language
RPA	Robotic Process Automation
SAP	SystemAnalyse und Programmentwicklung
HANA	HochleistungsANalyseAnwendung
SQL	Structured Query Language
TUM	Technical University of Munich
XES	eXtensible Event Stream
XML	eXtensible Markup Language

CHAPTER – 1

INTRODUCTION

Process mining is a family of techniques relating the fields of data science and process 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 desire to improve processes.

Process mining is a revolutionary approach that unveils the hidden dynamics of business processes through data analytics. In a world where every digital interaction generates a trail of data, process mining harnesses this information to create a vivid map of how processes truly unfold. Process mining techniques are often used when no formal description of the processes can be obtained by other approaches, or when the quality of existing documentation is questionable.

- 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 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 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.

CHAPTER – 2

ABOUT CELONIS

2.1 What is Celonis?

Celonis is a German data processing company that offers software as a service (SaaS) to improve business processes.

2.2 History of Celonis

Celonis was founded in 2011 by Alex Rinke, Bastian Nominacher, and Martin Klenk as a spin-off from the Technical University of Munich (TUM). In 2012, Celonis joined the SAP Startup Focus program, an accelerator for analytics startups building new applications on the SAP HANA platform.

In July 2015, Celonis signed a reseller agreement with SAP. Celonis has since been offered by SAP as Celonis Process Mining by SAP. Celonis was the first company from the SAP Startup Focus program to sign a reseller agreement with SAP. In October 2020, Celonis launched its Execution Management System (EMS), a process analysis tool.



Fig. 2.1: Celonis Logo

CHAPTER – 3

FUNDAMENTALS OF PROCESS MINING

3.1 What does Process Mining do?

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 techniques use event data to show what people, machines, and organizations are really doing. Process mining techniques are often used when no formal description of the process can be obtained by other approaches, or when the quality of existing documentation is questionable.

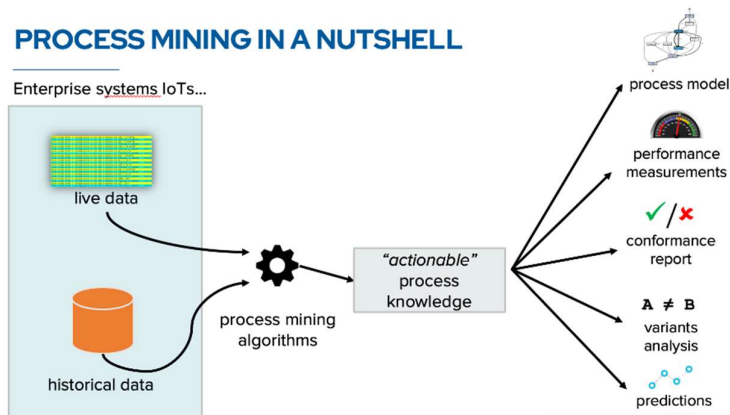


Fig. 3.1: Process Mining in a Nutshell

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 (1) a unique identifier for a particular process instance (called case id), (2) an activity (description of the event that is occurring), and (3) a timestamp. There may be additional event attributes referring to resources, costs, etc., but these are optional. Process mining uses these event data to answer a variety of process-related questions.

3.2 What are processes?

A process is a series of actions taken in order to achieve a particular end. It can be thought of a procedure which is needed to be followed to achieve a required outcome.

An example for a process is the recipe for fried chicken:

- i. Marinate your chicken in buttermilk and hot sauce for at least 3 hours.
- ii. Season some all-purpose with spices of your choice.
- iii. Coat the marinated chicken in the flour well.
- iv. Set the flour coated chicken aside for about 5 mins.
- v. Heat up some vegetable oil in a shallow pan in the meanwhile.
- vi. Once the oil is hot enough, fry your chicken until it's golden-brown and crispy.
- vii. Serve hot and sprinkle some additional seasoning if needed, with a side of your preferred sauce.

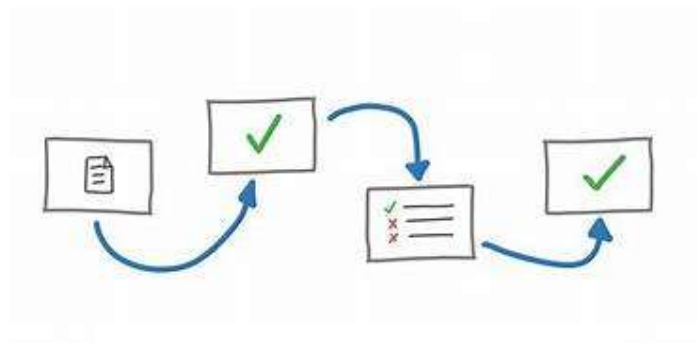


Fig. 3.2: Illustration of a Process

3.3 Why are processes important?

As discussed above in the definition of processes, they describe a step-by-step procedure which is to be followed to obtain a required set goal. Processes in process mining describe and explain about these various procedures in establishments, firms, companies, applications, institutions, organizations etc. Identifying these processes is very crucial because these processes happening can be identified and further analyzed and conclusions can be written. And inferences can be drawn upon these conclusions which can be used for the growth of the organization.

CHAPTER – 4

PROCESS MINING TECHNIQUES AND TOOLS

4.1 What are Process Mining techniques?

The techniques which enable us to perform process mining on various process in industries, institutions, organizations, establishments or companies are called as process mining techniques.

4.2 How are Process Mining techniques grouped?

According to the Process Mining Manifesto, there are three main groups of process mining techniques:

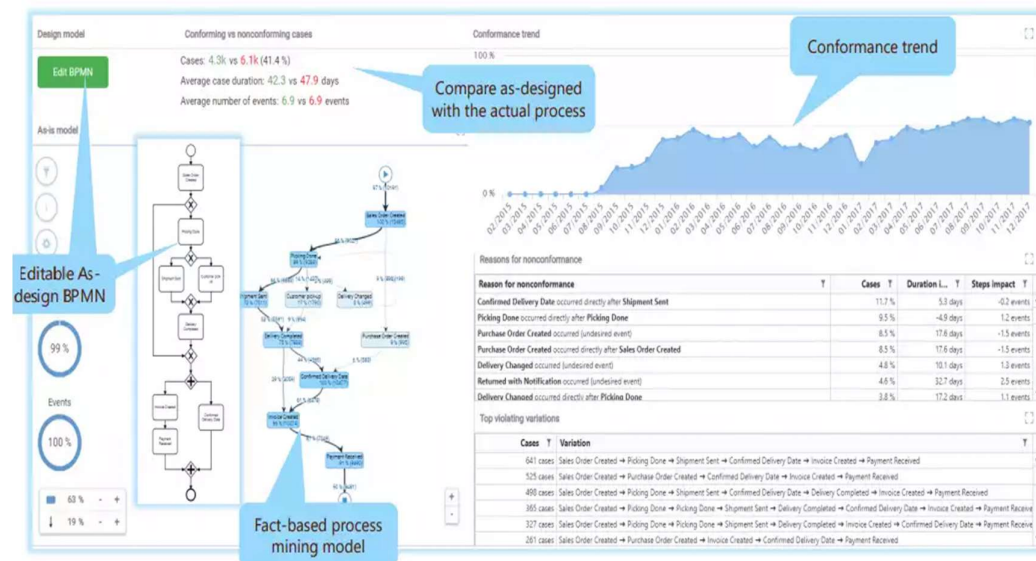


Fig. 4.1: Conformance Analysis

- (Automated) Process discovery is a primary technique and implies extracting and visualizing process models from an event log like we described above.
- Conformance checking compares the actual process with a predefined model to discover deviations. So, it's used to check if the reality conforms to an existing pattern.
- Enhancement goes beyond measuring and comparing. It is focused on extending the model with additional information, such as location data, costs, timing, etc. Enhancing the model with such attributes helps improve its performance and conduct more advanced analysis.

4.3 Tools used for Process Mining

Even though process mining is a relatively new discipline, there's already a number of vendors that offer comprehensive software solutions. Some of the most popular providers of such tools are:

- Celonis
- QPR
- UiPath
- Fluxicon
- ProM Tools
- Apromore

There are many other tools too to use for process mining, but these are the most used and most popular tools on the market. Whatever you choose, remember that you'll have to establish a seamless integration between your software and the process mining solution to enable flawless data extraction and ensure accurate analysis. Process mining software helps organizations analyze and visualize their business processes based on data extracted from various sources, such as transaction logs or event data. This software can identify patterns, bottlenecks, and inefficiencies within a process, enabling organizations to improve their operational efficiency, reduce costs, and enhance their customer experience.



Fig. 4.2: Celonis



Fig. 4.3: QPR



Fig. 4.4: UiPath



Fig. 4.5: fluxicon



Fig. 4.6: ProM Toolkit



Fig. 4.7: apromore

CHAPTER – 5

STAGES OF PROCESS MINING

5.1 Main stages of Process Mining

As people (and software) interact with business IT systems, their actions are captured by these systems and can then be transformed into event logs and visualized with the help of process mining. That's how it happens.



Fig. 5.1: Process Mining Stages

- I. The activity or interaction with the system takes place, creating a digital record. Some examples of such activities are receiving an order, submitting a piece of documentation, approving a loan, entering information into a health record, etc.
- II. Process mining software transforms the digital records into event logs. The most common format for these event logs is an XML-based format XES (eXtensible Event Stream) which was adopted by IEEE Task Force on Process Mining. Event logs have at least three main attributes: case ID, activity, and timestamp.

- III. The visualization of a process is automatically created using event logs. It's important to understand that unlike traditional BPM techniques, process mining shows the real process as it's actually done, not the ideal model as it was meant to be.

order number	activity	timestamp	user	product	quantity
9901	register order	22-1-2014@09.15	Sara Jones	iPhone5S	1
9902	register order	22-1-2014@09.18	Sara Jones	iPhone5S	2
9903	register order	22-1-2014@09.27	Sara Jones	iPhone4S	1
9901	check stock	22-1-2014@09.49	Pete Scott	iPhone5S	1
9901	ship order	22-1-2014@10.11	Sue Fox	iPhone5S	1
9903	check stock	22-1-2014@10.34	Pete Scott	iPhone4S	1
9901	handle payment	22-1-2014@10.41	Carol Hope	iPhone5S	1
9902	check stock	22-1-2014@10.57	Pete Scott	iPhone5S	2
9902	cancel order	22-1-2014@11.08	Carol Hope	iPhone5S	2

...

case id **activity name** **timestamp** **resource** **other data**

Fig. 5.2: A Sample Event Log of Order Handling

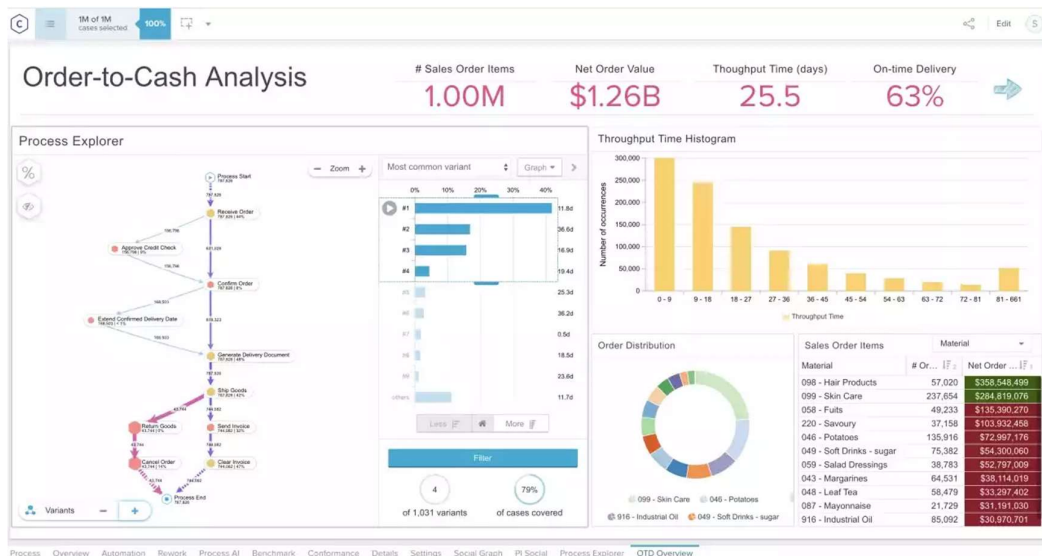


Fig. 5.3: Process Visualization and Analytics

CHAPTER – 6

PROCESS VISUALIZATION IN PROCESS MINING

6.1 What is Process Visualization?

Process visualization provides an end-to-end overview of a process flow including all the steps taken, sub-processes, activities, data inputs and outputs, and risk objects. Analysts can benefit from visualization to illustrate a customer journey or a supply chain as it is, plan a new process, or ensure that the process complies with the ideal model.

6.2 How to visualize processes?

Process visualization can be done manually or automatically, with the help of a software. Manual process visualization can be useful for simple processes in a small organization. However, processes are often complicated and include several intersectionalities even in a small-scaled business, turning manual process visualization into a long project. By using process mining, you can visualize your process model without going through these visualization steps of:

1. Selecting the relevant technique and diagram
2. Understanding the workflow
3. Collecting data
4. Designing and optimizing visuals

6.3 Process Visualization techniques

The various process visualization techniques are:

1. Flowchart mapping
2. Process mapping
3. Journey mapping
4. Process blueprinting
5. Value stream Mapping

Process Flowchart – Employee Onboarding Process

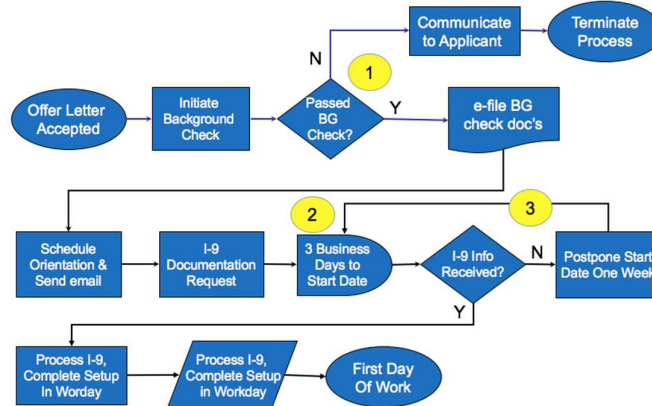


Fig. 6.1: Flowchart Mapping Example

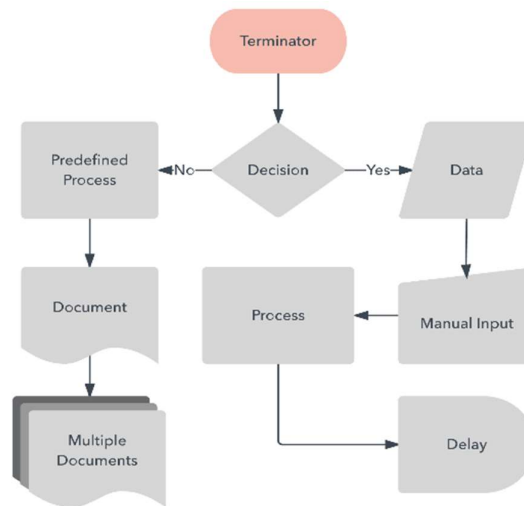


Fig. 6.2: Process Mapping Example

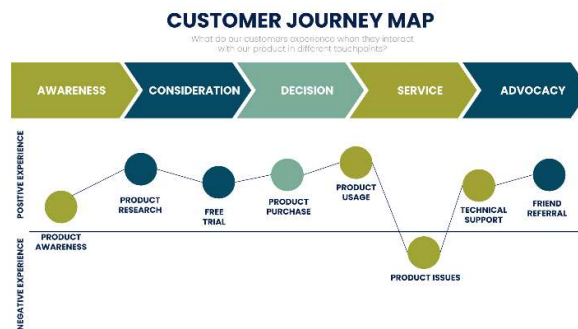


Fig. 6.3: Journey Mapping Example

CHAPTER – 7

MODULES

7.1 Module – 1: Introduction to Process Mining

This module introduces the process mining field and its vast variety of applications. Process Mining is the combination of two disciplines: Data Science and Business Process Management.

Process Mining is the leading new technology when it comes to talking about algorithmic businesses - in other words, businesses that use algorithms and large amounts of real-time data to create business value. This has only become possible through the advent of information systems and administrative tools (e.g. Enterprise Resource Planning or Customer Relationship Management systems) which provide a good data source for process analytics.

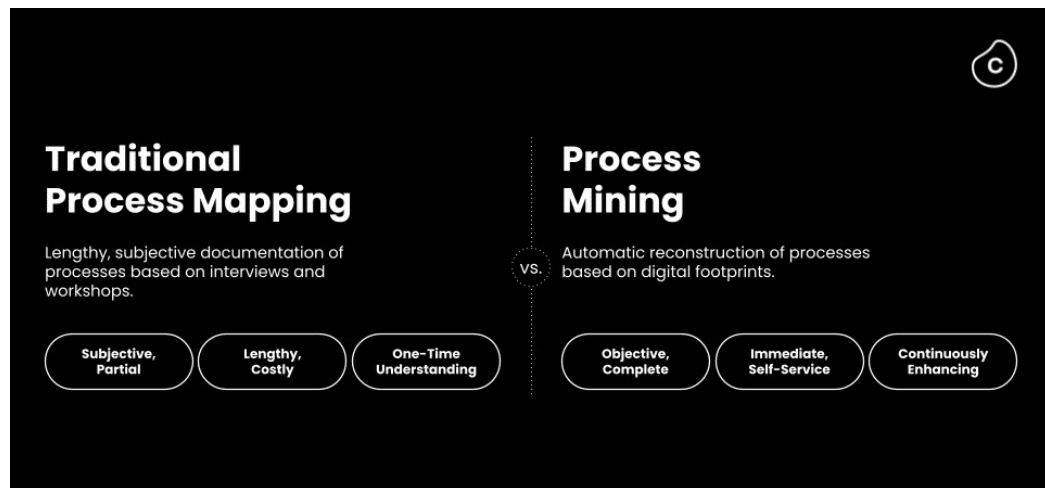


Fig. 7.1: Traditional Process Mapping vs Process Mining

Compared to the traditional process mapping approaches, Process Mining technology solves the complexity and visibility problem.

7.1.1 Five Common Use Cases of Process Mining:

- I. Improve processes with process discovery and analysis
- II. Improve auditing and compliances with process comparison and validation
- III. Improve IT operations with resource optimization
- IV. Improve automation by discovering, validating and monitoring opportunities
- V. Support digital transformation initiatives to achieve success

7.2 Module – 2: Process Mining Fundamentals

The key fundamentals of process mining refer to the core concepts and principles that underlie the analysis and improvement of business processes using process mining techniques. It involves extracting insights from event data to understand how processes function, identifying bottlenecks, inefficiencies, and opportunities for optimization.

Key components include data extraction, process discovery, conformance checking, and process enhancement. Fundamentals of process mining refer to the core concepts and principles that underlie the analysis and improvement of business processes using process mining techniques.

7.2.2 Variant Explorer

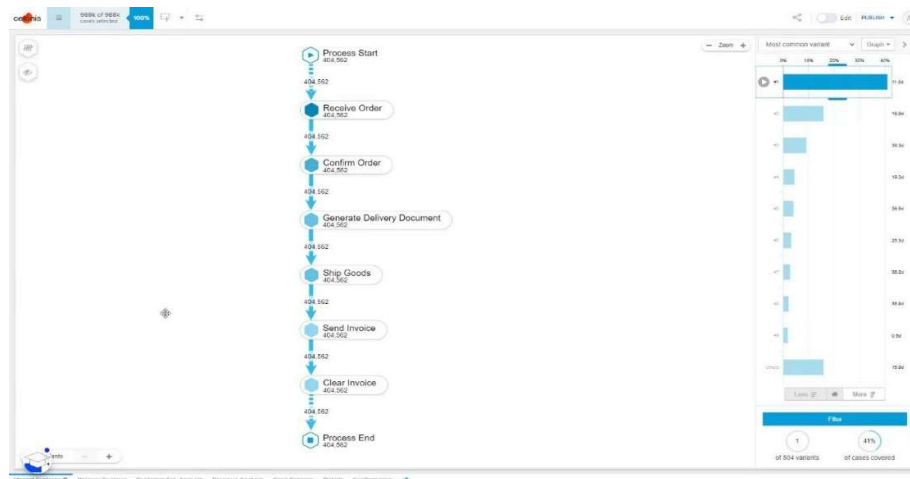


Fig. 7.2: Variant Explorer

Variant is an object which have connections of products or object each other in the portal space. Variant Explorer gives you a quick way to see whether most process cases follow an acceptable flow of activities or not and helps you develop your first analysis questions.

7.2.3 Process Explorer

The Process Explorer is another analysis tool to use when taking an exploratory approach. It's 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're going to. You'll be nicely surprised by all that you can accomplish with it.

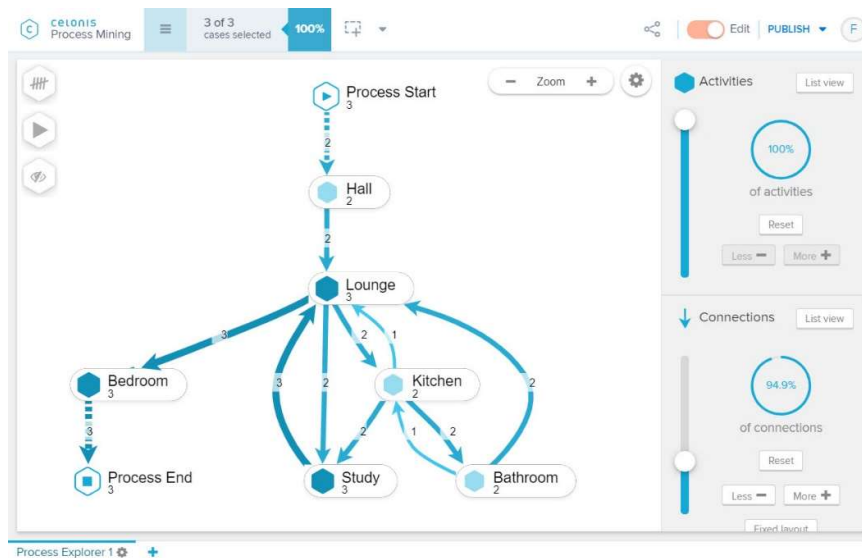


Fig. 7.3: Process Explorer

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. In these Process Explorer and Variant Explorer, KPIs are used to measure the performance of the processes.

7.2.4 Case Explorer

The Case Explorer is useful once you've narrowed down the analysis to a few cases that you want to investigate further. You can view specific case details such as timestamp of activities, user type (manual or automatic), possibly even user name (depending on your setup), and other useful pieces of info.

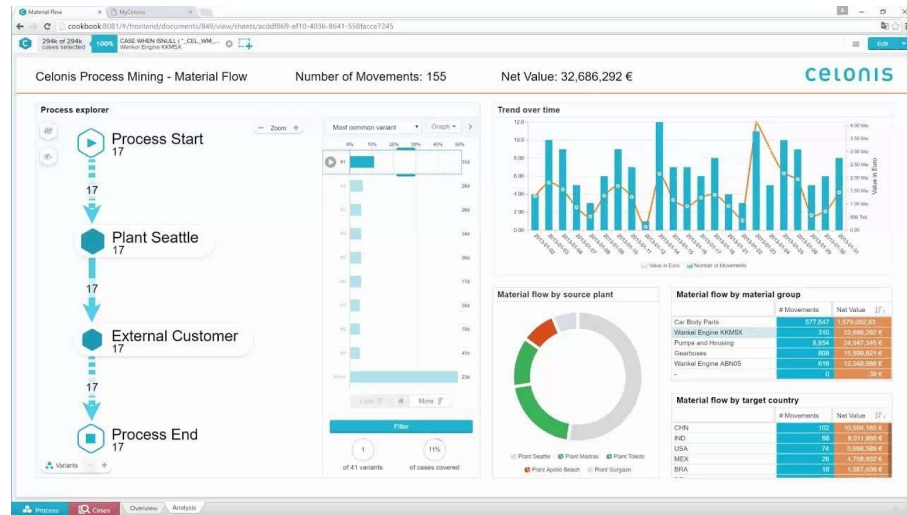


Fig. 7.4: Case Explorer

7.2.5 Charts and Tables, Review KPIs

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. 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.

Examples of such charts and tables are as shown:

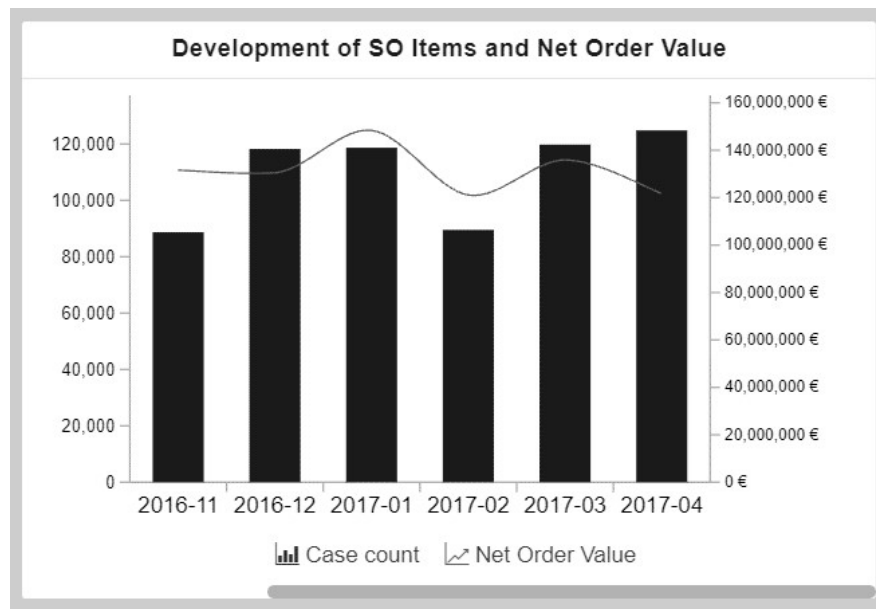


Fig. 7.5: Example of a Chart

Drilldown by Company Code			Sales Organization ▼
Sales Organization	# Sales ... ⚙️ 1	Avg. C... ⚙️ 3	Order Value ⚙️ 2
Vertimode Germany 2	791,770	10 days	\$560M
Vertimode Netherlands 1	120,315	11 days	\$68M
Stratodex Germany 2	54,197	9 days	\$330M
Stratodex China 8	10,837	9 days	\$140M
Stratodex USA 1	6,957	6 days	\$130M
Stratodex France 3	2,368	9 days	\$15M
Stratodex Italy 6	891	9 days	\$7.6M
Stratodex Israel 4	565	9 days	\$6.1M
Stratodex UK 5	201	7 days	\$930k

Fig. 7.6: Example of a Table

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.

7.2.6 Conformance Checker

Every organization has an optimal process in mind that they want to achieve. With the conformance checker, you can see how far away the organization is from reaching that goal and investigate common patterns for inefficiency.

The conformance checker evaluates each case against the process model your organization has specified to determine whether it conforms to it or not. More specifically, the conformance checker is looking at each case's set and sequence of activities to the one defined in the process model.

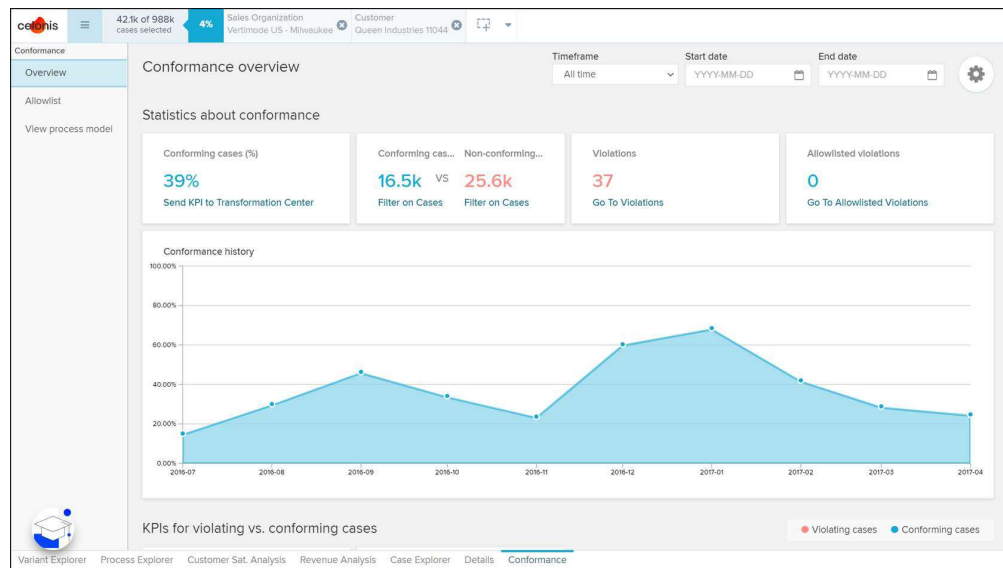


Fig. 7.7: Conformance Overview

7.3 Module – 3: Rising Star Technical

7.3.1 Celonis PQL Engine

Celonis PQL is an integral component of the Celonis Software Architecture. All Celonis applications use this language to query data from a data model.

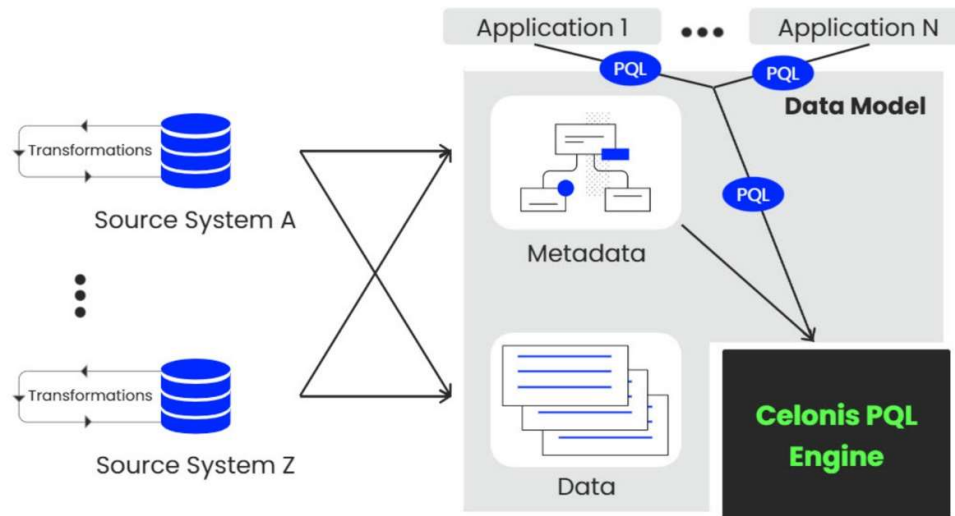


Fig. 7.8: Celonis PQL Engine

7.3.2 SQL vs Celonis PQL

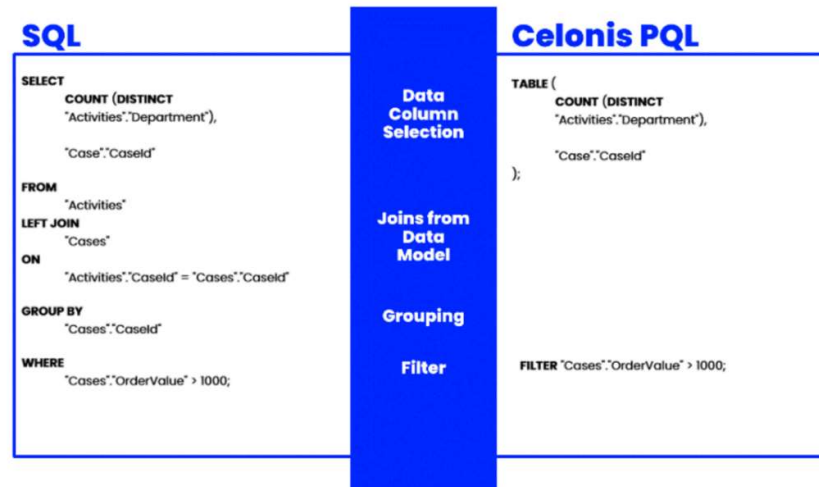


Fig. 7.9: SQL vs PQL

- Celonis PQL does not support all operators that are available in SQL. This is because customer requirements drive the development of the language, and only operators needed for the target use cases are implemented
- Second, Celonis PQL is not supported by a data manipulation language (DML). As all updates in the Process Mining scenario should come from the source systems, there is no need to manipulate and update the data through the query language directly.
- Furthermore, Celonis PQL does not provide any data definition language (DDL). As the data model is created by a visual data model editor and stored internally, there has not been any need for creating and modifying database objects.
- In contrast to SQL, Celonis PQL is domain-specific and offers a wide range of Process Mining operators not available in SQL. Consequently, Celonis PQL seamlessly integrates the data with the process perspective.

Celonis PQL provides more than 150+ different operators that can be combined to process data and answer complex business questions.

Across all classes of operators, Celonis PQL follows four language features:

- i. 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.
- ii. The supported data types comprise STRING, INT, FLOAT, and DATE.
- iii. Boolean values are not directly supported, but can be represented as integers.
- iv. Each data type can hold NULL values. Celonis PQL operates as follows: In aggregations: treats NULL values as non-existing and ignores them. In row-wise operations: returns NULL if one of its inputs is NULL.

7.3.3 Applications in the EMS

PQL is the query language to formulate your process questions and calculate KPIs. This is why you can apply it in a multitude of applications in the EMS.

The various applications include:

- Data Explorer

It offers not only the possibility to use a visual editor and switch between visual editor and Code editor, but also visualizations to validate your queries and to make sure that this is exactly the query you need.

- Knowledge Model

The central place for all your records, KPIs, filters, etc. is the Knowledge Model. This is where you can add new PQL queries and save them for future usage across the different assets.

- Analysis

Although your analysis should ideally be connected to your Knowledge Model, you can also write individual PQL queries inside analyses.

- Action Flow

When building Action Flows, sometimes you have to filter on specific subsets in your data or build a logic based on values in your data model.

- Machine Learning Workbench

When working with ML Workbench and the Pycelonis package, you can interact with Celonis objects as native objects, e.g. copy an analysis, pull&push data, reload data models, etc.

7.3.4 Get Data into the EMS

The process of setting up of a data pipeline is divided into 5 subparts, namely:

- i. Data Integration
- ii. Connecting to Systems
- iii. Extracting Data
- iv. Transforming Data
- v. Loading a Data Model

ii. 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.

Without this data, no other activities can take place within the EMS and you won't be able to mine, improve, act on, or automate your processes. Data Integration helps connect systems, extract relevant data, transform it to required needs, and load it into a polished data model.

ii. Connecting to Systems

Without this data, no other activities can take place within the EMS and you won't be able to mine, improve, act on, or automate your processes.

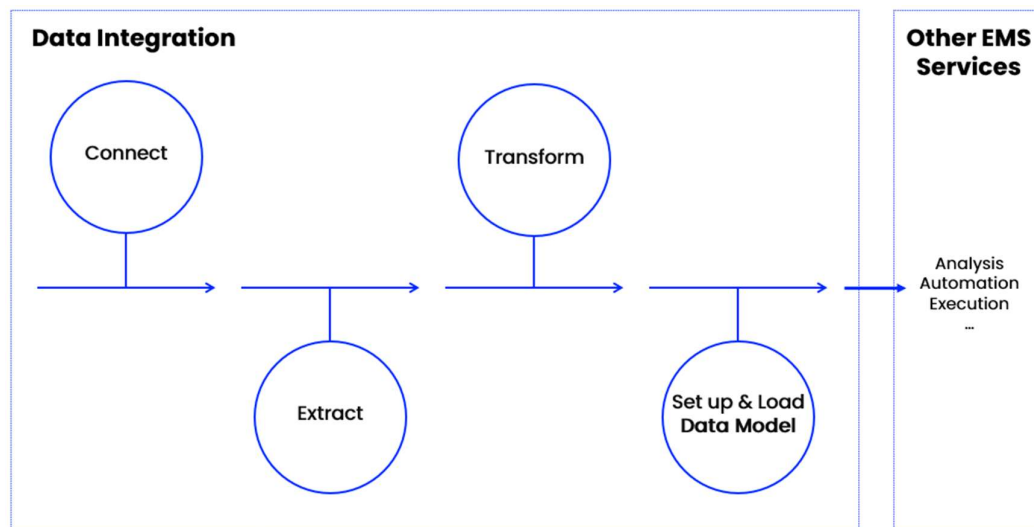


Fig. 7.10: Data Integration

Structuring data in data integration:

Looking at this structure, you should do your best to keep all related processes and systems within one Data Pool. Likewise, you should keep unrelated processes in separate Data Pools. For example, you can create one Data Pool for all your SAP processes and another Data Pool for your IT Service Management processes.

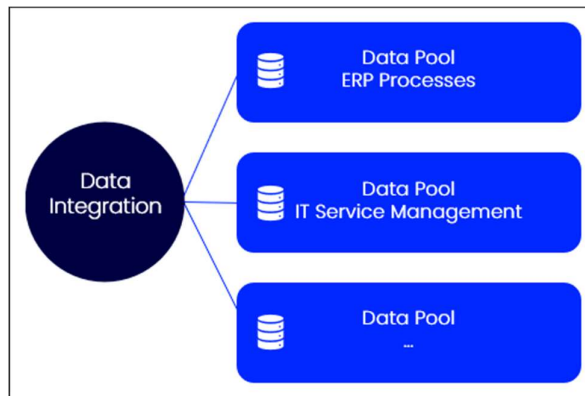


Fig. 7.11: Data Pool - 1

In some cases, it also makes sense to create separate Data Pools for different regional or legal entities to restrict access to the data for certain users or user groups.

iii. Extracting Data

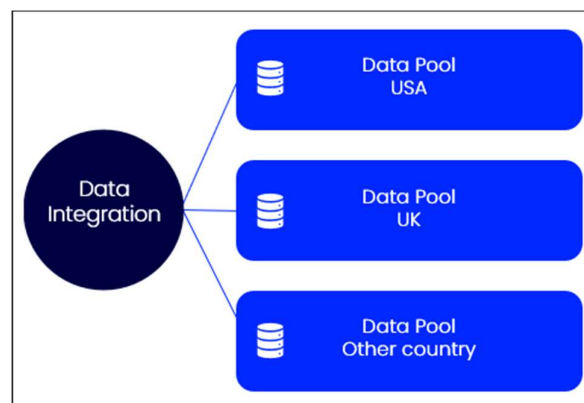


Fig. 7.12: Data Pool - 2

The process of extracting data from the given system is very crucial and can be done using these tools:

- Data Jobs
- Replication Cockpit

Why don't we extract entire databases and make our lives simple?

For simple reasons—entire database extractions would:

- take too long,
- be taxing on source systems,
- take up unnecessary cloud storage,
- and be expensive!

Celonis Data Integration uses Vertica SQL.

iv. Transforming Data

Any process always contains three columns that map them:

- The object ID or case key,
- the process steps or activities that took place for the different case keys
- and the timestamps or event time of each activity

In the Purchase-to-Pay process, the Purchase Order Item Number is the central case key we follow. Every Purchase Order Item goes through different activities such as creating the request, creating the item, receiving goods, and paying the invoice. And every activity has a corresponding event time.

In short, every Purchase Order Item has a unique case key that goes through various activities at different points in time. Together these three columns build the core of your process flow. Every Purchase Order Item goes through different activities such as creating the request, creating the item, receiving goods, and paying the invoice. And every activity has a corresponding event time. This step is very crucial because it

lets us transform the data into our required form and apply the various process mining techniques on the data to get a desired and expected end result.

v. Loading a Data Model

Once your Data Model is set up and you click on "load", the data from the Celonis Data Storage is loaded into a Data Model for the Query Engine. Together the Data Storage and the Query Engine form what is called the "Process Data Engine". This is also what you see in the Data Flow Diagram:

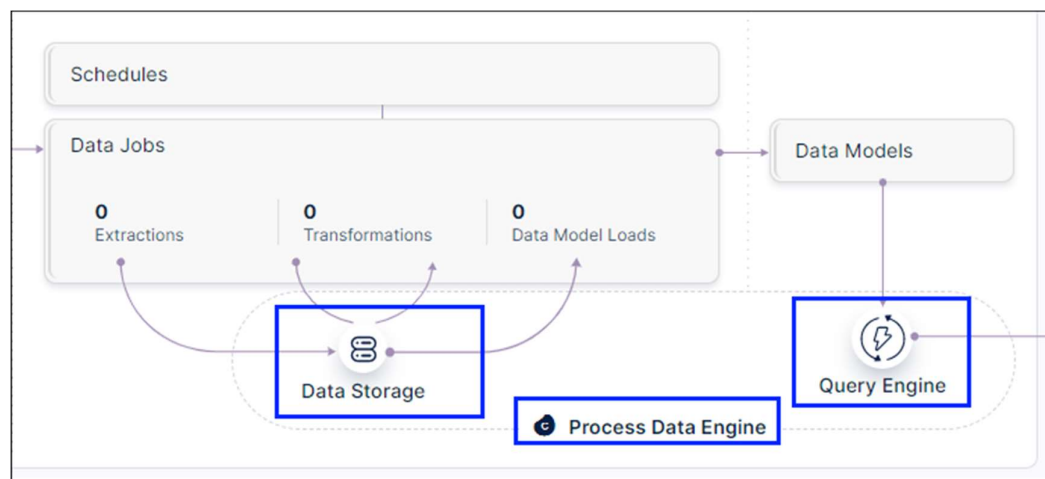


Fig. 7.13: Data Flow Diagram

Once you've defined your Data Model and load it, Celonis pulls the data from Data Storage, transforms it into several parquet files, and using metadata information on your Data Model's schema, stores it into the Query Engine. The Query Engine is an analytical database best suited for Analyses. Aside from normal Data Model load, you can also run partial Data Model loads when loading Data Models using tasks in Data Jobs.

If running your Data Model Load manually as a task in a Data Job, you can always choose which tables to load regardless of whether you set up your Data Model Load task as partial or full. Another important time you can use partial Data Model Loads is with schedules. As with your testing work, the partial load makes sense when

you know only a few tables need to be updated and the rest are unaffected. In this way, a Data Model can be loaded into our Celonis workspace and the required methods can be applied on it.

7.3.5 Troubleshooting

Occasionally, your Data Model may fail or display some warnings. Especially in the early stages, or when making changes, it's a good idea for you to stay aware of the most common issues that can happen here.

Here is a short list of the most common errors:

- Mismatch of cases between Activity Table and Case Table
- No distinct order with timestamp and sorting columns
- Invalid associations between Activity Table and Case Table
- Day-based activities detected

The warnings here reflect a few important points for your Data Model. You need:

- clearly sorted timestamps for your activities
- accurate case keys
- accurate joins between your tables
- clean and matching data in your activity and case table
- no data job running on your tables while you are loading your data model

7.3.6 Permissions

Another important aspect of Data Models are permissions. There are two types:

1. Usage Permissions show who can use the model for downstream activities in the Studio such as an Analyses.
2. Data Permissions show who can see which part of the data in the Data Model.

CHAPTER – 8

APPLICATIONS

Italian management consulting company HSPI publishes a database of process mining projects and case studies annually. The purpose of this project is to create the most complete list of process mining adoptions, increase awareness, and explore business potential. In the 2020 application database, there are 551 case studies from 27 countries around the world, proving the spread of process mining adoption and growth of interest in these techniques.

Some of the many useful applications of Process Mining are:

i. Process Mining and RPA

Robotic process automation or RPA is focused on automating repetitive business processes to increase efficiency. QPR, one of the largest providers of process mining software, claims that process mining can reduce RPA implementation time by 50 percent and RPA project risk by 60 percent. And according to UiPath, one of the leading RPA companies, “78 percent of people who automate say process mining is key to enabling their RPA efforts.”

ii. Financial Services

Because of the rise in transaction volume and the digitization of more industries, aberrant activity is harder to detect using manual methods. Companies in the financial services sector have the chance to continually and thoroughly identify issues within high-volume processes thanks to process mining, which is a solution to the increased regulatory and audit requirements.

For financial institutions, such as banks, insurance companies, or loan associations, the biggest operational priorities are security, accuracy, and speed of transactions. However, just like in any other industry, a lot of organizations suffer from

inefficiencies, bottlenecks, and poor process management. In 2018, Piraeus Bank, a Greek multinational financial services company, encountered problems after adopting RPA to automate the consumer loan process. Since traditional analytical methods proved to be of little value, they implemented QPR Process Analyzer to gain a deeper understanding of their process problems and identify bottlenecks.

iii. Supply Chain

Process Mining offers a lot of optimization opportunities to the complex, multifaceted supply chain industry, including such aspects as manufacturing, warehousing, transportation, inventory management, retail management, etc. Some of the typical processes that are often performed ineffectively and require improvement include:

- Procure to pay
- Order to cash
- Production process
- Warehouse operations
- Accounts payable/accounts receivable management

iv. Healthcare

In healthcare, process mining can be used to analyze patient pathways, identify treatment variations, and optimize hospital processes to enhance patient care and resource utilization. The risks associated with preserving population health and achieving individual patient journey objectives rise as data about patient experiences and results keep growing.

Process mining supports the delivery of effective and high-quality end-to-end patient journeys for healthcare organizations dealing with the exponential growth of data, from before a first doctor appointment through treatment regimens to closed treatment cases.

v. Information Technology:

IT processes can be analysed using Process Mining to improve system deployment, software development and deployment, and help desk operations. Process Mining aids in identifying opportunities for automation, reducing downtime, and optimizing IT service management.

vi. Digital Transformation and Telecommunication:

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

Process Mining also helps analyse customer interactions, billing processes, and network operations. It optimizes service delivery and enhances customer satisfaction thus leading to improved telecommunications.

CHAPTER – 9

LEARNING OUTCOMES

By the completion this internship, one should have successfully

- Got a basic idea about the fundamentals of Process Mining.
- Learnt about Celonis software, various tools it provide for Process Mining and how to use them.
- Achieved a general idea about what are the different Process Mining techniques and when to use them.
- Gained knowledge on the different stages of Process Mining and how to implement/execute each stage.
- Gained the ability to comprehend and implement the PQL for our use in Celonis and to distinguish it from SQL.
- Understood the different process visualization methods and where to use them appropriately.
- Learnt how to use the different explorers in Celonis i.e., Variant Explorer, Process Explorer and Case Explorer.
- Understood the vast applications and the importance of Process Mining.

CONCLUSION

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

The advancement of technology everyday only leads to the development of Process Mining and it may even give rise to new branches of Process Mining which are more domain-specific and have relevant techniques and applications, enhancing the ability of institutes, organizations, establishments or companies to use these techniques and methodologies and get them growing.

However, to use these various Process Mining techniques, one must have a thorough knowledge on how to implement them, what are the errors and challenges that may occur and how to troubleshoot them, modify the techniques according to their needs and so on. Nevertheless, it is not understatement to say that the field of Process Mining has a huge role to play in tomorrow's, or even today's world and thus it is almost mandatory that one in a computer science field must possess at least some basic knowledge about it Process Mining.

INTERNSHIP CERTIFICATE



REFERENCES

1. <https://internship.aicte-india.org/>
2. <https://www.celonis.com>
3. <https://www.analyticssteps.com/blogs/types-and-applications-process-mining>
4. <https://www.ibm.com>