

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

K.Karuna

(214G1A32C9)



**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 Virtual Internship**” is the bonafide work carried out by **K.Karuna** bearing Roll Number **214G1A32C9** 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 2023 to September 2023.>>

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

Celonis is actually a German software company, but it is strongly internationalised, with headquarters in Munich and New York. Founded in 1811 by three university students, Celonis market leading platform (Gartner, 1819) has helped companies achieve process excellence by removing operational friction points with its Intelligent Business Cloud platform.

Celonis has developed software that does process mining. Process mining is a technology that allows processes to be analyzed based on data. Processes such as deliveries, productions, or even orders can be recorded and evaluated. The aim behind this is to uncover inefficiencies and uncover potential areas for improvement. Especially in huge companies with cross-departmental processes, it is almost impossible to know, let alone understand, the entire process at a certain point.

Celonis software is called an "Execution Management System". In concrete terms, this means that digital processes are automatically recorded and mapped. This is to create a completely objective and data-driven picture of the company's internal processes. The data for this is usually already available in software such as SAP.

Celonis offers a comprehensive suite of tools that enable businesses to: Discover processes, Diagnose Issues, Optimize Performance, Monitor in Real-Time, Predict and Simulate.

"Process mining delivers insights that enable faster, smarter decisions and stronger performance on an organization's most critical priorities."

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.

K.Karuna
214G1A32C9

Contents	Page No
List of Figures	vi
List of Abbreviations	vii
Chapter 1 : Introduction	1
Chapter 2 : Process Mining Fundamentals	2
Chapter 3 : Review & Interpret Analyses – I Variant Explorer, Process Explorer Charts & Tables, Review KPIs	3-4
Chapter 4 : Review & Interpret Analyses – II Selection views, Case Explorer Conformance Checker Save & Share Analysis Selection	5-6
Chapter 5 : Build Analyses Analysis Sheet, Configure Tables, Charts & KPIs, Background Filters	7-8
Chapter 6 : Basic Queries in PQL Celonis PQL Engine, Basic PQL Queries, Standard Aggregation Functions, Data Flow Functions, String Modifications	9-10
Chapter 7 : Joining & Aggregating Data Joins & Filters PU_Functions	11-12
Chapter 8 : Get Data Into EMS - I Data Pipeline Data Integration in EMS	13-14
Chapter 9 : Get Data Into EMS - II Schedule Data Jobs Troubleshoot & Monitor Extract Necessary Data Custom Processes	15-16
Chapter 10 : Conclusion	17
Internship Certificate	18
References	19

List of Figures

Fig.No	Description	Page No
3.1	Analysis Sheet	3
4.1	Selection Views	5
5.1	New Analysis Sheet	7
6.1	Celonis PQL Engine	9

LIST OF ABBREVIATIONS

AICTE	All India Council for Technical Education
AI	Artificial Intelligence
KPI	Key Performance Indicators
EMS	Execution Management System
PQL	Process Query Language
PU	Pull-Up Functions

Chapter 1

Introduction

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

According to MarketWatch ([opens in a new tab](#)), Global Process Mining Software Market is valued approximately at USD 322.02 Million in 2018 and is anticipated to grow more than 50.1% by 2027. The strong development of Celonis - from a student startup to a company with over 3000 employees in 2022 and a customer base of the biggest enterprises like Coca-Cola, Unilever, Vodafone or Uber - paints the same picture. Process Mining is in high demand, which is further backed up by current hypes around automation and other performance acceleration measures (cf. Gartner 2018).

Process Mining is achieved by taking the digital footprints that are created in IT systems and using them to reconstruct and visualize process flows. From here, Process Mining technology can identify patterns and deviations and ultimately eliminate bottlenecks. Process mining is a technique to analyze and track processes. In traditional business process management, it is done with process workshops and interviews, which results in an idealized picture of a process. Process mining, however, uses existing data available in corporate information systems and automatically displays the real process.

Chapter 2

Process Mining Fundamentals

Process mining is an analytical discipline for discovering, monitoring, and improving processes as they actually are and not as you think they might be. Process Mining works by extracting knowledge from event logs (also called digital footprints) readily available in today's information systems, in order 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. The Review and Interpret Analyses (opens in a new tab) training track is designed for data and business analysts, process experts, and process improvement specialists. . In traditional business process management, it is done with process workshops and interviews, which results in an idealized picture of a process. Process mining, however, uses existing data available in corporate information systems and automatically displays the real process. Process Mining helps to build the analyses by using the available tools like variant explorer, process explorer, charts and tables, selection views. It provides Analysis Sheets where the user can extract useful data. 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.

The basic terms in this course includes:

- ✓ **Process** is a series of linked steps taken in order to achieve a particular goal.
- ✓ **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** 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.

Chapter 3

Review & Interpret Analyses - I

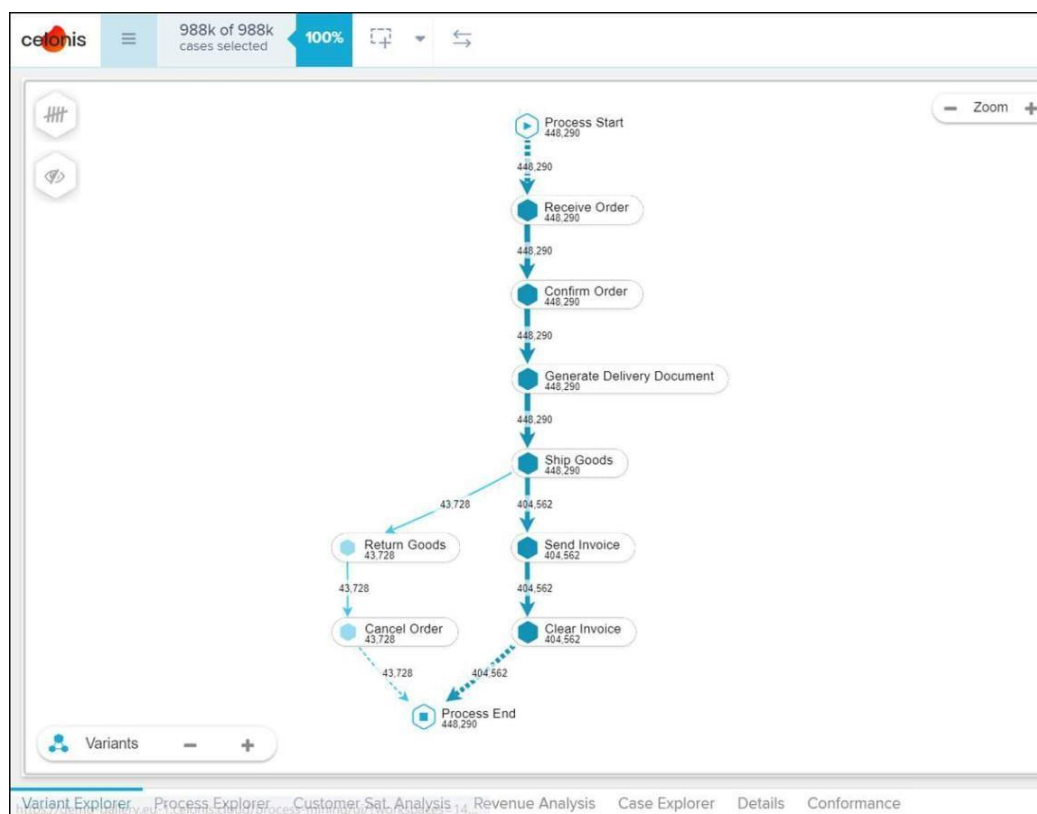
Variant Explorer

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 in actual for the provided series of steps i.e a process.

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.

Fig 3.1 Analysis Sheet



Clicking on an activity in the Process Explorer, you can see a list of the predecessor and successor activities.

Process Explorer is the List view of activities and connections. This way, you can quickly scroll through a list of the most common activities and connections. And from there, you can even reveal the predecessor and successor activities. With both the Variant and Process Explorers, you got a good sense of what it's like to take an exploratory approach to understand your process; to detect inefficiencies and to raise your first analysis questions.

Charts, Tables & KPIs

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.

KPIs may also appear as standalone numbers as seen below with the On-Time Delivery and Net Promoter Score (NPS) examples; we call these analysis components "Single KPIs." Represented by a number, a Key Performance Indicator (KPI) allows you to quickly assess how your process is performing. Other times, you want to look at the KPI over time or in combination with process dimensions. That's where charts and tables come in. Charts and tables are critical components of analyses. They help us understand the process and are the go-to tools to drill down on root causes of process inefficiencies. In charts, you can select one or more of the displayed timeframes to narrow the scope of the analysis.

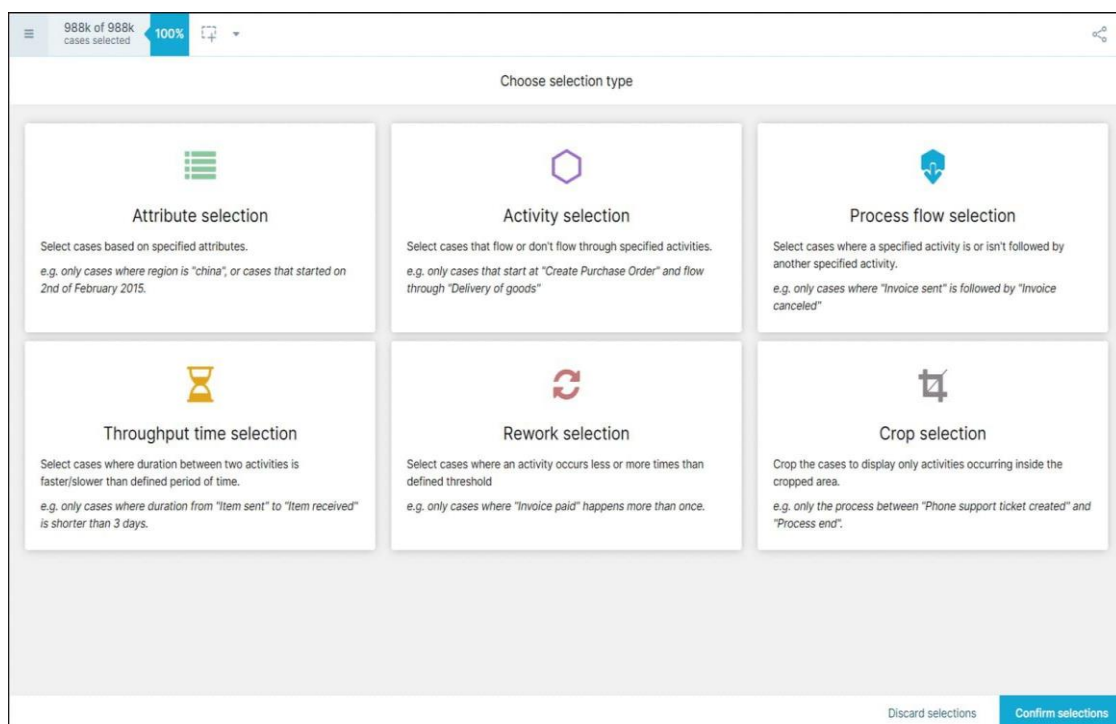
Chapter 4

Review & Interpret Analyses - II

Selection Views

Selection Views offer a more comprehensive set of options to filter on cases as compared to what you can do in the components in analysis sheets. You can access the six Selection Views from anywhere in the analysis by clicking on the Selection Views button located in the analysis toolbar. We use these views out of six available views to help us to analyze the data more accurately and precisely.

Fig 4.1 Selection Views



Case Explorer

The Case Explorer, Process Mining tool available is useful once you've narrowed down the analysis to a few cases that you want to investigate further. Case Explorer allows us to 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.

Conformance Checker

You can use the Conformance checker to:

Get perspective on the percentage of cases that conform to the idea flow of the list activities specified in the Analysis (target process model). Regardless of the process, Conformance Check takes into account two types of criteria in the Violations list: Activities not reflected in the process model, Order of activities not reflected in the process model.

Get automated insights into potential root causes of inefficiency by reviewing the most common process violations and the attributes they're associated with.

Mark certain process violations as acceptable to include them in the conforming cases statistics. The Conformance checker evaluates each case against the process model your organization has specified to determine whether it conforms to it or not.

Conformance Check takes into account two types of criteria in the Violations list:

- ✓ Activities not reflected in the process model
- ✓ Order of activities not reflected in the process model

Save & Share Analysis Selection

In Celonis Analysis, you can export data and even the process visualization, if enabled by the person building the analysis for the given set of data by the users.

Right-click on the component to see your options as they differ depending on the component. For example, in the Variant Explorer, you may see options to export the variant(s) as a PNG image (visual graphic) or an XML file. You might also see the option to save either format to your Process Repository. This is one of the ways by which you can define your to-be process to get automatic insights in the Conformance Checker.

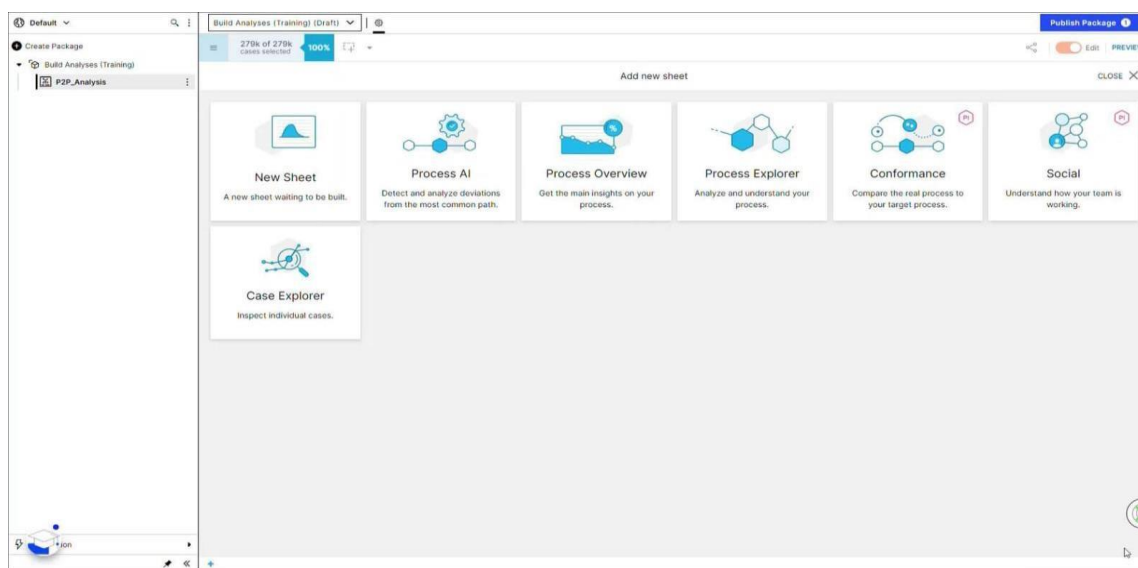
Chapter 5

Build Analyses

Analysis Sheet

You'll build analyses in Studio and users will view the published analyses in Apps.

Fig 5.1 New Analysis Sheet



Configure Tables, Charts & KPIs

Three data tables we used to configure

- ✓ OLAP Tables
- ✓ Column and Line Charts,
- ✓ Pie Charts.

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. Sometimes you might save yourself time by copying a component and updating the copy instead of creating one from scratch. You can just copy and paste the component in the analysis sheet. Standard Process KPIs are generic KPIs that apply across any process you analyze. They are accessible in all Celonis Analyses, regardless of process, in the KPIs selection window. Anytime you add a table or chart to the analysis, you'll need to select the dimension(s) and KPI(s) to display.

You can customize it to calculate the throughput time between specific activities in the process shown in analyses sheet.

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. Aside from the Number, you might choose other Single KPI components such as Gauge, Fill, and Radial, depending on what you need to display. We can add the Number (Single KPI component) as well as Dropdown and Date Picker (Selection components). You also saw that you can enhance the user interface of the analysis with Design components such as Text with KPI, and Line, Logo, and Image. Standard Process KPIs that are applicable to all processes and therefore available as a standard set of KPIs regardless of the process. We saw how after selecting KPIs such as "Total throughput time in days" and "Ratio of cases with a certain process flow," you could customize them using the Visual Editor. table will contain at minimum three columns: Case ID, Activity, and Event Time. Then, we took it up a notch with the Visual Editor. Using the Activity Table in our demo analysis data, you were able to build custom KPIs such as automation rate in the Visual Editor.

Background Filters

It's true that analysis users can apply filters when interacting with components and one of six Selection Views. But as the analysis builder, you might want to set background filters that are applied before the end user begins their analysis. For example, you might want to remove certain data from the Analysis due to privacy reasons. End users cannot remove these background filters.

Background filters can be applied at three levels:

- ✓ A component (such as a Process Explorer or an OLAP Table)
- ✓ A sheet
- ✓ The entire Analysis

To apply static background filters, we use PQL (Process Query Language). PQL is a Celonis proprietary language. The syntax of PQL is very close to the SQL standard. PQL is simpler and specialized for process-related queries. We can add a dropdown to OLAP tables, such that the displayed dimension updates according to the selection in the dropdown.

Chapter 6

Basic Queries in PQL

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.

Celonis PQL follows four language features:

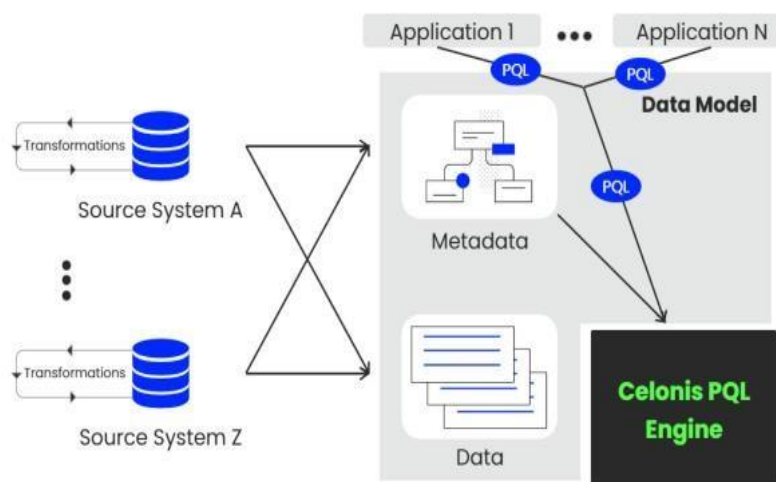
First, operators usually create and return a **single column** that is either added to an existing table (e.g., the case or activity table) or to a new, temporary result table.

Second, the **supported data types** comprise STRING, INT, FLOAT, and DATE.

Third, **Boolean values** are not directly supported, but can be represented as integers.

Fourth, each data type can hold **NULL values**. Celonis PQL operates

Fig 6.1 Celonis PQL Engine



Basic 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. The Data Explorer is an excellent tool that allows you to easily validate your data and build your first PQL queries.

Standard Aggregation Functions:

Besides counting, distinct counting, average and sum, many more standard aggregations such as summary statistics (min, max, median, quartiles, standard deviation). 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.

Data Flow Functions:

Data Flow operators cover a variety of functions to return or change particular values based on conditions defined by the user. **Case_when**: Evaluates a list of conditions and returns results based on these conditions. **Coalesce**: Returns the first input value which is not NULL in a list of values. **Remap_values**: Maps values of a column to other values.

String Modifications contain a set of combinations, transformations and cropping to get the most out of your Celonis queries as **lower**, **upper**, **reverse**, **ltrim**, **rtrim**, **substring**. **Data Type Conversions**: Different types (integer, float, string or date) **Regex** or **Regular Expressions** are a way to search within a text according to a particular search pattern. **Regular Expressions** are built from a sequence of characters that define this search pattern. The **PQL Function Library** is the complete and most detailed collection of articles about all **Process Query Language** functions available in Celonis. **Comparison Operators** are important and just as easy as the **BETWEEN** operator when working in **PQL** the operator **IN** maps the value of the current case to either true or false which determines whether the condition is fulfilled or not. Instead of using single values in the match list of your statement, you can also list another column here. **IN** and **NOT IN** can be used for the following value types: string, int, float, date. **Date Time rounding functions** which are very straightforward as well. You can think of those functions like they “cut off” everything after the Date Time element corresponding with your chosen function.

Chapter 7

Joining & Aggregating Data

Joins & Filters

The set of tables containing all the data is called the **Data Model**.

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

A PQL query is executed in the following order:

Joins and regular PQL functions (not aggregations). The common table is defined after joining the required tables. Filters are applied (if there are filters defined). Standard Aggregations (AVG, COUNT, SUM, etc.).

Filters:

Center the attention on a given subset of data in order to better understand the information or view it from different angles. In PQL, the filter statement allows you to remove some rows of the result table of a query, according to a specified condition. Only rows that meet the condition remain, while the rest are discarded.

The syntax is very simple. You start with the keyword **FILTER**, followed by the condition. An easy condition starts with the table and column you'd like to apply the filter to, followed by the predicate operator, which could be **BETWEEN**, **!=**, **<**, **>=**, **<=**, **=**, or many others. To complete the condition, you add the respective value that you would like to filter on. Finally, the **FILTER** statement is terminated with a semicolon. We use PQL (Process Query Language). PQL is a Celonis proprietary language. The syntax of PQL is very close to the SQL standard. PQL is simpler and specialized for process-related queries.

PU_ Functions

PU-functions can be applied in many use cases and are often the only way to calculate a certain KPI or dimension.

- **PU_SUM**
- **PU_AVG**
- **PU_COUNT, PU_COUNT_DISTINCT**
- **PU_MAX, PU_MIN**
- **PU_MEDIAN**
- **PU_QUANTILE**
- **PU_FIRST, PU_LAST**
- **PU_STRING_AGG**

The first 8 functions should be quite intuitive to understand. The syntax starts always the same: `PU_X(target_table, source_table.column)` with X being the aggregation you want to perform. In addition, there are 3 PU-functions that don't correspond to a regular aggregation: `PU_FIRST`, `PU_LAST` and `PU_STRING_AGG`.

PU-functions are key when working across multiple tables that are connected with a 1:N relationship. One of the advantages over standard aggregations is that PU-functions can be used inside filters, whereas standard aggregations can't be used in this context. PU-functions are one of the best possibilities for solving the **“no common table”** error. In some cases it is not possible to solve the problem directly with a single PU-function. As seen throughout this course, it is possible to aggregate from the N to the one side only. If we have a case where we need to use tables that are on the one-side, so a 1:N:1 relationship, we need to use the command `BIND`.

Chapter 8

Get Data Into EMS - I

Data Pipeline

It helps you connect to source systems, extract the relevant data, transform it to your needs, and load it into a polished Data Model. The four major steps include:

Validation, Scheduling, Monitoring, Optimizing Performance

Celonis offers real-time extraction and transformation capabilities. In simple terms, real-time means the EMS can track and frequently replicate incremental changes in data from source systems. The main objective behind real-time pipelines is for users to operationally act on the data. To speed up your work in Data Integration, Celonis provides "Process Connectors" for the most common processes and systems. Process Connectors contain templates and scripts that support you in the connect, extract, transform, load, and scheduling steps of building your data pipeline. When building a data pipeline, Celonis offers real-time extraction and transformation capabilities. In simple terms, real-time means the EMS can track and frequently replicate incremental changes in data from source systems (e.g. SAP, Salesforce, Databases, Azure Event Hub, Azure Service Bus). This way, new data comes in every few minutes—depending on how often you want the EMS to check for changes.

Data Integration in EMS

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. Data Integration is where you set up connections and your data pipeline. The main ways you can bring data are:

Process Connectors:

Process Connectors are the fastest and most common way for you to connect and cover the most established process use cases.

Extractors (Data Connections): Extractors are blank data connections with no reference to a process. With Extractors, you connect to source systems and then have to build your data pipeline from scratch—i.e. your extractions, transformations, and Data Model.

Extractor Builder:

This tool helps you quickly build an Extractor using REST API and supports the OData V2 and V4 standards as well. It allows you to quickly connect to source systems and extract the data you need by identifying the right tables, filters, and joins in your extraction.

File Uploads:

When connecting to a new data source to upload data directly to the EMS. For additional static data, for a list of translation terms, for a list of permissions outside of your source systems, for historical data that you know will never change, for one-time uploads.

Data Push API:

The Data Push API uses micro batching (data chunks) to provide a scalable and reliable integration point and allows Parquet and CSV file formations. The Data Push API is a set of sequential API calls to create, execute and monitor "Push Jobs"

Celoxtractor:

The Celoxtractor is a Python package designed to let you develop your own EMS Extractor easily. It gives you: complete control over your data, feature parity to native Celonis extractors, and full flexibility in adjusting all aspects of your extractions.

Once you've established a connection, and set up your extractions, your next step is to transform the data with one or more of these tools: Data jobs, Replication Cockpit. The Replication Cockpit can currently handle full and delta loads for extractions and transformations. Prioritize it if a real-time connection is possible as it supports operational use cases—i.e. acting on data in close to real-time. After extracting raw data with a Data Job, you need to transform it to get it ready for process mining. 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. Without it, no process mining or other activities are possible. Building the Activity Table is also often referred to as consolidating the digital footprint.

Chapter 9

Get Data Into EMS - II

Schedule Data Jobs

Extractions and Transformations and Data Model Loads in a continuous manner.

For a continuous and automated data load you can use the scheduling functionality.

Schedules allow you to sequentially execute Data Jobs on a regular basis.

Troubleshoot & Monitor

It's important for you to know the ins and outs of troubleshooting your Data Pipeline from the moment you connect to source systems to when you set up your Data job schedules or replications using the Replication Cockpit.

If your Data Pipeline is set up end to end, then it is theoretically time for you to sit back and enjoy the beauty of your work. That said, you should make sure to keep an eye on things to ensure everything keeps running smoothly.

Extract Necessary Data

An optimized data pipeline extracts and transforms only necessary data. Skipping an optimization of your extractions can have very detrimental effects on your data pipeline. It leads to: Higher storage needs (affects your license's APC) .Heavier consequences of bad practices in transformations (e.g. SELECT *) . Negative performance impact on load times in extractions, transformations, and data model loads. For each query submitted to Vertica, the Vertica query optimizer assembles a query execution plan—a sequence of steps and required operations to access data and calculate the result. Prioritize it if a real-time connection is possible as it supports operational use cases—i.e. acting on data in close to real-time. After extracting raw data with a Data Job, you need to transform it to get it ready for process mining. The goal is to transform your data so it can become a Process Data Model. Table statistics are analytical summaries of tables that assist the query optimizer in making better decisions. Table statistics significantly improve query performance, often reducing the query execution time by over 50%.

Custom Processes

A unique identifier makes it possible to distinguish between different cases and correctly assign the activities and timestamps. In the Event Log, we call such a unique identifier a “Case ID.” Single process can be parallelly executed in multiple systems or sequentially executed in multiple systems.

Four important points to consider when defining data requirements.

Activity Data - Used as the main ingredient for process mining to generate an event log.

Dimensions - Show the process/metrics for specific attributes such as vendors and product categories.

Key Metrics - Allow us to align on the most important calculations prior to data extraction.

Translation & Name Mappings - Convert certain technical terms into meaningful text fields. To explore you data with the Data Explorer, you simply need to: extract all potentially relevant tables,load all extracted tables into a Data Model,connect your Data Model to a Knowledge Model in the Studio,create an instance of the Data Explorer,start exploring.

With our Project, we'd like to improve sales productivity, accelerate delivery fulfillment & quality, optimize working capital. In many systems such as SAP, there are translation and name mapping tables that help us translate certain code fields into meaningful text. These translation tables serve as a lookup table, which is simply used to look for a related value from some other table. Common translations are reason codes, customer IDs, company codes, and document types. It's vital to understand if such tables exist in your data set because they will allow us to build analysis, processes, and dimensions that are easy to understand for our users. Finally ,run a **quality assurance check** on your Data Pipeline before it goes live. The checklist is based on project best practices and currently used in implementations.

Chapter 10

Conclusion

Process Mining is achieved by taking the digital footprints that are created in IT systems and using them to reconstruct and visualize process flows. From here, Process Mining technology can identify patterns and deviations and ultimately eliminate bottlenecks.

Process mining techniques have been used to improve process flows across a wide variety of industries.

- ✓ **Education:** Process mining can help identify effective course curriculums by monitoring and evaluating student performance and behaviors, such as how much time a student spends viewing class materials.
- ✓ **Finance:** Financial institutions have used process mining software to improve inter-organizational processes, audit accounts, increase income, and broaden its customer base.
- ✓ **Public works:** Process mining has been used to streamline the invoice process for public works projects, which involve various stakeholders, such as construction companies, cleaning businesses, and environmental bureaus.
- ✓ **Software Development:** Since engineering processes are typically disorganized, process mining can help to identify a clearly documented process. It can also help IT administrators monitor the process, allowing them to verify that the system is running as expected.

Internship Certificate



References

- [1] <https://academy.celonis.com/learn/dashboard>
- [2] <https://academy.celonis.com/pages/academic-rising-star-technical>
- [3] <https://academy.celonis.com/learn/learning-path/process-mining-fundamentals-for-students>
- [4] <https://academy.celonis.com/learn/learning-path/get-data-into-the-ems-1>