IBM Cloud Pak for Business Automation Demos and Labs Fall 2021

IBM Process Mining

Use Process Mining to Create and Explore Process Models

V 3.5

Paul Pacholski
pacholsk@ca.ibm.com

NOTICES

This information was developed for products and services offered in the USA.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
United States of America

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

TRADEMARKS

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

IT Infrastructure Library is a Registered Trade Mark of AXELOS Limited.

ITIL is a Registered Trade Mark of AXELOS Limited.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

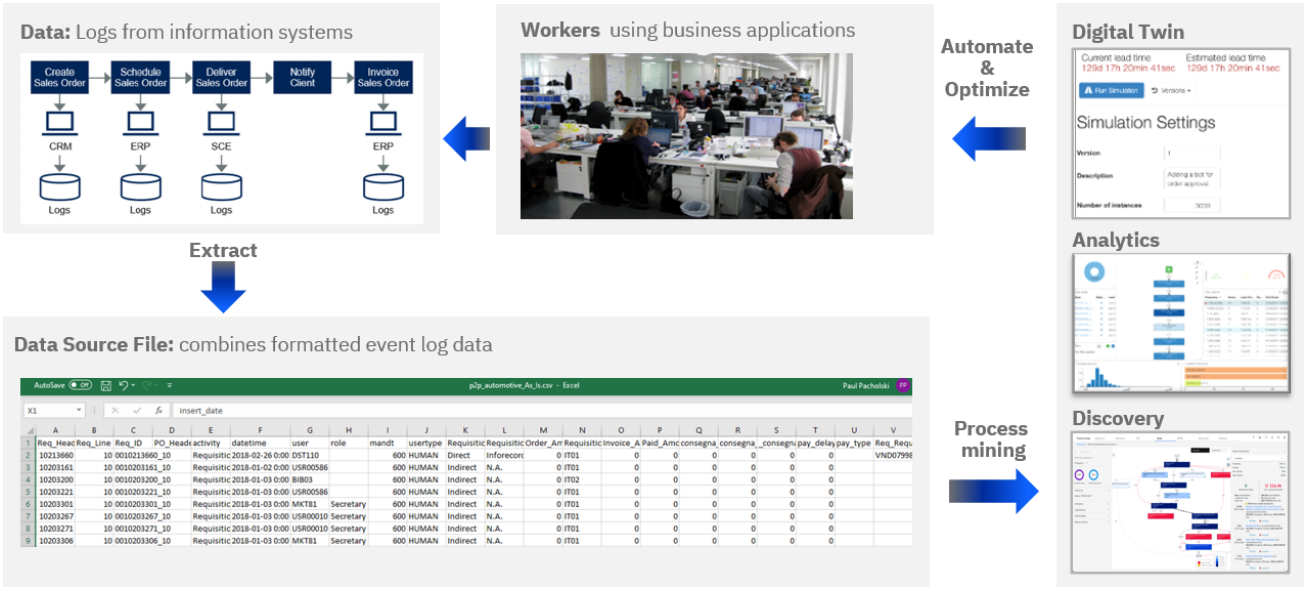
UNIX is a registered trademark of The Open Group in the United States and other countries.

Table of Contents

1 Introduction

1.1 Process Mining

Process mining is a family of techniques in the field of process management that support the analysis of real business processes based on event logs. During process mining, specialized data mining algorithms are applied to identify trends, patterns, and details contained in event logs recorded by an information system. Process mining aims to improve process efficiency and understanding of processes.



1.2 Lab Objectives

In this lab you will learn the basics of IBM Process Mining tools. Specifically you will:

- learn how to create a process from mined process data,
- get a high-level introduction of key process mining features and their value.

2 Lab Setup

2.1 Import Lab Files

_2. Download the following file, you will them in this lab:

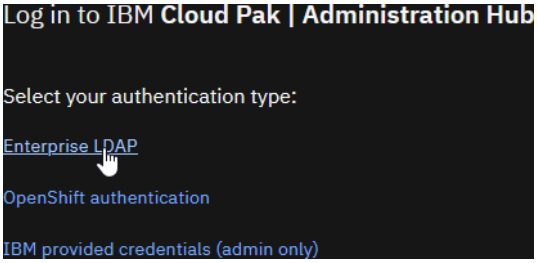
File	Link
Hands_On_tutorial_exercise.zip	https://ibm.box.com/v/PM-LAB-1-DATASET
Order Processing_2021-10-28_112716.idp	https://ibm.box.com/v/PM-LAB-1-IDPFILE

2.2 Open IBM Process Mining Application

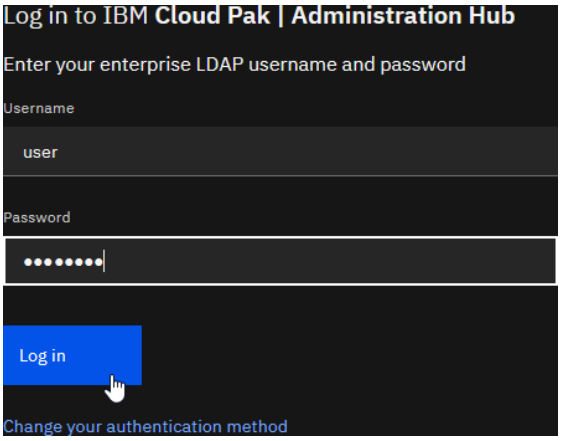
_3. If you are performing this lab as a part of an IBM event, access the document that lists the available systems and URLs along with login instructions. For this lab, you will need to access **IBM Business Automation Studio**.

_1. Start your browser and use the **IBM Business Automation Studio** link

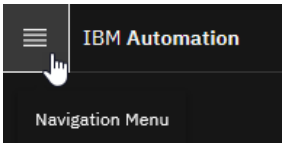
_2. Click **Enterprise LDAP**



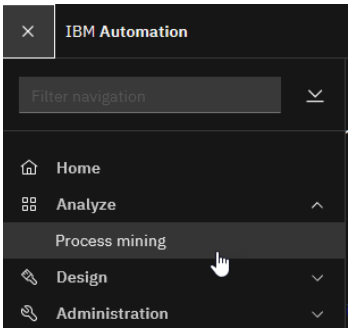
_3. Enter your username and password and then click **Log in**



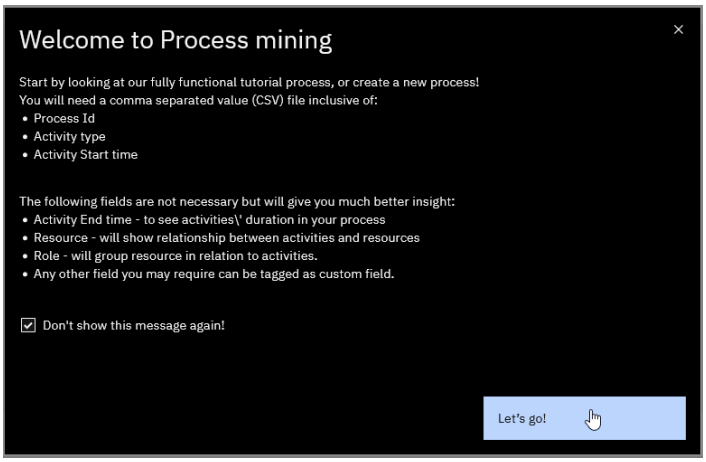
_4. Click **Navigation Menu**



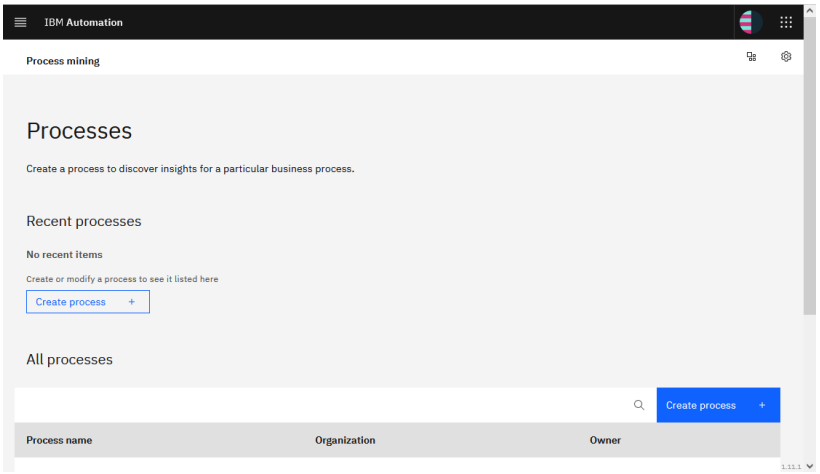
_5. Select **Process mining**



_6. If you see the Welcome window, check **Don't show this message again** check box and click **Let's go!**



You should now see IBM Process Mining web UI



_7.

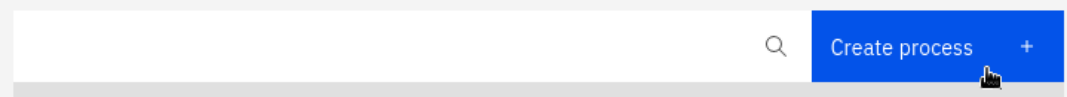
3 Lab Instructions

3.1 Create a New Process

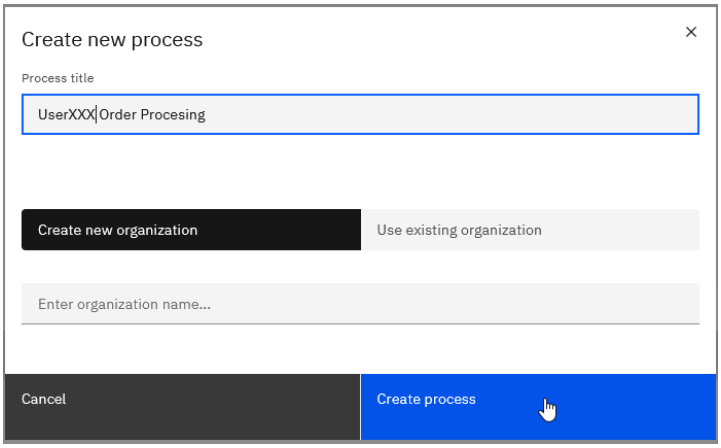
When you open a workspace you can create a Process and Organization or a New Process in an Exiting Organization (in our case you should see, and an organization called Tutorials).

> Note: if you are using a shared environment, when creating new process please use your user name prefix in the process name.

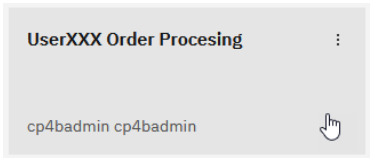
_1. Click in **Create process +**



_2. For *Process Title* enter **<Your User id> Order Processing** and click **Create process**



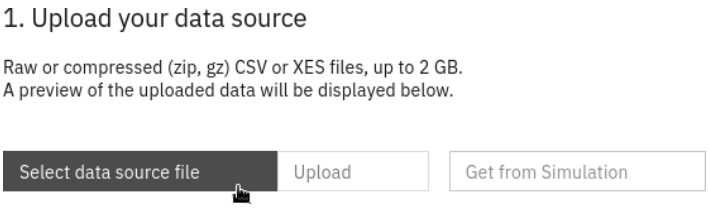
_3. Click **<Your User id> Order Processing**



3.1.1 Upload Process Data

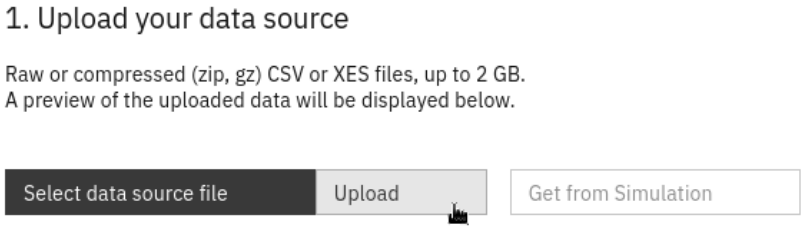
To analyze the process, you will need to upload a log file (.CSV or .XES) containing mined process data into the Data Source.

_1. Click **Select data source file** to upload a CSV data which was captured from an existing Oder Processing process



_2. Select **Hands_On_tutorial_excercise.zip** then click **Open**

_3. Click **Upload**



3.1.2 Map Data Columns

After uploading the log file, you will need to identify data columns in your log file and map them to fields used by process mining algorithms:

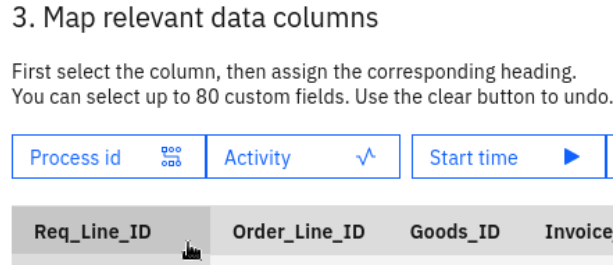
- System Data : Process ID, Activity, Start time, End Time, Resource and Role
- Business Data: to map Custom Fields

To be able to visualize your process, it is mandatory to map:

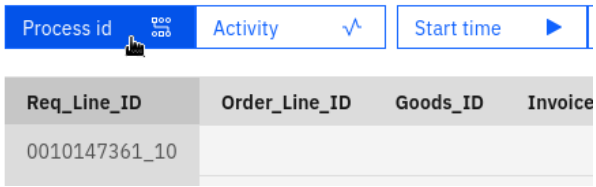
- at least one process id,
- the activity field and
- a datetime field (as start time)

3.1.2.1 Map Columns to Process ID

_1. Select **Req_Line_ID** column



_2. Click **Process ID** button to complete the mapping



_3. Note the icon and number 1 appearing on the data column heading to indicate that mapping is now in effect.

_4. Repeat the above two steps to map **Order_Line_ID**, **Goods_ID** and **Invoice_ID** columns to **Process ID**.

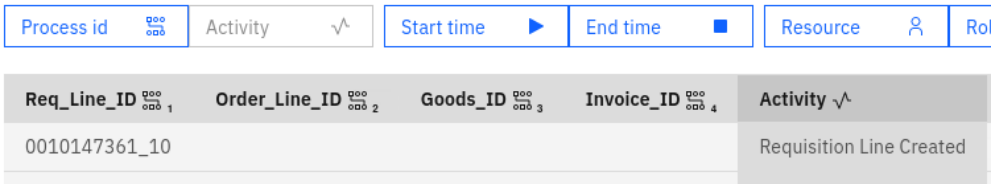
_5. Your columns headings should look like this:

3.1.2.2 Map Column to Activity

_1. Select **Activity** column

_2. Click **Activity** button to complete the mapping

_3. You should now see the activity icon on the Activity column



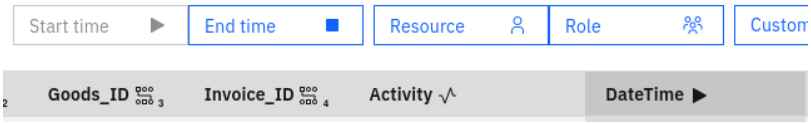
3.1.2.3 Map Column to Start time

_1. Select **DateTime** column

_2. Click **Start time** button to complete the mapping

_3. On Field mapping click **OK**

_4. You should now see the right arrow icon on the DateTime column



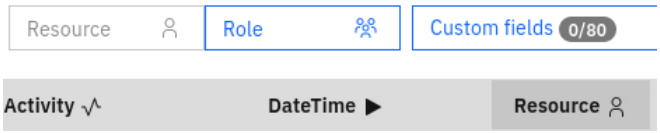
3.1.2.4 Map Column to Resource

_1. Select **Resource** column

_2. Click **Resource** button to complete the mapping

_3. On Field mapping click **OK**

_4. You should now see the person icon on the Resource column

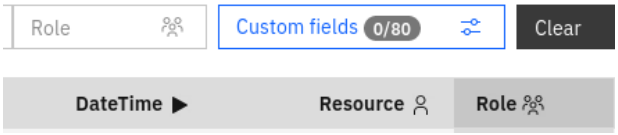


3.1.2.5 Map Column to Role

_1. Select **Role** column

_2. Click **Role** button to complete the mapping

_3. You should now see the people icon on the Role column



3.1.2.6 Map data columns to custom fields

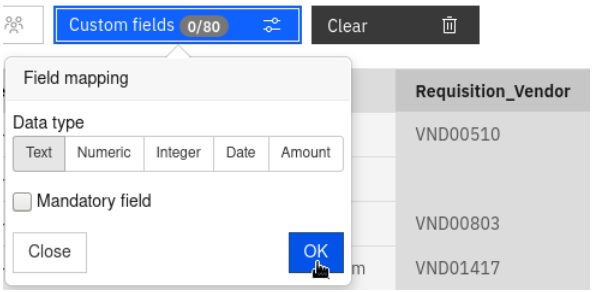
Custom Fields include additional relevant process data. They are commonly referred to as the business data. When you map a Custom field, you must specify the type of data contained in the respective column and whether the field is mandatory for every event (every line of the log file).

3.1.2.7 Map Requistion_Vendor Column to Custom Field

_1. Select Requisition_Vendor column

_2. Click **Custom Fields** button to complete the mapping

_3. On *Filed mapping* select *Data type* of **Text** and keep *Mandatory field* **un-selected**, then click **OK**



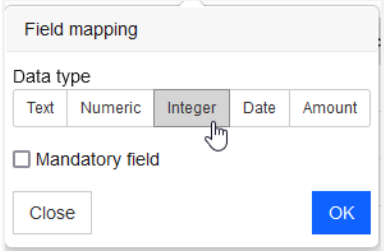
_4. You should now see the *Custom Field tag* icon on the *Requisition_Vendor* column.

3.1.2.8 Map other columns to Custom Field

_1. Repeat the above steps to map the following columns as Custom Fields:

- Requisition_Type
- Requisition_Header
- UserType

Note: when mapping business data columns as Custom fields make sure you match the data types. For example if a custom field contains integers select Data type of Integer

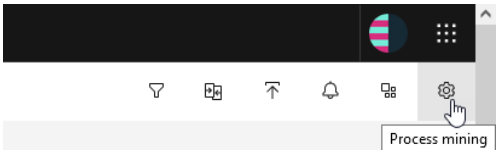


3.1.3 Import Project Settings

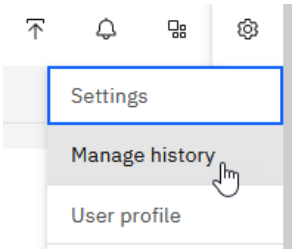
You will now import project settings that include the reference model.

Note: The reference model of a process describes its expected standard behavior, in terms of activities and workflow. The reference model is usually designed in BPMN language by a process owner and can be imported to a process Mining project.

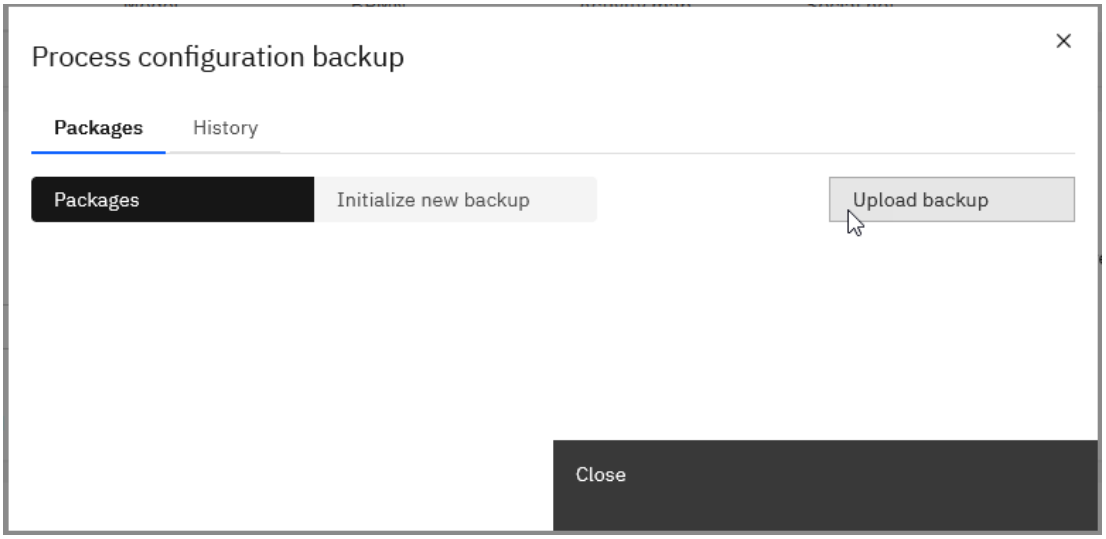
_1. Click **Process mining** icon on the toolbar



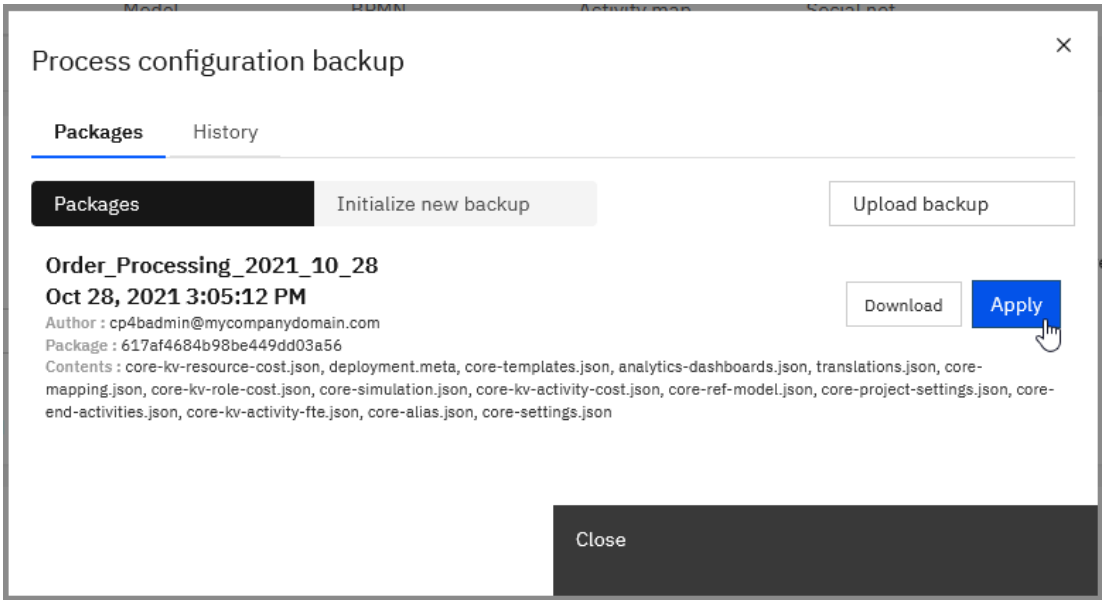
_2. Select **Manage history**



_3. Select **Upload backup**

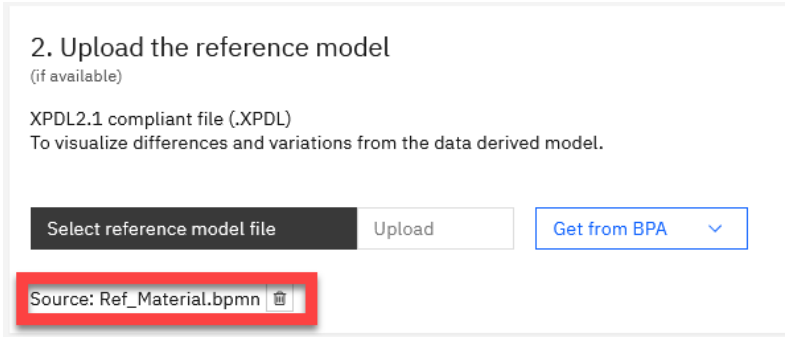


_4. Select **Order Processing_2021-10-28_112716.idp** file and click **Apply**



_5. Click **Close**

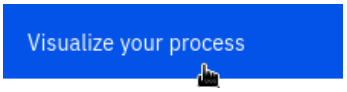
You should now see the reference model



3.1.4 Create Process Model

After mapping the log file, you can visualize the process by creating process Model. Every time something changes in the Workspace or more data is added you will need to recreate the process Model.

_1. To create or update your visualization click in **Visualize your process**



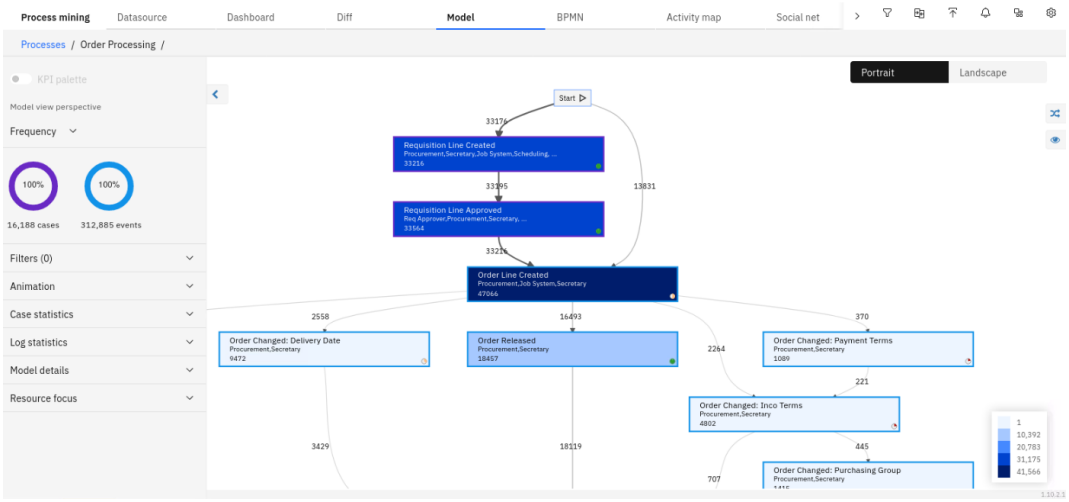
You should now see

3.2 Explore the Process Model

We will now explore the major process mining information derived from the process data: Model, Dashboard, BPMN, Activity map and Social net.

3.2.1 Explore Model View

You should now see the process Model displayed.

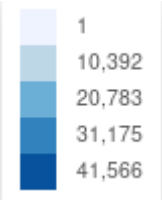


The Model automatically displays the frequency analysis. The dark blue color highlights the most frequent activities, whilst the bold arrows highlight the most frequent transitions. In this way, the most frequent paths between activities of the process can be identified.

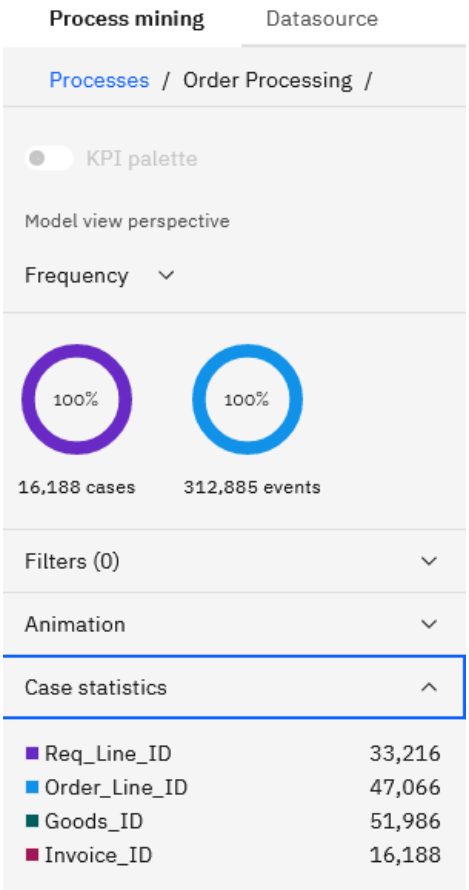
- The numbers next to the lines shows how many times that specific process flow has been followed.
- The numbers within the rectangles shows the number of times that the activity is performed
- The description in the rectangles indicates the name of the activity and the roles by which the activity is carried out. They could be more than one role (multiple roles followed by dots are displayed).

- The green circle at the bottom right corner of the activity rectangle indicates the Model coverage (100% indicates that the Model details cover all the possible relationships of that activity. The percentage indicates how many possible relationships you are currently visualizing. The level of relations is adjustable)

- The color saturation of Activity reflects how often an activity was invoked (the frequency). The legend gives you the frequency coloring detail



- The Activity border reflects the multilevel nature of the process. See the *Case statistics* legend to decipher the color schema

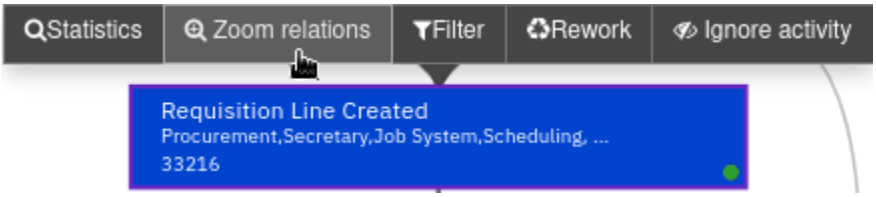


3.2.1.1 Explore Model View – Control View Complexity

You can control how much data is used in the process model

Note that, by default, the visualization does not show all the relationships and activity instances. This is to reduce unnecessary complexity that can impair visualizing and exploring the process. There are two ways to get a more detailed view: Zoom Relations and Model Details.

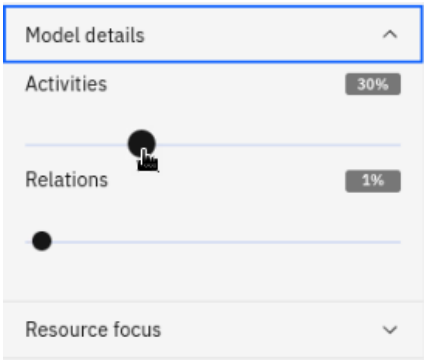
_1. You can select **Zoom relations** after you click on an **Activity**



_2. Click **X** to close Full relationship of Requisition Line Created window.

You can also control % of Activities (occurrences) and % of Relations by expanding Model details section. Selecting 100% for each setting will result in a “spaghetti” style visualization!

_3. Navigate to **Model details** and change **Activities** and **Relations** sliders.

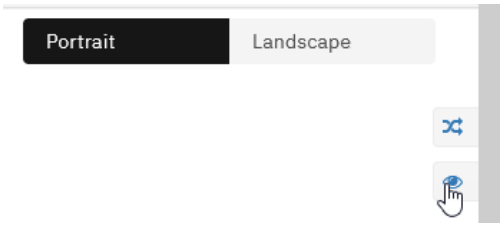


_4. Revert back to **Activities: 30%** and **Relations: 1%**

3.2.1.2 Explore Model View – Model Conformance

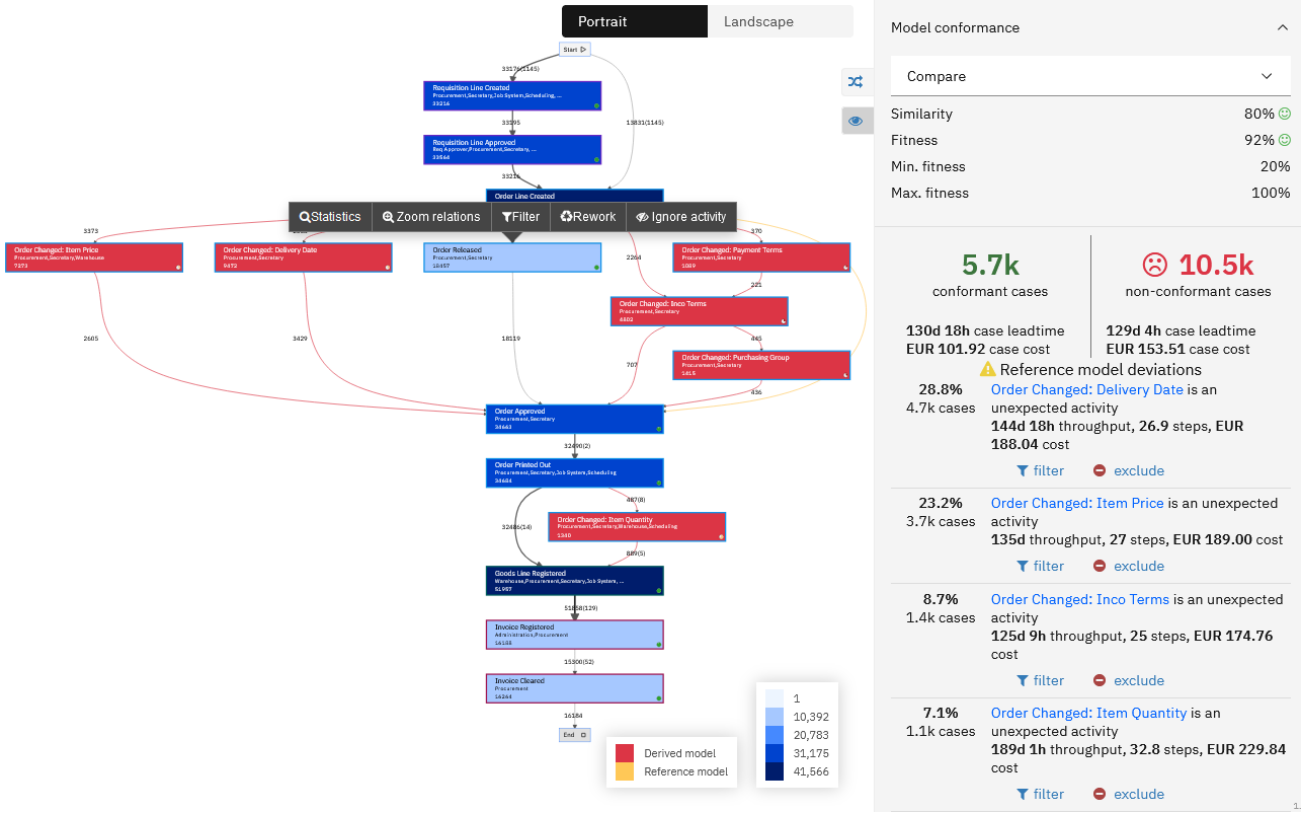
You can evaluate conformance of the derived from log data model to the BPMN reference model you imported.

_1. Click the **eye** icon

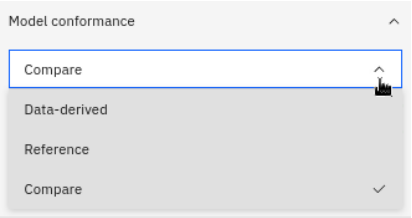


_2. Note that the Model Conformance view. The color coding in this view indicates the degree of divergence between derived and reference models:

- **Red box** or arrow indicate that the activity or transition are only present in the Derived model
- **Yellow box** or arrow indicate that the activity or transition are only present in the Reference model
- **Blue box** (dark or light blue depending on the frequency) indicates that the activity is present in both models
- **Black arrow** indicates that the transition is present in both models



_3. Click the **drop-down** (but keep Compare selected)



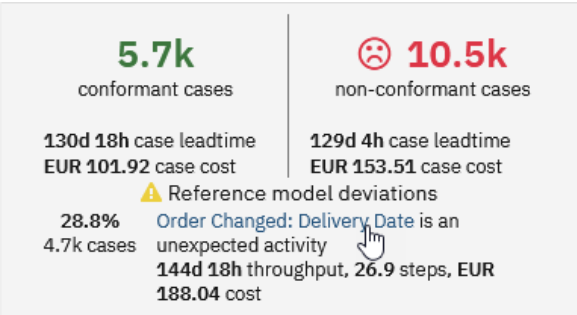
There three views avilable:

- Data-derived – IBM Process Mining visualizes the Data-Derived Model only
- Reference - IBM Process Mining visualizes the Reference Model only
- Compare - BM Process Mining automatically visualizes the resemblance between the two models.

3.2.1.3 Explore Model View – Model Conformance – Root Cause Analysis

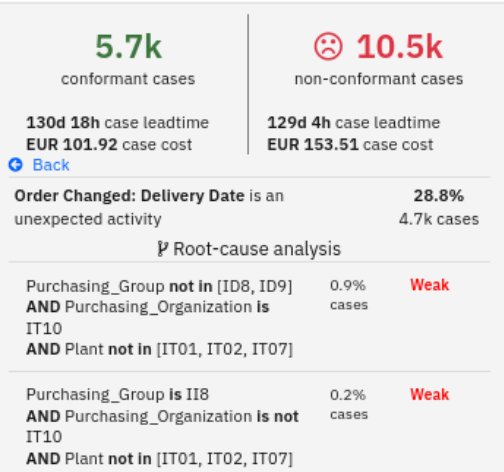
Now let's explore Root Casue Abnalysis in Model Conformace View. You can also select a specific unexpected activity or process flow, obtaining some information about the root cause. Based on the custom fields entered, you can have information about how many cases (where the deviation is present) involve a particular resource, role, supplier, product, company, and so on.

_1. Click **Order Changed: Delivery Date**

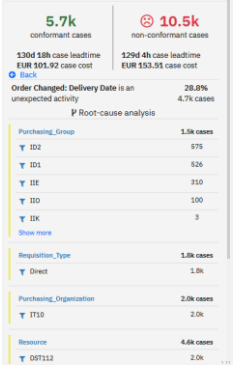


You should now see Machine Learning powered view showing:

1. What business specific data has the most significant influence of the deviation
2. A number of rules sets
3. The number proportion of cases influenced by the rule set and



Note, If you using a Porcess Mining version priot to 1.12 you will have a different (non ML-powered) view.



There are many other ways to get insights into the processes using the Model view…
We will cover this in other hands-on labs!

3.2.2 Explore Dashboard View

_1. Click **Dashboard** tab

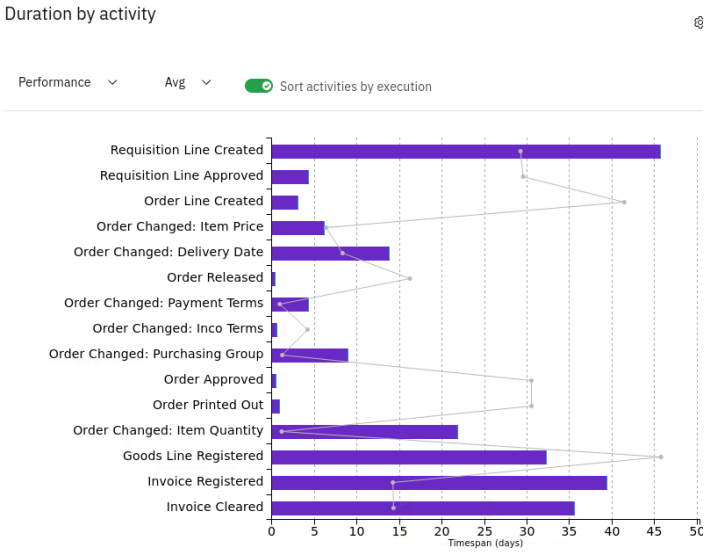
Let's explore each dashboard...

3.2.2.1 Process Details

This dashboard contains process and case statistics as well as high level performance statistics related to time

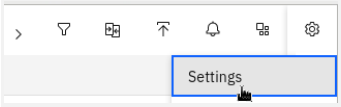
Notice the Conformance column. It contains summary of conformance with the reference model you imported.

3.2.2.2 Duration by activity

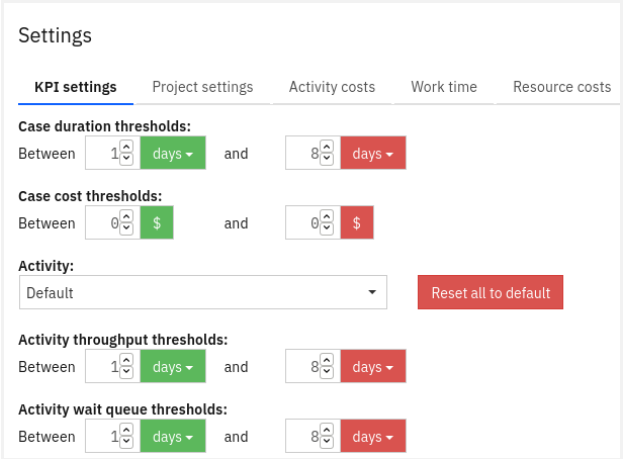


This dashboard displays activities duration (average or median) either absolute or how the durations compare with the KPIs defined in the project settings.

Note to view the KPI, select Settings from the top menu



And select the KPI tab. Note the KPI settings tab. This is where you define the KPI that are used thought the different views in the model.



3.2.2.3 Performance Drilldown

You can use the **Performance drilldown** dashboard to identify most critical activities (in terms of frequency and performance), and most critical resources (resources who are most frequently and are involved with critical activities)

This dashboard enables you to drill down on most critical activities and resources (users performing the tasks).

You can also examine activity and resource performance over time. For example you can see an *Activity waiting* queue durations or a *Resource load* variation over time.

The **Timespan** makes you choose a period of time to focus your analysis on.

In Case duration and count the colored line shows average/median remaining lead time of cases running in the selected date. And the grey bar shows number of cases running in the selected date.

Average duration and count represents, for the selected activity: average/median remaining service time, based on the cases running on the activity in the selected date (colored line), and number of activities running in the selected date (gray line).

Activity waiting queue represents, for the selected activity: average/median remaining waiting time, based on the cases waiting for the activity in the selected date (colored line), and number of cases waiting for the activity in the selected date (gray line).

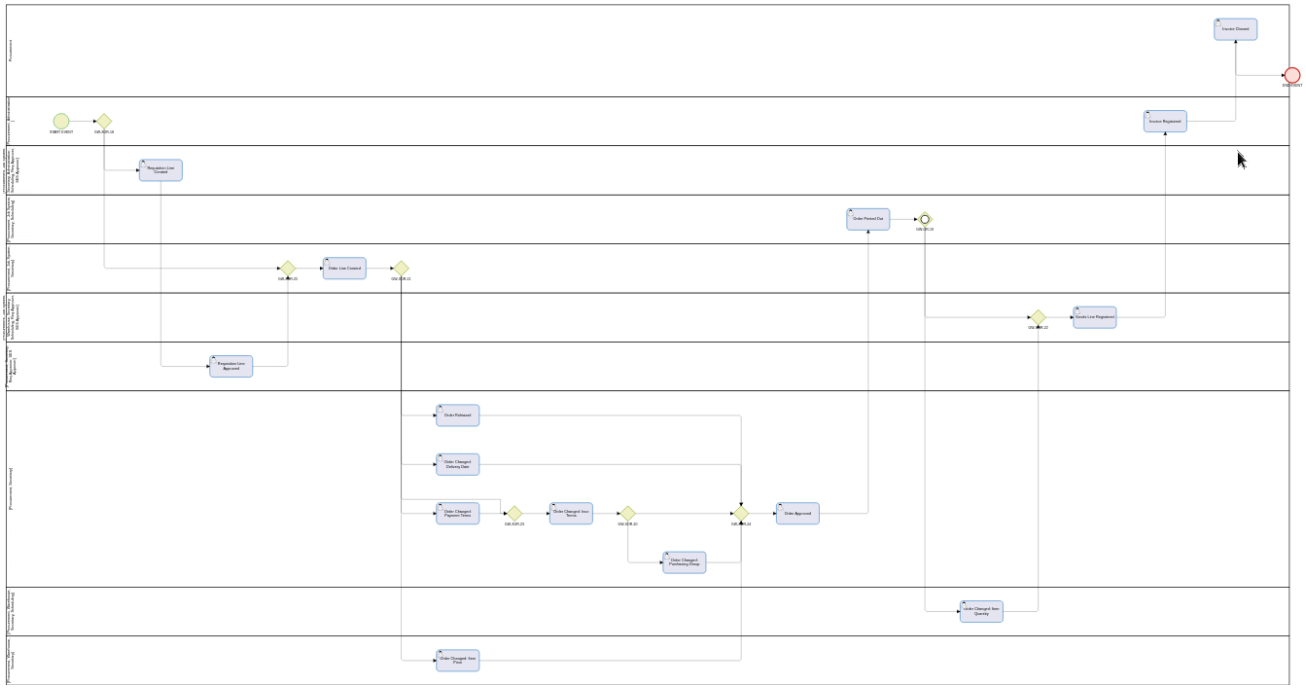
Active load allows to understand what resources are involved in each activity. By selecting a specific activity from the drop-down menu, you will obtain a dotted chart visualization that shows: the name of the resource that have carried out the activity in every swim lane, and the exact moment (relative to the timespan) in which a specific resource carried out the activity (represented by the small bullet points identify)

Resource load displays the workload of every resource. By selecting a specific resource from the drop-down menu you will obtain a dotted-chart visualization that shows: the name of the activity that had been carried out by the resource in every swim lane, and the exact moment (relative to the timespan) in which a specific activity had been carried out by the resource (identified by small the bullet points identify

3.2.3 Explore BPMN View

_1. Click **BPMN** tab

_2. You will see the BPMN diagram on the left:

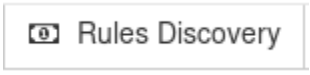


_3. Note the BPMN related features on the right:

- Rules Discovery - The Decision Rules Mining capability can automatically discover the correlations within the data mined data and enables automatic detection of the decision rules governing the process.
- Overview – Provides an overview of the rules discovery results
- Settings – Used to configure and refine the decision rules mining settings
- Publish – Enables to save a snapshot of the current model in the BPA tool. Snapshots can then be published as BPMN2 files.
- Edit a Copy – Use it to edit a copy of the actual model. Both the Publish and Edit a Copy commands will load the BPA tool.
- Simulation – Enables to create a simulated scenario based on the current BPMN model, derived from the process data.

3.2.3.1 Rules Discovery

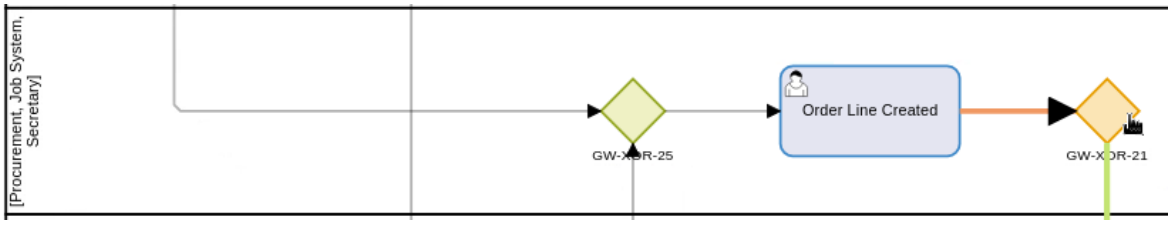
_1. Click **Rules Discovery** tab



_2. Wait for the "**Discovering decision rules...**" pop up to disappear

_3. Find the **GW-XOR-21** gate on the BPMN diagram and **click** it

Hint: Use the mouse wheel to zoom and mouse right-button to move the BPMN diagram.



_4. Note the decisions generated form the process mining data:

Gateway info: GW-XOR-21

Rules overview

Rules overview | Reduce | Elaborate

Order Line Created → Order Changed: Delivery Date

```
Purchasing_Group in ["II6", "IIB", "IID"]
AND Material_Number not in [Missing, "S150020101", "S150021101"]
AND Order_Header not in ["4500189397", "4500199545", "4500200794"]
OR
Purchasing_Group in ["II3", "II8", "IIE"]
AND attr-resource in ["ACQ01", "USR00600", "USR02069"]
AND Order_Vendor in ["VND05362", "VND05955", "VND06216"]
AND Order_Header in ["4500248874", "4500251352", "4500258513"]
OR
Purchasing_Group in ["II3", "II8", "IIE"]
AND attr-resource not in ["ACQ01", "USR00600", "USR02069"]
AND Order_Vendor in ["VND06322", "VND07617", "VND07995"]
AND Order_Header not in ["4500196291", "4500200761", "4500215864"]
```

Order Line Created → Order Changed: Inco Terms

```
Purchasing_Group in ["II3", "II8", "IIE"]
AND attr-resource not in ["ACQ01", "USR00600", "USR02069"]
AND Order_Vendor in ["VND04890", "VND06230", "VND06922"]
OR
Purchasing_Group in ["II3", "II8", "IIE"]
AND attr-resource in ["ACQ01", "USR00600", "USR02069"]
AND Order_Vendor not in ["VND05362", "VND05955", "VND06216", "VND06546", "VND06863", "VND07704"]
```

For each rule, the target transition is indicated (in the form “Gateway activity → Target activity”), and the rule’s conditions are shown underneath. Note the business data is used on the conditions!

The green circle indicates the Coverage of the decision rule: percentage of events in which the rule’s condition was the specified one, out of the total number of occurrences of that transition.

The Coverage decreases when the transition occurs in the cases, but the rule’s condition is not the specified one. These transitions are called “immigrants”.

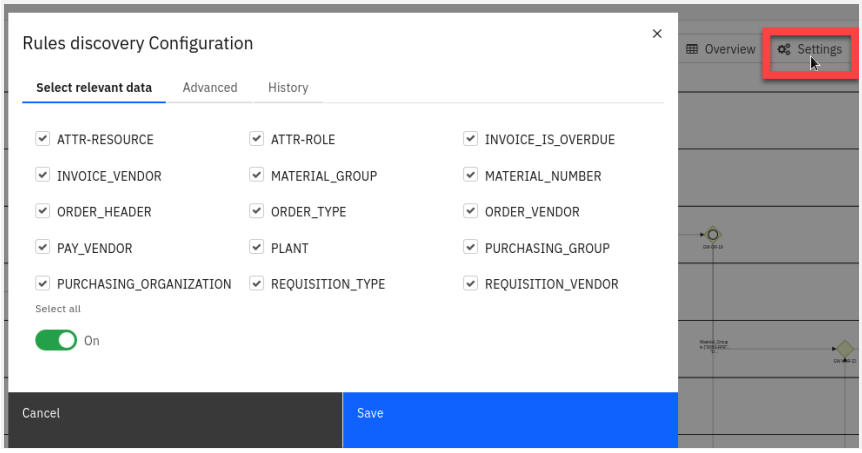
The orange circle indicates the Precision of the decision rule: percentage of events in which the rule’s condition was met, out of the total number of events in which the rule’s condition was the specified one.

The Precision decreases if the condition was specified in the cases but instead of the expected transition, another transition occurs. These transitions are called “emigrants”.

The Reduce / Elaborate button allow you to reduce / increase the rule complexity. Use the depth-reduction when you want a more concise and summarized view of the gateway’s rules.

_5. Click **x** to close the *Gateway info: GW-XOR-21* window

Note that you can use the Setting button launch Rules discovery Configuration which you can use to select what variables are used in decision definitions

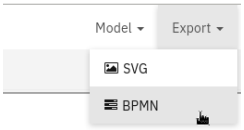


3.2.3.2 Export BPMN File

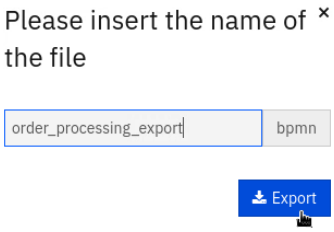
_1. Click **Publish**

Note that this action takes you away from Process Mining to the BPA (Business Process Analysis) environment.

_2. Click **Export > BPMN**

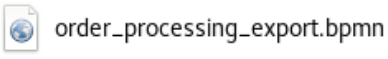


_3. Change file name to **order_processing_export** and click **Export**



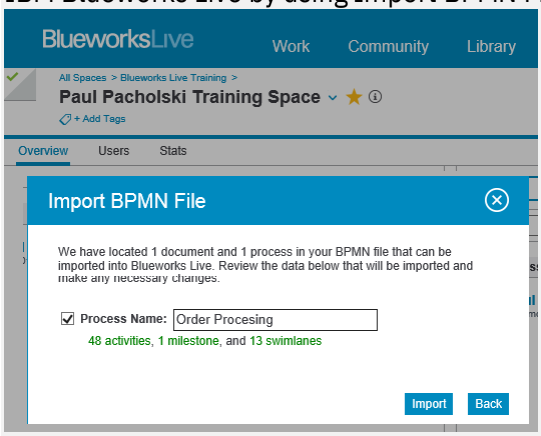
_4. Select a directory of your choice.

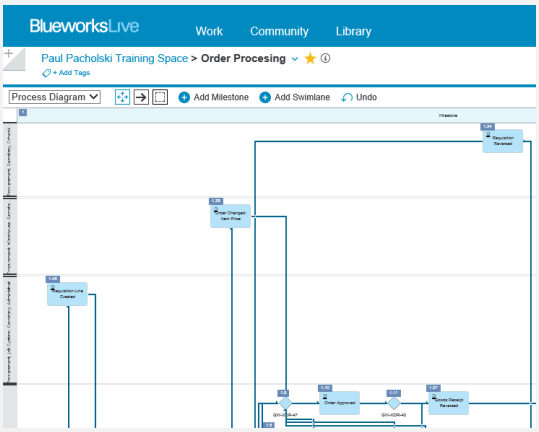
_5. Note the generated BPMN file:



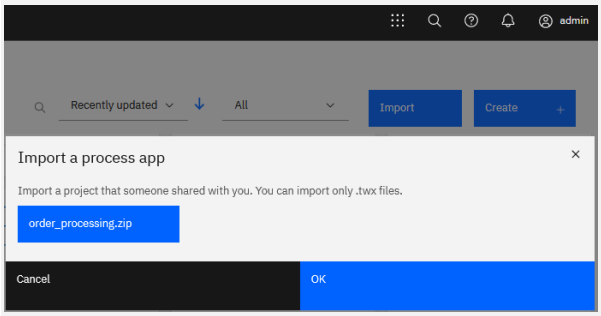
Note that once you create a BPMN file you can import it to the following Digital Business Automation process tools:

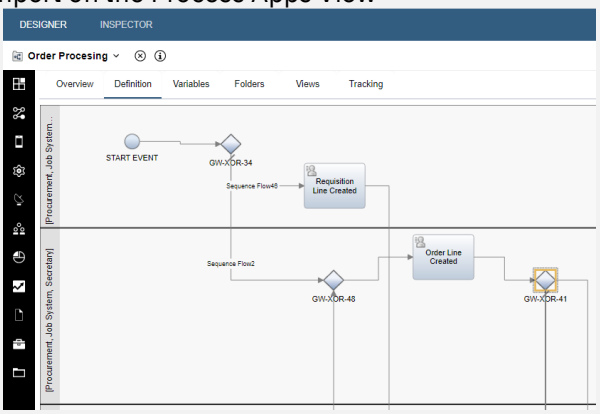
IBM Blueworks Live by using Import BPMN File feature:





IBM Business Automation Workflow by selecting Import on the Process Apps View



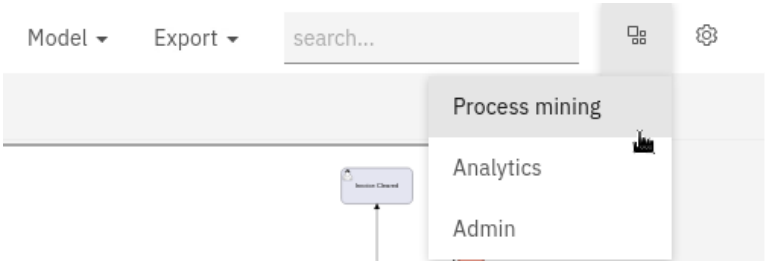


Note Importing these files to above tools is not part of this lab. You can do this later after completing this lab.

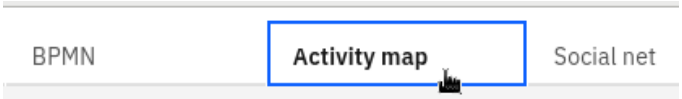
3.2.4 Explore Activity Map View

Activity Map can be used to analyze human resources by name and job title. It highlights if employees are doing what they are supposed to be doing. For example, by using Activity map, we can discover that the Procurement Team is managing activities not included in their duties.

_1. Click the **chess-board icon** and then **Process mining**

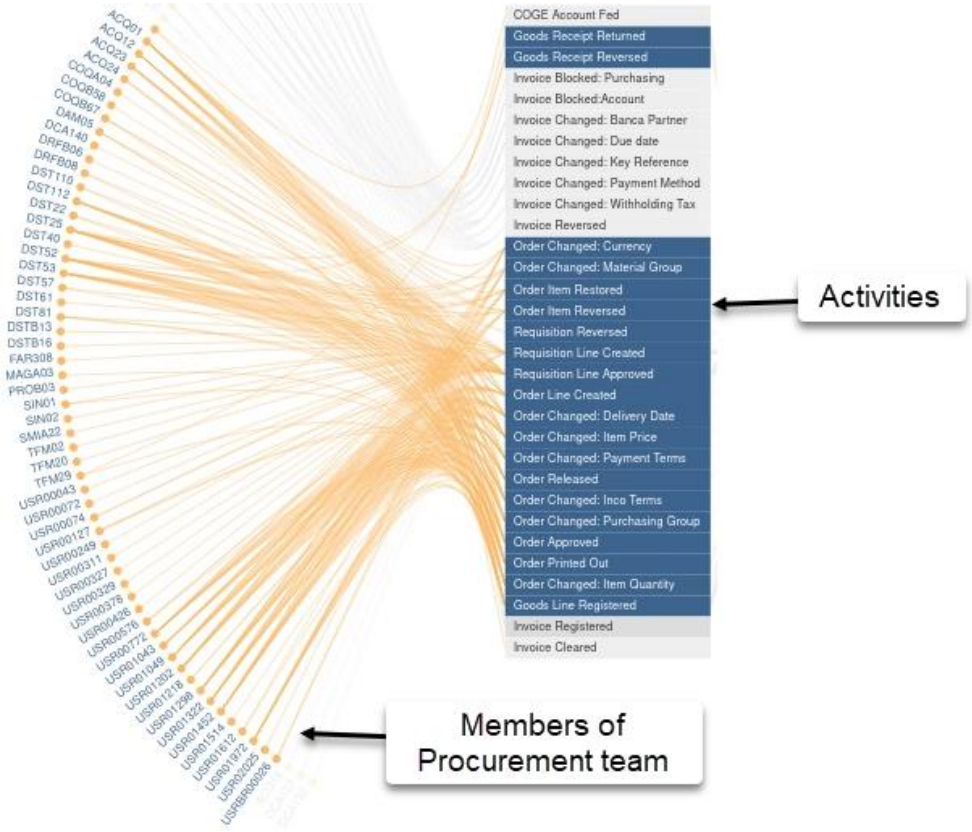


_2. Click **Activity map**



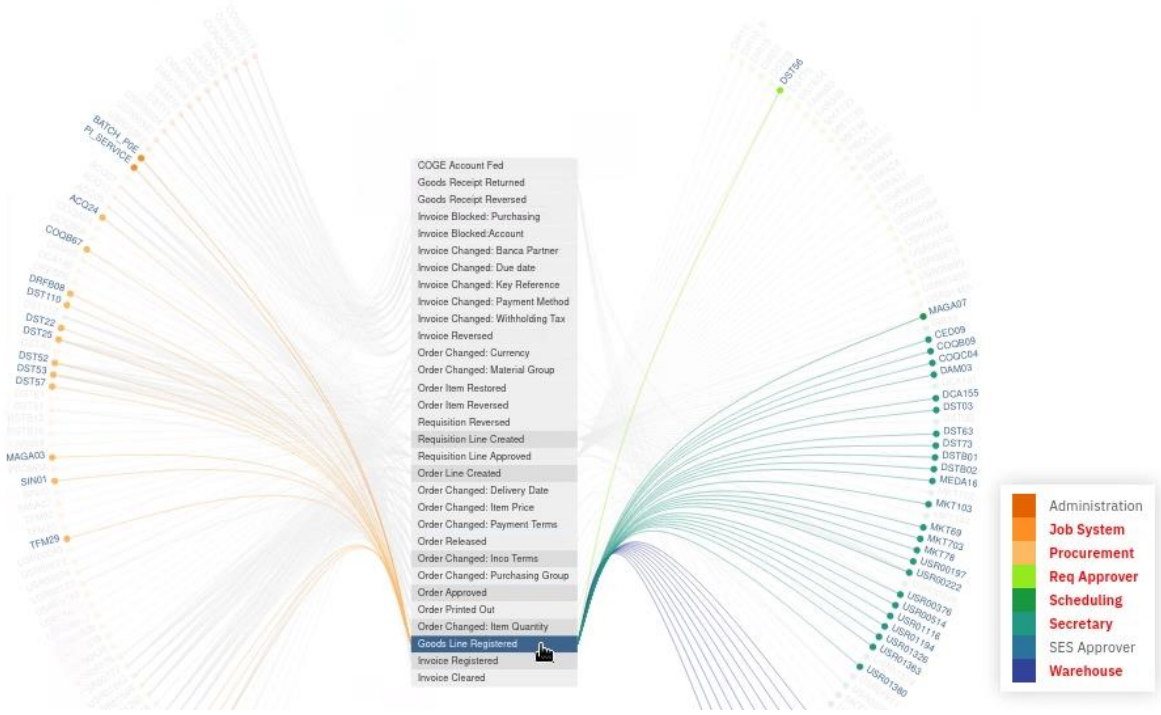
_3. Click **Procurement**

_4. Note all the activities that Procurement Team members are involved in! Also note the large number of users (user ids) that are part of the Procurement team.



_5. **Click Goods Line Registered**. Notice the six Teams (highlighted in red) are involve in completing this task.

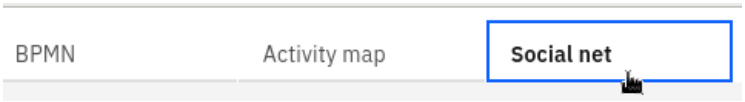
Hint: Use the mouse wheel to zoom and mouse right-button to move the Activity map diagram.



3.2.5 Explore Social Net View

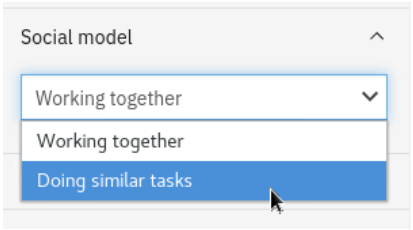
The Social net allows you to discover and analyze the relationships between users and groups that are formed within a process. We can divide these relationships into two social models, selectable from the dropdown menu.

_1. Click **Social net**

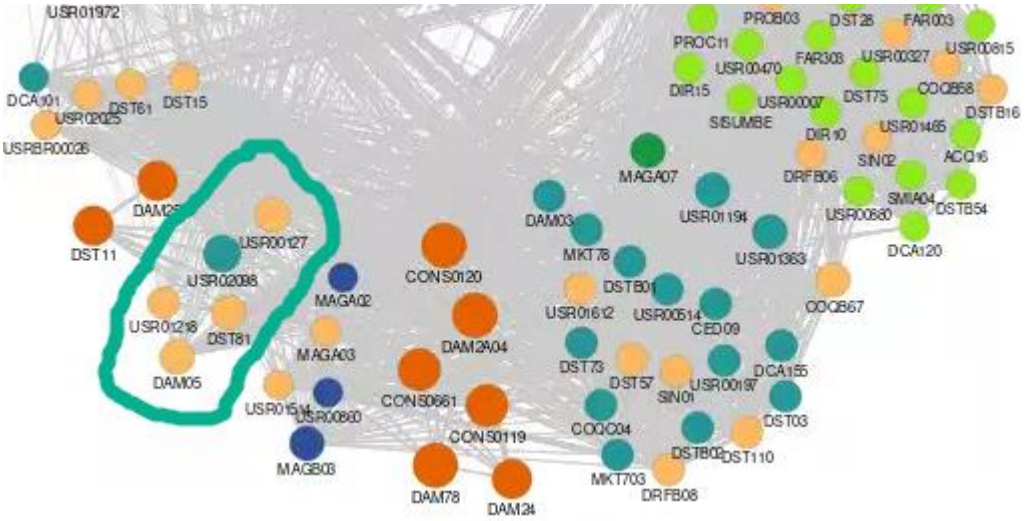


Let's explore *Doing similar tasks* view...

_2. Click **Social model > Doing similar tasks**



_3. Notice a cluster (highlighted in green) discovered by *Doing similar tasks* view

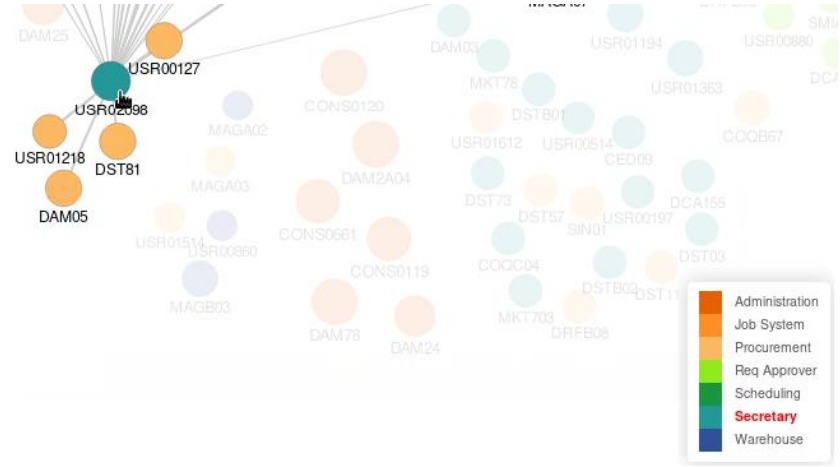


In the *Doing similar tasks* view:

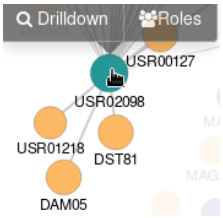
- The bullets represent resources
- Resources are clustered by common activities carried out
- The bullet will be bigger for those resources who are sharing more activities
- Different colors identify different roles

_4. Hover the pointer over the **blue green USR02098** dot.

_5. Notice that the user USR02098 is a Secretary while all the other users performing similar activities belong to the Procurement team



_6. Select *USR02098* dot and click **Drilldown**



Notice that this user is typically performing many *Order Approved* Activities!

USR02098

Req_Line_ID	Order_Line_ID	Goods_ID	Invoice_ID	Activity	DateTime	Resource	Role	Requisition
0010234119_10	4500240631_10			Order Line Created	2018-07-09 17:50:03	USR02098	Secretary	
0010237169_10	4500243295_10			Order Line Created	2018-07-23 17:54:11	USR02098	Secretary	VND02905
	4500240631_10			Order Released	2018-07-09 17:51:04	USR02098	Secretary	
	4500243147_10			Order Approved	2018-07-23 14:47:15	USR02098	Secretary	
	4500243266_10			Order Approved	2018-07-23 16:29:48	USR02098	Secretary	
	4500243266_20			Order Approved	2018-07-23 16:29:48	USR02098	Secretary	
	4500243166_10			Order Approved	2018-07-23 16:30:57	USR02098	Secretary	
	4500243241_10			Order Approved	2018-07-23 16:32:20	USR02098	Secretary	
	4500243177_10			Order Approved	2018-07-23 16:32:39	USR02098	Secretary	

This finding is confirmed by the size of user's USR0298 dot which indicates that the users shares large number of similar activities in other users in the cluster.

3.3 Lab Summary

In this lab you have learned the basics of IBM Process Mining tools. You should now know:

- how to create a process from mined process data,
- be familiar with key process mining features and their value.