Business Process Management

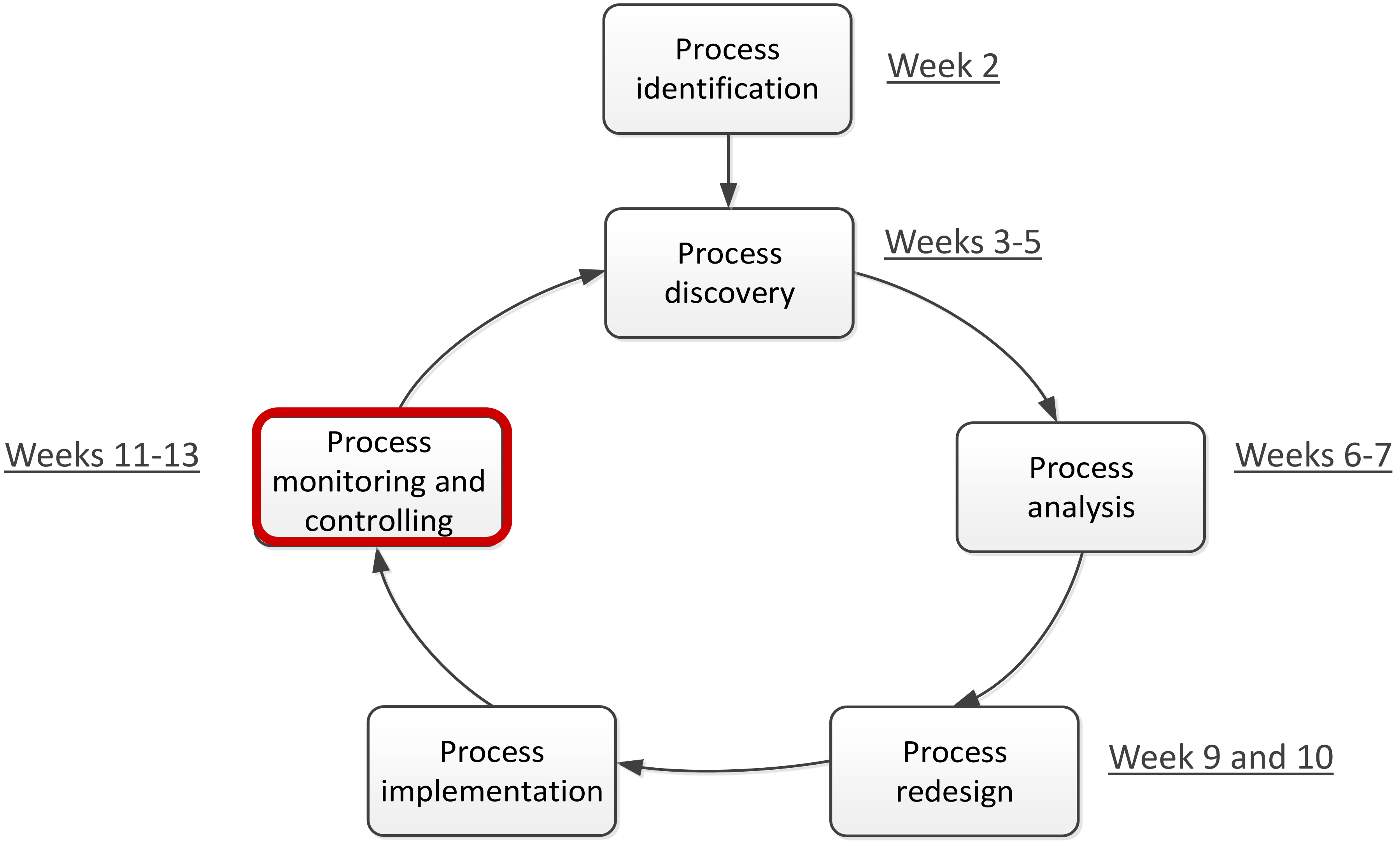
Lecture 12
Process Mining II

Prof. Josué Obregón

Department of Industrial Engineering - ITM

Seoul National University of Science and Technology

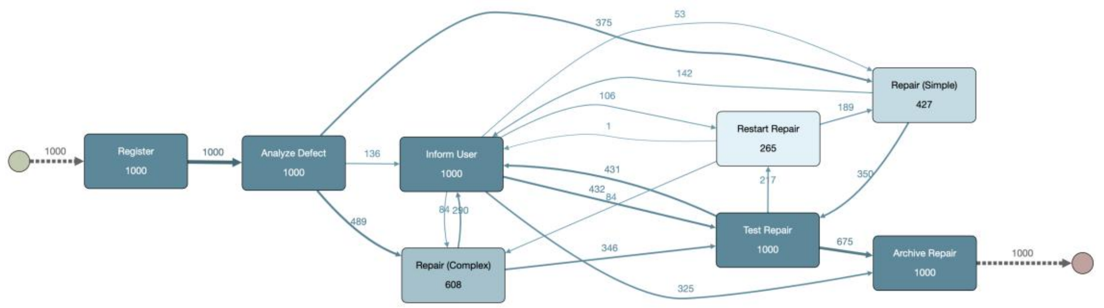
Course structure

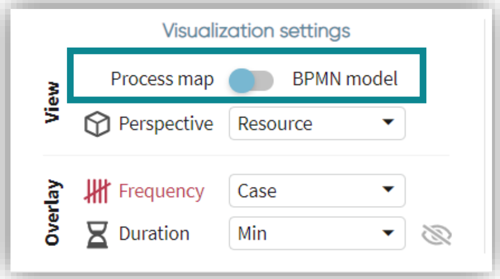


Contents

- More on Process Mining
    - Discovering BPMN models
    - Conformance Checking
    - Performance Mining
    - Variant Analysis

Discovery of BPMN models





Visualization settings

View

Process map / BPMN model

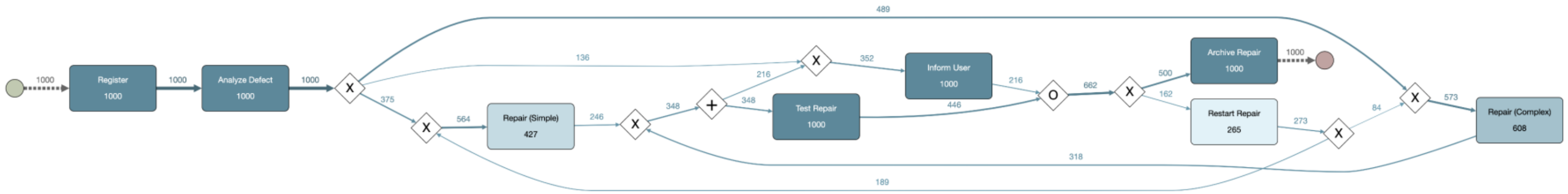
Perspective: Resource

Overlay

Frequency: Case

Duration: Min

A given configuration of node/arc sliders, and parallelism slider



*Note: Apromore uses an algorithm called **Split Miner** to turn process maps into process models*

Process Mining Using BPMN

- To get an intuition of how BPMN models are discovered from an event log, let's draw the process maps of the following logs and let's try to imagine what is the BPMN process model corresponding to these logs:
  - [abc, abd]
  - [abcd, acbd]
  - [abcd, abcbcd]

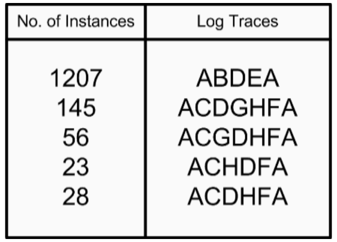
1. [abc, abd]
2. [abcd, acbd]
3. [abcd, abcbcd]

Automated Process Discovery: Analysis Template

What?	How?
Flow Analysis 1. Analyze the process structure & main case variants 2. Identify parallelism, branching points & rework loops 3. Analyze case entry and exit points, and check for incomplete cases	• Visualize the most frequent case variant(s) using the Case inspector • Use abstraction sliders to focus on the most frequent activities and dependencies. • Switch to BPMN view; inspect the behavior around the gateways • Check the dotted lines in the process map (arcs emanating from the start event or leading to the end event). Unexpected dotted lines indicate some cases are incomplete
Filtered Flow Analysis Analyze different components or slices of the process separately	• Use event filtering to retain/remove subsets of activities, e.g. separate automated or non-core activities, or separate activities belonging to two different systems.
Frequency analysis Analyze the most frequent activities and relations	• Use the Activity inspector • Use the color-coding to find most frequent arcs • Consider switching between frequency metrics (max, avg, …) • Use the Activities tab in performance dashboard (Apromore EE)
Handoff analysis Analyze handoffs between workers, teams, groups, org units	• Switch between perspectives • Use the Resource and Other Attributes tabs in the Dashboard

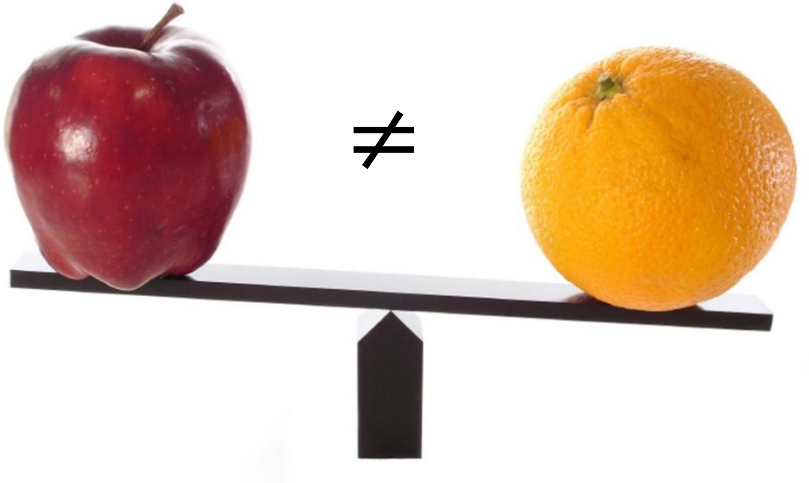
Conformance checking

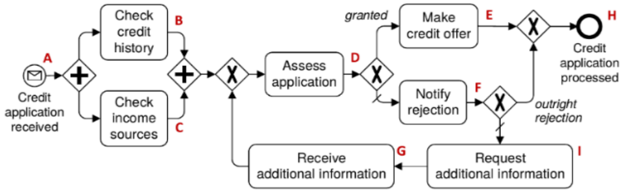
Given an event log and a set of **business rules** or a **process model**, find, describe, and measure the impact of *differences* between the rules/model and the log.



No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA

Log





Rules, process model

Types of Conformance Checking

1. **Rule-Based Conformance Checking**
   - Checking that a process follows certain rules
2. Model-Based Conformance Checking
   - Comparing the behaviour captured in the process model against the one in the event log

Rule-Based Conformance Checking

Checking whether regulations and control measures embedded in processes are followed in practice.

Common types of compliance rules:

Flow constraints

Certain activities may be mandatory e.g. a specific approval to proceed with an application.

SLA constraints

Customer agreements on service quality define timeframes at the case, milestone or activity transition level.

Resource constraints

Certain activities must be executed by different employees, e.g. creation and approval of an invoice.

Conformance Checking Levels

Accordingly, we can identify the following *compliance checking levels*:

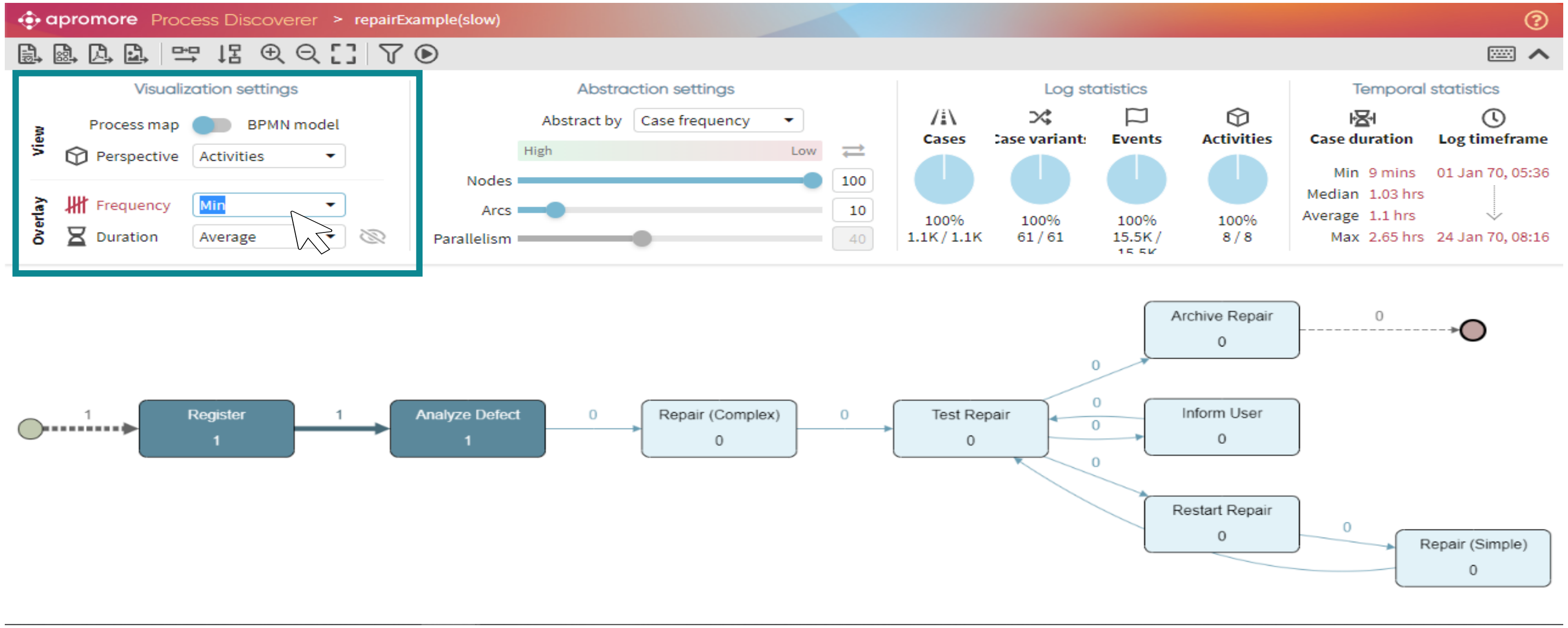
- **Flow:** have all mandatory activities been performed? Have activities been performed in the right order and with the right frequency?
- **Temporal:** has a given timeframe between activities, process milestones or for the end-to-end process been adhered to?
- **Resource:** did everyone obey the segregation of duties?
- **Exception:** explore highly-infrequent behavior to identify possible deviations from the standard process

Procedure:

1. Define compliance rules (based on attribute-value pairs, activity orders or performance targets)
2. Apply one or more filters to keep cases that violate the rules in question
3. If needed, switch perspective to highlight non-compliance issues (e.g. show resources that violate a segregation of duties principle)
4. Assess impact on process performance

Flow Compliance Checking

1 Mandatory tasks: Have all mandatory tasks been completed?



apromore Process Discoverer > repairExample(slow)

Visualization settings

View: Process map / BPMN model

Perspective: Activities

Overlay: Frequency: Min

Duration: Average

Abstraction settings

Abstract by: Case frequency

Nodes: 100

Arcs: 10

Parallelism: 40

Log statistics

Cases: 100% 1.1K / 1.1K

Case variants: 100% 61 / 61

Events: 100% 15.5K / 15.5K

Activities: 100% 8 / 8

Temporal statistics

Case duration: Min 9 mins, Median 1.03 hrs, Average 1.1 hrs, Max 2.65 hrs

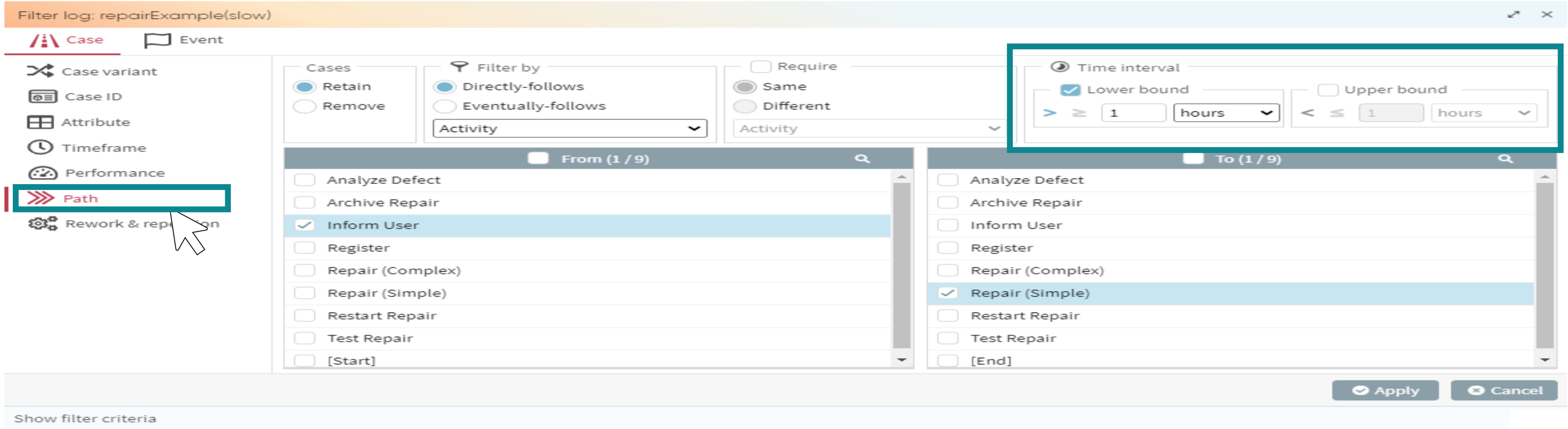
Log timeframe: 01 Jan 70, 05:36 – 24 Jan 70, 08:16

To check whether all your tasks have been executed at least once, change the frequency view to “Min.” in the visualization settings of your Process Discoverer.

The numbers in your activities now indicate the minimum amount of times it has been completed in all process instances.

Temporal Compliance Checking

2 Prescribed time frame: Has the given time frame been adhered to?



Filter log: repairExample(slow)

Case | Event

- Case variant
- Case ID
- Attribute
- Timeframe
- Performance
- Path
- Rework & repetition

Cases: ● Retain ○ Remove

Filter by: ● Directly-follows ○ Eventually-follows — Activity

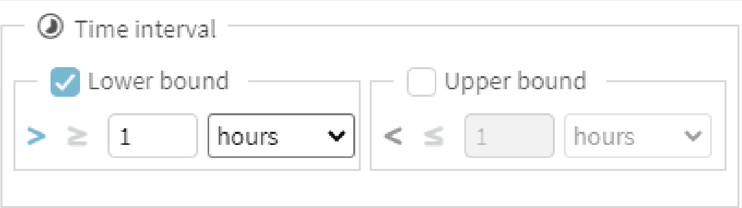
□ Require: ● Same ○ Different — Activity

Time interval: ☑ Lower bound > ≥ 1 hours — □ Upper bound < ≤ 1 hours

From (1 / 9): □ Analyze Defect, □ Archive Repair, ☑ Inform User, □ Register, □ Repair (Complex), □ Repair (Simple), □ Restart Repair, □ Test Repair, □ [Start]

To (1 / 9): □ Analyze Defect, □ Archive Repair, □ Inform User, □ Register, □ Repair (Complex), ☑ Repair (Simple), □ Restart Repair, □ Test Repair, □ [End]

Select the "Path" filter and define the steps that must be performed in a defined time frame. Use "Time interval" to then set your temporal figures.



Time interval: ☑ Lower bound > ≥ 1 hours — □ Upper bound < ≤ 1 hours

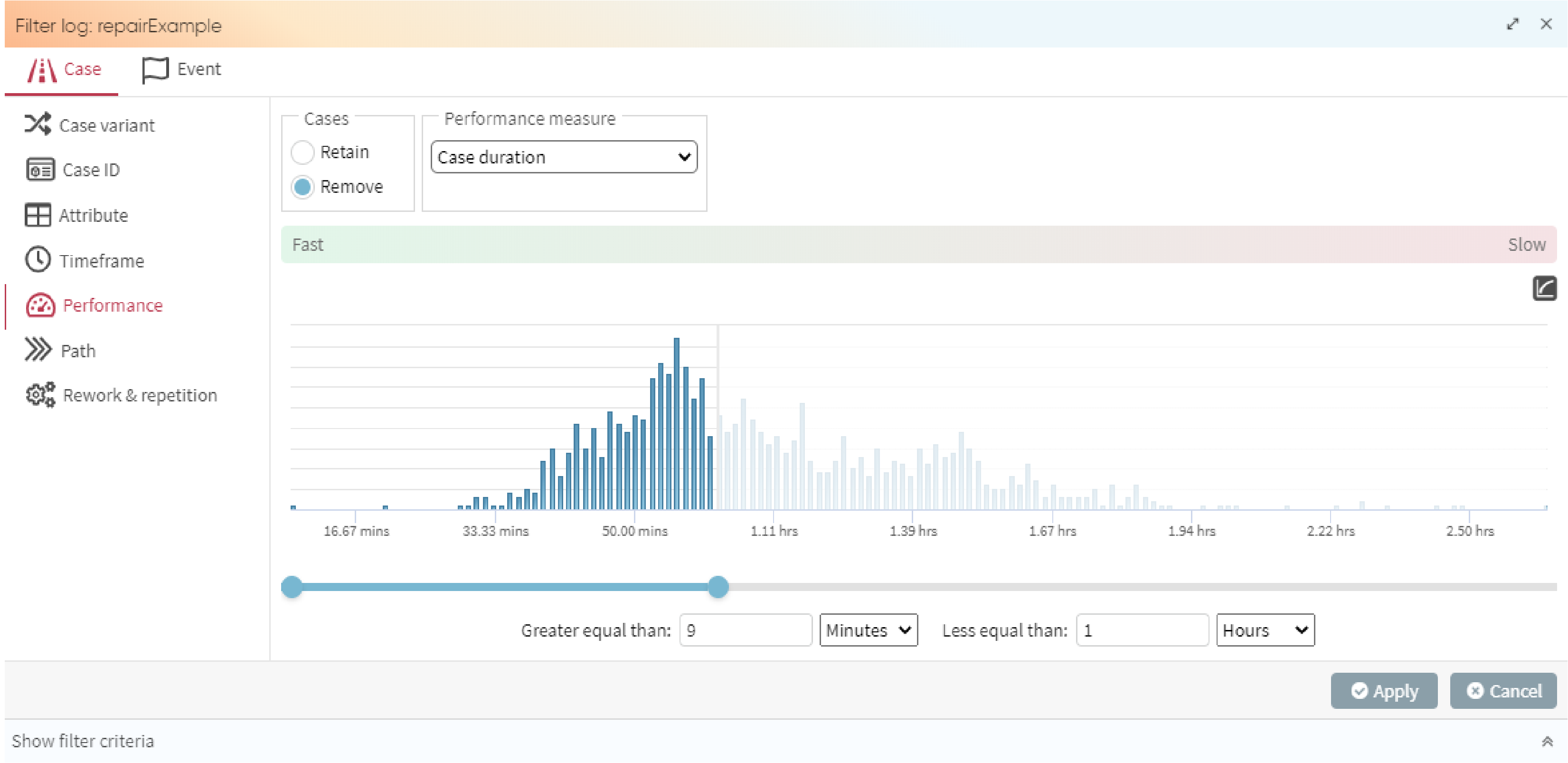
Example:

Compliance rule:
Waiting time between "Inform User" and "Repair (Simple)" should not exceed 1 hours.

"Time Interval" Filter:
> 1 hour, so that only those cases are retained that do not comply with the compliance rule.

Temporal Compliance Checking

1 Prescribed time frame: Has the given time frame been adhered to?



Cases: ○ Retain ● Remove

Performance measure: Case duration

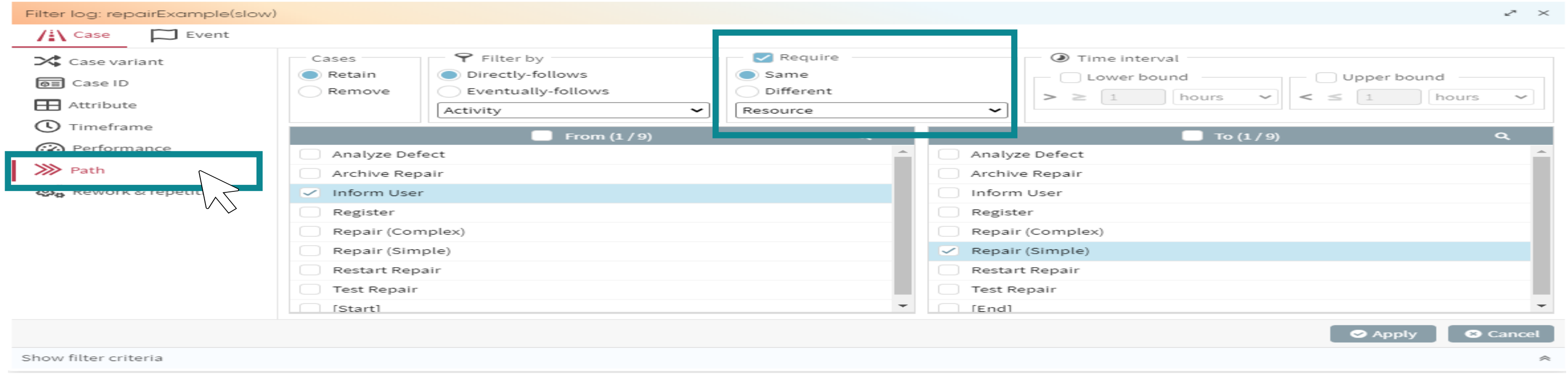
Greater equal than: 9 Minutes Less equal than: 1 Hours

Performance filter can help to check for end-to-end SLA violations.

For example, retain only those cases that gets completed in at most 1 hour.

Resource Compliance Checking

1 Segregation of duties: Did everyone obey the segregation of duties?


Filter log: repairExample(slow)

Cases: ● Retain ○ Remove

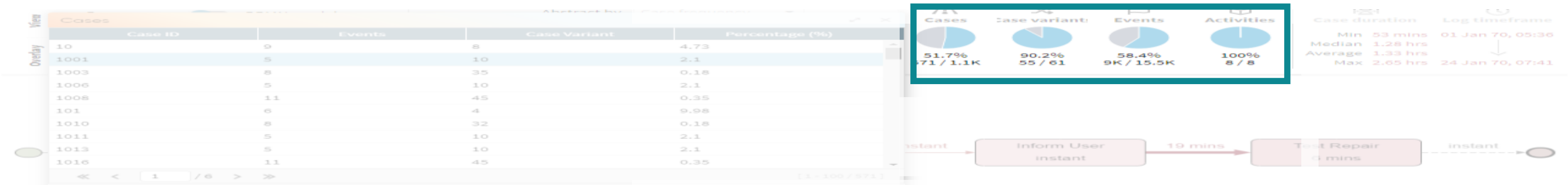
Filter by: ● Directly-follows ○ Eventually-follows — Activity

☑ Require: ● Same ○ Different — Resource

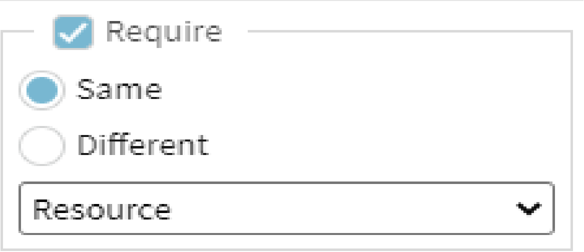
Time interval: □ Lower bound > ≥ 1 hours; □ Upper bound < ≤ 1 hours

From (1 / 9): □ Analyze Defect, □ Archive Repair, ☑ Inform User, □ Register, □ Repair (Complex), □ Repair (Simple), □ Restart Repair, □ Test Repair, □ [Start]

To (1 / 9): □ Analyze Defect, □ Archive Repair, □ Inform User, □ Register, □ Repair (Complex), ☑ Repair (Simple), □ Restart Repair, □ Test Repair, □ [End]



To check whether the segregation of duties has been breached, select a “Path” filter. Now, set another filter which indicates that the same resource must have completed the tasks (e.g. creating & approving an invoice)


☑ Require

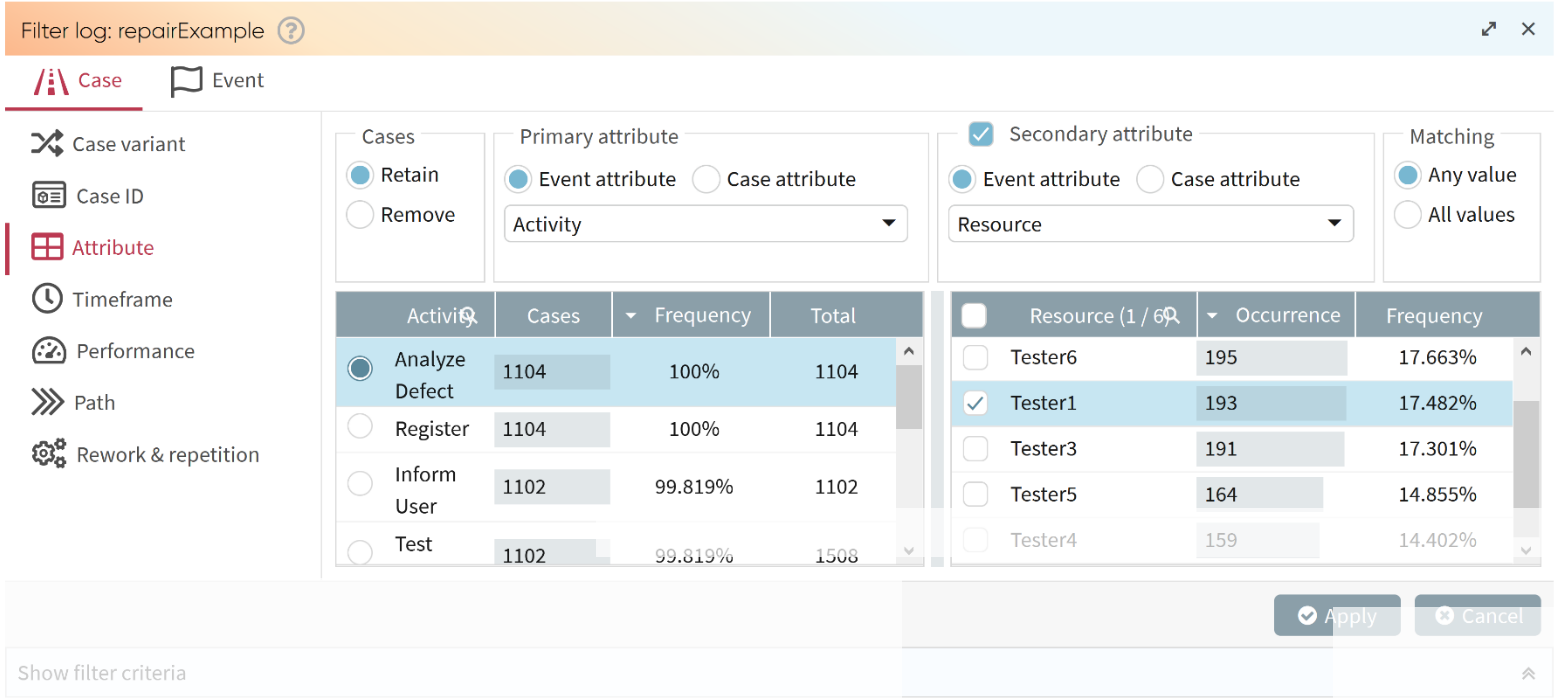
● Same

○ Different

Resource

Resource Compliance Checking

1 Unexpected resource assignments



Filter log: repairExample

Cases: ● Retain ○ Remove

Primary attribute: ● Event attribute ○ Case attribute — Activity

☑ Secondary attribute: ● Event attribute ○ Case attribute — Resource

Matching: ● Any value ○ All values

	Activity	Cases	Frequency	Total
●	Analyze Defect	1104	100%	1104
○	Register	1104	100%	1104
○	Inform User	1102	99.819%	1102
○	Test	1102	99.819%	1508

□	Resource (1 / 6)	Occurrence	Frequency
□	Tester6	195	17.663%
☑	Tester1	193	17.482%
□	Tester3	191	17.301%
□	Tester5	164	14.855%
□	Tester4	159	14.402%

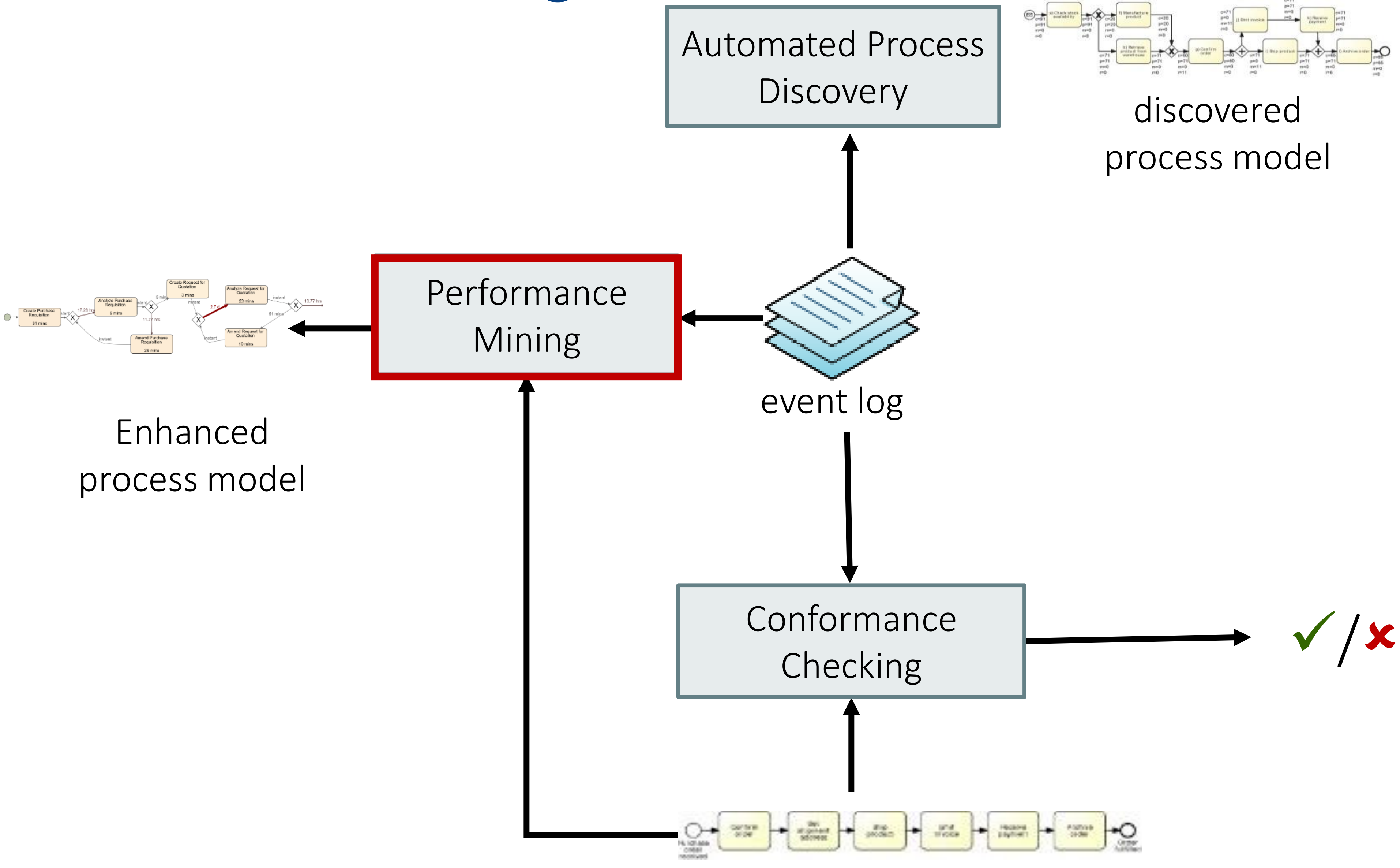
User the Attribute filter with two attributes (primary and secondary) to retrieve cases where a given activity was performed by a given resource (or any other combination of attributes)

Conformance Checking: Analysis Template

What?	How?
Flow compliance checking 1. Check skipped mandatory activities 2. Check forbidden repetitions. 3. Check activity exclusion or co-occurrence relations	• Identify skipped activities in the process map or BPMN model • Visualize process map with the min/max frequency metrics • Visualize min/max statistics in Activity tab of the dashboard • Use Path filter (directly-follows or eventually-follows)
Temporal compliance checking Check temporal constraints violations	• Use performance filter (to check for end-to-end SLA violations) • Use Path filter (eventually-follows) with a duration constraint to identify violations of temporal constraints between two activities
Resource compliance checking Check four-eyes principle* or same-case-hander principle Check if an activity has been performed by a resource who should not do so.	• Use Path filter (eventually-follows) and require the same Resource (or different resources) • Use the Attribute filter and search with “Activity” as the primary attribute and “Resource” as secondary attribute to check if a resource performed an activity they were not supposed to do.

* a process that requires at least two authorized people to review and approve a task or decision before it's considered finalized

Process Mining



Automated Process Discovery

discovered process model

Performance Mining

event log

Enhanced process model

Conformance Checking

✓ / ✗

input process model

Process performance mining: Analytics & Analysis types

Extract process execution performance in an aggregate or detailed level (i.e. at the level of individual activities and directly-follows relations)

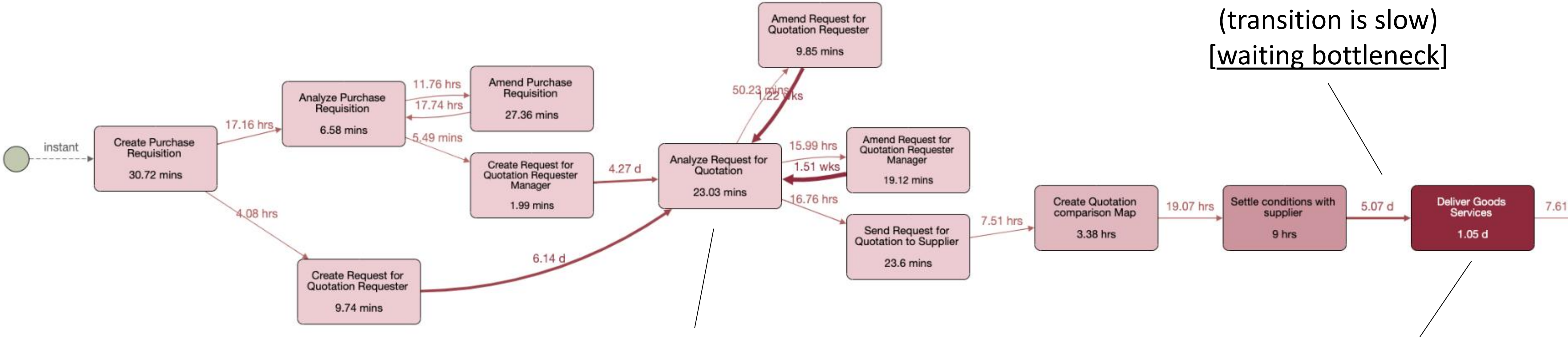
Analytics:

- Duration-enhanced process maps (activity, resource perspective…)
- Duration-enhanced BPMN models
- Process performance dashboards

Analysis types:

- Bottleneck analysis
- Workload analysis
- Rework analysis

Bottleneck analysis in the Activity perspective

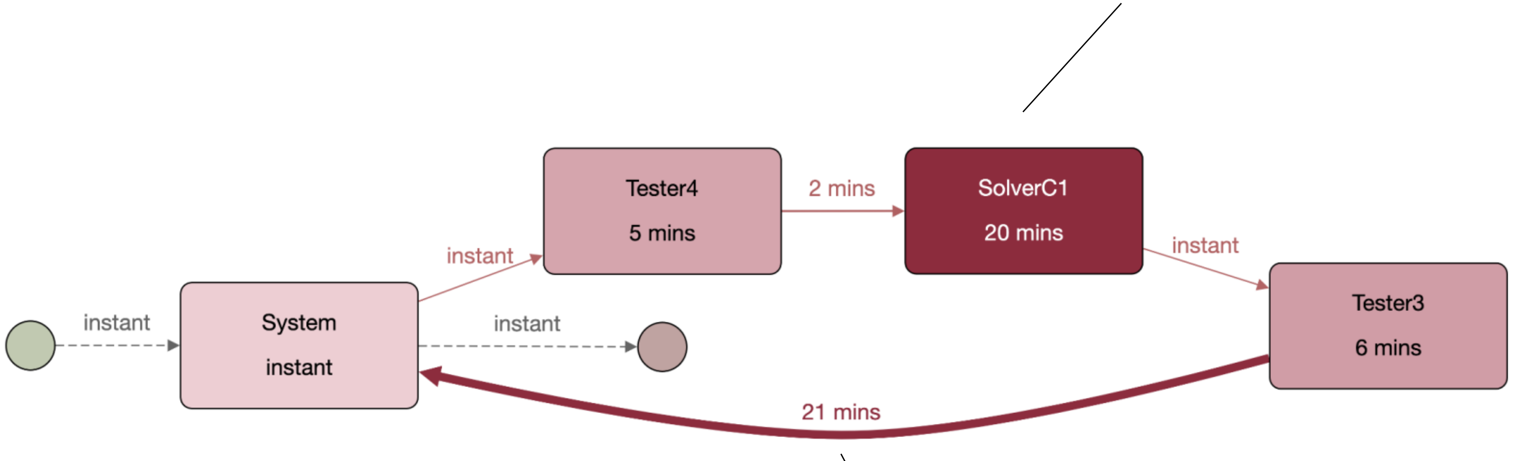

instant
Create Purchase Requisition
30.72 mins
17.16 hrs
4.08 hrs
Analyze Purchase Requisition
6.58 mins
11.76 hrs
17.74 hrs
5.49 mins
Amend Purchase Requisition
27.36 mins
Create Request for Quotation Requester Manager
1.99 mins
4.27 d
Create Request for Quotation Requester
9.74 mins
6.14 d
Analyze Request for Quotation
23.03 mins
50.23 mins
1.22 wks
Amend Request for Quotation Requester
9.85 mins
15.99 hrs
1.51 wks
Amend Request for Quotation Requester Manager
19.12 mins
16.76 hrs
Send Request for Quotation to Supplier
23.6 mins
7.51 hrs
Create Quotation comparison Map
3.38 hrs
19.07 hrs
Settle conditions with supplier
9 hrs
5.07 d
Deliver Goods Services
1.05 d
7.61

Slow handoff
between activities
(transition is slow)
[waiting bottleneck]

Resource-capacity
bottleneck
(all incoming
arcs are slow)
[waiting bottleneck]

Activity
bottleneck
(activity is slow)

Bottleneck analysis in the Resource perspective

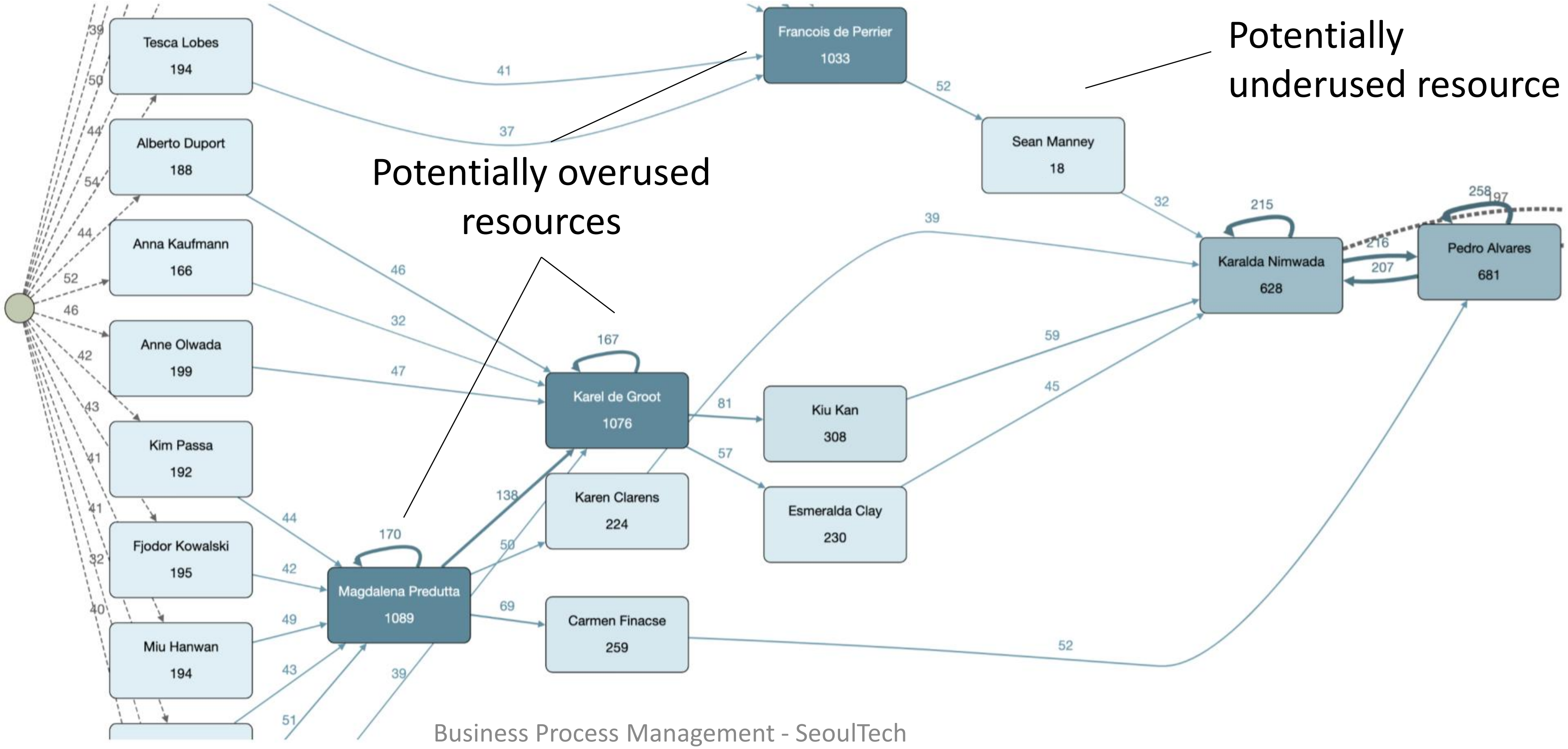


Resource bottleneck

Slow handoff between
specific resources
[waiting bottleneck]

Workload analysis: overused vs underused resources

- **Potentially overused resources** are those with high total frequency and/or high total duration
- **Potentially underused resources** are those with low total frequency and/or low total duration



Potentially overused resources

Potentially underused resource

Rework analysis

Rework means repetition of activities within the same case:

- **Self loop:** an activity takes place n times in a row
- **Short loop ("ping-pong effect"):** two activities are repeated one after the other. Note: <u>this may be confused with *parallelism*</u>
- **Indirect repetition (loop):** several activities are executed several times in sequence.

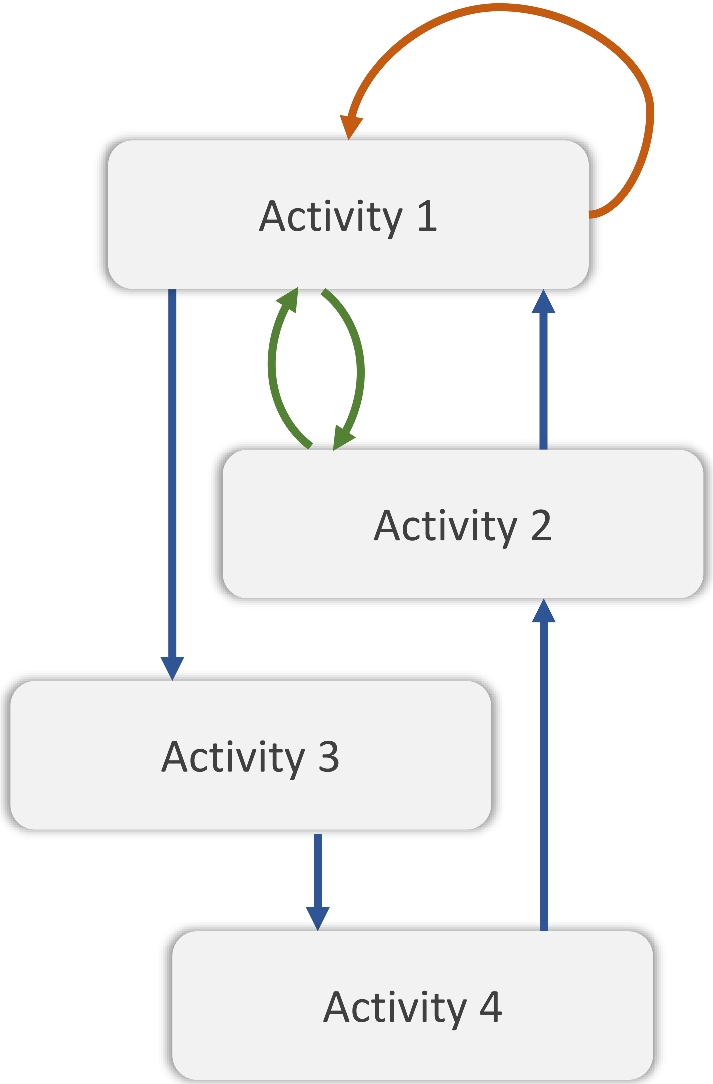
To analyze repetitions, one can answer the following questions:

(1) What type of repetition is present?

(2) How often is an activity / series of activities repeated?

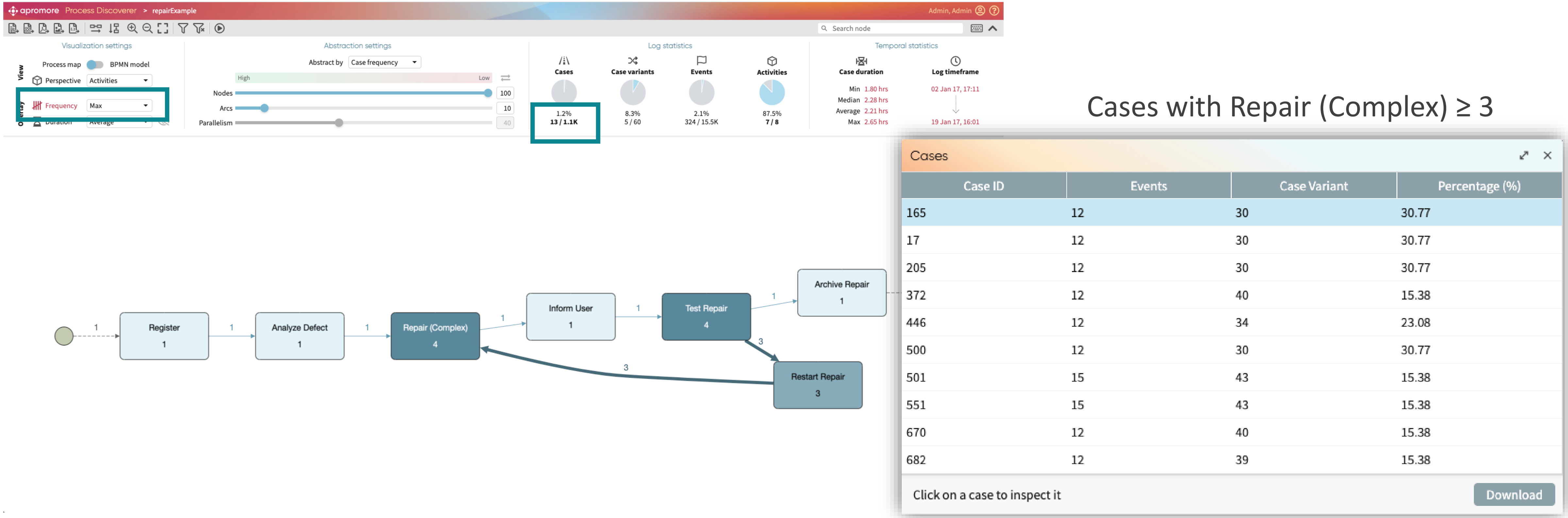
(3) How many cases showcase this repetition? How often does it occur?

(4) <u>Which cases are affected exactly? And how do they differ from cases where there is no repetition?</u>



Rework analysis

To **assess the impact of rework on process performance (time and cost)**, we should apply filtering to isolate cases that are affected by rework.



apromore Process Discoverer > repairExample

Admin, Admin

Search node

Visualization settings

View: Process map / BPMN model

Perspective: Activities

Overlay: Frequency: Max

Duration: Average

Abstraction settings

Abstract by: Case frequency

Nodes: 100

Arcs: 10

Parallelism: 40

Log statistics

Cases: 1.2% **13 / 1.1K**

Case variants: 8.3% 5 / 60

Events: 2.1% 324 / 15.5K

Activities: 87.5% **7 / 8**

Temporal statistics

Case duration: Min 1.80 hrs, Median 2.28 hrs, Average 2.21 hrs, Max 2.65 hrs

Log timeframe: 02 Jan 17, 17:11 – 19 Jan 17, 16:01

Cases with Repair (Complex) ≥ 3

Cases

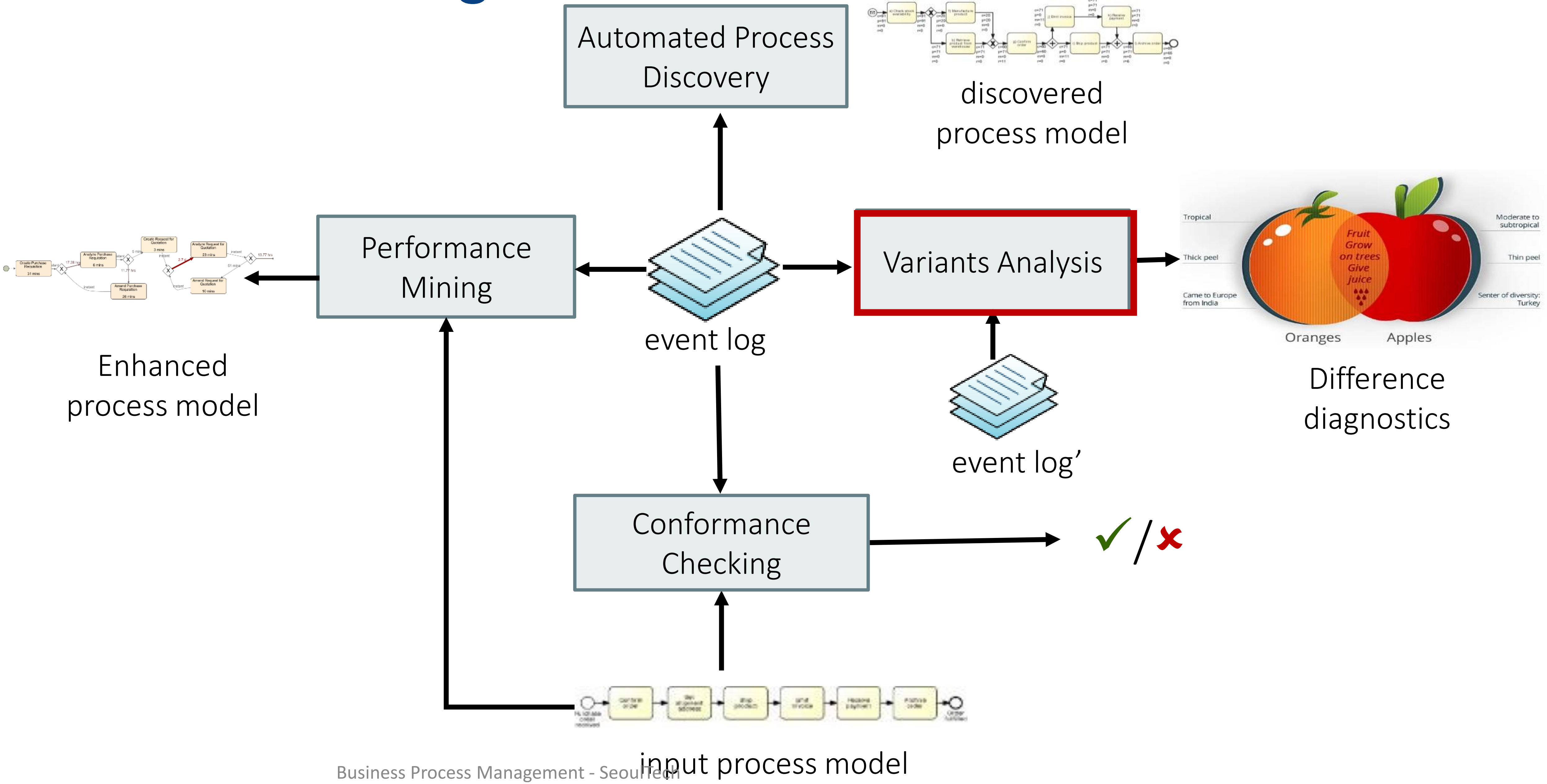
Case ID	Events	Case Variant	Percentage (%)
165	12	30	30.77
17	12	30	30.77
205	12	30	30.77
372	12	40	15.38
446	12	34	23.08
500	12	30	30.77
501	15	43	15.38
551	15	43	15.38
670	12	40	15.38
682	12	39	15.38

Click on a case to inspect it

Performance Mining: Analysis Template

What?	How?
Bottleneck analysis • Identify activity bottlenecks (high-effort activities) • Identify waiting bottlenecks (slow handoffs and capacity bottlenecks) • Identify resource bottlenecks	• In the process map, select the duration overlay (average or total duration): activity bottlenecks have long processing time; slow handoffs between activities are arcs with long waiting times; resource-capacity bottlenecks relate to activities with all/most incoming arcs with long waiting times • Switch to the Resources perspective in the map: slow handoffs between resources are arcs with long waiting times; resource bottlenecks are resources with long avg duration (slow resources) • Use slider with Average duration to focus on slowest or fastest nodes/arcs • Use Activities tab in Dashboard to visualize activity processing times • Use Animation plugin to visualize the build-up of bottlenecks.
Workload analysis • Identify overused vs underused resources/idleness waste • Analyze resource workload	• In the Resources perspective: overused resources have high total frequency or high “total duration”; underused resources have low total frequency and low total duration • Use also the Resources tab in Dashboard to visualize resource workload distribution • Use Filtering to focus on cases where the resource(s) with the highest workload are involved.
Rework analysis Identify defects/errors and rework waste	• Use the Max frequency overlay in the process map to visualize the max amount of repetitions of each activity / transition • Use the Case > Rework & repetition filter to retain cases with activity repetition; use the Case > Path filter with eventually-follows relation between an activity and itself to isolate cases with specific loops • Use the chart in the Activities Dashboard to visualize the distribution of max activity frequencies per case.

Process Mining



Automated Process Discovery

discovered process model

Performance Mining

event log

Variants Analysis

Enhanced process model

event log’

Difference diagnostics

Conformance Checking

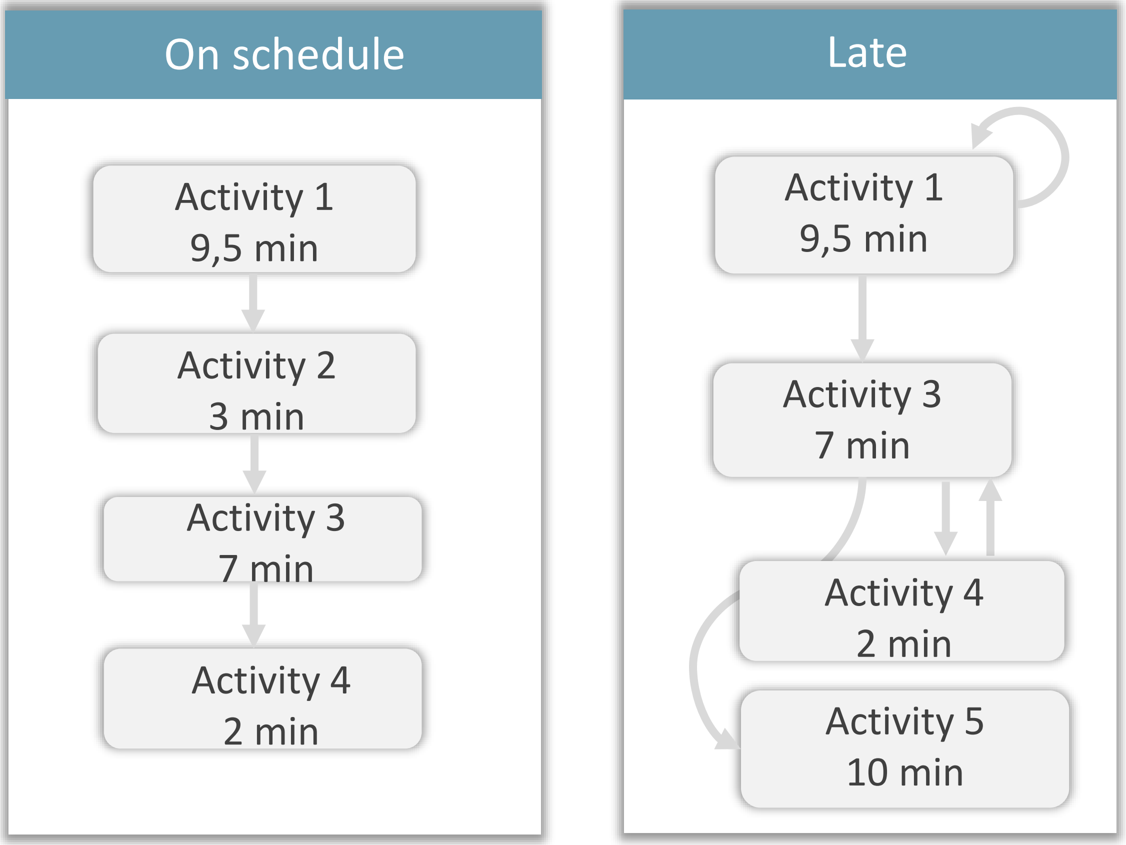
✓ / ✗

input process model

Tropical
Thick peel
Came to Europe from India
Fruit Grow on trees Give juice
Moderate to subtropical
Thin peel
Senter of diversity: Turkey
Oranges
Apples

Variant analysis

Given two or more logs representing variants of the same process, find the *differences* and *root causes* for variation or deviance between the logs.



The comparison of process variants offers insights into the the "why": Why do certain process cases differ from others?

- Why some take longer than others? (fast vs. slow cases)
- Why some end successfully, while others don't? (positive vs. negative outcomes)
- Why is employee XY more efficient and more successful when it comes to their activities than others? (resource XY vs. remaining resources)

Variant analysis: drivers

Performance

Identify and compare variants based on performance measures and their targets (e.g. slow vs fast cases)

Logical

Identify and compare variants based on case attributes (e.g. product type, customer segment, geographic area, loan amount, claim outcome)

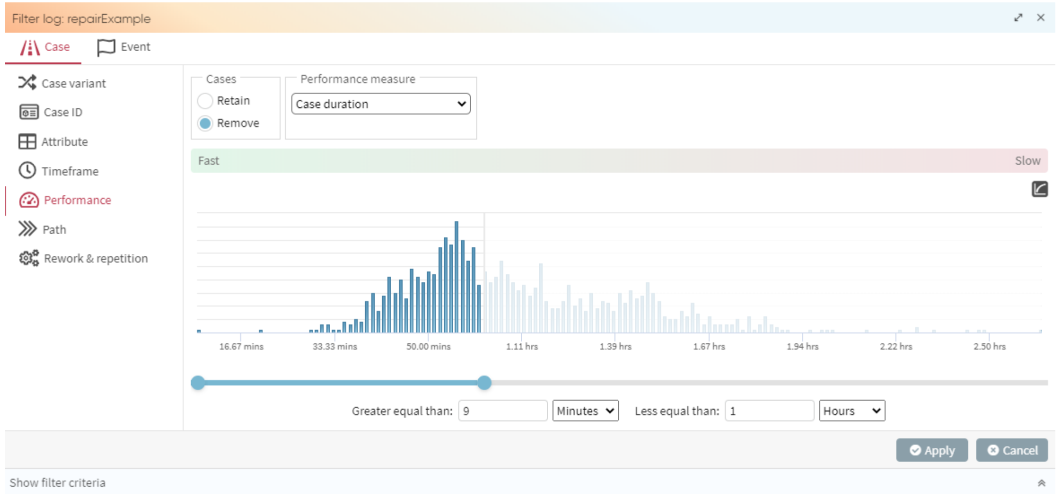
Temporal

Identify and compare variants based on different log timeframes (e.g. before and during COVID)

Variant analysis: approach

1. Apply the same filter one per variant, to isolate the respective sublog:
   - **Performance:** use Case > Performance filter
   - **Logical:** use Case > Attribute filter
   - **Temporal:** use Case > Timeframe filter
2. Save filtered process map and BPMN model back to the repository
3. Analyze the variants using one or more analysis techniques.

Apply filter



Filter log: repairExample

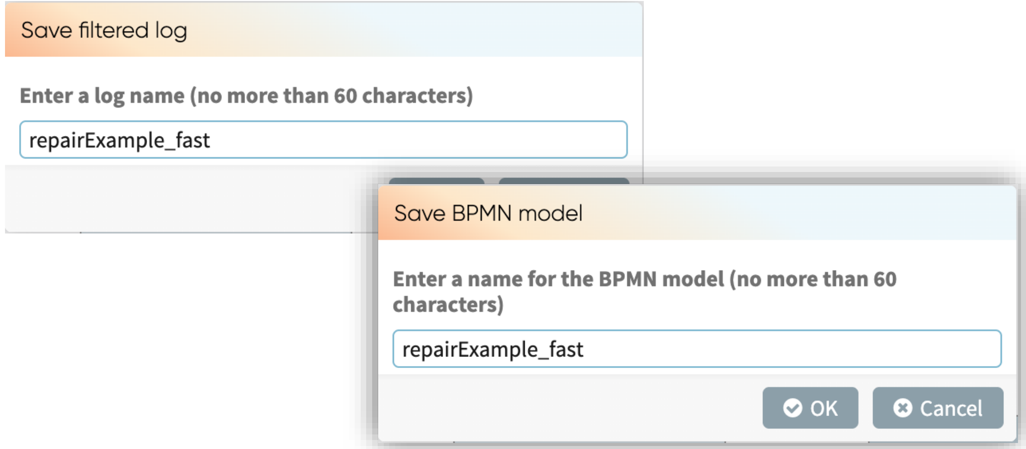
Cases: ○ Retain ● Remove

Performance measure: Case duration

Greater equal than: 9 Minutes

Less equal than: 1 Hours

Save filtered log & BPMN model



Save filtered log

Enter a log name (no more than 60 characters): repairExample_fast

Save BPMN model

Enter a name for the BPMN model (no more than 60 characters): repairExample_fast

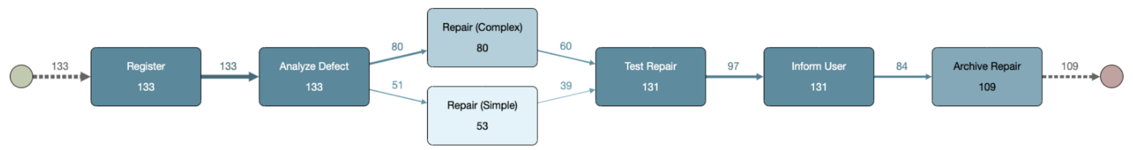
Variant analysis: comparison via maps/models

Visual comparison of process maps/BPMN models, to identify structural differences:

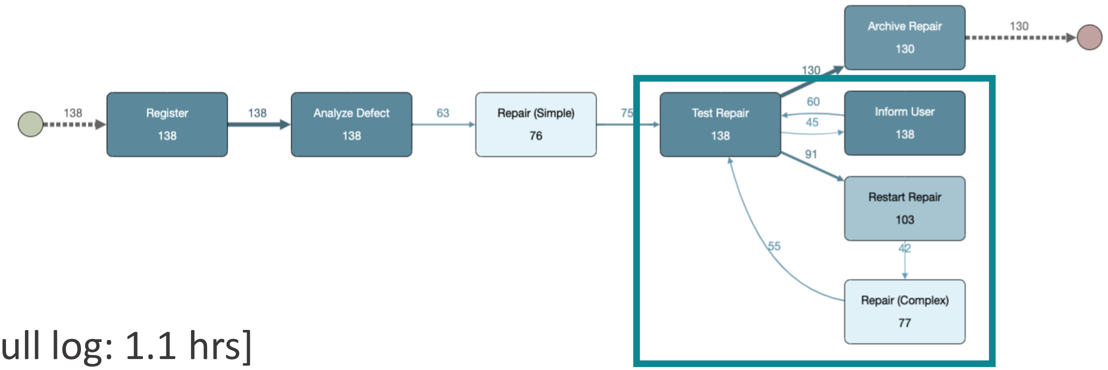
- Flow
- Frequency & rework
- Bottlenecks
- etc.

Consider different perspectives (Activities, Resources…)

Variant 1: “fast” (up to 45 min)



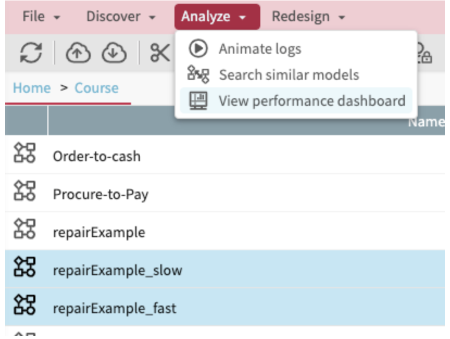
Variant 2: “slow” (more than 1.5 hrs)



[Avg duration for full log: 1.1 hrs]

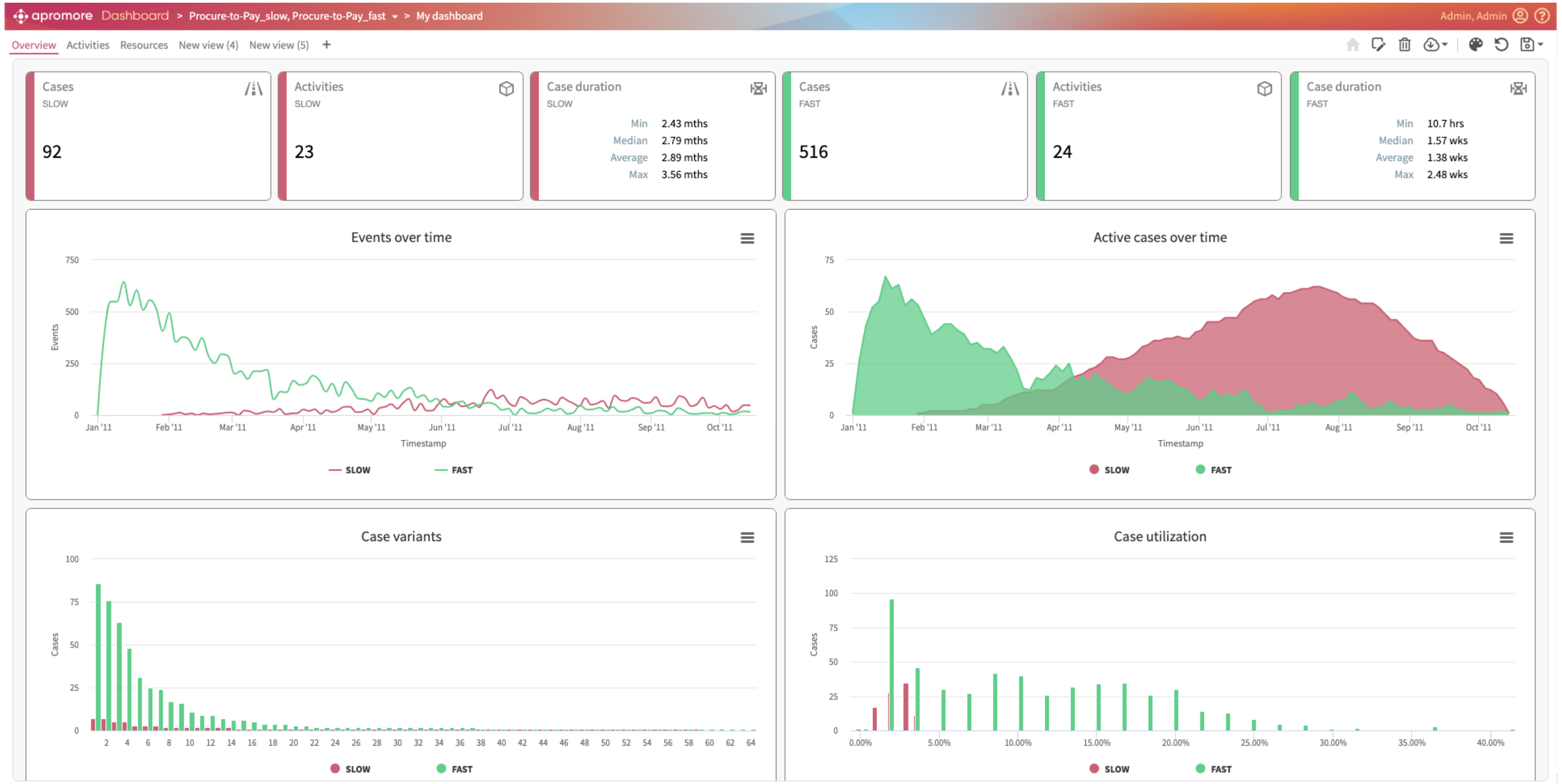
Variant analysis: comparison via dashboard

Statistical comparison via dashboard charts and tables. Variants are color-coded (e.g. red for slow, green for fast cases)



- Order-to-cash
- Procure-to-Pay
- repairExample
- repairExample_slow
- repairExample_fast

Select your filtered logs and launch the Dashboard. All dashboard functionalities are also available when comparing multiple variants.

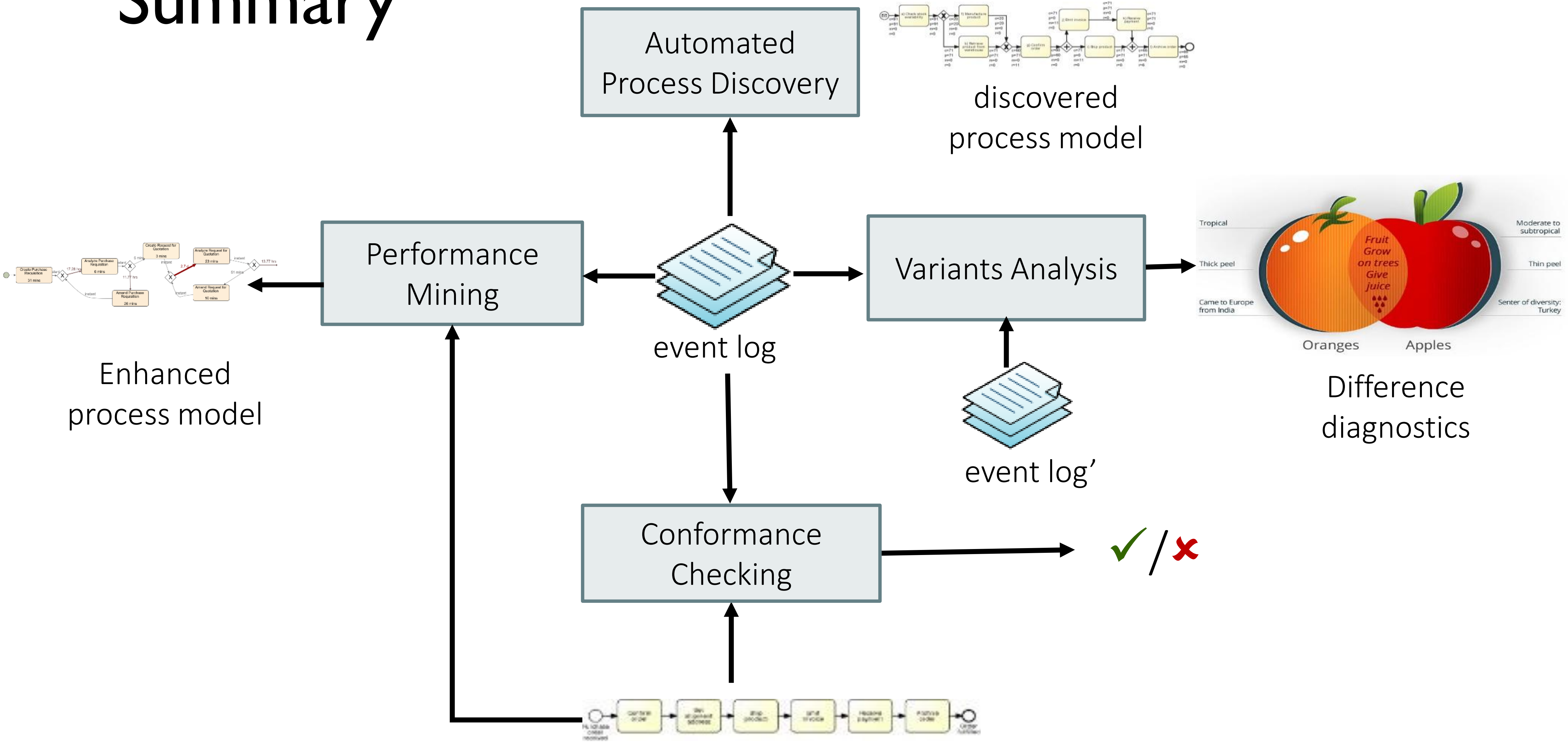

apromore Dashboard > Procure-to-Pay_slow, Procure-to-Pay_fast > My dashboard
Admin, Admin
Overview
Activities
Resources
New view (4)
New view (5)
+

Cases SLOW	Activities SLOW	Case duration SLOW	Cases FAST	Activities FAST	Case duration FAST
92	23	Min 2.43 mths Median 2.79 mths Average 2.89 mths Max 3.56 mths	516	24	Min 10.7 hrs Median 1.57 wks Average 1.38 wks Max 2.48 wks

Variant Analysis Template

What?	How?
Flow comparison Identify notable diffs in activity flow	• Discover a BPMN process model from the log of each variant. Visualize them side-by-side. Look for the gateways around each task to detect for example situations where two tasks are in parallel in one variant, but sequential in the other.
Frequency & rework comparison Compare activity or handoff frequency and rework	• Compare the maps of the process variants side-by-side (as above), but using the frequency overlay instead of the duration overlay. This allows us to detect the most frequent transitions between tasks. Look also for “thick” loop-backs. • Use a multi-low dashboard to compare frequencies of activities, resources or other attributes.
Bottleneck comparison Compare the location and magnitude of bottlenecks across variants	• Use filtering interface to retain all traces that fulfill the condition to belong to a variant (then repeat for the other). • Open the maps of the process variants side-by-side and compare the dependencies (arcs) with the highest waiting times. If the maps are too complex, use the abstraction slider with the abstraction metric “average duration” and the ordering from slow (right) to fast (left). Use the arc slider to retain the slowest dependencies. • Compare activity durations using side-by-side comparison of process maps or a multi-log dashboard (open the logs of 2+ variants in a single dashboard). • Consider comparing the two variants using the “resource” perspective to identify hand-offs between resources.

Summary



Automated Process Discovery

discovered process model

Performance Mining

Enhanced process model

event log

Variants Analysis

event log’

Difference diagnostics

Conformance Checking

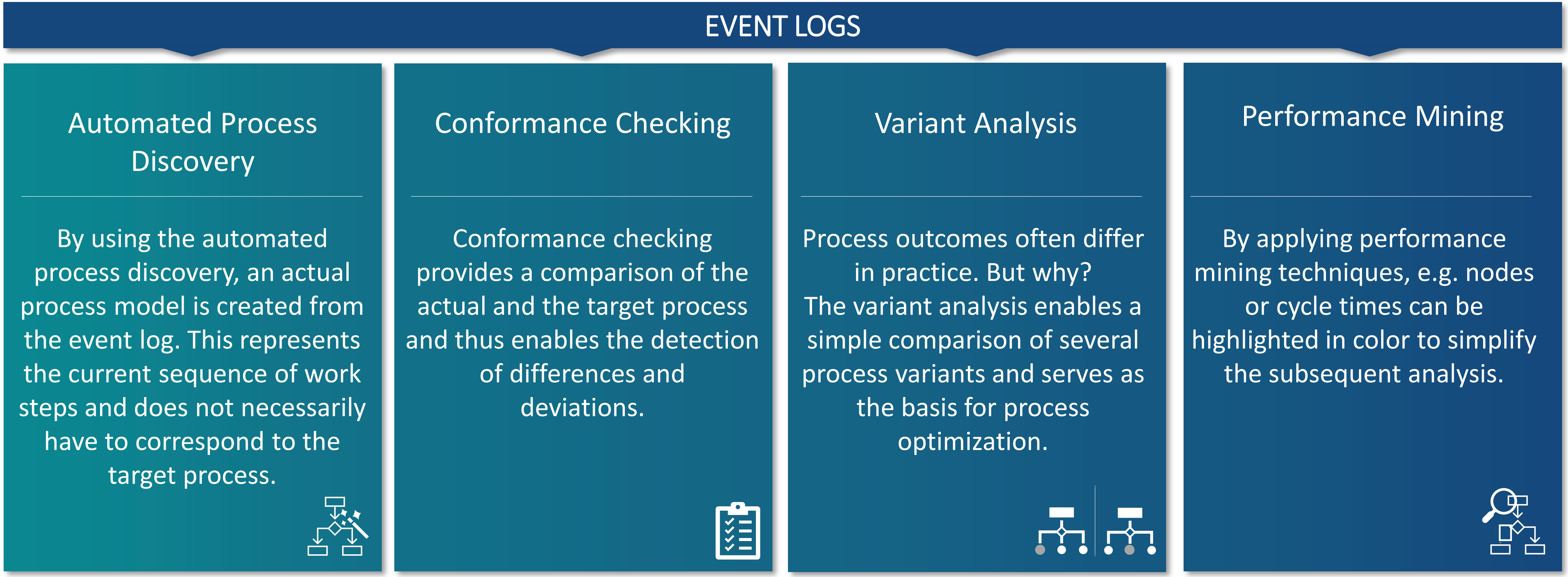
✓ / ✗

Tropical
Thick peel
Came to Europe from India
Fruit Grow on trees Give juice
Moderate to subtropical
Thin peel
Senter of diversity: Turkey
Oranges
Apples

input process model or compliance rules

Summary: Four Process Mining Capabilities

The bulk of process mining techniques can be divided into four capabilities.



EVENT LOGS

Automated Process Discovery

By using the automated process discovery, an actual process model is created from the event log. This represents the current sequence of work steps and does not necessarily have to correspond to the target process.

Conformance Checking

Conformance checking provides a comparison of the actual and the target process and thus enables the detection of differences and deviations.

Variant Analysis

Process outcomes often differ in practice. But why?
The variant analysis enables a simple comparison of several process variants and serves as the basis for process optimization.

Performance Mining

By applying performance mining techniques, e.g. nodes or cycle times can be highlighted in color to simplify the subsequent analysis.

Exercise 1

- Consider the purchase-to-pay sample log.
  1. Are there any cases where the supplier's invoice is released and authorized by the same resource? And if so, who is doing this most often?
  2. What's the *on-time completion rate for this process*, assuming that the service-level objective for completing this process is 21 days from the start to end?

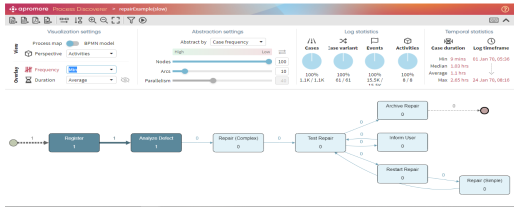
Exercise 2

- Consider the purchase-to-pay sample log.
  - Which task has the longest waiting time and which one has the longest processing time?
  - How much time is spent in rework waste in this process?
  - Which resource spends the highest amount of time (on average) performing work in the process?
  - Which handoffs between resources are responsible for the highest amount of waiting time?

Exercise 3 - Variant analysis

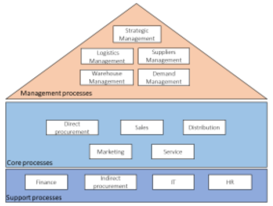
- Consider the purchase-to-pay sample log.
  - Is there a difference in cycle time for the cases that had to settle a dispute with the purchasing agent, compared to the ones that did not? Make sure you only compare cases that actually reach the endpoint 'Pay invoice'
  - What differences can you observe between cases that completed in 21 days or less and those that took more than 21 days (again only consider cases where "Pay invoice" occurs).

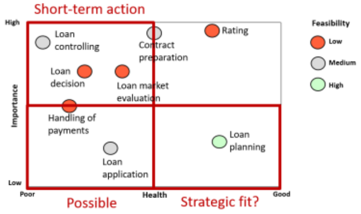
Business Process Management Course



Process identification

Process architecture

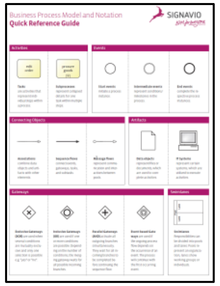


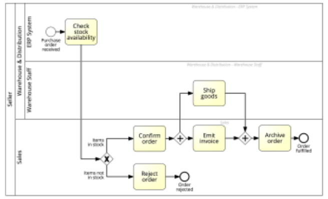


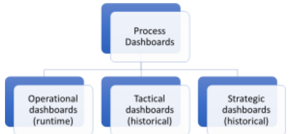
Process discovery

Conformance and performance insights

As-is process model

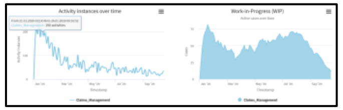


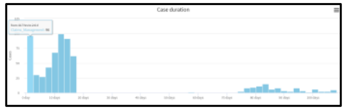


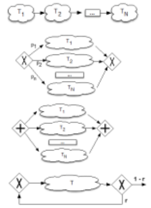

Process Dashboards
Operational dashboards (runtime)
Tactical dashboards (historical)
Strategic dashboards (historical)

Process monitoring and controlling

Process analysis







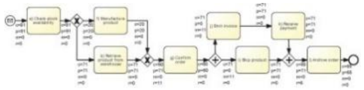
Executable process model

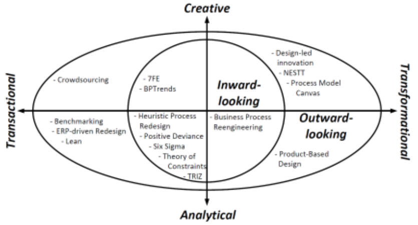
Insights on weaknesses and their impact

Process implementation

To-be process model

Process redesign





Acknowledgements

- The content notes for this lecture feature content borrowed with or without modification from the following sources:
  - "Source: M. Dumas, M. La Rosa, J. Mendling and H. Reijers, *Fundamentals of Business Process Management*, 2nd edition, Springer, 2018".
  - Chapter 11