

**Univerzita Pardubice**  
**Fakulta ekonomicko-správní**  
**Ústav systémového inženýrství a informatiky**

**Metody a nástroje Process Miningu**

**Bc. Marek Dvořák**

**Diplomová práce**

**2018**

Univerzita Pardubice  
Fakulta ekonomicko-správní  
Akademický rok: 2017/2018

## ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Marek Dvořák**  
Osobní číslo: **E15682**  
Studijní program: **N6209 Systémové inženýrství a informatika**  
Studijní obor: **Informatika ve veřejné správě**  
Název tématu: **Metody a nástroje process miningu**  
Zadávatel katedra: **Ústav systémového inženýrství a informatiky**

### Zásady pro vypracování:

Cílem práce je realizovat analýzu vybraných metod a nástrojů využívaných v oblasti process miningu.

Práce bude obsahovat:

- úvod do problematiky process miningu, objasnění základních pojmů;
- popis a charakteristika používaných metod a nástrojů pro process mining;
- na příkladu realizovaná analýza vybraných metod a nástrojů pro process mining;
- zhodnocení, doporučení, závěr.

Rozsah grafických prací:

Rozsah pracovní zprávy:

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

**DIŠEK, M., MUSIL, D., KOLESÁR, J.** *Analýza výrobního procesu a předpříprava dat pro process mining.* [cit. 2017-10-06]. Dostupné z

<http://www.opf.slu.cz/uds/konference/sbornik16>

**AALST, W.** *Process Mining: Data Science in Action. 2nd edition.* Berlin: Springer, 2016.

**AALST, W.** *Process Mining: Discovery, Conformance and Enhancement of Business Processes.* Berlin: Springer, 2011.

**FEREIRA, D. R.** *A Primer on Process Mining: Practical Skills with Python and Graphviz.* Berlin: Springer, 2017.

**Introduction to Process Mining with ProM.** [cit. 2017-10-06]. Dostupné z <https://www.futurelearn.com/courses/process-mining>

Vedoucí diplomové práce:

  
**doc. Ing. Pavel Petr, Ph.D.**

Ústav systémového inženýrství a informatiky


Konzultant diplomové práce:

**doc. Ing. Stanislava Šimonová, Ph.D.**


Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce: **1. září 2017**

Termín odevzdání diplomové práce: **30. dubna 2018**

  
doc. Ing. Romana Provančíková, Ph.D.  
děkanka

L.S.

  
doc. Ing. Pavel Petr, Ph.D.  
vedoucí ústavu

V Pardubicích dne 1. září 2017

## **PROHLÁŠENÍ**

Prohlašuji, že jsem tuto práci vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

V Pardubicích dne 30. 4. 2018

Bc. Marek Dvořák

## **PODĚKOVÁNÍ:**

Tímto bych rád poděkoval svému vedoucímu práce doc. Ing. Pavlu Petrovi, Ph.D. za jeho odbornou pomoc, věcné připomínky, vstřícnost při konzultacích a poskytnuté materiály, které mi pomohly při zpracování diplomové práce.

Děkuji také své rodině za podporu, které se mi dostávalo po celou dobu mých vysokoškolských studií.

## **ANOTACE**

*Diplomová práce se zabývá Process Miningem jakožto poměrně novou vědeckou disciplínou, která propojuje vědu o datech a vědu o procesech. Obsahem této práce je vysvětlení Process Miningu a jeho využití. Dále jsou v práci uvedeny podrobnosti o event logu jako o vstupu analýzy. Poté jsou uvedeny a porovnány nejznámější Process Miningové nástroje a nejvyužívanější metody pro analýzu procesu. V závěru práce je realizována vlastní analýza a je vytvořen reálný model procesu ve vybraných nástrojích.*

## **KLÍČOVÁ SLOVA**

*Dolování procesů, protokol událostí, věda o procesech, nástroje, metody, algoritmy, analýza, ProM, Disco, Celonis.*

## **TITLE**

Methods and tools of Process Mining

## **ANNOTATION**

*This diploma thesis deals with Process Mining as a relatively new scientific discipline that links Data Science and Process Science. The content of this work is an explanation of the Process Mining and its use. Further, the thesis is giving details of the event log as the input of the analysis. Then, the paper lists and compares the most well-known Process Mining tools and the most widely used process analysis methods. At the end of the thesis is realized own analysis and creation of real model of the process in selected tools.*

## **KEYWORDS**

*Process Mining, Event log, Process Science, tools, methods, algorithms, analysis, ProM, Disco, Celonis.*

# OBSAH

<b>ÚVOD .....</b>	<b>1</b>
<b>1 PROCESS MINING .....</b>	<b>2</b>
1.1 Historie Process Miningu .....	3
1.2 Definice Process Miningu .....	3
1.3 Rozdíl mezi Data Science a Process Science .....	4
1.3.1 Data Science .....	4
1.3.2 Process Science .....	6
1.3.3 Spojení Process a Data Science .....	7
1.4 Využití Process Miningu .....	8
1.5 Kroky v Process Miningu .....	9
1.5.1 Proces prozkoumávání .....	10
1.5.2 Kontrola shody .....	10
1.5.3 Vylepšení modelu .....	11
1.6 Budoucnost Process Miningu .....	11
<b>2 EVENT LOGY .....</b>	<b>13</b>
2.1 Zdroje dat .....	13
2.2 Struktura event logu .....	16
2.3 eXtensible Event Stream .....	19
2.3.1 Meta model XES formátu .....	21
2.3.2 Benefity XES .....	22
<b>3 NÁSTROJE .....</b>	<b>24</b>
3.1 Typy užití Process Miningu .....	25
3.1.1 Typ AD-HOC .....	25
3.1.2 Typ Opakování .....	25
3.1.3 Typ Standard .....	25
3.2 Nekomerční nástroje .....	27
3.2.1 ProM .....	27
3.2.2 PMLab .....	28
3.2.3 CoBeFra .....	29
3.2.4 RapidProM .....	29
3.2.5 Další nástroje .....	29
3.3 Komerční nástroje .....	30
3.3.1 Disco .....	31
3.3.2 Celonis .....	31
3.4 Kritéria .....	32
3.4.1 Import dat .....	32
3.4.2 Výstup modelu .....	33

3.4.3	Perspektiva výkonu.....	33
3.4.4	Organizační perspektiva.....	33
3.4.5	Podpora XES.....	34
3.4.6	Získání dat o událostech z jiných zdrojů.....	34
3.4.7	Filtrování.....	35
3.4.8	Automatické shlukování .....	36
3.4.9	Reporting a animace .....	36
3.5	Shrnutí.....	36
<b>4</b>	<b>METODY .....</b>	<b>38</b>
4.1	Heuristic miner.....	40
4.2	Inductive miner.....	40
4.3	Fuzzy Miner .....	41
<b>5</b>	<b>REALIZOVANÁ ANALÝZA.....</b>	<b>42</b>
5.1	Data .....	42
5.2	ProM.....	44
5.2.1	Import event logů v ProM.....	45
5.2.2	Statistika v ProM .....	46
5.2.3	Prozkoumávání reálného modelu procesu v ProM .....	48
5.2.4	Animace reálného modelu procesu v ProM.....	50
5.3	Disco.....	51
5.3.1	Import event logů v Disco.....	51
5.3.2	Statistika event logů v Disco.....	52
5.3.3	Prozkoumávání reálného modelu procesu v Disco .....	54
5.3.4	Animace reálného modelu procesu v Disco .....	56
5.4	Celonis.....	56
5.4.1	Import event logů v Celonisu.....	57
5.4.2	Statistika event logů v Celonisu.....	58
5.4.3	Prozkoumávání reálného modelu procesu v Celonisu .....	59
5.4.4	Animace reálného modelu procesu v Celonisu .....	62
5.5	Shrnutí.....	62
	<b>ZÁVĚR .....</b>	<b>65</b>
	<b>POUŽITÁ LITERATURA .....</b>	<b>69</b>



## SEZNAM TABULEK

Tabulka 1: Příklad event logu	16
Tabulka 2: Vysvětlení aktivit	17
Tabulka 3: Datové typy XES	21
Tabulka 4: Přehled komerčních nástrojů	30
Tabulka 5: Shrnutí nástrojů	37
Tabulka 6: Náhled event logu	42
Tabulka 7: Aktivity v event logu	43
Tabulka 8: Role v event logu	43

## SEZNAM ILUSTRACÍ

Obrázek 1: Disciplíny Data Science	5
Obrázek 2: Disciplíny Process Science	6
Obrázek 3: Spojení Data a Process Science	7
Obrázek 4: Kroky v Process Miningu	9
Obrázek 5: Průběh Process Miningu	15
Obrázek 6: Stromová struktura event logu	18
Obrázek 7: Fragment souboru XES	20
Obrázek 8: Meta model XES	22
Obrázek 9: Typy užití v Process Miningu	26
Obrázek 10: Prostředí v ProM	44
Obrázek 11: Import event logu v ProM	45
Obrázek 12: Konverze event logu v ProM	46
Obrázek 13: Statistika v ProM	46
Obrázek 14: Explorer event logu v ProM	47
Obrázek 15: Průvodce Fuzzy minerem v ProM	48
Obrázek 16: Reálný model procesu v ProM	49
Obrázek 17: Animace reálného modelu procesu v ProM	50
Obrázek 18: Úvodní stránka v Discu	51
Obrázek 19: Import event logů v Discu	52
Obrázek 20: Statistika event logů v Discu	53
Obrázek 21: Frekvence a výkon reálného modelu procesu v Discu	54
Obrázek 22: Reálný model procesu v Discu	55
Obrázek 23: Animace reálného modelu procesu v Discu	56
Obrázek 24: Úvodní stránka v Celonisu	57
Obrázek 25: Import event logů v Celonisu	57
Obrázek 26: Happy path v Celonisu	58
Obrázek 27: Počet případů ve dnech v Celonisu	58
Obrázek 28: Celková doba trvání aktivit v Celonisu	59
Obrázek 29: Frekvence aktivit a průměrná doba trvání případů v Celonisu	60
Obrázek 30: Model reálného procesu v Celonisu	61
Obrázek 31: Animace reálného modelu procesu v Celonisu	62

## **SEZNAM ZKRATEK**

ABPD	Automated Business Process Discovery
API	Application Programming Interface
BPM	Business Process Modeling
BPMN	Business Process Model and Notation
CSV	Comma-Separated Values
DSC	Disco Project Files
EDS	Enterprise Discovery Suite
EPC	Event-Driven Process Chain
ERP	Enterprise Resource Planning
FXL	Disco Log Files
GMT	Greenwich Mean Time
HTML	Hyper Text Markup Language
IBM	International Business Machines
IEEE	Institute of Electrical and Electronics Engineers
IT	Information Technology
KPI	Key Performance Indicator
MXML	Mining eXtensible Markup Language
OLAP	Online Analytical Processing
PNML	Petri Net Markup Language
PPM	Process Performance Manager
RACI	Responsibility Assignment Matrix
RDBMS	Relational Database Management System
SAP	Systems – Applications – Products
SQL	Structured Query Language
UML	Unified Modeling Language
XES	eXtensible Event Stream
XML	eXtensible Markup Language
YAWL	Yet Another Workflow Language

# ÚVOD

V dnešní době informační systémy generují mnohem více dat, než jak tomu bylo v minulosti. Jejich objem roste exponenciálním tempem a nejenak je tomu u dat, která se pojí s konkrétními podnikovými procesy. Zájem o Data Science rychle roste a mnozí považují tuto disciplínu za povolání budoucnosti. Stejně jako Computer Science se v 70. letech stala disciplínou, nyní jsme svědky rychlého vytváření výzkumných středisek v oblasti Data Science. Data, ať už ve velkém či malém měřítku, jsou především zásadní pro firmy a jejich význam a důležitost se bude postupně zvyšovat. Nestací se však soustředit pouze na ukládání a analýzu dat. Datový analytik musí také umět spojovat data s operačními procesy a být schopen klást správné otázky. [4]

Process Mining je nově se rozvíjející disciplínou, která překlenuje propast mezi Data Science (například Data Mining nebo Machine Learning) a Process Science (Business Process Management nebo Work-flow Management) a poskytuje tak nové prostředky pro zlepšení podnikových procesů v různých aplikačních oblastech. Většina podnikových informačních systémů dnes generuje tzv. „Event logy“, což jsou informace o událostech. Analýza event logů, tj. protokolů událostí, umožňuje pomocí Process Miningu organizacím diagnostikovat problémy, které jsou založené na faktech, a ne na fikci. Firmy tak skrze vytvořené modely mohou na procesy nahlížet úplně z jiného úhlu pohledu a získat tak přesný obraz skutečného stavu. Otevírá se jim nová možnost, jak zvýšit efektivitu dennodenně vykonávaných činností a uspořit tak čas i náklady firem. [26]

Process Mining se využívá především ve chvíli, kdy je potřeba ověřit, zda reálné procesy opravdu probíhají tak, jak byly původně navrhnuty nebo naopak mohla být fáze navrhnutí původních procesů podceněna a firma nemá aktuální obraz o svém každodenním chodu. Často je také Process Mining využíván ve chvíli, kdy je v prováděných činnostech neefektivita či slabina, ale nejsme si jisti, kde přesně nežádoucí jev nastává a jaké jsou jeho příčiny. Podoba reálného procesu se tedy zjišťuje na základě skutečných dat – event logů, nikoli pomocí rozhovorů se zaměstnanci a investigací napříč firmou nebo otestováním pouhého vzorku položek. [4], [26]

Cílem diplomové práce je definování Process Miningu jako disciplíny a objasnění základních pojmů. Dále se tato práce zabývá popisem vstupních dat tzv. event logů. V dalších kapitolách jsou představené přední nástroje pro Process Mining a nejpoužívanější metody pro analýzu procesu. Závěrem je realizována analýza pomocí metod Process Miningu.

# 1 PROCESS MINING

Nedávné průlomové výzkumy Process Miningu umožňují objevovat, analyzovat a zlepšovat podnikové procesy založené na protokolech události, tzv. event logů. Jak již bylo řečeno v úvodu, tyto event logy se využívají pro analýzu a tvorbu reálného modelu procesu. Aktivity prováděné lidmi, stroji a softwarem zanechávají stopy v již zmíněných event logech. Události, jako například zadání zákaznické objednávky do systému SAP, kontrola odbavení cestujících pro cestu letadlem, změna dávky léku pro pacienta určená doktorem nebo stavební agentura, která odmítá stavební povolení mají společné, že všechny jsou zaznamenány v informačních systémech. [5]

Objem dat a kapacita úložiště vzrostly obdivuhodně za poslední desetiletí. Obchodní procesy by tedy měly být řízeny, podporovány a vylepšeny spíše na základě údajů o událostech než na subjektivních názorech nebo zastaralé zkušenosti. Na základě stovek analýz podnikových procesů v organizacích na celém světě je dokázáno, že manažeři a uživatelé mají tendenci přeceňovat své znalosti svých vlastních procesů a výsledné objevené reálné modely na základě event logů se značně rozcházejí s původně navrhnutými modely. Výsledky Process Miningu mohou být tedy považovány za jakési „rentgenové záření“, které odhaluje to, co se konkrétně děje v nějakém procesu, diagnostikuje problémy a navrhuje správné vylepšení. Praktické využití Process Miningu a jeho spojitost s jinými zajímavými vědeckými výzvami dělají z něj horké téma v Business Process Managementu (BPM). [5], [2]

Cílem Process Miningu je zjistit, monitorovat a zlepšovat skutečné procesy získáním znalostí z event logů, které jsou v dnešní době k dispozici z jakéhokoliv informačního systému. Ačkoliv jsou event logy všude přítomné a management by je spíše měl využít k analýze reálných procesů, manažerská rozhodnutí jsou obvykle založená na grafech, místní politice nebo na management dashboardu, jenž často poskytuje rychlý pohled na klíčové ukazatele výkonnosti, které jsou relevantní pro konkrétní cíl nebo obchodní proces. [5], [26]

Znalosti skryté v event logech nelze změnit na informace, které lze spustit. Pokrok v oblasti Data Miningu umožnil nalézt cenné vzory ve velkých datových sítích a podporoval komplexní rozhodnutí založená na datech. Nicméně klasické problémy s Data Miningem jako je klasifikace, shlukování, regrese, učení s asociačními pravidly nejsou procesní. Proto přístupy BPM mají tendenci uchylovat se k modelům a Process Mining má za cíl překlenout propast mezi Data Miningem a BPM. Metaforicky, Process Mining můžeme vidět jako „rentgenové záření“, které pomáhá diagnostikovat, předvídat problémy a doporučit řešení. [5]

## 1.1 Historie Process Miningu

Process Mining představuje důležitý most mezi Data Miningem a Business Process Modellingem. Výzkum Process Miningu začal na Technologické Univerzitě v Eindhovenu v roce 1999. V té době existovalo jen málo údajů o event logech a počáteční techniky byly extrémně naivní, a proto v praxi nepoužitelné. Během posledního desetiletí se event logy staly snadno dostupnými a techniky Process Miningu se zdokonalily. Navíc algoritmy pro Process Mining byly implementovány v různých akademických a komerčních softwarech. Kromě toho dochází k rychle rostoucímu zájmu průmyslu a stále více dodavatelů softwaru začalo do svých nástrojů přidávat funkčnost Process Miningu. [31]

## 1.2 Definice Process Miningu

Procesy jsou nedílnou součástí dnešního světa, poskytují služby a vnitřní funkce v podnicích, vládních orgánech a organizacích po celém světě. Zatímco existuje dostatek systémů k podpoře provádění takových procesů, stávající postupy pro sledování a analýzu procesů nejsou tak dobré. Process Mining je unikátní disciplína, která je schopna zaplnit tyto mezery a poskytnout revoluční algoritmy pro analýzu a sledování reálných procesů. [5]

Podle Wil van der Aalsta lze definovat Process Mining jako techniku v oblasti Process Managementu, která podporuje analýzu podnikových reálných procesů na základě event logů. Během Process Miningu se na event logech zaznamenaných informačním systémem aplikují specializované algoritmy, které zjišťují trendy, vzory a informace s cílem zlepšit efektivitu a lépe porozumět procesům. [36]

Process Mining je také známý jako Automated Business Process Discovery (ABPD). Nicméně, některé zdroje tento termín uvádí v užším smyslu, aby se konkrétně týkal disciplín, u kterých jsou vstupem event logy a výstupem model podnikového reálného procesu. Termín Process Mining se používá v širším kontextu, aby odkazoval nejen na metody pro objevování procesních reálných modelů, ale také na disciplíny pro shodu původního modelu podnikového procesu s reálným modelem podnikového procesu, který je založen na event logech. [17], [12]

Spektrum Process Miningu je poměrně široké a zahrnuje techniky, jako je zjišťování procesů, kontrola shody, vylepšení modelu, zjišťování rolí, analýza slabých míst, predikce zbývajících průtoků času a doporučování dalších kroků. [5] V této diplomové práci je především zaměřeno na zjišťování procesů.

Process Mining se snaží v kontextu s procesy zodpovědět na tyto otázky [5]:

- Co se reálně stalo v minulosti?
- Proč k tomu došlo?
- Co se pravděpodobně stane v budoucnosti?
- Kdy a proč se organizace a lidé liší?
- Jak řídit procesy efektivněji?
- Jak přepracovat proces ke zlepšení výkonu?

### **1.3 Rozdíl mezi Data Science a Process Science**

Data Mining, Predictive Analysis nebo Machine Learning patřící do Data Science rostou v astronomické míře a firmy nyní hledají odborníky, kteří dokážou skrze firemní data pomoci ke zefektivnění obchodních rozhodnutí. Například IBM předpovídá, že do roku 2020 se počet pracovních míst pro všechny experty v oblasti dat v USA zvýší o 364 000 na celkových 2 720 000. [15]

Na druhé straně disciplíny jako Business Process Management, které patří do skupiny Process Science, si za cíl taktéž kladou efektivnější práci, avšak na základě optimalizovaných procesů. Práce se rychle mění v důsledku technologického, ekonomického a demografického vývoje. Nové digitální možnosti, pracovní postoje a modely spolupráce přinášejí revoluci ve způsobu práce. Na Process Mining se dá nahlížet jako na chybějící článek mezi Data Science a Process Science. Proto je v dnešní době mezi experty tolik populární, i když se zatím nedokázal tolik prosadit. [5], [29]

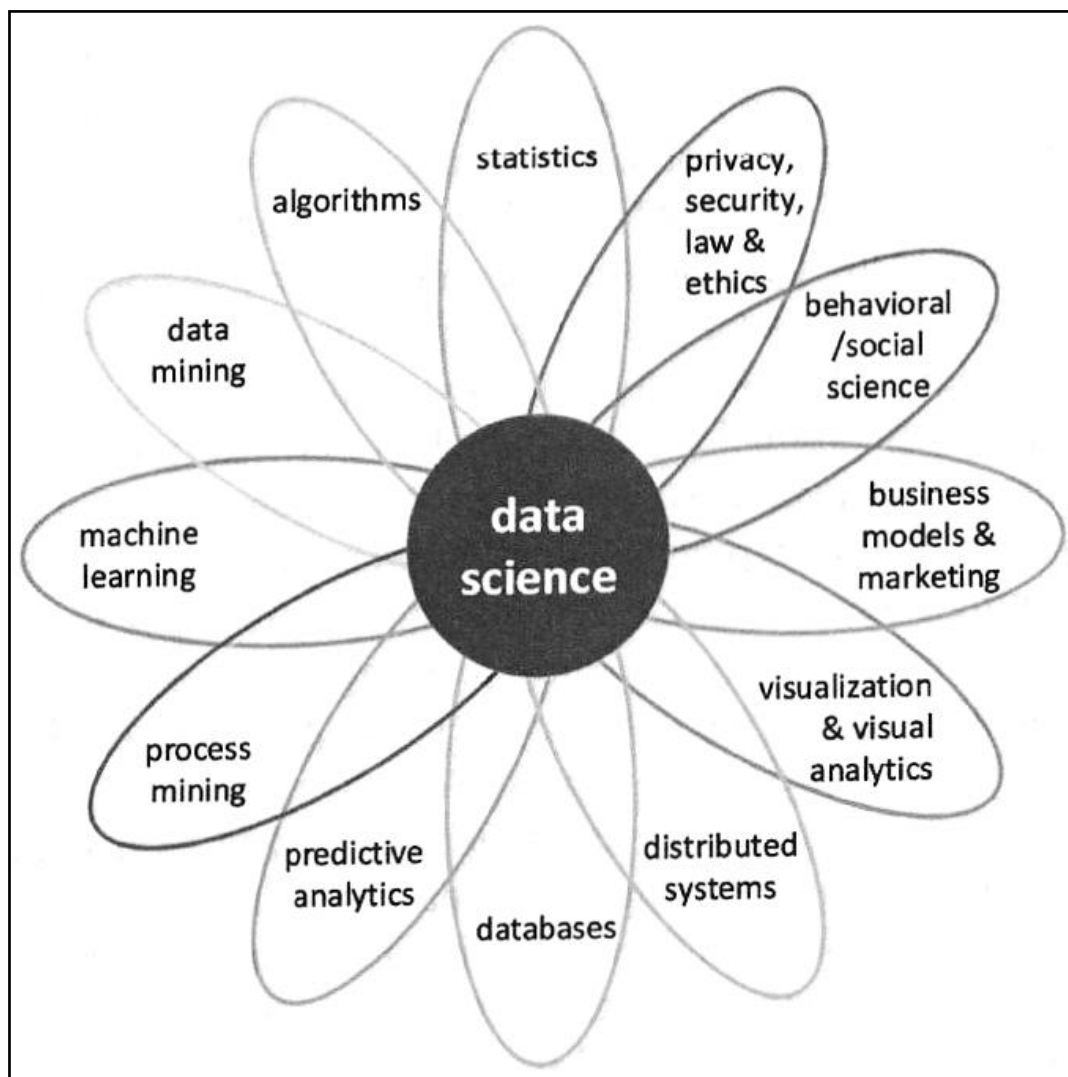
V dalších kapitolách budou vysvětlené Data Science a Process Science a umístění Process Miningu.

#### **1.3.1 Data Science**

Data Science, taktéž věda o datech, se v posledních letech ukázala jako nová vědecká disciplína a bylo pro ni navrženo mnoho definic. Wil van der Aalst specifikuje tuto vědu jako interdisciplinární oblast, jejímž cílem je přeměnit data na skutečnou hodnotu. Hodnota může být poskytována ve formě předpovědí, automatizovaných rozhodnutí, modelů načtených z dat nebo jakéhokoli typu vizualizace dat poskytující informace. Data Science zahrnuje získávání, přípravu, průzkum, transformaci nebo ukládání dat a různé typy dolování a učení. Dále prezentaci výstupů a využití výsledků s ohledem na etické, sociální, právní a obchodní aspekty. [5], [28]

Výše uvedená definice znamená, že Data Science je z kategoriálního hlediska širší pojem a nachází se na křižovatce stávajících disciplín. Jedná se tedy o sloučení různých částečně se překrývajících disciplín. Například původ mezi Data Miningem, Machine Learningem nebo statistikou jsou velmi rozdílné. Data Mining vznikl na základě potřeb z databázové komunity. Machine Learning se objevil z komunity umělé inteligence a statistika ze statistické komunity. Všechny tyto disciplíny mají tedy různý původ, ale přesto mají společné znaky a rozhodně se překrývají. Obrázek 1 znázorňuje hlavní složky Data Science. Překrývající disciplíny nemají jasné hranice, mění se a relativně vyvíjí podle potřeb. Jedním z nich je i Process Mining. [5]

Data Science se nyní využívá kvůli všudypřítomnosti a hojnosti údajů, a proto je vysoká poptávka po datových vědcích, kteří mohou přeměnit data na hodnotu a odpovědět na řadu otázek. Ty se dají seskupit do následujících čtyř hlavních kategorií – Reporting, Diagnóza, Predikce a Doporučení. [13]



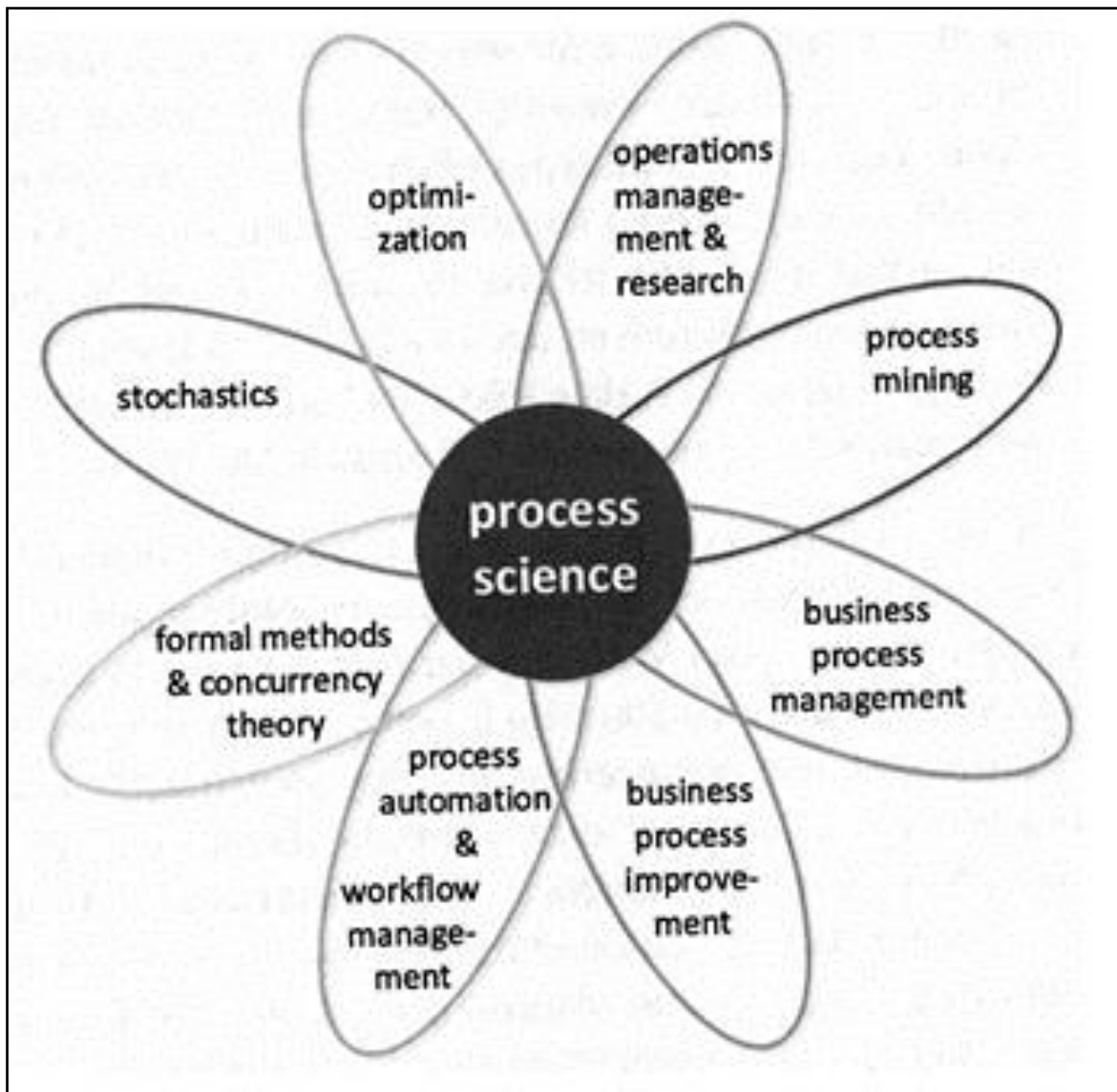
Obrázek 1: Disciplíny Data Science

Zdroj: [5]

### 1.3.2 Process Science

Process Mining je, jak již bylo zmíněno, prostředek k překlenutí propasti mezi Data Science a Process Science. Přístupy v oblasti věd o Data Science mají tendenci přeměnit data na skutečnou hodnotu, zatímco přístupy v Process Science mají tendenci být modelově řízené bez ohledu na informace skryté v datech. Proto se využívá zastřešovacího termínu Process Science jako širší oblasti, která kombinuje znalosti z informačních technologií a znalosti z vedoucích oborů pro zlepšení a provozování operačních procesů. Oproti Process Miningu, chybí ostatním disciplínám perspektiva v oblasti zpracování procesu. [5]

Obrázek 2 zobrazuje složky v Process Science, které se stejně jako v Data Science překrývají, nemají jasné hranice, mění se a relativně vyvíjí podle potřeb. [5]



Obrázek 2: Disciplíny Process Science

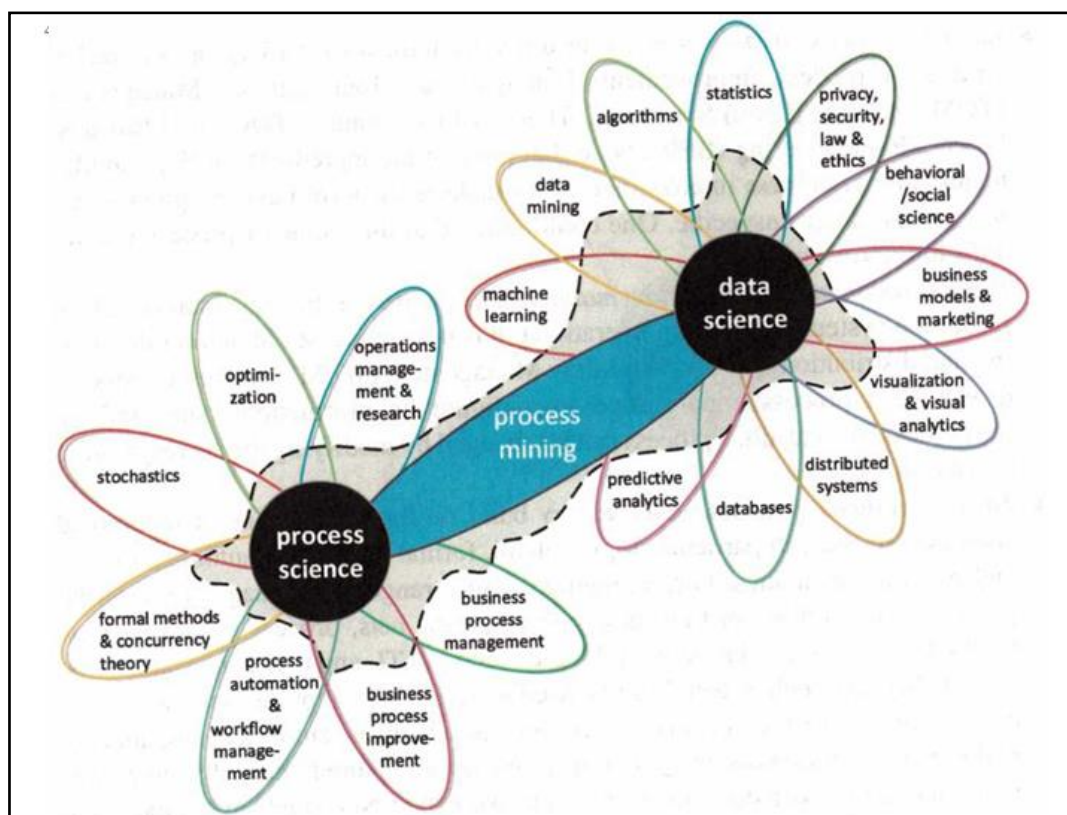
Zdroj: [5]



### 1.3.3 Spojení Process a Data Science

Process Mining sdružuje tradiční modelově založenou analýzu procesů a techniky analýz event logů. Jak je ukázáno na obrázku 3, Process Mining lze považovat za spojení mezi Data Science a Process Science. Úkolem Process Miningu je hledat spojení mezi event logy a modely procesů. Techniky Data Miningu, statistiky a Machine Learningu nezohledňují modely procesů typu end-to-end. Průmyslové přístupy procesů jsou procesně orientované, ale často se zaměřují spíše na modelování než na získání informací o událostech. Jedinečné umístění Process Miningu, jak je zobrazeno na obrázku 3 z něj činí silný nástroj pro využívání rostoucí dostupnosti dat pro zlepšení procesů. [5]

Je obtížné kombinovat všechny různé dovednosti potřebné pro jednu osobu, a proto je velkou výzvou najít nebo vzdělávat univerzální vědce v oblasti Data Science a Process Science schopné pokrýt celé spektrum. Tito univerzální výzkumní pracovníci jsou ve výsledku velmi žádaní a extrémně hodnotní pro organizace zabývající se daty a procesy. Jako alternativu je také možné vytvořit multidisciplinární týmy pokrývající spektrum z obrázku 1 a obrázku 2 a je důležité, aby členové týmu měli schopnost nadhledu a navzájem se doplňovali, zvláště pak co se dovedností týče. [32]



Obrázek 3: Spojení Data a Process Science

Zdroj: [5]

## 1.4 Využití Process Miningu

Process Mining, jak již bylo zmíněno dříve, se jako disciplína objevil teprve nedávno jak v oboru Data Science, tak i Process Science, ale příslušné techniky lze aplikovat na různé typy procesů.

Příklady využití [5], [26]:

- analýza léčebných procesů ve zdravotnických zařízeních,
- optimalizace procesů zákaznických služeb v nadnárodních korporacích,
- analýza chování spotřebitele v rezervačních systémech,
- analýza poruch zavazadlového manipulačního systému,
- zlepšení uživatelského rozhraní průmyslového stroje,
- zvýšení provozní účinnosti,
- řízení rizik,
- snížení chyb,
- kontrola firemních partnerů,
- úprava a přehodnocení stávajících procesů,
- změny v informačních systémech,
- zavedení dodatečných pravidel a kontrol,
- školení zaměstnanců eliminující neefektivní kroky.

Tato využití mají společné to, že jejich dynamické chování souvisí s procesními modely. Process Mining není pouze omezen na automatizované zjišťování procesů, ale lze ho také využít ke kontrole shody, diagnostikovat odchylky, zjistit nedostatky, zlepšit výkon, předpovědět tok času a doporučit akce. [5], [26]

Technologie Process Miningu jsou také generické, stejně jako software tabulkového procesoru. Event logy a operační procesy lze nalézt všude a analytické techniky nejsou omezeny na konkrétní aplikační domény. Stejně jako aplikace Excel může být využita v oblasti financí, výroby, prodeje, vzdělávání a sportu, může být Process Mining využit v různých aplikačních oblastech. [5]

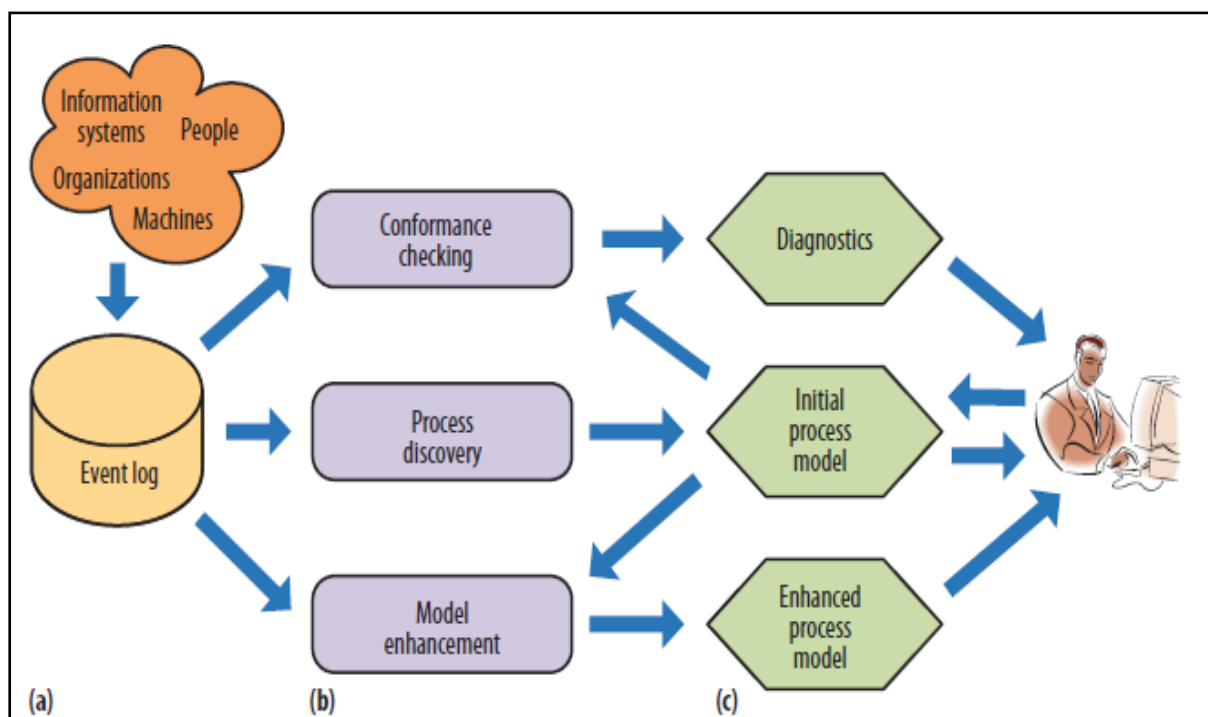
Process Mining je mocný nástroj, který mohou moderní organizace využít při řízení netriviálních provozních procesů. Rozvíjející se algoritmy a techniky umožňují analyzovat stále složitější event logy a sladit procesy s informacemi tak, aby splňovaly různé požadavky na dodržování předpisů, efektivitu a požadavky na služby zákazníkům. [36]

## 1.5 Kroky v Process Miningu

Kroky v Process Miningu se dají rozdělit do třech hlavních kategorií [6]:

- process discovery – jako proces prozkoumávání,
- conformance checking – jako kontrola shody,
- model enhancement – jako vylepšení modelu.

Obrázek 4 zobrazuje tyto 3 kroky a jejich fungování.



Obrázek 4: Kroky v Process Miningu

Zdroj: [5]

Obrázek 4 je rozdělen na 3 části – (a), (b), (c).

V první části (a) analytici používají techniky zjišťování procesů k extrapolaci, tj. bližšímu nastínění reálného modelu z event logů. [5]

Dále v části (b) analytici používají techniky kontroly shody pro diagnostiku odchylek mezi reálným modelem vytvořeným z event logů a původním procesním modelem. [6]

Nakonec v části (c), během vylepšování modelu, analytici používají informace z event logů k opravě nebo rozšíření reálného modelu. Například mohou přidat do modelu časové stopy pro časové informace (čekací doby a servisní časy). Výsledný vylepšený model procesu může zefektivnit rozhodování. [5], [6]

### 1.5.1 Proces prozkoumávání

Prvním krokem, a jedním z nejnáročnějších úkolů Process Miningu, je zjišťování procesů. Tato technika využívá event logy z informačního systému jako vstup a pomocí algoritmů generuje reálný model procesu bez použití jakékoliv předchozí informace. [5]

Pokud event log obsahuje informace o zdrojích, lze pomocí různých algoritmů vytvořit také modely, pomocí kterých je možné zjistit role, mechanismy distribuce práce a charakteristiky zdrojů. Analýza dalších atributů v event logu může přinést pravidla rozhodování, která vysvětlují možnosti, které se v procesu uskutečňují, například případy, kdy pracovník jako expert, důkladně kontroluje, zdali je činnost správně prováděna a případně ji zamítne. [5], [8]

### 1.5.2 Kontrola shody

Druhým krokem Process Miningu je kontrola shody. Kontrola shody se používá pro zjišťování odchylek, predikci, rozhodování a k systémovému doporučení. Při kontrole shody je reálný model vytvořený z event logů porovnán s původně navrženým modelem a odhalují se odchylky. [3]

Kontrola shody signalizuje potřebu lepší kontroly procesu. Nesrovnalosti mohou odhalit odchylky. Zaměstnanci se například mohou odchýlit od původního modelu, aby lépe vyhověli zákazníkům nebo aby zvládli jiné situace, které proces modelu nepředpokládá. Ve skutečnosti pak flexibilita a neshoda často korelují pozitivně. Srovnáním toho, co se skutečně stalo a co se mělo stát můžou odhalit žádoucí, ale i nežádoucí odchylky od původního procesu. [6]

Kontrola shody se nejčastěji využívá například při auditu, kdy se hodnotí organizace a jejich procesy za účelem ověření platnosti a spolehlivosti informací a s nimi související procesy. Tato kontrola se provádí, aby se ověřilo, zda jsou podnikové procesy prováděny v rámci určitých mezí, které stanovují manažeři, vlády a další zúčastněné strany. Je zřejmé, že Process Mining může pomoci odhalit podvody, špatné praktiky, rizika a neefektivitu. [5], [26]

Při kontrole shody je důležité se na odchylky dívat ze dvou úhlů pohledu [5]:

- a) model neodráží realitu, nabízí se otázka „Jak zlepšit model?“,
- b) případy se odchylují od modelu a je potřeba nápravných opatření, zde se ptáme „Jak zlepšit kontrolu s cílem prosadit lepší shodu?“.

Jako další příklad lze uvést bankovní sektor, kde můžeme zjistit, které případy se neřídí podle původního procesního modelu a proč tomu tak je. Analyzování event logů pomocí technik

kontroly shody zvyšuje průhlednost a vede k odhalování podvodů a je tím velmi užitečným nástrojem pro auditory. [5], [26]

### **1.5.3 Vylepšení modelu**

Třetí krok Process Miningu je vylepšování modelu, během kterého analytici zdokonalují nebo rozšiřují původní model pomocí informací o aktuálním procesu zaznamenaném v některém event logu. Zatímco kontrola shody měří sladění původního modelu s reálným modelem, tento třetí typ Process miningu se zaměřuje na změnu nebo rozšíření původního modelu. [5]

Jednou možností pro vylepšení je oprava, tj. úprava modelu tak, aby lépe odrážel skutečnost. Například pokud jsou dvě činnosti modelovány sekvenčně, ale ve skutečnosti mohou stát v libovolném pořadí, pak nastavení tohoto modelu může být změněno tak, aby tuto skutečnost odrážel. [5]

Dalším typem vylepšení je rozšíření, tj. přidání nové perspektivy do modelu procesu a jeho vzájemné propojení s event logem. Příkladem je rozšíření modelu procesu s údaji o výkonu přidáním časových stop souvisejících s případem za účelem změření doby trvání mezi dvěma následujícími aktivitami. Tyto informace mohou odhalit slabá místa a mohou být použity k predikci průběhu již běžících případů. Podobně lze i rozšířit informace o zdrojích a jiných attributech. [6]

## **1.6 Budoucnost Process Miningu**

Není pochyb o tom, že Process Mining má velký potenciál a dokáže se široce uplatnit v různých oborech a průmyslových odvětvích, a nejen v akademické sféře jako je tomu ve většině případů doteď. Firmy dnes potřebují řešení, které umožní rychle reagovat na dynamicky se měnící trh a růst dat. [5], [21]

Podle studie zmíněné v [5] se odhaduje, že množství digitálních informací generovaných například pomocí osobního počítače, digitálních fotoaparátů, serverů, senzorů atd. v roce 2014 již překročilo 4 Zettabyty a předpovídá se, že "Digital Universe" poroste exponenciálně nadále a to na 44 Zettabytů do roku 2020. To dokládá, že dlouho očekávaná exploze dat se stala nepopíratelnou realitou. Většina dat uložených v „Digital Universe“ je nestrukturovaná a organizace mají problémy s tak velkým množstvím dat pracovat. Jedním z hlavních úkolů dnešních organizací je získávání informací a hodnot z dat uložených v jejich informačních systémech. Význam informačních systémů není odrazem jen velkolepého růstu dat, ale také role, kterou tyto systémy hrají v dnešních obchodních procesech.

Například "stav banky" je určen především daty uloženými v informačním systému banky, protože se peníze staly převážně digitální entitou. Při rezervaci letu přes internet zákazník interaguje s mnoha organizacemi, například leteckou společností, cestovní kanceláří nebo bankou, přičemž o tom často ani neví. Pokud je rezervace úspěšná, obdrží zákazník e-ticket, což je v podstatě číslo. Jiný příklad, pokud systém SAP velké korporace ukáže, že určitý produkt není na skladě, není možné produkt prodat ani dopravit, přestože je k dispozici ve fyzické podobě. Tyto příklady jenom dokazují, že budoucnost Process Miningu je především v exponenciálním růstu dat. Současná situace se však nejeví příliš optimisticky, protože kritika se většinou soustřeďuje na to, jak je Process Mining příliš technický a složitý, aby byl samostatným nástrojem bez poradenských služeb. [5]

## 2 EVENT LOGY

Jak již bylo zmíněno v předchozí kapitole, předpokladem pro využití Process Miningu je existence protokolu událostí, tzv. event logů. Tato kapitola se proto bude věnovat získáváním a popisem event logů.

Každý případ v event logu odkazuje na událost, aktivitu a bod v čase. Event log může být chápán jako soubor případů a případ jako sled událostí. Event logy mohou obsahovat další doplňující údaje o událostech, například zdroj (osoba nebo zařízení), který provádí nebo iniciuje aktivitu, časovou stopu události a další datové atributy, jako například náklady nebo role. Události patřící k případu jsou uspořádány a popisují jeden "průběh" v procesu. Po nalezení příslušných dat (event logů) je jejich extrakce a konverze poměrně přímočará. [9]

### 2.1 Zdroje dat

Velkou výzvou před samotnou analýzou dat je extrakce dat. Extrahovaná data však obsahují spoustu nechtěných informací a je potřeba je očistit, transformovat, zredukovat a vyfiltrovat pro jeden zkoumaný proces. Důležitou roli při transformaci hraje syntaxe a sémantika, a protože Process Mining je disciplína, která vychází z vědy o datech je taktéž důležité věnovat pozornost kvalitě dat. [5], [18]

Event logy se nacházejí prakticky všude a mohou mít různou formu, například [18]:

- databázový systém – například data ve formě tabulek,
- různé protokoly – například protokol transakcí, zpráv,
- systém ERP – například, SAP, Oracle,
- API – data z webových stránek nebo sociálních médií.

Výchozím bodem při získávání event logů jsou "raw" (surové) údaje skryté ve všech typech forem dat. Event logy mohou mít podobu jednoduché tabulky nebo databázové tabulky. Jedná se tedy o naprosto surová data a nelze očekávat, že všechna budou v jediném dobře strukturovaném zdroji. Realita spočívá v tom, že event logy jsou obvykle rozptýleny v různých zdrojích a často je třeba vynaložit značné úsilí na shromažďování příslušných údajů. [5]

Například plná implementace SAP může obsahovat více než 10 000 tabulek a event logy mohou být rozptýleny na určitých úrovních oddělení z technických nebo organizačních důvodů. Nejsou výjimkou ani event logy rozptýlené v několika organizacích. [18]

Zdroje dat mohou být strukturovány a dobře popsány pomocí metadat. Bohužel v mnoha situacích jsou data nestrukturované nebo chybí důležitá metadata. I když jsou event logy strukturovány a popsány metadaty, mohou být podnikové informační systémy velice komplikované. Nemá smysl vyčerpávajícím způsobem extrahovat event logy z tisíců tabulek a dalších zdrojů dat. Extrakce dat by měla být motivována spíše otázkami než dostupností dat. [5]

V kontextu s Business Intelligence a Data Miningem se výraz "Extract, Transform and Load" (ETL) používá k popisu procesu, který zahrnuje [5]:

1. extrahování dat z vnějších zdrojů,
2. přeměnu dat na provozní potřeby (řešení syntaktických a sémantických problémů při zajištění předem definovaných úrovní kvality),
3. načtení dat do cílového systému, například datového skladu nebo relační databáze.

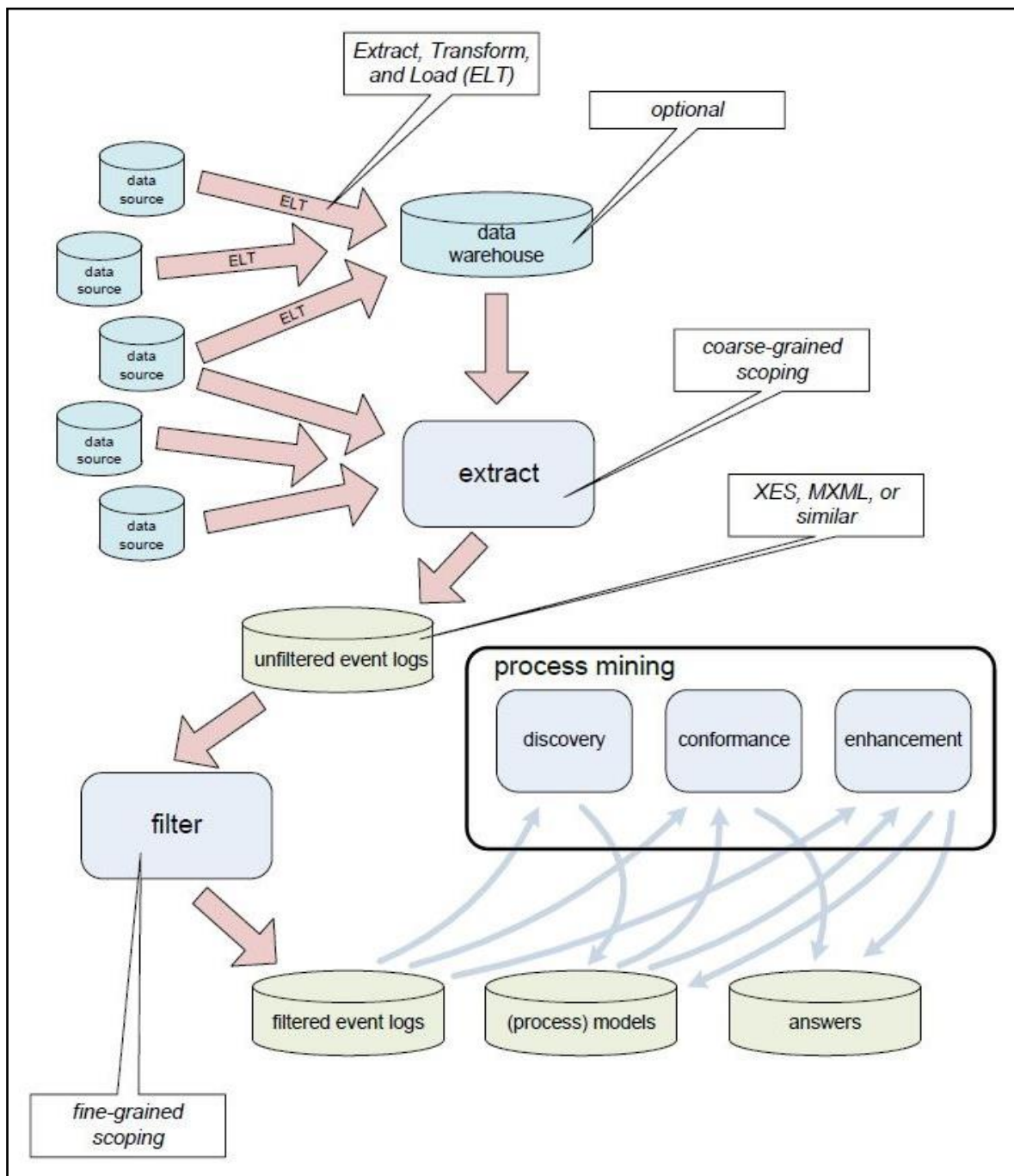
Obrázek 5 zobrazuje celkový průběh Process Miningu, který zdůrazňuje úlohu event logů. Datový sklad je jediným logickým repositářem transakčních a provozních dat organizace. Datový sklad nevytváří data, ale jednoduše odděluje data od operačních systémů. Cílem je sjednotit informace tak, aby mohly být použity pro reportování, analýzu, prognózu apod. Obrázek 5 ukazuje, že aktivity ETL lze použít k naplnění datového skladu, což může být značně složité, protože datový sklad by měl fungovat na základě ucelených pravidel, například klíčů nebo konvencí. [5]

Například jeden zdroj dat může identifikovat pacienta podle jeho příjmení a data narození, zatímco jiný zdroj dat používá své číslo sociálního zabezpečení. Jeden zdroj dat může používat formát data "31-12-2010", zatímco jiný používá formát "2010/12/31". Pokud datový sklad již existuje, je s největší pravděpodobností cenným vstupem pro Process Mining. Mnoho organizací však nemá dobrý datový sklad – často obsahuje pouze podmnožinu informací potřebných pro Process Mining, například bývají uložena pouze data týkající se zákazníků, a ne celého procesu. Navíc, pokud firmy datovým skladem disponují, nemusí být orientovány na proces. Ať už firmy mají datový sklad nebo nikoliv, musí být data extrahována a převedena do event logů. [5]

Typické formáty pro ukládání event logů jsou XES (eXtensible Event Stream) a MXML (Mining eXtensible Markup Language). Ty budou popsány v další kapitole 2.3. [5] V této diplomové práci předpokládáme, že jeden event log odpovídá jednomu procesu, tj. při stanovení rozsahu údajů v extrakčním kroku by měly být zahrnuty pouze události důležité pro jeden analyzovaný proces.



Event log je po vytvoření obvykle filtrován. Filtrování je iterativní, tj. opakovaný proces. Při extrakci dat do event logu je využíván „Coarse-grained scoping“. „Fine-grained scoping“ koresponduje s filtrováním na základě prvotních výsledků analýzy. Na základě filtrovaného event logu jsou popsány kroky Process Miningu, které již byly vysvětleny v předchozí kapitole 1.5. Je nutné poznamenat, že výsledky mohou vyvolat nové otázky, které povedou k podrobnějším extrakcím dat a k opakováním fází Process Miningu. [5]



Obrázek 5: Průběh Process Miningu

Zdroj: [5]

## 2.2 Struktura event logu

Po poskytnutí přehledu, průběhu a zařazení Process Miningu se tato kapitola zaměří na strukturu event logu a objasnění některých důležitých pojmů. Event logy by měly být uloženy ve formátu XES především kvůli jeho výhodám, které jsou popsány v kapitole 2.3.2. Formát XES dnes využívá většina Process Miningových nástrojů. Tento formát založený na schématu XML bude později vysvětlený v další kapitole 2.3. Tabulka 1 ilustruje náhled na část event logu z procesu vyřizování žádostí o odškodnění. Tabulka 1 zobrazuje v záhlaví sloupců [5]:

- case ID – ID případu,
- event ID – ID události,
- properties (vlastnosti) – timestamp (časová stopa), activity (aktivita), resource (zdroj) a cost (náklady).

**Tabulka 1: Příklad event logu**

Case ID	Event ID	Properties			
		Timestamp	Activity	Resource	Cost
1	35654423	30-12-2010:11.02	register request	Pete	50
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400
	35654425	05-01-2011:15.12	check ticket	Mike	100
	35654426	06-01-2011:11.18	decide	Sara	200
	35654427	07-01-2011:14.24	reject request	Pete	200
2	35654483	30-12-2010:11.32	register request	Mike	50
	35654485	30-12-2010:12.12	check ticket	Mike	100
	35654487	30-12-2010:14.16	examine casually	Pete	400
	35654488	05-01-2011:11.22	decide	Sara	200
	35654489	08-01-2011:12.05	pay compensation	Ellen	200
3	35654521	30-12-2010:14.32	register request	Pete	50
	35654522	30-12-2010:15.06	examine casualv	Mike	400
	35654524	30-12-2010:16.34	check ticket	Ellen	100
	35654525	06-01-2011:09.18	decide	Sara	200
	35654526	06-01-2011:12.18	reinitiate ivuuest	Sara	200
	35654527	06-01-2011:13.06	examine thoroughly	Sean	400
	35654530	08-01-2011:11.43	check ticket	Pete	100
	35654531	09-01-2011:09.55	decide	Sara	200
	35654533	15-01-2011:10.45	pay compensation	Ellen	200
4	35654641	06-01-2011:15.02	register request	Pete	50
	35654643	07-01-2011:12.06	check ticket	Mike	100
	35654644	08-01-2011:14.43	examine thoroughly	Sean	400
	35654645	09-01-2011:12.02	decide	Sara	200
	35654647	12-01-2011:15.44	reject request	Ellen	200

*Zdroj: [5]*

V případě aktivit jsou jednotlivé činnosti vysvětleny v tabulce 2.

**Tabulka 2: Vysvětlení aktivit**

Register request	Žádost o registraci
Examine thoroughly	Důkladné prověření
Examine casually	Obvyklé prověření
Check ticket	Kontrola ticketu
Decide	Rozhodnutí
Reinitiate request	Opakování žádosti
Reject request	Odmítnutí žádosti
Pay compensation	Peněžní kompenzace

*Zdroj: Vlastní zpracování*

Každý řádek v tabulce 1 prezentuje jednu událost. V tomto případě jsou již události spojené s jednotlivými případy. Například k případu s ID 1 je již přidruženo celkem 5 událostí s časovou známkou, aktivitou, zdrojem a náklady. Tyto vlastnosti události jsou její atributy. První událostí případu 1 je provedení aktivity žádosti o registraci Petrem ze dne 30. prosince 2010. Tabulka 1 také zobrazuje unikátní identifikační číslo této události: 35654423. ID se používá pouze k identifikaci události, například k jejímu odlišení od události 35654483, která rovněž odpovídá provedení žádosti o registraci. [5]

Tabulka 1 zobrazuje datum a časové razítko pro každou událost. V některých případech nemá event log doplňující atributy, v jiných mohou být podrobnější informace o času. Například, kdy byla aktivita zahájena, kdy byla dokončena nebo jak dlouho trvala. Časová známka uvedená v tomto příkladu je interpretovaná jako doba dokončení. [5]

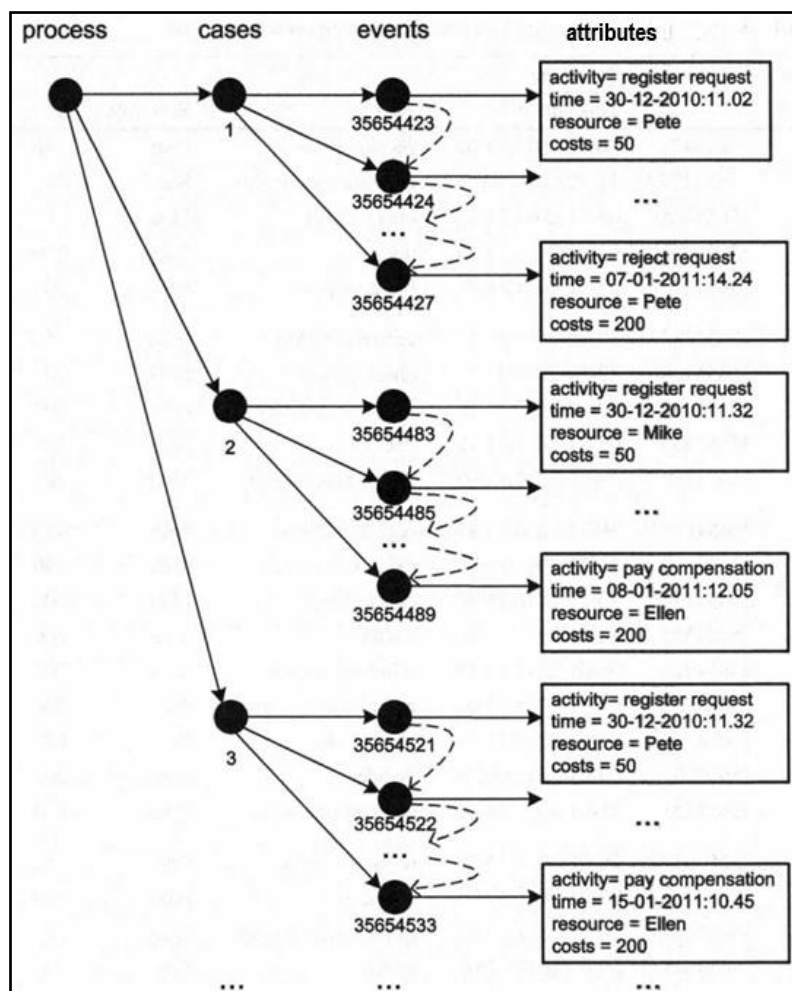
Dále lze z tabulky 1 vyčíst, že každá událost je přidružena ke zdroji, tedy zaměstnancům vykonávající aktivity. Je zde rovněž uveden doplňující atribut náklady, který je spojen s událostmi. Může existovat mnoho dalších datových atributů. Například v tomto konkrétním příkladu by bylo zajímavé zaznamenat výsledek různých typů zkoušek a kontrol. Dalším datovým prvkem, který by mohl být pro analýzu užitečný, je požadovaná výše kompenzace. Tabulka 1 zobrazuje typické informace obsažené v event logech. V závislosti na použité technologii Process Miningu a na dotyčných otázkách se používá pouze část těchto informací. Minimální požadavky na Process Mining spočívají v tom, že jakákoliv událost může souviset jak s případem, tak s aktivitou. [5]

V tabulce 1 se události vztahují k aktivitám jako je žádost o registraci, kontrola ticketu nebo zamítnutí žádosti. Tyto předpoklady jsou zcela přirozené v kontextu Process Miningu. Všechny hlavní procesní modelovací notace specifikuje proces jako sbírku aktivit tak, aby byl

popsán jeho životní cyklus. Proto sloupce ID události a aktivity představují minimum pro Process Mining. Kromě toho je třeba věnovat pozornost událostem v daném případě. Například událost 35654423 (aktivita žádost o registraci pro případ 1) nastane před událostí 35654424 (aktivita důkladné prověření pro stejný případ). Bez informací o procesu je nemožné odhalit kauzální závislosti v procesních modelech. [5]

Na event log zobrazený v tabulce 1 lze nahlížet ještě z jiného úhlu pohledu. Obrázek 6 ilustruje rafinovanější stromovou strukturu. Pomocí této ilustrace můžeme uvést předpoklady o event logech [5]:

- proces se skládá z případů,
- případ se skládá z takových událostí, že každá událost se týká přesně jednoho případu,
- události mohou mít atributy, příklady typických názvů atributů jsou aktivita, čas, náklady a zdroje.



Obrázek 6: Stromová struktura event logu

Zdroj: [5]

## 2.3 eXtensible Event Stream

Do roku 2010 byl de facto standardem pro ukládání a editaci protokolů událostí MXML (Mining extensible Markup Language), který je založen na syntaxi XML. MXML měl standardní notaci pro ukládání časových stop, zdrojů a typů transakcí a bylo možné přidávat další atributy k případům. Postupem času si situace žádala další přidávání atributů a odkazování prvků v event logech a začaly se objevovat nedostatky. I když tento přístup v praxi fungoval poměrně dobře, byl MXML nahrazen novým standardem XES. [5]

XES (eXtensible Event Stream) je tedy v dnešní době standardní formát pro Process Mining podporovaný většinou nástrojů a byl přijat v roce 2010 pracovní skupinou IEEE pro Process Mining jako standardní formát pro event logy. [46]

Jedním z cílů pracovní skupiny IEEE pro Process Mining, ve kterém je členem i společnost Fluxicon s jejím Process Miningovým nástrojem Disco, který v této práci bude také zmíněn, je podpora používání technik a nástrojů Process Miningu. Proto je důležitým aspektem existence společných a široce přijímaných standardů. Norma poskytuje uživatelům jistotu, že se nebudou muset vyrovnávat s různými formáty a nebudou tak odkázáni na jednoho dodavatele. [45]

Tento formát definuje gramatiku pro jazyk založený na tagu, jehož cílem je poskytnout návrhářům informačních systémů jednotnou a rozšiřitelnou metodologií pro zachycení chování systémů pomocí event logů a toků událostí definovaných ve standardu XES. [45], [46]

Jazyk založený na tagu slouží k definování prvků v dokumentu nebo webové stránce vložením kódů (tagů), které jsou obklopeny kolem prvku. Nejčastěji se ke spuštění tagu využívá v HTML nebo XML symbol (<) a (>) [44]. Například událost je v XES formátu zachycena jako „<event>“.

Obrázek 7 zobrazuje část event logu transformovaného do XES, na kterém lze vidět strukturu.

Navíc je v této normě zahrnuta základní sada tzv. prototypů rozšíření XES, které poskytují sémantiku určitým atributům zaznamenaným v event logech nebo toku událostí. [45]

Účelem této normy je poskytnout obecně uznávaný formát XML pro výměnu event logů mezi informačními systémy v mnoha aplikačních oblastech na jedné straně a analytickými nástroji pro tyto údaje na straně druhé. Cílem této normy je napravit syntaxi a sémantiku event logů, které se například přenášejí z informačního systému do Process Miningového nástroje, který analyzuje tato data. V důsledku tohoto standardu, pokud jsou event logy přenášeny pomocí

syntaxe, jak je popsáno v této normě, její sémantika bude dobře srozumitelná a jasná na obou místech. [45]

Obrázek 7 ilustruje schéma fragmentu souboru XES.

```
<?xml version="1.0" encoding="UTF-8" ?>
<extension name="Concept" prefix="concept" uri="http://.../concept.xesext"/>
<extension name="Time" prefix="time" uri="http://.../time.xesext"/>
<extension name="Organizational" prefix="org" uri="http://.../org.xesext"/>
<global scope="trace">
  <string key="concept:name" value="name"/>
</global>
<global scope="event">
  <date key="time:timestamp" value="2010-12-17T20:01:02.229+02:00"/>
  <string key="concept:name" value="name"/>
  <string key="org:resource" value="resource"/>
</global>
<classifier name="Activity" keys="concept:name"/>
<classifier name="Resource" keys="org:resource"/>
<classifier name="Both" keys="concept:name org:resource"/>
<trace>
  <string key="concept:name" value="1"/>
  <event>
    <string key="concept:name" value="register request"/>
    <string key="org:resource" value="Pete"/>
    <date key="time:timestamp" value="2010-12-30T11:02:00.000+01:00"/>
    <string key="Event_ID" value="35654423"/>
    <string key="Costs" value="50"/>
  </event>
  <event>
    <string key="concept:name" value="examine thoroughly"/>
    <string key="org:resource" value="Sue"/>
    <date key="time:timestamp" value="2010-12-31T10:06:00.000+01:00"/>
    <string key="Event_ID" value="35654424"/>
    <string key="Costs" value="400"/>
  </event>
  <event>
    <string key="concept:name" value="check ticket"/>
    <string key="org:resource" value="Mike"/>
    <date key="time:timestamp" value="2011-01-05T15:12:00.000+01:00"/>
    <string key="Event_ID" value="35654425"/>
    <string key="Costs" value="100"/>
  </event>
  <event>
    <string key="concept:name" value="decide"/>
    <string key="org:resource" value="Sara"/>
    <date key="time:timestamp" value="2011-01-06T11:18:00.000+01:00"/>
    <string key="Event_ID" value="35654426"/>
    <string key="Costs" value="200"/>
  </event>
  <event>
    <string key="concept:name" value="reject request"/>
    <string key="org:resource" value="Pete"/>
    <date key="time:timestamp" value="2011-01-07T14:24:00.000+01:00"/>
    <string key="Event_ID" value="35654427"/>
    <string key="Costs" value="200"/>
  </event>
</trace>
<trace>
  <string key="concept:name" value="2"/>
  <event>
    <string key="concept:name" value="register request"/>
    <string key="org:resource" value="Mike"/>
    <date key="time:timestamp" value="2010-12-30T11:32:00.000+01:00"/>
    <string key="Event_ID" value="35654483"/>
    <string key="Costs" value="50"/>
  </event>
  ...
</trace>
...
</log>
```

Obrázek 7: Fragment souboru XES

Zdroj: [5]

### 2.3.1 Meta model XES formátu

Obrázek 8 zobrazuje meta model XES vyjádřený pomocí diagramu UML 2.0. V této kapitole budou podrobněji představeny základní komponenty XES.

Meta model XES obsahuje [5]:

- záznam – Log, který může mít libovolný počet stop,
- stopa – Trace, která popisuje sekvenční záznam událostí,
- událost – Event, která odpovídá konkrétnímu případu (case).

Záznam, stopy a události mohou mít libovolný počet atributů, které se v průběhu mohou přidávat.

Ve formátu XES existuje 5 datových typů, které popisuje tabulka 3.

**Tabulka 3: Datové typy XES**

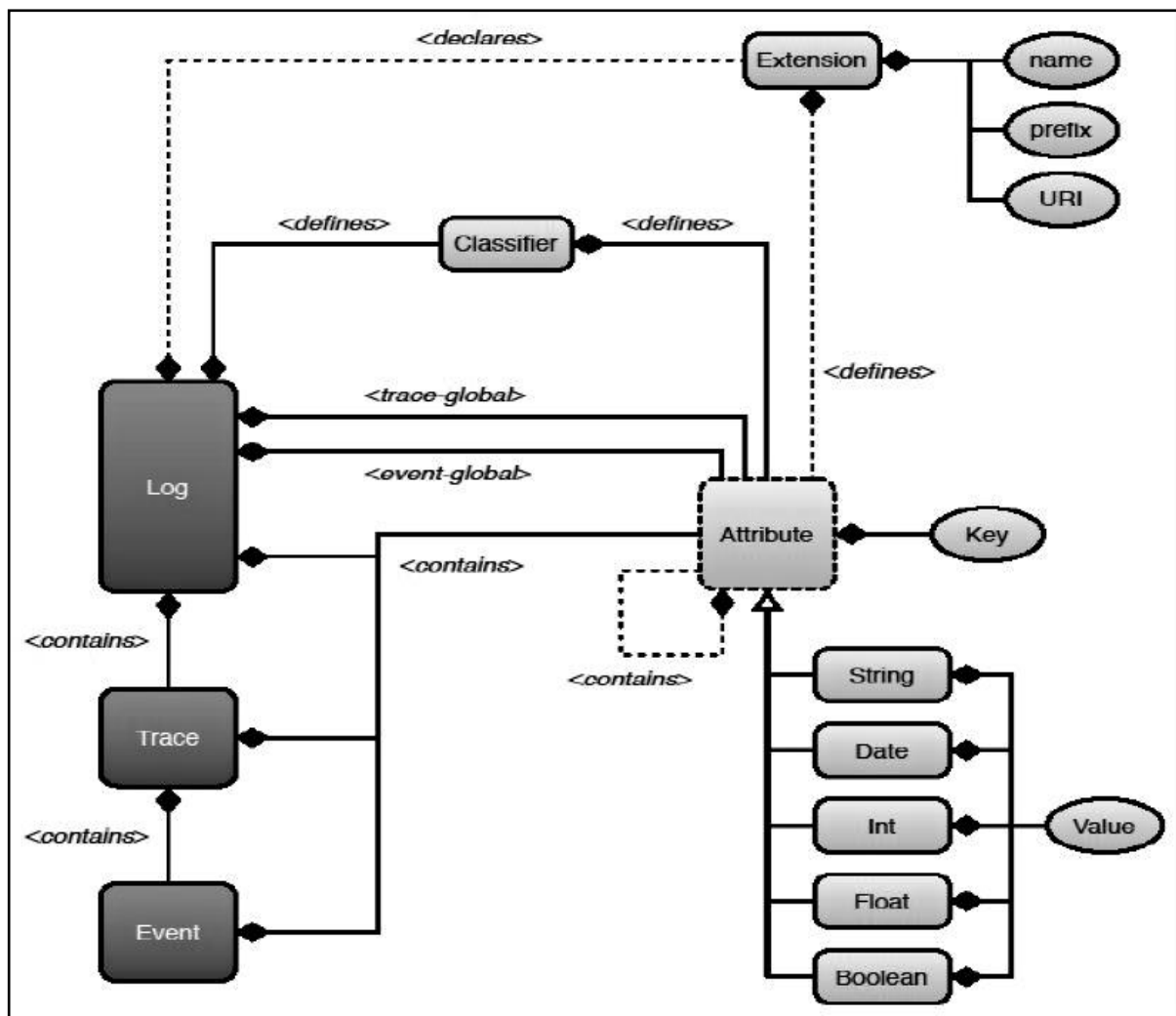
Hodnota XES	Hodnota XML	Význam
string	xs: string	textový řetězec
date	xs: dateTime	datum
int	xs: long	celé číslo
float	xs: double	reálné číslo
boolean	xs: boolean	logická hodnota

*Zdroj: vlastní zpracování*

Například „2011-12-17T21:00:00.000+02:00“ je hodnota typu xs: dateTime představující devět hodin večer 17. prosince 2011 v časové zóně GMT + 2. XES nepředepisuje fixní množinu atributů pro každý prvek (log, trace, event). Událost může mít libovolný počet atributů. Avšak pro poskytnutí sémantiky pro takové atributy se k protokolu vztahuje tzv. rozšíření. Rozšíření dává sémantiku určitým atributům. XES podporuje koncept klasifikátoru. Log XES definuje libovolný počet klasifikátorů, přičemž každý klasifikátor je určen seznamem atributů. Jakékoliv dvě události, které mají stejné hodnoty vzhledem k těmto atributům, se pro tento klasifikátor považují za stejné. Tyto atributy by měly být povinnými atributy události. Například pokud je klasifikátor zadán jak atributem „name“, tak atributem „resource a tyto atributy se shodují, jsou dvě události namapovány do stejné třídy. [5]

V současné době je nástroj XES podporován Process Miningovými nástroji jako ProM, XESame, Disco, Celonis, nebo Minit. ProM je pravděpodobně nejpoužívanějším nástrojem pro Process Mining poskytující širokou škálu technologických postupů. ProM může načíst jak soubory MXML, tak soubory XES. [5]

Nástroje jako například Disco mohou pomoci pro rychlou konverzi event logů do XES nebo MXML. [5]



Obrázek 8: Meta model XES

Zdroj: [5]

### 2.3.2 Benefity XES

Důležitým aspektem je existence společných a široce využívaných standardů, které poskytují uživatelům jistotu, že se nebudou muset vypořádávat s problémy různorodých formátů a nebudou se tak muset přizpůsobit dodavatelům. Při navrhování standardu XES, byly použity níže uvedené čtyři hlavní zásady, které také shrnují jeho hlavní výhody. [45], [46]

**Jednoduchost:** XES využívá nejjednodušší způsob, jak prezentovat informace a jeho protokoly by měly být snadno analyzovány a generovány a zároveň by měly být také snadno čitelné a pochopitelné. Při navrhování tohoto standardu byla věnována velká pozornost pragmatické stránce věci, aby se mohl XES jednoduše implementovat. [23]



**Flexibilita:** Standard XES by měl být schopen zachytit event logy z libovolného prostředí bez ohledu na to, jakou aplikační doménu nebo IT podporuje sledovaný proces. Proto si XES klade za cíl dívat se za hranice Process Miningu a Business Processu a snaží se být obecným standardem pro event logy. [23]

**Rozšiřitelnost:** V budoucnu musí být tento standard snadno doplnitelný o další rozšíření. Rozšíření standardu by mělo být co nejtransparentnější a zároveň by si mělo zachovat zpětnou kompatibilitu. Ve stejném duchu musí být možné rozšířit standard pro zvláštní požadavky, například pro konkrétní aplikační domény nebo pro konkrétní implementaci nástroje. [23]

**Expresivita:** Při snahách o generický formát by se event logy v XES měly setkat s co možná nejmenší ztrátou informací. Proto musí být všechny informační prvky jednoznačně definovány a musí existovat obecná metoda, která jim umožňuje připojit sémantiku, kterou je možné interpretovat. [23]

### 3 NÁSTROJE

Úspěšná aplikace Process Miningu závisí především na dobré podpoře funkcí. Tradiční nástroje pro Business Intelligence jsou zaměřené na data a na jejich zjednodušené formy analýzy. Známé nástroje pro Data Mining a Machine Learning poskytují sofistikovanější formy analýzy, ale na druhou stranu nejsou přizpůsobeny k analýze a zdokonalení procesů. Proto existují vyhrazené nástroje pro Process Mining, které dokáží transformovat event logy do procesních souvislostí. [5]

Při výběru vhodného nástroje pro analýzu podnikového procesu se nabízí otázka, který software je nejlepší. Avšak odpověď na tuto otázku je poněkud nejednoznačná a rozhodně to není o tom, který software je nejvyužívanější, protože to často nemusí znamenat, že je taky nejlepší. Lepší odpověď je spíše, který nástroj nejvíce vyhovuje požadavkům pro analýzu a je proto nejvhodnější. Záleží také, kým je Process Miningový nástroj využíván. Smysl spočívá v tom, že některé nástroje jsou spíše orientované pro akademický přístup, jsou tedy z pravidla komplexnější a složitější a některé najdou větší uplatnění v podniknu v ruce koncového uživatele a měly by být spíše uživatelsky a praktičtěji orientované. [40]

Vzhledem k tomu, že data jsou předpokladem pro Process Mining, jedním z prvních úvah by mělo být, zda má software odpovídající integraci dat. Tedy jestli si umí software poradit s příslušným formátem dat. Dále je důležité se zaměřit na funkčnost a výběr analýz a jejich použitelnost. Nástroj by určitě měl mít možnost reportu procesů nebo flexibility v nastavení a úpravě KPI (Klíčový indikátor výkonu) [33]. Klíčový indikátor výkonu je měřitelná hodnota, která ukazuje, jak efektivně společnost dosahuje klíčových obchodních cílů. Organizace používají KPI k vyhodnocení svého úspěchu při dosahování cílů [24]. Mnozí odborníci poukazují, že pro některé případy je velkou výhodou, když mají nástroje implementovanou funkci pro predikci procesu. Analyzovat již uzavřené případy není tolik efektivní, jako analyzovat případy v reálném čase, protože pro tyto uzavřené případy již neexistují žádná opatření ke zlepšení. Dalšími úvahami mohou být také spolupráce, zabezpečení a možnosti nasazení. [33]

Tento výčet funkcí je pouze nastíněním toho, co by měl uživatel brát v úvahu před výběrem vhodného softwaru. Kritérií je samozřejmě mnohem více a budou v kapitole 3.4 více rozepsány. Tato kapitola se bude zabývat přehledem dostupných nástrojů ať už komerčních či nekomerčních, jejichmi výhodami a nevýhodami a v závěru této kapitoly budou nástroje shrnuty na základě vybraných kritérií.

### **3.1 Typy užití Process Miningu**

Před popisem konkrétních nástrojů pro Process Mining je nejprve důležité se zmínit o jejich různých způsobech charakterizace. Potenciálně existuje mnoho různých skupin uživatelů, kteří pracují s nástrojem Process Miningu.

Na jednu stranu to mohou být experti, kteří musejí být schopni se vypořádat s unikátním problémem vyžadující datové extrakce ad-hoc a složité transformace dat nebo musejí využít sofistikované analytické techniky. Na druhou stranu existují koncoví uživatelé, kteří se chtějí pouze podívat na celkový vygenerovaný přehled, například pomocí dashboardu zaměřený na proces. [5], [6]

Spektrum Process Miningových případů lze charakterizovat těmito otázkami [5]:

- Jak často se opakuje stejná analýza?
- Může koncový uživatel volně rozhodnout, zda má být provedena analýza?

Na obrázku 9 jsou diagonálně uvedené tři typy případů použití na základě odpovědí na tyto dvě otázky.

#### **3.1.1 Typ AD-HOC**

Použití případů typu AD-HOC vyžaduje odbornou znalost. Otázky jsou ad-hoc a uživatel potřebuje mít úplnou způsobilost a kvalifikaci k provádění analýz. Proces analýzy je nedefinovaný. Výsledky jednoho analytického kroku mohou vést k neočekávané dodatečné extrakci dat nebo transformacím k umožnění dalšího kroku analýzy. Analytické výstupy jsou jedinečné a zřídka se opakují. [5]

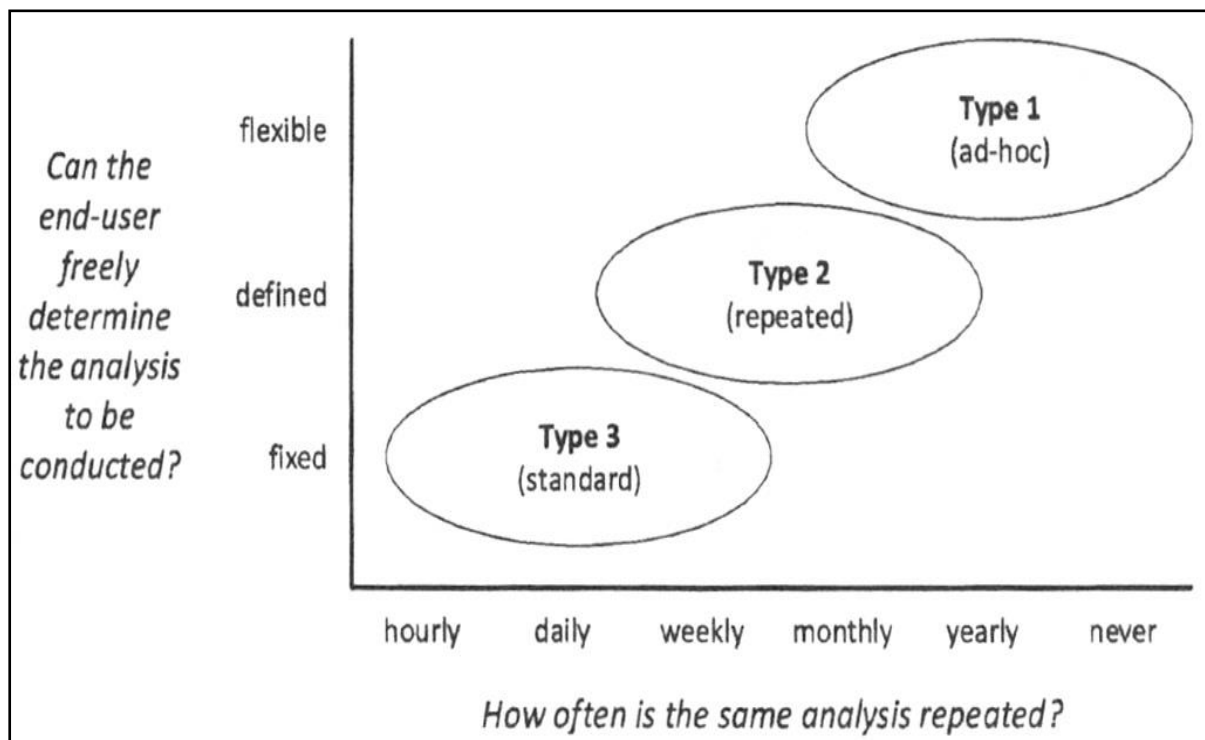
#### **3.1.2 Typ Opakování**

Použití případů typu Opakování zahrnuje opakované otázky, ale s nižší frekvencí. Analytické výstupy mohou být předdefinovány, ale nejsou zcela ustáleny. Může být zapotřebí určité přizpůsobení a interpretace výsledků vyžaduje znalost Process Miningu a porozumění dat. [5]

#### **3.1.3 Typ Standard**

Použití případů typu Standard zahrnuje rutinní otázky, které se často opakují. Analytické výstupy jsou jednoznačné a není možné je přizpůsobovat. Uživatel potřebuje pouze pochopit předdefinované pohledy typu dashboard. [5]

Na obrázku 8 jsou na diagonále uvedeny tři typy případů, které již byly výše vysvětleny. Případy, které nejsou na diagonále, nemají smysl. Například nemůžeme poskytnout předdefinovaný dashboard pro otázky typu ad-hoc, které odpovídají kombinaci "nikdy" a "fixní". To samé platí například při kombinaci „pružné“ a „hodinové“.



Obrázek 9: Typy užití v Process Miningu

Zdroj: [5]

Process Miningové nástroje mohou být přizpůsobeny jednomu ze tří typů na obrázku 8. Je složité určit, který z nich je nejlepší, protože všechny mají své výhody i nevýhody.

Nástroje, které se zaměřují spíše na případy typu AD-HOC, kde uživateli například stačí pouze načíst požadovaná data, vybrat si konkrétní výstup a získat okamžité výsledky bez systémové konfigurace, mohou být vhodné pro efektivní prozkoumávání a rychlé získávání výsledků. [5]

Na druhou stranu, takovýto styl není úplně vhodný pro koncové uživatele, kteří nerozumí datům a analytickým technikám, jako je tomu u typu Standard. Proto se vyplatí zainvestovat a zaškolit uživatele tak, aby byli schopni nakonfigurovat data potřebným způsobem a definovat výstup. Potom budou mít uživatelé znalosti a přehled v analýzách, které budou moci aplikovat opakovaně. Typ Opakování je něco mezi typem AD-HOC a typem Standard. [6]

Výhodou typu Opakování je, že má částečně předdefinovaný work-flow, který ale není kompletní. [5], [6]

## 3.2 Nekomerční nástroje

Nekomerčních softwarů není tolik na trhu jako těch komerčních, nicméně obecně mají komplexní nabídku funkcí pro kvalitní analýzu procesu. Předním představitelem nekomerčních nástrojů je bezesporu ProM. Existuje však i několik dalších nástrojů, jako PMLAB nebo CoBeFra, nicméně většina výzkumu v Process Miningu je implementována v ProM. [5]

V této kapitole jsou vysvětleny nástroje jako ProM, PMLab a CoBeFra.

### 3.2.1 ProM

ProM (Process Mining Frame-work) je rozšiřitelný nástroj s otevřeným kódem a patří mezi nejznámější a nejvíce využívané nástroje pro Process Mining nekomerční licence. Jedná se o nezávislý software implementovaný v Javě a lze ho bezplatně stáhnout. [5]

Poskytuje platformu uživatelům s podporou široké škály algoritmů ve formě zásuvných modulů, kterých je v dnešní době přes 1500 a neustále díky vývojářům přibývají nové. ProM je de facto standardní Process Miningovou platformou v akademickém světě a má aktivní a uznávanou komunitu vývojářů a koncových uživatelů firem. [10]

ProM umožňuje import a export do nejrůznějších formátů a systémů a poskytuje služby od podnikových informačních systémů a produktů work-flow, až po software pro sociální sítě, klasické nástroje pro Process Mining nebo pokročilé možnosti vizualizace a ověřování. [5]

#### Plug-iny ProM

V této části je poskytnut přehled plug-inů, jaké jsou v současné době implementovány v kontextu s ProM.

Existuje pět druhů pluginů v ProM [7]:

##### 1. Plug-iny pro dolování

Implementuje nějaký dolovací algoritmus, například algoritmy dolování, které vytvářejí Petriho síť založenou na event logech.

##### 2. Plug-iny pro export

Implementují některé funkce umožňující uložit výstup jako jiný formát. Například existují plug-iny pro uložení EPC, Petriho sítí nebo různých grafů apod.

### 3. Plug-iny pro import

Implementují funkci "otevřít" pro exportované objekty, například načíst instance EPC z ARIS-PPM.

### 4. Plug-iny pro analýzu

Obvykle provádějí analýzu vlastností na výstupu Process Miningu. Například existují analytické plug-iny pro porovnání reálného modelu na základě event logů a původního modelu (tj. testování shody).

### 5. Plug-iny pro konverzi

Implementují konverze mezi různými datovými formáty, například z EPC do Petriho sítí. Kvůli úspěchu ProM v akademické sféře existuje jen několik málo dalších nástrojů pro Process Mining nekomerční licence.

Několik pozoruhodných výjimek lze však zmínit, například:

- PMLab,
- Cobefra,
- RapidProM.

Použitelnost stovek dostupných plug-inů se však liší a složitost nástroje může být pro koncové uživatele nadměrná. V uplynulých letech několik softwarových dodavatelů vydalo vyhrazené nástroje pro Process Mining (například Celonis, Disco, EDS, Fujitsu, Minit, Mylnvenio, Perceptive, PPM, QPR, Rialto nebo SNP). Tyto nástroje zpravidla poskytují méně funkčnosti než ProM, ale jsou snadněji použitelné při zaměření na extrakci dat, analýzu výkonu a škálovatelnost. [5]

#### 3.2.2 PMLab

PMLAB (Process Mining Lab) je interaktivní programovací prostředí pro Process Mining vyvinuté skupinou Josepa Cannonyho na Katalánské univerzitě v Barceloně. PMLAB se inspiroval MATLABem, Mathematicou či RapidMinerem a je napsaný v iPythonu určený pro skriptování v Pythonu. PMLAB také dokáže implementovat plug-iny ProM pomocí skriptů nebo rozšiřovat o své plug-iny. Stejně jako ProM může načíst soubory XES, MXML a CSV. Zabývá se AND/OR výzkumem založeným na procesně orientovaném jazyku. [5], [30]

PMLab obsahuje algoritmy jako například Genet, Petrify, Rbminer nebo Dbminer a jsou většinou založeny na tzv. state-based regionech. Tyto algoritmy jsou vhodné pro kompaktní a čitelné procesní modely, transformování modelu mezi různými formáty atd. Výhodou je také možnost paralelizovat složitější výpočty z dostupných výpočetních zdrojů. [5]

### 3.2.3 CoBeFra

V oblasti Process Miningu lze postupovat dvěma různými kroky. Za prvé, zjišťování procesů, které se zabývá automatickou konstrukcí procesního modelu z event logu. Za druhé, kontrola shody, která se zaměřuje na posouzení kvality navrženého modelu procesu s ohledem na skutečný model procesu vytvořený z event logů. [5], [14]

CoBeFra je právě porovnávací platforma pro kontrolu shody, který byl vyvinut na katedře Management Informatiky na univerzitě Leuven v Belgii. Používá se hlavně pro posouzení kvality procesního modelu z kvantitativního hlediska. CoBeFra nabízí architekturu pro rozšiřitelný rámec v ProM, což umožňuje konzistentní, srovnávací a opakovatelné výpočty. Pro vývoj, hodnocení a zjišťování procesů i metod shody je takový rámec velmi cenný. CoBeFra podporuje event logy ve formátu souborů XES nebo MXML a modely procesů ve formátu PNML. [5], [14]

### 3.2.4 RapidProM

Mnoho nástrojů pro analýzu dat podporuje analýzu work-flow. Například široce využívaný nástroj RapidMiner podporuje spojovat činnosti ve formě bloků tak, aby vytvářely work-flow. Další příklady nástrojů mohou být také KNIME a R. [5], [35]

RapidMiner je systém, který podporuje návrh a dokumentaci celkového procesu Data Miningu. Jako součást této funkce nabízí komplexní soubor operátorů pro techniky Data Miningu, Machine Learningu a statistických metod. Dále obsahuje mnoho struktur pro definování řízení toku procesu. [35]

Avšak původně tento nástroj nepodporuje Process Mining. Naopak zase ProM neposkytuje podporu work-flow. Proto byla společnost RapidMiner rozšířena o moduly pro Process Mining od společnosti ProM. Výsledný nástroj se nazývá RapidProM a nejvíce se využívá v rozsáhlých experimentech při analyzování work-flow. [5]

Zatím co tyto zmíněné 3 nekomerční nástroje se nejvíce hodí na případy typu opakování, ProM je spíše určený pro případy typu AD-HOC. [5]

### 3.2.5 Další nástroje

Podpora formátu zdrojových dat patří mezi hlavní kritéria. Především u komerčních nástrojů se tato podpora odlišuje, proto stojí za zmínění několika málo nástrojů pro jednoduchý převod mezi formáty. Mezi nástroje patří například XESame, OpenXES nebo ProMImport. [34], [45], [47],

### 3.3 Komerční nástroje

V posledních letech se na trhu objevilo několik komerčních nástrojů pro Process Mining. Ve srovnání s programem ProM, zmíněným v předchozí kapitole, jsou tyto nástroje jednodušší a přívětivější pro koncového uživatele, ale poskytují méně funkcí než ProM, který s přibližně 1500 rozšiřujícími plug-iny je nejobsáhlejší open-sourcový nástroj. Tím se výrazně snižuje komplexnost a univerzálnost těchto nástrojů. [5]

Tabulka 4 zobrazuje přehled komerčních nástrojů pro Process Mining, které jsou v současné době na trhu. Nástroje Disco, Fujitsu, QPR a PPM jsou na trhu již pár let. Minit, myInveo a Rialto vznikly teprve nedávno v roce 2015. [5]

Tabulka 4: Přehled komerčních nástrojů

Zkratka	Název softwaru	Společnost	Web
<b>Celonis</b>	Celonis Process Mining	Celonis GmbH	<a href="http://www.celonis.de">www.celonis.de</a>
<b>Disco</b>	Disco	Fluxicon	<a href="http://www.fluxicon.com">www.fluxicon.com</a>
<b>EDS</b>	Enterprise Discovery Suite	StereoLOGIC Ltd	<a href="http://www.stereologic.com">www.stereologic.com</a>
<b>Fujitsu</b>	Interstage Business Process Manager Analytics	Fujitsu Ltd	<a href="http://www.fujitsu.com">www.fujitsu.com</a>
<b>Minit</b>	Minit	Gradient ECM	<a href="http://www.minitlabs.com">www.minitlabs.com</a>
<b>myInvenio</b>	myInvenio	Cognitive Technology	<a href="http://www.my-invenio.com">www.my-invenio.com</a>
<b>Perceptive</b>	Perceptive Process Mining	Lexmark	<a href="http://www.lexmark.com">www.lexmark.com</a>
<b>QPR</b>	QPR ProcessAnalyzer	QPR	<a href="http://www.qpr.com">www.qpr.com</a>
<b>Rialto</b>	Rialto Process	Exeura	<a href="http://www.exeura.eu">www.exeura.eu</a>
<b>SNP</b>	SNP Business Process Analysis	SNP Schneider-Neureither & Partner AG	<a href="http://www.snp-bpa.com">www.snp-bpa.com</a>
<b>PPM</b>	webMethods Process Performance Manager	Software AG	<a href="http://www.softwareag.com">www.softwareag.com</a>

*Zdroj: [5]*

Process Miningové nástroje se na trhu rychle mění a většina zmíněných nástrojů v tabulce 4 ještě před několika lety neexistovala. Schopnosti nástrojů se mění s každým vydáním nové verze. Mění se použitelnost a škálovatelnost, tj. rozšiřitelnost o nové funkce a vlastnosti tak jak si žádá trh. Je zbytečné tyto změny vyjádřit v kontrolním seznamu. Proto jsou vyzývány



organizace, které vybírají komerční nástroj pro Process Mining, aby vyhodnotily nástroje založené na konkrétních otázkách a jejich datových setech. [5]

Cílem této práce není poskytnout podrobné informace o všech nástrojích, spíše poukázat na jejich existenci a vybrané z nich na základě kritérií podrobněji popsat a porovnat. Pro účely práce byly zvoleny dva komerční nástroje a jeden nekomerční nástroj, které budou blíže představeny. Jsou to Disco, Celonis a ProM.

### **3.3.1 Disco**

Disco je komerční nástroj využívaný pro Process Mining, který představila společnost Fluxicon. Je licencovaný na desktopové verze na platformě Windows a Mac. Disco se podle odborníků především prosazuje díky příjemnějšímu uživatelskému prostředí, jednoduchosti a rychlosti zpracování event logů. Je tak vhodný jak pro začátečníky, tak pro odborníky. [41]

Zatímco ProM je komplexní a nesmírně efektivní nástroj, na základě požadavků společností, vlastních zkušeností z projektů a poradenstvím a v neposlední řadě díky mnoha rozhovorům s analytickými odborníky se Fluxiconu podařilo vyvinout v rámci možností stejně funkční Process Miningový nástroj. [16]

### **3.3.2 Celonis**

Společnost Celonis má svůj původ v oblasti univerzitního výzkumu. V roce 2011 založili tři studenti na technické univerzitě v Mnichově společnost zaměřenou plně na Process Mining. Dnes Celonis využívá stovky zákazníků ve více než 25 zemích. Mezi zákazníky patří jak velké společnosti, které se každoročně umisťují v americkém žebříčku časopisu Fortune na základě hrubého obratu, tak i malé a střední firmy. Vzhledem k rozsáhlému využívání softwaru Celonis má stejnojmenná společnost v Process Miningu hluboké znalosti a poskytuje tak cenný přínos pro výzkumné výzvy. [5]

Celonis představuje technologii pro zjišťování procesů v reálném čase, zatímco jsou informační systémy v provozu. Podporuje databáze SAP a další různé RDBMS (Relational Database Management System). Pro import event logů do systému Celonis se musí vytvořit zdroj datového spojení se specializovanou službou PostgreSQL. Databáze, jako jsou SAP a RDBMS, jsou tedy podporovány společností Celonis. Mezi funkce tohoto nástroje patří automatizovaná integrace zdrojových dat, sledování všech obchodních transakcí v reálném čase, provádění procesních analýz, různé filtrovací mechanismy, hlášení procesů atd. [5]

### 3.4 Kritéria

Nutno říci, že žádný z produktů v komerční sféře neobsahuje takový rozsah schopností v Process Miningu jako ProM podporovaný stovkami dostupných plug-inů. ProM však vyžaduje expertní znalosti a není podporován komerčními organizacemi. Proto má výhody a nevýhody, které jsou běžné pro open-source software. [5]

Na jednu stranu je mezi komerčními nástroji a ProM poměrně hodně společných funkcí. Na druhé straně existují velké rozdíly především v použitelnosti a škálovatelnosti. Některé nástroje se zaměřují více na případy typu AD-HOC. Zástupcem tohoto typu je již zmiňovaný ProM a Disco, ale i Minit nebo myInveo. Jiné nástroje se více zaměřují na případy typu Standard. Příkladem jsou Celonis nebo PPM. Organizace, které se rozhodují mezi komerčními nástroji, jsou vyzývány k tomu, aby provedly nejdříve pilotní projekt, kde je několik produktů aplikováno na specifické údaje a otázky týkající se organizace. I přes rozdíly mezi nástroji, lze obecně vymezit kritéria, podle kterých by se měly organizace rozhodovat. Tyto kritéria jsou popsány v níže. [5]

#### 3.4.1 Import dat

Každý proces v Process Miningu začíná daty (event logy), které by měly být analyzovány. Nástroje by měly být navrženy tak, aby byl import snadný. Neměly by tedy chybět schopnosti jako automatické zjišťování časových razítek, zapamatování si konfiguračních nastavení a načítání datových setů s vysokou rychlostí. Samozřejmostí je podpora co nejvíce možných formátů dat. [41]

**ProM** může načíst soubory XES, MXML a CSV. CSV se, avšak před samotnou analýzou musí skrz plug-iny převést na XES nebo MXML. Pro Extrahování souborů z jiných zdrojů lze použít nástroje jako XESame nebo ProMImport. [5]

**Disco** může při importu otevřít soubor MXML, XES, CSV, XLS nebo FXL a uživatel nakonfiguruje, které sloupce obsahují ID případu, časová razítka, názvy aktivit a které další atributy by měly být zahrnuty do analýzy. Data jsou importována v režimu pouze pro čtení, takže původní soubory nelze změnit, což je důležité například pro auditora. Disco je také plně kompatibilní s akademickými sadami nástrojů ProM 6 a díky tomu se mohou pokročilí uživatelé bez problémů pohybovat tam a zpět mezi Disco a ProM, pokud chtějí využít nových výzkumných technologií a algoritmů v akademickém prostředí. [3]

**Celonis** podporuje formáty jako CSV, XLSX, XES nebo DBF. Po importu dat se pouze potvrdí již automaticky přiřazené Case, Event, Timestamp a jiné atributy. Dále se musí analogicky přiřadit datové typy.

### 3.4.2 Výstup modelu

Množství dostupných modelovacích notací, které jsou dnes k dispozici, ilustruje relevanci modelování procesů. Některé organizace mohou používat pouze neformální procesní modely pro strukturovanou diskuzi a dokumentování postupů. Organizace, které pracují s vyšším stupněm BPM, však používají modely, které lze analyzovat a používat k tvorbě operačních procesů. [5]

**ProM** umí vytvořit modely s různou notací. Například lze vytvořit BPMN diagram, Petriho síť, Heuristická síť, UML diagram, YAWL, EPC, přechodový systém, induktivní model a model pomocí Fuzzy mineru. [5]

**Disco a Celonis** jako svůj výstup modelu využívá pouze Fuzzy miner k zobrazení reálného modelu.

### 3.4.3 Perspektiva výkonu

Primárním zaměřením komerčních nástrojů pro Process Mining je výkonnost. Každý z nástrojů může odhalit slabiny v procesu. Většina komerčních nástrojů umožňuje rychle nalézt slabiny, zbytečné vytížení a zpoždění. [5]

**ProM, Disco a Celonis** podporují širokou škálu inteligentních grafů v různých oblastech, které pomáhají odhalit slabiny, zvýraznit podezřele vytížené aktivity a zpoždění. V tomto ohledu tyto nástroje jsou velice podobné. [5], [16]

### 3.4.4 Organizační perspektiva

Některé nástroje jsou schopné vytvořit analýzu sociální sítě a tím zjistit využití zdrojů a jaká je jejich role. Téměř všechny nástroje považují informace o zdrojích, rolích a dalších organizačních entitách za obyčejné datové prvky. Organizační perspektiva může být řešena stejným způsobem jako data (například při filtrování dat). Tímto způsobem lze zkontrolovat rozdělení rolí. [5]

**ProM** má několik typů technik pro vytváření analýz sociálních sítí. Pro analýzu k tomu lze využít tečkovaný graf s různými nastaveními v kombinaci s Inductive Minerem. ProM nabízí celkem 4 plug-iny. [43]

První typ, „Hand-over of work“, je algoritmus, pomocí kterého lze zjistit předávání práce mezi zdroji v procesu. V praxi to pak znamená, že zdroje si mezi sebou předávají aktivity a ve výsledku jsou tyto skutečnosti vykresleny v grafu. [43]

Druhý typ, „Reassignment“, znamená přeřazení činnosti jinému zdroji. Například když jeden zdroj přeřadí aktivitu jinému zdroji, který ale tuto aktivitu nemůže předat dále nebo zpět. Znamená to, že první zdroj je hierarchicky nadřazený. Avšak tyto informace většinou nejsou uloženy v event logu, a proto analýzu nelze provést. [43]

Třetí typ, „Subcontract“, je podobný druhému typu až na rozdíl, že jeden zdroj vytvoří aktivitu mezi jinými dvěma aktivitami, které vytvořil jiný zdroj. [43]

Čtvrtý typ, „Working together“, analyzuje, které zdroje pracují spolu. [43]

**Celonis** vytváří mapu aktivit jako zjednodušenou matici RACI (matice odpovědnosti). Matice odpovědnosti RACI je metoda pro přiřazení a zobrazení odpovědností jednotlivých osob či pracovních míst v nějakém úkolu, projektu, službě či procesu v organizaci. Pro poskytnutí tohoto typu informací je třeba obohatit event logy. [27]

**Disco** analýzu sociální sítě zatím nepodporuje.

### 3.4.5 Podpora XES

Nekomerční nástroje jako ProM, RapidProM, PMLAB a CoBeFra standard XES podporují. Z komerčních nástrojů v současnosti tento formát podporují Disco, Celonis a Minit. Jak již bylo řečeno, trh s nástroji pro Process Mining je velice dynamický a je mezi nimi velká konkurence, proto ostatní firmy jako například QPR a myInvenio oznámily podporu XES pro další verzi. PPM, EDS a Fujitsu zatím XES nepodporují, ale očekává se brzká integrace. [5]

XES usnadňuje kombinování různých nástrojů, například používáním komerčního nástroje ve spojení s ProM, RapidProM, PMLAB nebo CoBeFra. [5]

### 3.4.6 Získání dat o událostech z jiných zdrojů

Dodavatelé komerčních nástrojů si uvědomují, že na získávání dat z informačních systémů je věnována značná část času. [5]

Process Miningové nástroje mohou načíst event logy čtyřmi způsoby [5]:

- **soubor** – event logy jsou uloženy ve formátu xes, mxml, excel nebo csv,
- **databáze** – event logy jsou načteny z databázového systému, například přes JDBC,

- **adaptér** – event logy jsou načteny z určité aplikace, například SAP nebo Sharepoint skrz vyhrazený software (ve většině případů lze data načíst postupně),
- **streaming** – nástroj pro process mining pracuje na proudu událostí vysílaných prostřednictvím sběrnice událostí nebo webové služby a události jsou zachyceny, jelikož se vyskytují a nejsou načteny ze souboru, databáze nebo aplikace později.

**ProM** kromě podpory importu souborů XES, MXML a CSV také podporuje extrakci dat z databází pomocí JDBC. Události mohou být často načteny ze systémů, jako jsou MySQL, IBM DB2, Oracle DB, SQL Server, PostgreSQL a. Data sety často mohou být také postupně aktualizovány. [5]

**Disco** může načíst data ze serveru s takzvaným rozhraním Airlift. Na straně serveru připojení Airlift lze připojit libovolné databáze a výrobní systémy. [41]

**Celonis** poskytují další podporu pro získávání dat ze systémů SAP. Díky partnerství mezi SAP a Celonis je zajištěna integrace s produkty SAP, jako je SAP HANA. Ve skutečnosti většina nástrojů pro Process Mining podporuje aplikační adaptéry, ale rozsah systémů a kvalita těchto adaptérů se liší podle jednotlivých nástrojů. [5]

### 3.4.7 Filtrování

Jakmile se extrahuje event log a otevře se v nástroji, je možné jej změnit různými způsoby. Filtrování se provádí ze dvou hlavních důvodů: čištění dat nebo zaměření analýzy. Někdy existují také technické důvody. [5]

Filtrování hraje zásadní roli a je důležité k tomu, aby bylo možné vytvořit kvalitní procesní modely nebo vytvořit analýzu kořenových příčin. Také se umožňuje soustředit na určitou část procesu nebo na konkrétní typ případu. [5]

**ProM** podporuje řadu filtrů, například [42]:

- filtry pro přemapování události mezi úrovněmi,
- duplicitní filtry,
- vkládací filtry,
- filtry atributů,
- filtr pro délku instancí,
- filtr pro frekvenci instancí.

**Disco** disponuje několika transparentními a rychlými záznamovými filtry, díky kterým lze vyčistit event logy a zaměřit se na analýzu. Umožňuje si vybrat mezi filtry zaměřené na

výkon, časový rámeček, varianty, atributy, případy nebo vztahy mezi událostmi. Například je možné odebrat všechny pomalé nebo výjimečné případy.

**Celonis** a jeho filtrování souvisí s protokolem OLAP. Rozměry v krychli OLAP rozdělují data na základě různých kritérií. Celonis ukládá události v multidimenzionálních kostkách, aby usnadnily výběr a porovnání jednotlivých skupin případů. [5]

### 3.4.8 Automatické shlukování

Filtrování a výběr rozměrů v krychli OLAP jsou založeny na uživatelsky definovaných kritériích. Nicméně, lze i používat shlukovací techniky, které automaticky seskupují případy, které si jsou podobné. [5]

**ProM** poskytuje několik způsobů seskupování podobných případů na základě vybraných kritérií. Standardní techniky jako k-means lze použít jako krok předzpracování pro Process Mining. Samotné klastry mohou již poskytnout nové poznatky. Tyto shluky lze navíc často použít k objevení více jednoduchých modelů procesů namísto jednoho komplexního modelu procesu. [5]

Překvapivě ale není shlukování podporováno současnou generací komerčních nástrojů pro Process Mining. [5]

### 3.4.9 Reporting a animace

Výsledky Process Miningu musí být srozumitelné a pochopitelné. Většina nástrojů umožňuje vytváření reportů, například ukládáním artefaktů, jako jsou grafy, tabulky a modely. Ve srovnání s nástroji pro BI jsou nástroje pro report poněkud limitované a omezené. [5]

**ProM, Disco i Celonis** vytvářejí animace a propracované statistiky založené na event logech. Takové animace pomáhají při přesvědčování managementu a jsou také prostředkem dobrým pro podporu řízení změn. Lze je použít k vytvoření pocitu naléhavosti a k budování vzájemného souhlasu managementu o základních příčinách. [5]

## 3.5 Shrnutí

Tato kapitola představuje shrnutí nástrojů na základě výše popsaných kritérií, které zobrazuje tabulka 5.

Podpora nástroje pro danou funkci je označena jako „√“ a funkce, které nejsou podporovány, jsou označeny jako „X“.

**Tabulka 5: Shrnutí nástrojů**

<b>Kritéria</b>	<b>ProM</b>	<b>Disco</b>	<b>Celonis</b>
<b>Typ použití</b>	Typ 1, 2, 3	Typ 1	Typ 2, 3
<b>Podpora formátu při importu</b>	mxml, xes	csv, xls, mxrnl, xes, fxl	csv, xls, xes, dbf
<b>Velikost logu při importu</b>	Neomezený	Do 5 milionů ze souboru/Omezeno databází	Omezeno databází
<b>Výstup modelu</b>	BPMN, Petriho síť, EPC, Transition systems, Heuristická síť nebo Fuzzy model	Modifikovaný Fuzzy model	Modifikovaný Fuzzy model
<b>Filtrování dat</b>	✓	✓	✓
<b>Prozkoumávání procesu</b>	✓	✓	✓
<b>Kontrola shody</b>	✓	X	✓
<b>Dolování sociální sítě</b>	✓	X	✓
<b>Vizualizace procesu</b>	✓	✓	✓
<b>Analýza výkonu</b>	✓	✓	✓
<b>Shlukování</b>	✓	X	X
<b>Podporovaná platforma</b>	Desktopová	Desktopová	Prohlížeč webu/Podle informačního systému
<b>Licence</b>	Open-source	Komerční	Komerční

*Zdroj: upraveno dle [5]*

## 4 METODY

Process Mining má za cíl získat informace ze zaznamenaných procesních dat (event logů), které lze využít k získání poznatků o procesu. To vyžaduje uplatnění algoritmu pro analýzu a eventuálně nastavení jeho parametrů, po němž by měl být vyhodnocen objevený reálný procesní model. Tyto kroky jsou iterativní a mohou se opakovat několikrát, dokud nebude nalezen uspokojivý model. Tento postup nazýváme jako průzkum procesů. [39]

Algoritmus je postup, vzorec nebo sada instrukcí určených k provádění specifického úkolu. Může to být jednoduchý proces, například vynásobení dvou čísel nebo složitá operace, například přehrávání komprimovaného videosouboru. [37], [11]

V kapitole 1 byly vysvětleny 3 hlavní kroky v Process Miningu, které zahrnují prozkoumávání reálného modelu procesu na základě event logů, kontrola shody původního modelu s reálným modelem procesu a vylepšení vytvořeného reálného modelu procesu. Algoritmus v procesu prozkoumávání je funkce, která mapuje event logy do procesních modelů, tak že model má reprezentativní chování, které odráží event logy. [5]

Programátoři obvykle usilují o vytvoření co nejefektivnějších algoritmů, aby jejich programy „běhaly“ co nejrychleji a používaly minimální systémové prostředky. Samozřejmě, ne všechny algoritmy jsou vytvořeny dokonale napoprvé. Proto vývojáři často vylepšují stávající algoritmy a zahrnují je do budoucích aktualizací softwaru nebo plug-inů, jako je tomu v případě Process Miningového nástroje ProM. Pokud je vydána nová verze softwarového programu nebo plug-inu, který byl optimalizován nebo má "rychlejší výkon", s největší pravděpodobností to znamená, že nová verze obsahuje efektivnější algoritmy. [11]

Potencionálně existuje několik přístupů a algoritmů pro vytvoření reálného procesu modelu. Akademický nástroj Process Miningu ProM jich podporuje ze všech ostatních, ať už komerčních či nekomerčních, suverénně nejvíce. Často jsou ale mnohé z nich pouze experimentální a dají se využít pouze na hrstku vybraných příkladů. V této práci bude zmíněno několik málo metod, které se běžně využívají v praxi.

Pro účely této diplomové práce byly vybrány celkem 3 nástroje, které již byly v kapitole 3 porovnány z hlediska podpory a funkčnosti. Zbývá tedy tyto nástroje ještě zhodnotit z hlediska tvorby reálného procesu modelu. Z Process Miningových nástrojů, jak již bylo řečeno, ProM disponuje několika algoritmy pro nalezení reálného modelu procesu, ale ostatní nástroje jako je Disco a Celonis, podporují pouze jeden výstup, a to modifikovaný Fuzzy



model. Proto budou tyto nástroje mezi sebou objektivně porovnány skrz výsledný Fuzzy model.

Ačkoliv je spousta algoritmů pro Process Mining, překvapivě jenom několik málo jich je použitelných pro analyzování reálných event logů. Pro praktické aplikace algoritmů pro proces prozkoumávání se nejčastěji využívají [39]:

- heuristic miner,
- inductive miner,
- fuzzy miner.

Reálné event logy se potýkají s řadou nedokonalostí, například nekompletnost (event log obsahuje příliš málo událostí, aby bylo možné vytvořit model procesu) nebo šum (event log obsahuje vzácné a občasné chování, které není typické pro sledovaný proces). Tyto nedokonalosti, avšak nic neříkají o kvalitě samotných event logů. Pro určení kvality event logů se proto využívají čtyři kritéria kvality (dimenze). [5]

### **1. Fitness (kondice)**

Model s dobrým koeficientem Fitness odráží chování zaznamenané v event logu. Model má dokonalý Fitness, pokud všechny stopy, tj. sekvence událostí v event logu, mohou být modelem přehrávány od začátku do konce. Fitness lze definovat z různých úhlů pohledu, například může být definována na úrovni případu nebo události. [5], [1]

### **2. Simplicity (jednoduchost)**

Simplicity se odkazuje na Occamovu břitvu. V kontextu s prozkoumáváním procesů to znamená, že nejjednodušší model, který dokáže vysvětlit chování zaznamenané v event logu, je nejlepším modelem. Složitost modelu by mohla být definována počtem uzlů a cest ve vytvořeném reálném modelu. Lze použít také sofistikovanější metriky, které berou v úvahu "strukturovanost" nebo "entropii" modelu. [5], [1]

### **3. Precision (přesnost)**

Kritérium kvality Precision souvisí s pojmem „underfitting“, tj. model je nedoučený. „Underfitting“ je vlastně takový problém, kdy model příliš zobecňuje příkladové chování v event logu. Jinými slovy, model se prakticky chová jinak, než jak by podle event logu měl. [5], [1]

### **1. Generalization (zobecnění)**

Model by měl zobecňovat, ale neomezovat chování event logů. Model, který nedokáže zobecnit chování v event logech má problém, který se odkazuje na pojem „overfitting“ tj.

přeučení. Znamená to, že je generován velmi specifický model, zatímco je zřejmé, že event log obsahuje pouze příkladné chování, tj. model vysvětluje konkrétní vzorek v event logu, ale jiný vzorek v event logu, který by měl model zobecnit stejným způsobem, může vytvořit zcela jiný model procesu. [5], [1]

#### **4.1 Heuristic miner**

Heuristic miner je nástupcem Alfa algoritmu, který byl vyvinut akademickými pracovníky v čele s Wilem Aalstem na technické univerzitě v Eindhovenu. Tento algoritmus měl právě problém s tím, že nedokázal vyřešit problémy týkající se šumu v reálných event logech. [5]

Heuristic miner, který vyvinul Dr. Ton Weijters, využívá Heuristický přístup k řešení mnoha problémů, které nebylo možné s alfa algoritmem provést, což činí tento algoritmus mnohem vhodnější v praxi pro reálné event logy. [39]

Podstatou Heuristic mineru je, že odvozuje logické členy XOR a AND od vztahových závislostí a bere v potaz frekvenci a sekvenci událostí při tvorbě reálného modelu. Hlavní myšlenka je taková, že příležitostné cesty by neměly být zahrnuty do modelu. [39], [22]

Jednou z výhod je, že výstupem je Heuristická síť, která může být převedena na jiné typy procesních modelů (například Petriho síť) pro další analýzu v nástroji ProM. Nejvíce se využívá v analýze reálných event logů, které neobsahují příliš mnoho různých událostí. [44]

#### **4.2 Inductive miner**

Inductive miner je vylepšený algoritmus v programu ProM, který vychází z alfa mineru a Heuristic mineru. Podstatou Inductive mineru je, že je založen na procesních stromech a netrpí problémy spojené s deadlockem, livelockem nebo jinými anomáliemi. [5]

Deadlock se v procesním řízení označuje pro situaci dvou případů ve stejném procesu, kdy úspěšné dokončení prvního případu je podmíněno předchozím dokončením druhého případu, přičemž druhý případ může být dokončen až po dokončení prvního případu. Jinými slovy oba případy čekají navzájem až každý z nich dokončí svoji úlohu, ale ta nikdy nenastane. [38]

Livelock se v procesním řízení označuje jako podmínka, která nastává, když dva nebo více programů (v process miningu případů) průběžně mění svůj stav, bez toho, aniž by dosáhly cílového stavu. Můžou nastat dva stavy, kdy se případy dokola opakují a nastane nekonečný cyklus anebo může případ postupovat proměnlivou cestou dále bez zacyklení. [25]

Inductive miner je jeden z několika algoritmů patřící do skupiny Inductive mining, jehož rámec je dobře rozšiřitelný a zahrnuje techniky, které zvládnou zpracovat rozsáhlé event logy a vytvořit zobecněné modely při zachování čtyř kritérií kvality. [5]

Výstupy vytvořené těmito technikami lze snadno převést na jiné notace, například Petriho sítě nebo BPMN. Inductive mining je v současnosti jedním z nejvyužívanějších technik k objevování procesů díky své flexibilitě, formálním zárukám a škálovatelnosti. [5]

### **4.3 Fuzzy Miner**

Fuzzy miner je jedním z novějších algoritmů pro zjišťování procesů a byl vyvinut spoluzakladatelem společnosti Fluxicon Christianem Güntherem v roce 2007. Jedná se o první algoritmus, který se přímo zabývá problémy velkého počtu aktivit a vysoce nestrukturovaného chování.[20], [41]

Fuzzy miner používá metriky významnosti/korelace, aby interaktivně zjednodušil procesní model na požadované úrovni a zdůraznil frekventované aktivity a cesty. Ve srovnání s Heuristic minerem, Fuzzy miner může také vynechat méně důležité aktivity (nebo je skrýt v klastrech), pokud jich je příliš. [5]

Výstupem je Fuzzy model, který nelze převést na jiné typy jazyků procesního modelování, ale lze jej použít k animaci event logů. Díky animaci event logů, lze získat pocit dynamicky se měnícího modelu. [39]

Nejvíce se využívá pro analýzu složitých a nestrukturovaných dat v event logu nebo pro zjednodušení interaktivní modelu. Modifikovaný Fuzzy miner v dnešní době využívá většina komerčních nástrojů. [39]

## 5 REALIZOVANÁ ANALÝZA

V této kapitole bude zpracována vlastní analýza event logů pomocí zvolené metody a bude vytvořen reálný model procesu. Vzhledem k objektivnímu posouzení výsledků metod ze zkoumaných nástrojů budou výstupy ze všech nástrojů mít podobu Fuzzy modelu.

Součástí analýzy bude zároveň představení prostředí nástrojů pro Process Mining, výsledky statistik event logu a animace reálného modelu procesu ve Fuzzy modelu.

### 5.1 Data

Data (event log) jsou stažena z internetu a týkají se blíže nespecifikovaného procesu nákupu nějaké firmy. Níže uvedená tabulka 6 zobrazuje náhled na část event logu. Event log obsahuje povinné atributy Case ID, Activity a Timestamp. Nechybí ani doplňkové atributy pro eventuálně komplexnější analýzu, jako jsou Resource a Role. Na základě údajů o zdrojích a rolích lze například vytvořit analýzu sociální sítě a získat tak přehled o personální struktuře ve firmě. Data jsou spíše nestructurovaná. V kapitolách 5.2, 5.3 a 5.4 budou tato data analyzována v jednotlivých nástrojích.

Tabulka 6: Náhled event logu

Case ID	Start Timestamp	Complete Timestamp	Activity	Resource	Role
1	2011/01/01 00:00:00.000	2011/01/01 00:37:00.000	Create Purchase Requisition	Kim Passa	Requester
1	2011/01/01 05:37:00.000	2011/01/01 05:45:00.000	Create Request for Quotation	Kim Passa	Requester
1	2011/01/01 06:41:00.000	2011/01/01 06:55:00.000	Analyze Request for Quotation	Karel de Groot	Purchasing Agent
1	2011/01/01 11:43:00.000	2011/01/01 12:09:00.000	Send Request for Quotation to Supplier	Karel de Groot	Purchasing Agent
1	2011/01/01 12:32:00.000	2011/01/01 16:03:00.000	Create Quotation Comparison Map	Magdalena Predutta	Purchasing Agent
1	2011/01/01 22:44:00.000	2011/01/01 23:13:00.000	Analyze Quotation Comparison Map	Immanuel Karagianni	Requester
1	2011/01/01 23:13:00.000	2011/01/01 23:13:00.000	Choose best option	Tesca Lobes	Requester
1	2011/01/02 01:22:00.000	2011/01/02 09:20:00.000	Settle conditions with supplier	Francois de Perrier	Purchasing Agent
1	2011/01/02 09:58:00.000	2011/01/02 10:10:00.000	Create Purchase Order	Karel de Groot	Purchasing Agent
...	...	...	...	...	...

*Zdroj: vlastní zpracování*

Tabulka 7 zobrazuje překlad aktivit.

**Tabulka 7: Aktivity v event logu**

<b>Activity</b>	<b>Aktivita</b>
Amend Purchase Requisition	Upravení nákupního požadavku
Amend Request for Quotation Requester	Upravení požadavku žadatele o kotaci
Amend Request for Quotation Requester Manager	Upravení požadavku žadatele o kotaci správcem
Analyze Purchase Requisition	Analýza nákupního požadavku
Analyze Quotation Comparison Map	Analýza komparační mapy nákupu
Analyze Request for Quotation	Analýza žádosti o kotaci
Approve Purchase Order for payment	Schválení platby objednávky
Authorize Supplier's Invoice payment	Autorizace platby faktury dodavatele
Choose best option	Výběr nejlepší možnosti
Confirm Purchase Order	Potvrzení objednávky
Create Purchase Order	Vytvoření objednávky
Create Purchase Requisition	Vytvoření nákupního požadavku
Create Quotation Comparison Map	Vytvoření komparační mapy o kotaci
Create Request for Quotation Requester	Vytvoření požadavku žadatele o kotaci
Create Request for Quotation Requester Manager	Vytvoření požadavku žadatele o kotaci správcem
Deliver Goods Services	Doručování objednávky
Pay invoice	Úhrada faktury
Release Purchase Order	Dokončení objednávky
Release Supplier's Invoice	Vyřízení faktury dodavatele
Send Request for Quotation to Supplier	Odeslání požadavku dodavateli o kotaci nabídky
Send invoice	Zaslání faktury
Settle conditions with supplier	Urovnání podmínek s dodavatelem
Settle dispute with supplier Financial Manager	Řešení sporu s finančním manažerem dodavatele
Settle dispute with supplier Purchasing Agent	Řešení sporu s obchodním zástupcem dodavatele

*Zdroj: vlastní zpracování*

Tabulka 8 zobrazuje překlad rolí v organizaci.

**Tabulka 8: Role v event logu**

<b>Role</b>	<b>Pozice</b>
Purchasing Agent	Nákupní agent
Requester	Žadatel
Financial Manager	Finanční manažer
Requester Manager	Správce žádosti
Supplier	Dodavatel

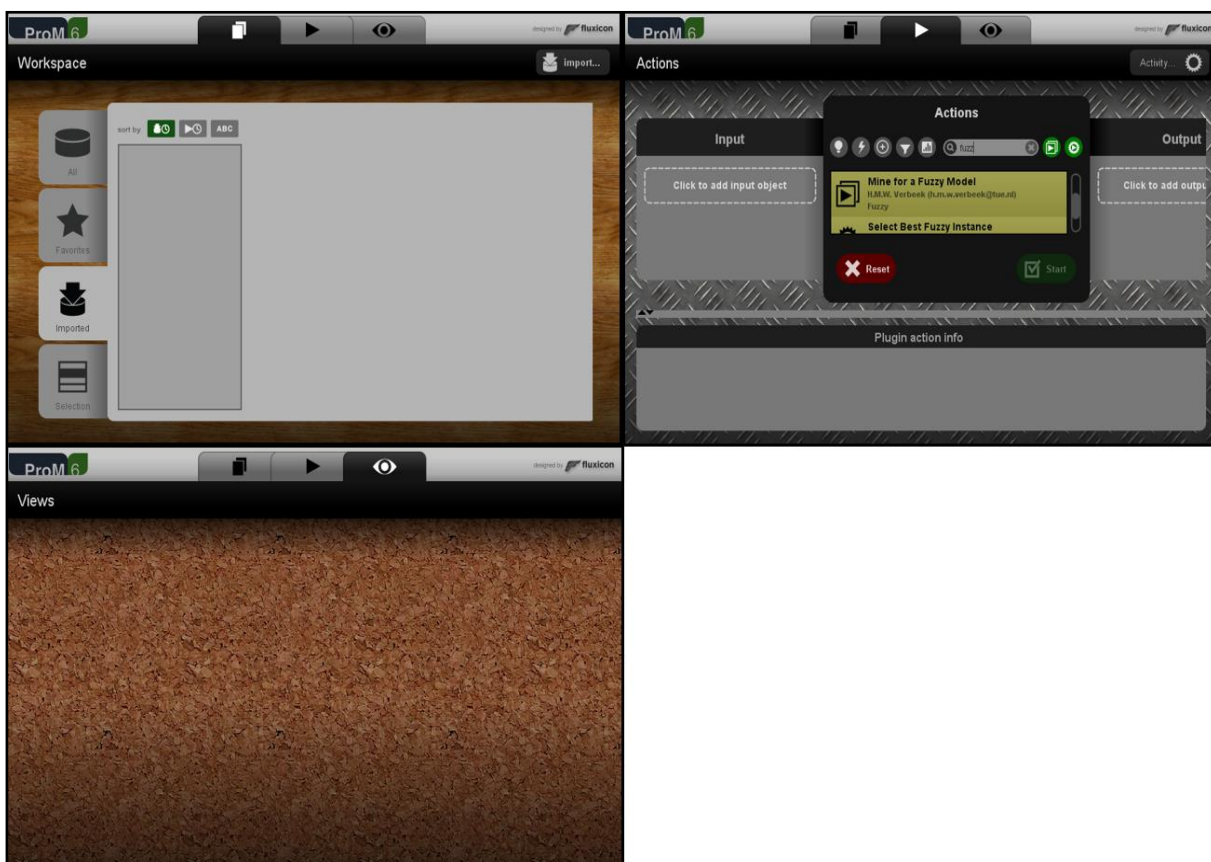
*Zdroj: vlastní zpracování*

## 5.2 ProM

Na obrázku 10 je zobrazeno prostředí programu ProM, které je rozděleno na 3 hlavní části – Workspace, Actions a Views. Ve Workspace je možné importovat event logy skrz tlačítko Import. Počet importovaných souborů je libovolný a je možné je řadit podle času importu, posledního používání nebo jména autora. Načteným event logům lze také přiřadit prioritu pomocí hvězdy, díky které se v seznamu nahraných event logů budou tyto označené soubory zobrazovat vždy mezi prvními.

Druhá část, Actions, se využívá pro analýzu event logů pomocí dostupných plug-inů. Aplikace zahrnují algoritmy pro dolování, filtrování, shlukování, formátování nebo jiné plug-iny pro dodatečné analýzy event logů. Na straně inputu je importovaný event log a na straně outputu možný výsledný model v určitém formátu.

Ve třetí části, Views, je možné procházet vytvořené statistiky event logu nebo si prohlédnout reálný model procesu, který je možné u některých plug-inů pomocí parametrů měnit podle potřeby a tím získat lepší pojem o tom, jak funguje. Výstupy mohou být například ve formátu Petriho sítí, BPMN, Fuzzy modelu atd.

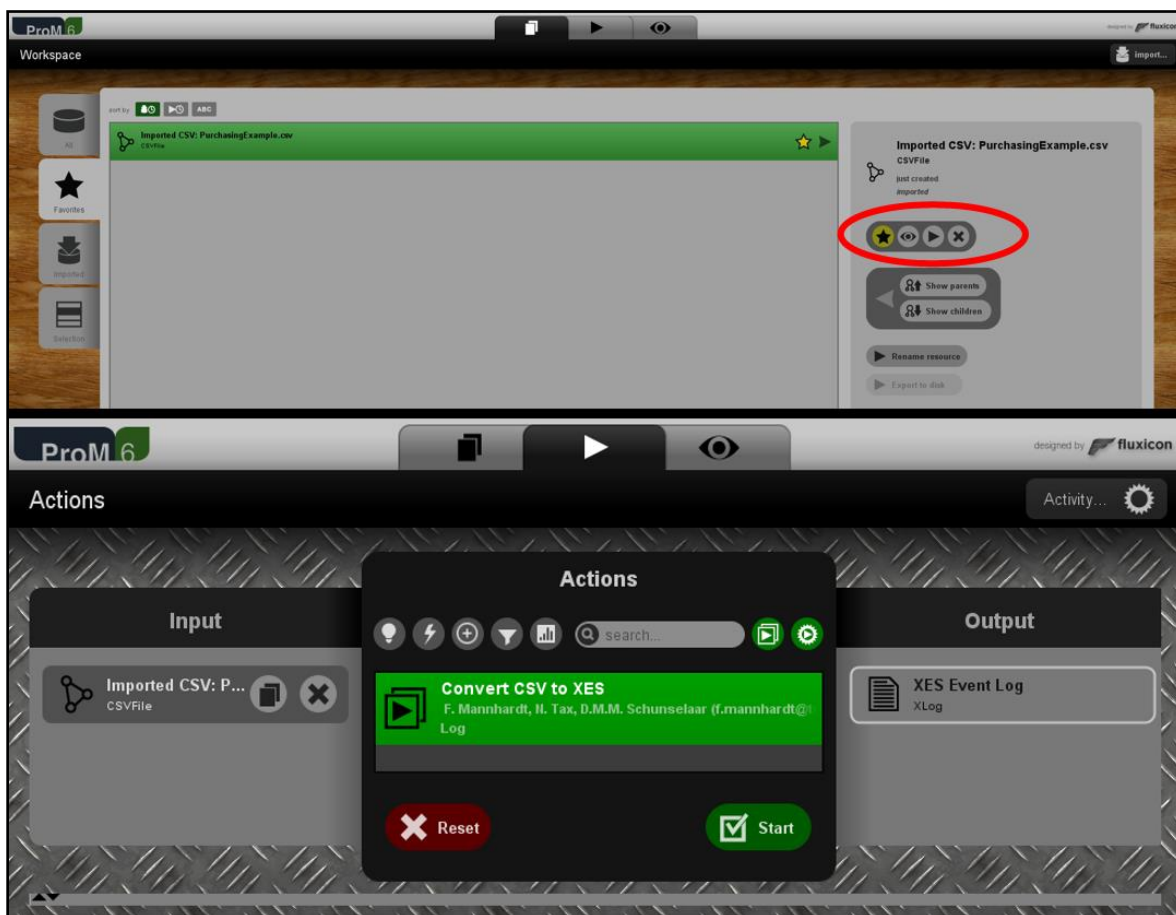


Obrázek 10: Prostředí v ProM

*Zdroj: vlastní zpracování*

## 5.2.1 Import event logů v ProM

Prvním krokem pro analýzu procesu je import event logu. Na obrázku 11 je zobrazen importovaný soubor „PurchasingExample“ ve formátu CSV. Před samotnou analýzou event logu je potřeba transformovat CSV do XES, protože jak již bylo řečeno dříve, ProM dokáže vytvářet pouze reálné modely z event logu ve formátu XES nebo MXML.



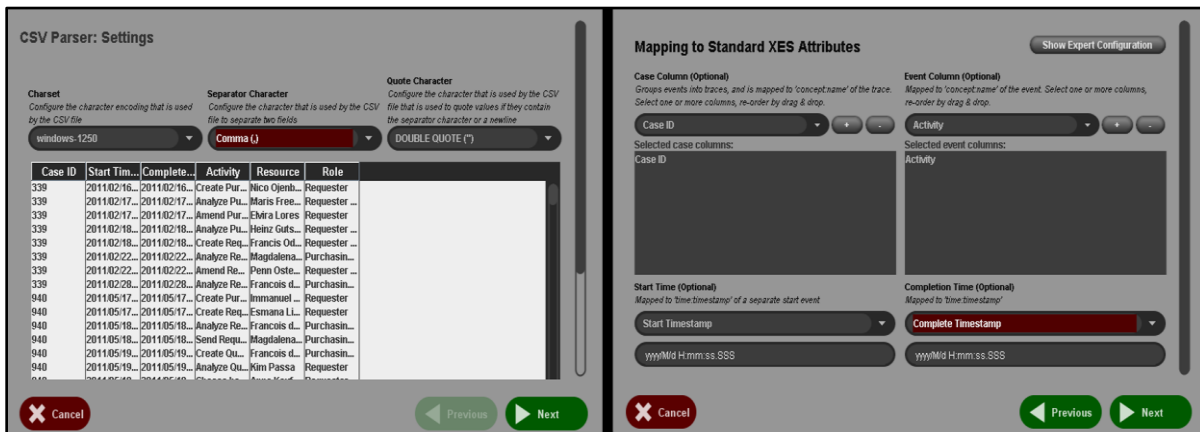
Obrázek 11: Import event logu v ProM

*Zdroj: vlastní zpracování*

Při konverzi event logu je potřeba v průvodci konverze ručně nakonfigurovat nastavení a datové typy jednotlivých atributů. Dále je možné nastavit kódování CSV dokumentu a jeho oddělovač. Většinou si s tím ProM dokáže poradit sám, stačí pouze potvrdit předdefinované nastavení. V dalším kroku se mapují standardní atributy XES, tedy case ID, aktivity a časová stopa pro začátek a konec. Tuto skutečnost zobrazuje obrázek 12.

Konverzi lze obejít v případě, že máme k dispozici event logy v XES nebo MXML. Ty je možné ještě před importováním transformovat přes jiné nástroje, než je ProM. Například v době, kdy plug-in pro konverzi formátu nebyl součástí ProM, bylo možné převést formáty pomocí open source nástroje XESame [47].

Obrázek 12 zobrazuje konverzi event logů a nastavení parametrů.

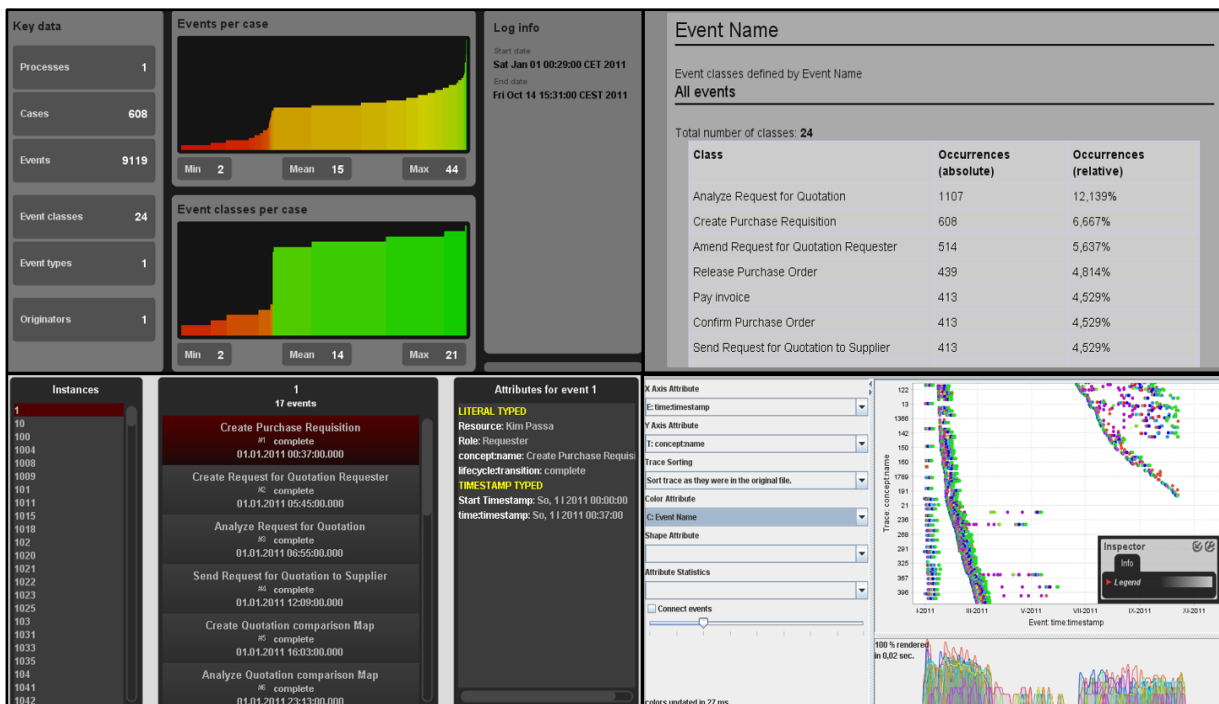


Obrázek 12: Konverze event logu v ProM

*Zdroj: vlastní zpracování*

## 5.2.2 Statistika v ProM

Po konverzi je možné zobrazit kompletní přehled statistiky event logu. Obrázek 13 ilustruje několik možných náhledů na event log, například pomocí Dashboardu, Inspektoru nebo Summary. Dále je možné změnit typ vizualizace, například na interaktivní tečkovaný graf, u kterého uživatel může měnit atributy na osách nebo měnit barvu teček na základě zvolených atributů.



Obrázek 13: Statistika v ProM

*Zdroj: vlastní zpracování*



Na dashboardu na obrázku 13 lze vidět souhrn základních informací o event logu, například počet procesů, případů, událostí nebo celková doba od zahájení první aktivity až po skončení poslední aktivity.

Je zde i vykreslený graf, který zobrazuje počet událostí na jeden případ s minimálním, průměrným a maximálním počtem událostí na jeden případ. Druhý graf se týká aktivit na jeden případ a význam počtů je analogický.

V sekci Inspector lze využít Browser k prozkoumávání jednotlivých instancí, tj. případů. Je možné se proklikat všemi případy a zobrazit události, které v tomto případě běží. Dále je možné se proklikat skrz aktivity, které obsahují přiřazené atributy, například zdroj nebo role.

Další zajímavou funkcí je Explorer, který seřadí všechny případy do seznamu, kde jsou v každém řádku za sebou seřazeny aktivity ve tvaru šipky. Každá šipka může mít jednu ze 4 možných barev, která vyjadřuje frekvenci aktivity.

Například na obrázku 14 lze vidět, že instance 10 obsahuje celkem 18 událostí. Červeně zvýrazněná aktivita „Settle dispute with supplier Financial Manager“ vykazuje anomálii na základě frekvence, která je velmi nízká, pouze 7.23 %. To znamená, že se tato aktivita v rámci celého procesu mockrát neobjevila.

Summary slouží především ke shrnutí absolutního a relativního počtu všech událostí, jak začínajících, tak končících.



Obrázek 14: Explorer event logu v ProM

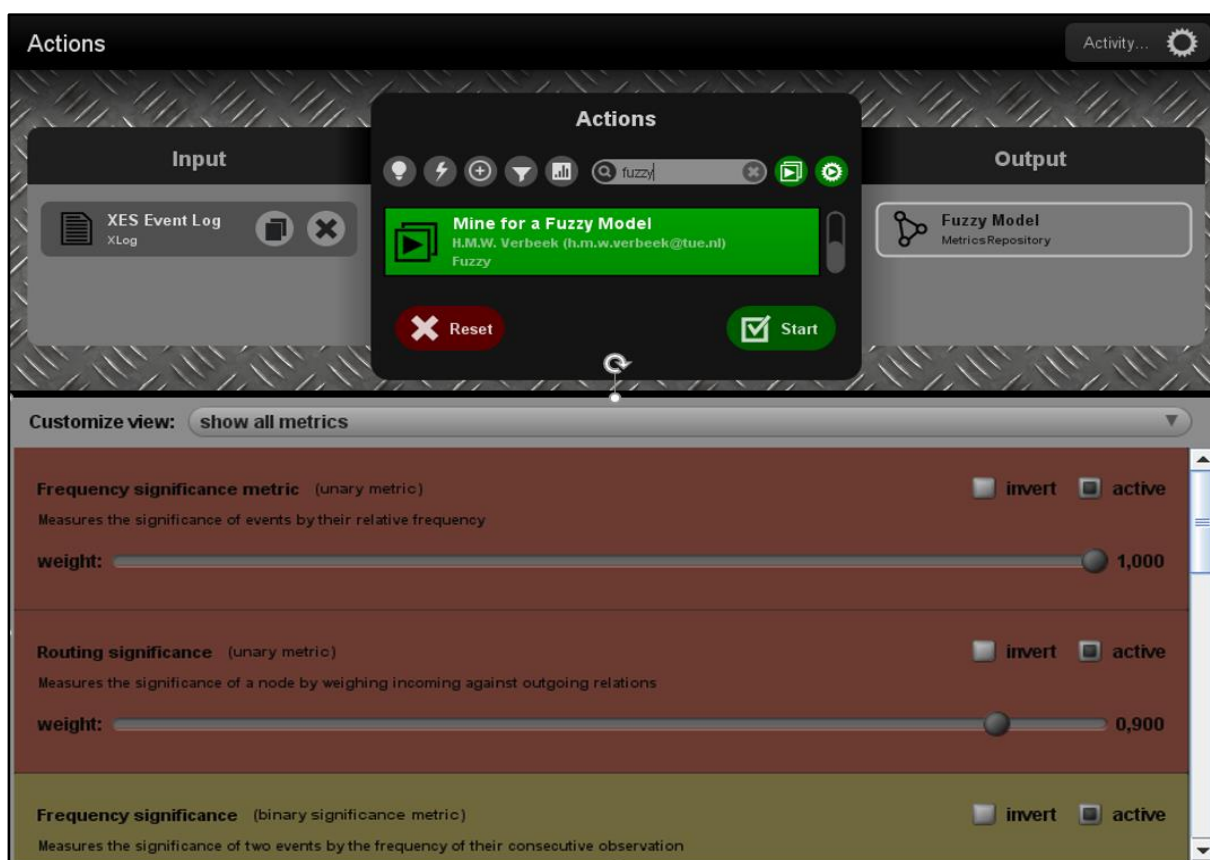
*Zdroj: vlastní zpracování*

### 5.2.3 Prozkoumávání reálného modelu procesu v ProM

ProM podporuje různé algoritmy pro dolování v event logech. Vzhledem k tomu, že je tato práce zaměřená na komparaci nástrojů a metod, bude reálný model procesu vytvořen pomocí Fuzzy mineru, protože Disco a Celonis taktéž využívají Fuzzy miner, i když v určité modifikaci. Každopádně, s využitím Fuzzy mineru, by mělo být docíleno objektivního porovnání.

Obrázek 15 zobrazuje plug-in – Fuzzy miner, který vytvoří reálný model procesu. Před spuštěním analýzy, průvodce plug-inem umožňuje nastavit metriky (unární, binární, binární korelační). Automaticky jsou všechny vybrané a podle [20] je to správná volba a měly by tak zůstat, protože změněním metrik se ve většině případů docílí prakticky stejného výsledku. Záleží na strukturovanosti dat a rozsáhlosti event logu.

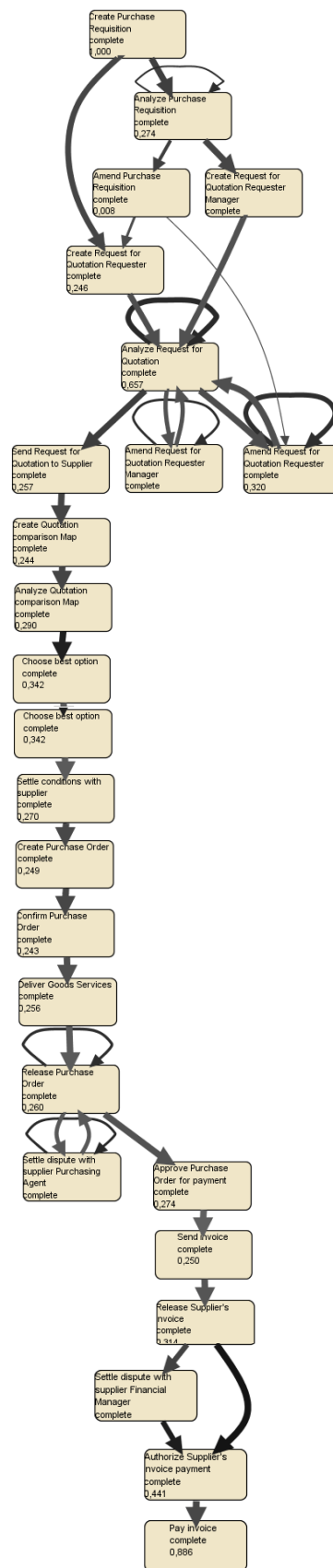
Navíc, i kvůli objektivní analýze zůstanou všechny parametry a metriky defaultně nastavené, protože v Disco a Celonisu není žádná možnost nastavování parametrů nebo metrik před vytvořením reálného modelu procesu. Na obrázku 15 v pravé části je také vidět, že výstupem analýzy je Fuzzy model.



Obrázek 15: Průvodce Fuzzy minerem v ProM

*Zdroj: vlastní zpracování*

Obrázek 16 zobrazuje výsledný model reálného procesu vytvořený v nástroji ProM.

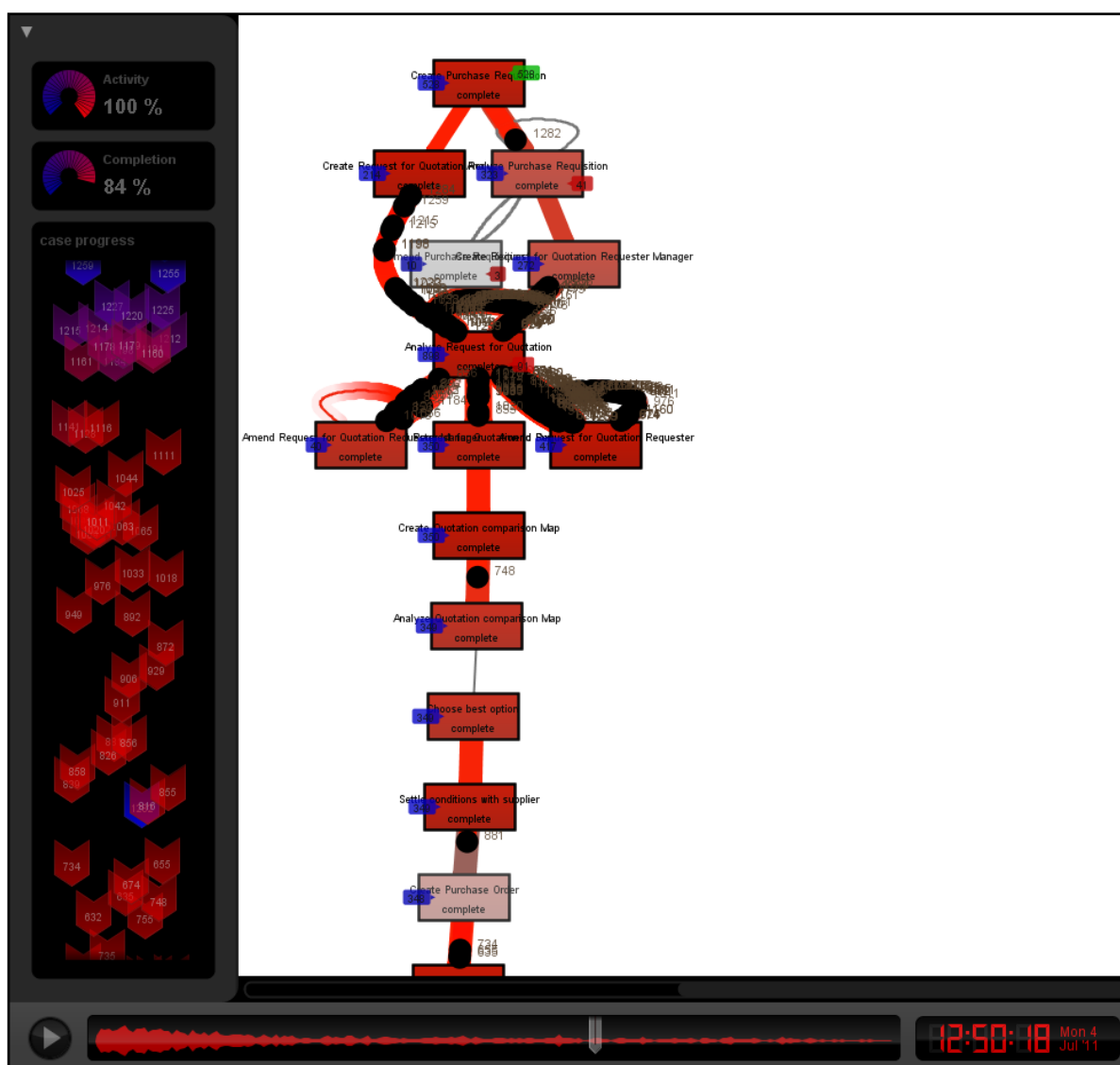


Obrázek 16: Reálný model procesu v ProM

Zdroj: vlastní zpracování

## 5.2.4 Animace reálného modelu procesu v ProM

Samotný plug-in Fuzzy Miner dokáže pouze vytvořit reálný model a dále pomocí parametrů měnit výsledek. K vytvoření animace je potřeba využít jiného plug-inu – Fuzzy Animation, který je napsaný stejnou skupinou vývojářů. Tento plug-in dokáže vytvořit animace založené na event logech v reálném modelu. Takové animace jsou velice užitečné pro pochopení chování procesu. Například na obrázku 17 lze vidět v reálném modelu černé tečky charakterizující jednotlivé případy. Na levé straně se paralelně pohybují šipky směrem dolů, které taktéž charakterizují jednotlivé případy. Pohybují se tak rychle, jak je případ vyřešen. Modré šipky se pohybují rychle, červené pomalu a fialové se pohybují průměrnou rychlostí. Na příkladu lze pozorovat, že v červnu roku 2011 je aktivita „Analyze Request for Quotation“ nejvíce přetížená a díky tomu dochází k brždění celého procesu.

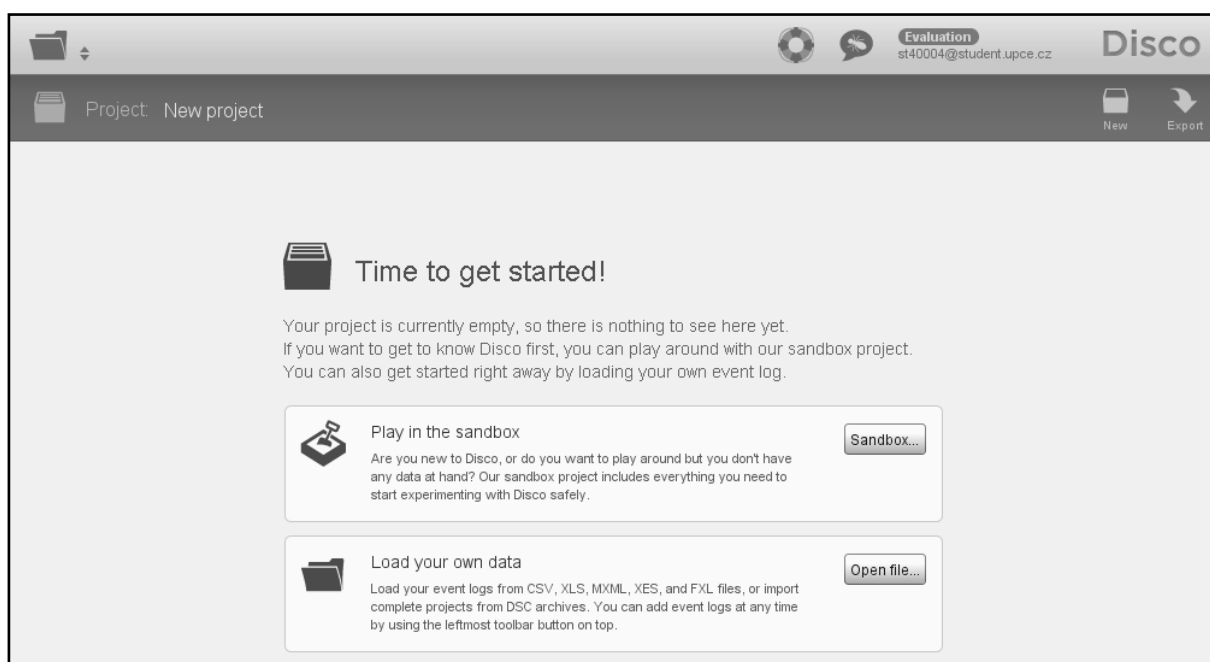


Obrázek 17: Animace reálného modelu procesu v ProM

*Zdroj: vlastní zpracování*

## 5.3 Disco

Obrázek 18 zobrazuje úvodní stránku v programu Disco. Hned na první stránce je možné importovat vlastní event logy nebo experimentovat s příloženými event logy. Disco je komerční nástroj a v základu nabízí „Demo“ verzi, která je limitovaná pro komerční využití, export modelu a import událostí v event logu je omezen na 100 událostí. Pro účely této diplomové práce mně byla poskytnuta „Evaluation“ licence, díky které program navíc nabízí neomezený import událostí a je tak možné relevantně porovnat výsledky ze všech nástrojů.



Obrázek 18: Úvodní stránka v Disco

*Zdroj: vlastní zpracování*

### 5.3.1 Import event logů v Disco

Disco bylo navrženo tak, aby usnadnilo import event logů pomocí automatického zjišťování časových stop, případů, událostí a dalších atributů. Načítání dat bylo navíc i rychlejší oproti programu ProM o několik vteřin, což nějak dramaticky nemění situaci. Event log v této práci obsahuje podle statistiky 9119 událostí, avšak pokud by event log obsahoval miliony událostí, mohl by se časový rozdíl pohybovat v řádech hodin. Záleží také na výpočetní technice.

V případě, že by se importoval event log ve standardním formátu XES, tak by se doba mohla zkrátit pro rozsáhlé data sety až 10x, protože XES má již informace o event logu nakonfigurované. Pro ještě rychlejší nahrávání event logů je možné transformovat XES nebo CSV do formátu FXL, který byl vytvořen společností Disco. Nejedná se o standardní formát, ale tento formát může rapidně urychlit import rozsáhlých data setů. Pro představu, rozsáhlý

event log ve formátu CSV obsahující počet událostí v řádech milionů, by se s průměrnou PC sestavou importoval 3.5 hodiny. Stejný soubor, ale formátu XES, by se importoval 20 minut a soubor formátu FXL přibližně 2 minuty. [41]

Obrázek 19 zobrazuje průvodce importováním event logu. Každý sloupec zastupuje atribut. Rozložení tabulky event logu v programu Disco je stejné jako rozložení tabulky ve formátu CSV. Dále je možné nastavit kódování stejně jako tomu bylo i v programu ProM. Funkce, kterou ale ProM postrádá je, že plug-in nemá podporu změny formátu času. Například formát „yyyy/MM/dd HH:mm:ss.SSS“ nelze změnit na „dd.MM.yyyy HH:mm“.

Po importu event logu se rovnou v prostředí Disco objeví 3 záložky – Maps, Statistics a Cases.

Case ID	Start Timestamp	Complete Timestamp	Activity	Resource	Role
7 339	2011/02/22 10:50:00.000	2011/02/22 11:03:00.000	Amend Request for Quotation Requester	Penn Osterwalder	Requester Mana
8 339	2011/02/28 08:10:00.000	2011/02/28 08:34:00.000	Analyze Request for Quotation	Francois de Perrier	Purchasing Ager
9 940	2011/05/17 06:31:00.000	2011/05/17 07:08:00.000	Create Purchase Requisition	Immanuel Karagianni	Requester
10 940	2011/05/17 09:58:00.000	2011/05/17 10:06:00.000	Create Request for Quotation Requester	Esmana Liubiata	Requester
11 940	2011/05/18 19:30:00.000	2011/05/18 19:56:00.000	Analyze Request for Quotation	Francois de Perrier	Purchasing Ager
12 940	2011/05/18 23:46:00.000	2011/05/18 23:59:00.000	Send Request for Quotation to Supplier	Magdalena Predutta	Purchasing Ager
13 940	2011/05/19 03:44:00.000	2011/05/19 08:31:00.000	Create Quotation comparison Map	Francois de Perrier	Purchasing Ager
14 940	2011/05/19 15:38:00.000	2011/05/19 15:52:00.000	Analyze Quotation comparison Map	Kim Passa	Requester
15 940	2011/05/19 15:52:00.000	2011/05/19 15:52:00.000	Choose best option	Anna Kaufmann	Requester
16 940	2011/05/20 23:31:00.000	2011/05/21 09:22:00.000	Settle conditions with supplier	Magdalena Predutta	Purchasing Ager
17 940	2011/05/21 18:48:00.000	2011/05/21 18:59:00.000	Create Purchase Order	Francois de Perrier	Purchasing Ager
18 940	2011/05/22 11:33:00.000	2011/05/22 11:44:00.000	Confirm Purchase Order	Esmeralda Clay	Supplier
19 940	2011/05/23 05:32:00.000	2011/05/24 13:46:00.000	Deliver Goods Services	Esmeralda Clay	Supplier
20 940	2011/05/24 20:59:00.000	2011/05/24 21:00:00.000	Release Purchase Order	Kim Passa	Requester
21 940	2011/05/26 07:41:00.000	2011/05/26 07:42:00.000	Approve Purchase Order for payment	Karel de Groot	Purchasing Ager
22 940	2011/05/28 01:11:00.000	2011/05/28 01:11:00.000	Send invoice	Kiu Kan	Supplier
23 940	2011/05/28 15:28:00.000	2011/05/28 15:28:00.000	Authorize Supplier's Invoice payment	Pedro Alvares	Financial Managi
24 940	2011/05/28 16:11:00.000	2011/05/28 16:19:00.000	Pay invoice	Karalda Nirwada	Financial Managi
25 1417	2011/07/23 12:53:00.000	2011/07/23 13:31:00.000	Create Purchase Requisition	Christian Francois	Requester
26 1417	2011/07/23 17:51:00.000	2011/07/23 17:59:00.000	Create Request for Quotation Requester	Immanuel Karagianni	Requester
27 1417	2011/08/02 07:02:00.000	2011/08/02 07:24:00.000	Analyze Request for Quotation	Karel de Groot	Purchasing Ager
28 1417	2011/08/02 08:17:00.000	2011/08/02 08:27:00.000	Amend Request for Quotation Requester	Anna Kaufmann	Requester Mana
29 1417	2011/08/08 04:03:00.000	2011/08/08 04:23:00.000	Analyze Request for Quotation	Magdalena Predutta	Purchasing Ager
30 159	2011/01/21 15:02:00.000	2011/01/21 15:42:00.000	Create Purchase Requisition	Elvira Lores	Requester
31 159	2011/01/22 02:58:00.000	2011/01/22 03:05:00.000	Analyze Purchase Requisition	Francis Odell	Requester Mana
32 159	2011/01/22 03:11:00.000	2011/01/22 03:14:00.000	Create Request for Quotation Requester Manager	Maris Freeman	Requester Mana
33 159	2011/01/22 03:15:00.000	2011/01/22 03:59:00.000	Analyze Request for Quotation	Francois de Perrier	Purchasing Ager

Obrázek 19: Import event logů v Disco

*Zdroj: vlastní zpracování*

### 5.3.2 Statistika event logů v Disco

Statistika v Disco je velice propracovaná a intuitivní. Poskytuje přehledné informace jak v tabulkách, které lze podle atributu filtrovat, tak v interaktivních grafech. Statistika je rozdělená na celkový přehled, aktivity v událostech a další atributy, které jsou uloženy v event logu. Je možné zobrazit různé grafy, u kterých můžeme měnit metriky. Na obrázku 20 jsou

například zobrazeny grafy, které ukazují celkový přehled event logu, aktivity v událostech a zdroje v událostech.

V celkovém přehledu je například na výběr, kolik událostí se vyskytuje v každém případě, kolik událostí proběhne za den, jak dlouho trvají jednotlivé události nebo i jednotlivé případy atd. Na obrázku 20 je zobrazen graf udávající počet událostí na případ. V tomto příkladu se v případě 106 nachází 18 událostí.

Graf, udávající aktivity v událostech na obrázku 20, především zobrazuje celkový počet aktivit ve všech případech.

Statistika taktéž nabízí přehled o zdrojích a rolích, pokud jsou k dispozici tyto informace v event logu. Například na obrázku 20 lze vidět, že frekvence zdroje „Kim Passa“ je 192. Znamená to, že Kim Passa 192x zadal aktivitu.



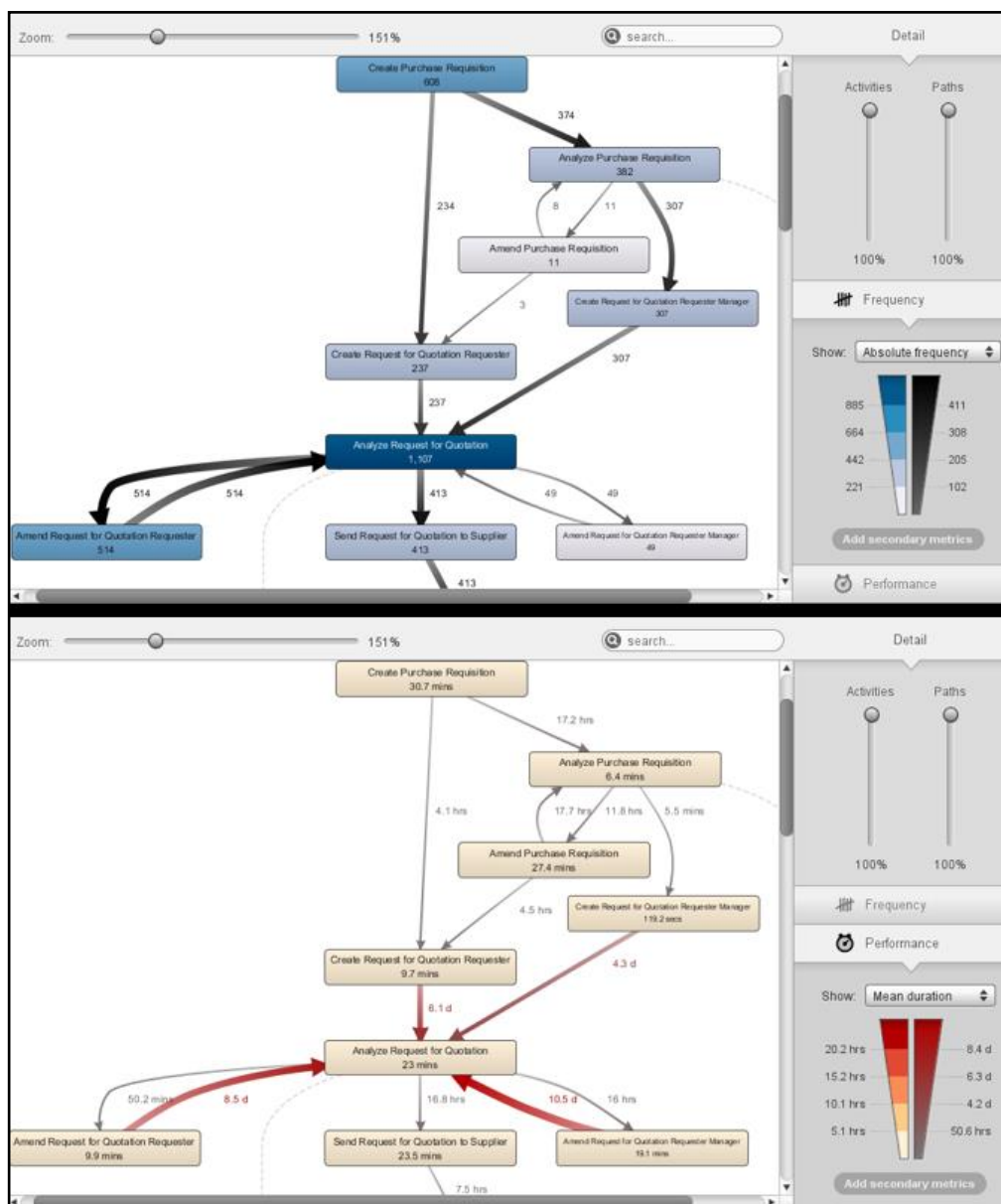
Obrázek 20: Statistika event logů v Discu

Zdroj: vlastní zpracování

### 5.3.3 Prozkoumávání reálného modelu procesu v Disco

Jak již bylo řečeno, po importu event logů se ihned vytvoří model, který je založený na modifikovaném Fuzzy mineru. V čem přesně je modifikovaný je neznámé, protože se jedná o obchodní tajemství společnosti Fluxicon.

Na obrázku 21 je vidět část vytvořeného reálného modelu procesu. Na pravé straně jsou k dispozici dvě posuvné lišty – Activities a Paths. Activities znamená, že při posouvání směrem dolů zmizí aktivity, které jsou málo frekventované a naopak. Paths analogicky zvýrazní cesty mezi aktivitami. Disco dále nabízí 2 důležité funkce, pomocí kterých lze zvýraznit slabiny – Frekvenci a Výkon.

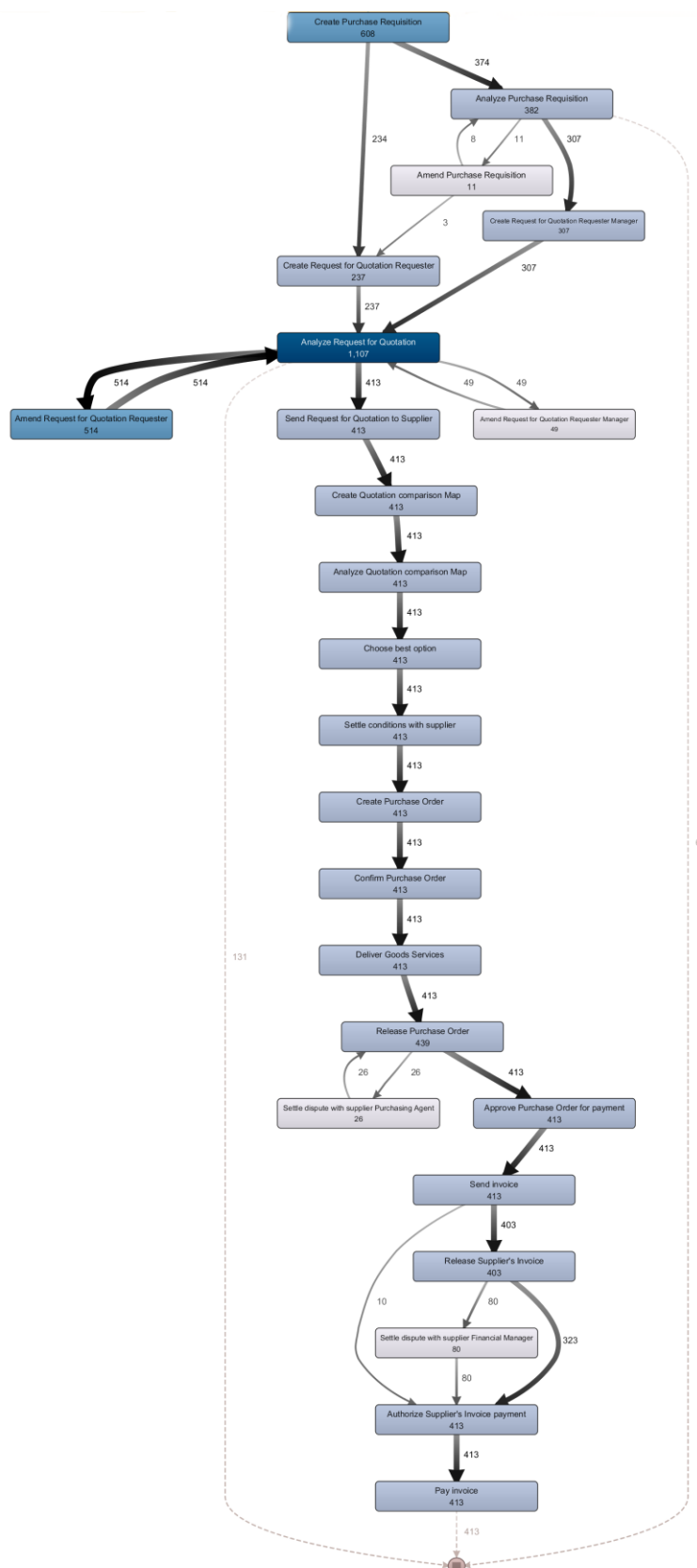


Obrázek 21: Frekvence a výkon reálného modelu procesu v Disco

Zdroj: vlastní zpracování



Obrázek 22 zobrazuje výsledný model reálného procesu vytvořený v programu Disco.



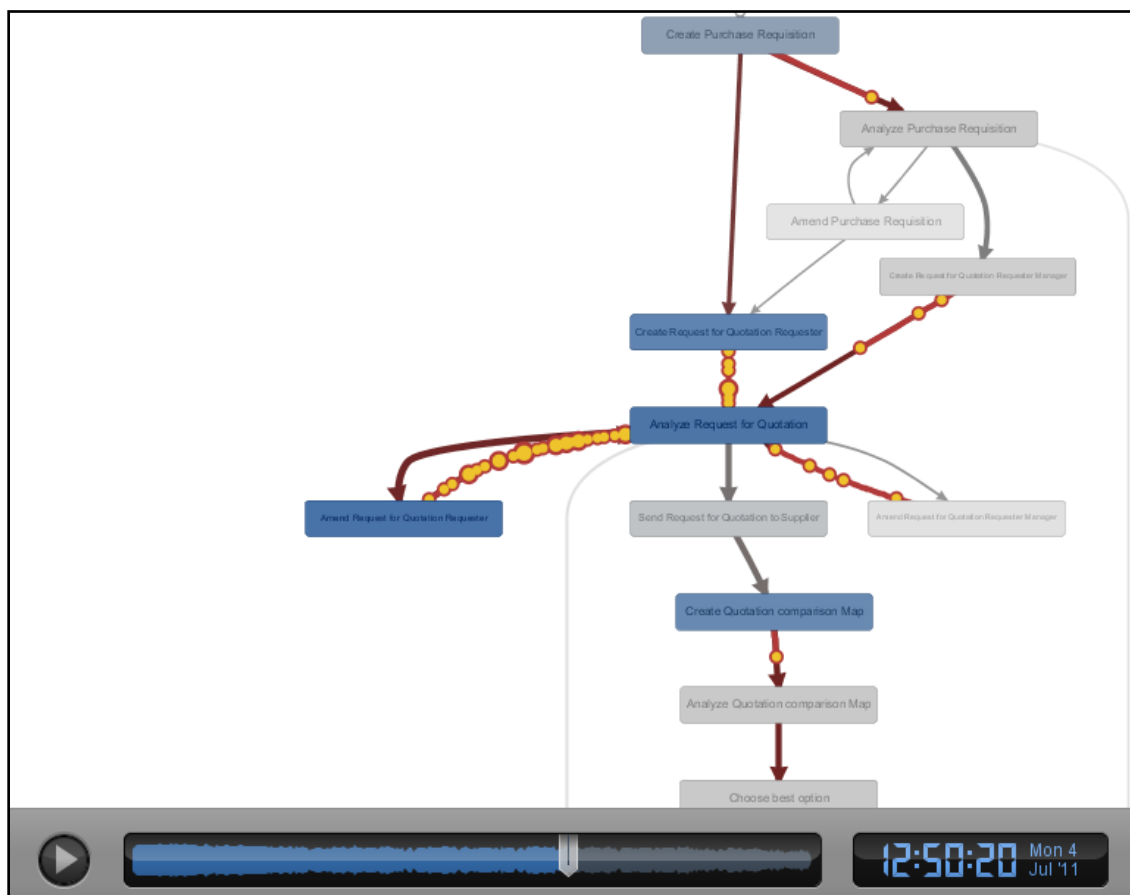
Obrázek 22: Reálný model procesu v Disco

Zdroj: vlastní zpracování

### 5.3.4 Animace reálného modelu procesu v Disco

Disco má již funkci animace reálného procesu modelu implementovanou.

Na obrázku 23 je zobrazena animace, která zachycuje skutečný průběh procesu na základě event logů, kde jednotlivé žluté tečky charakterizují probíhající případy. Průběh je velice podobný jako u nástroje ProM. Aktivita „Analyze Reuquest for Quation“ je stejně vytížená.



Obrázek 23: Animace reálného modelu procesu v Disco

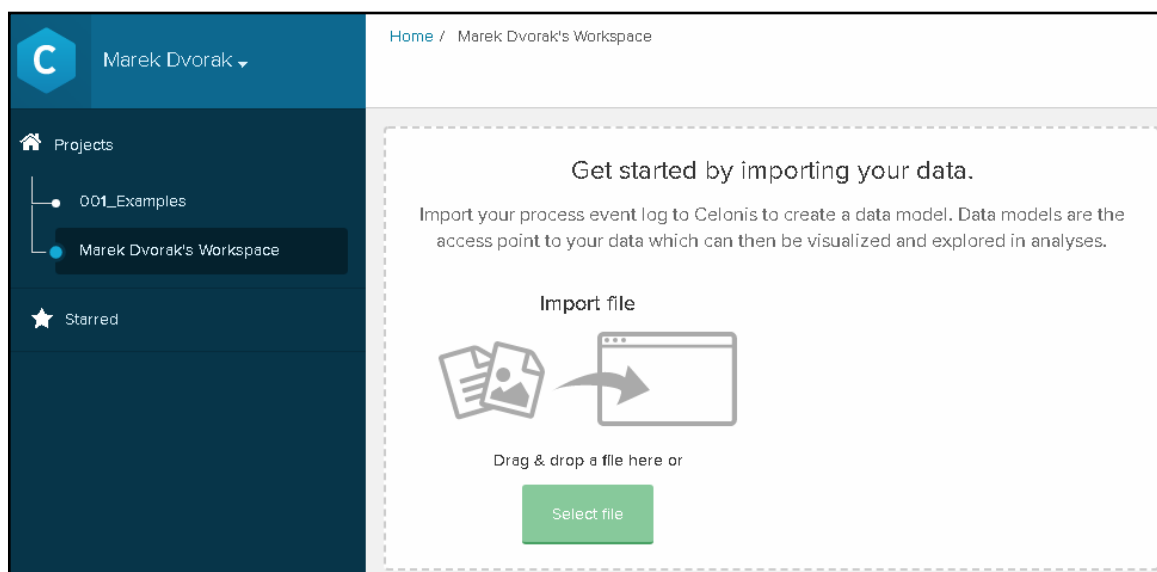
*Zdroj: vlastní zpracování*

## 5.4 Celonis

Celonis je komerční nástroj, který nemá desktopovou platformu, ale funguje na základě cloudového uložení, ke kterému se lze připojit přes webový prohlížeč. Velkou výhodou oproti konkurenci je jeho podpora desítek databázových management systémů, například SAP HANA, Oracle DB, MSSQL, MySQL, PostgreSQL nebo IBM DB2. Nejvíce se Celonis využívá právě ve spojení se SAPem. [5]

Celonis nabízí akademické licence a po ověření je možné dostat přístup k plné verzi po dobu 180 dní.

Obrázek 24 zobrazuje úvodní stránku v programu Celonis.

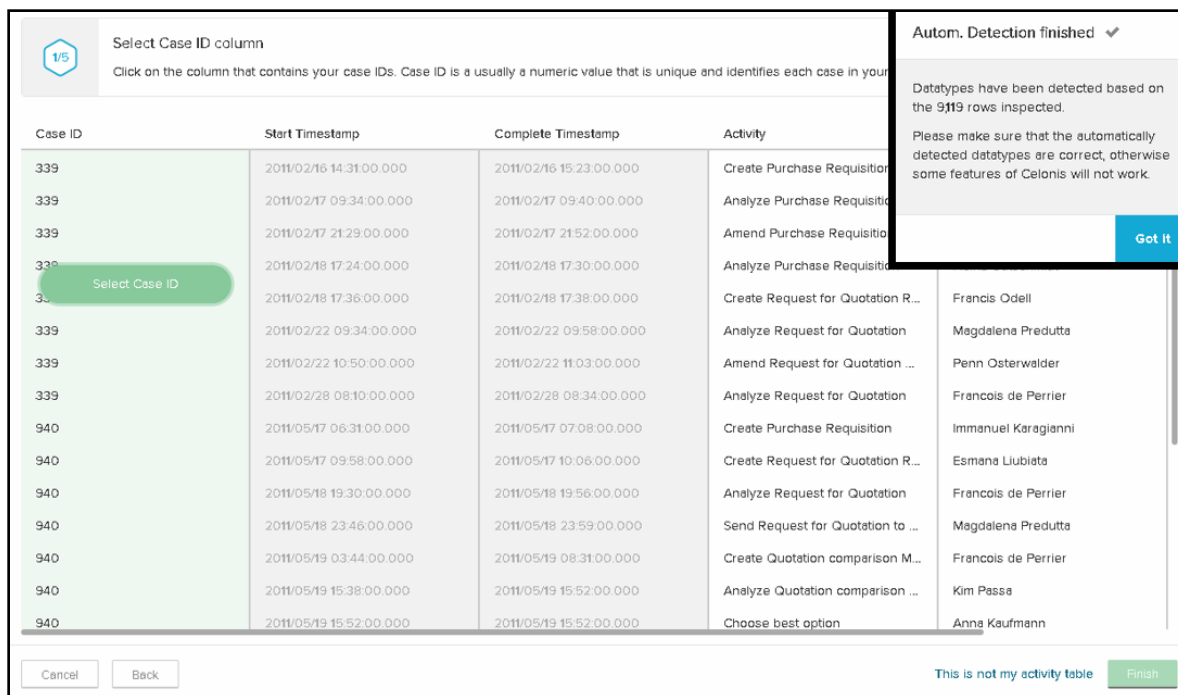


Obrázek 24: Úvodní stránka v Celonisu

*Zdroj: vlastní zpracování*

### 5.4.1 Import event logů v Celonisu

Tlačítkem import lze vložit event logy. Stejně jako v předchozích nástrojích, Celonis automaticky rozpozná atributy a přiřadí k nim odpovídající datové typy. Na obrázku 25 je náhled v průvodci importem.



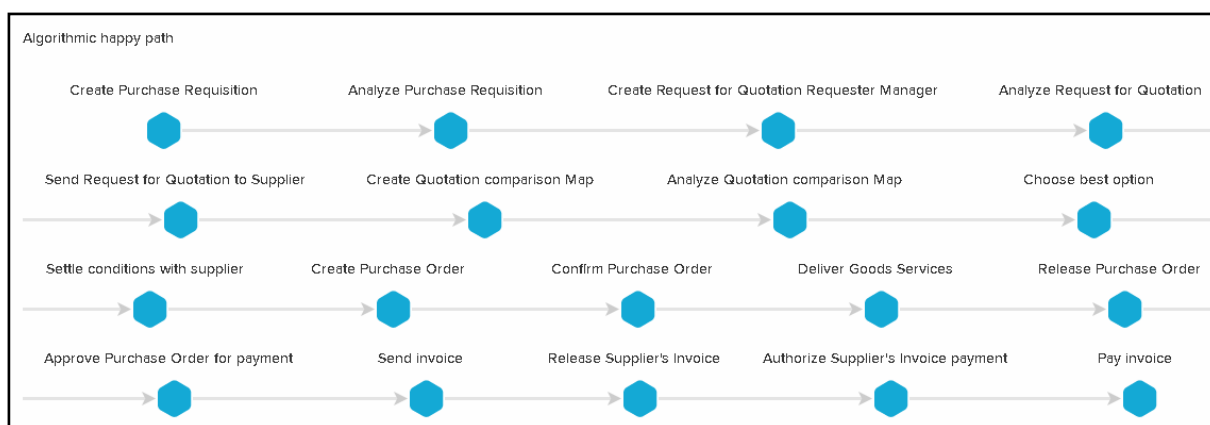
Obrázek 25: Import event logů v Celonisu

*Zdroj: vlastní zpracování*

## 5.4.2 Statistika event logů v Celonisu

Po importu lze nahlédnout do statistiky event logu. Je možné si vybrat mezi celkovým přehledem případů a událostí, průchodem reálného procesu a frekvencí aktivit.

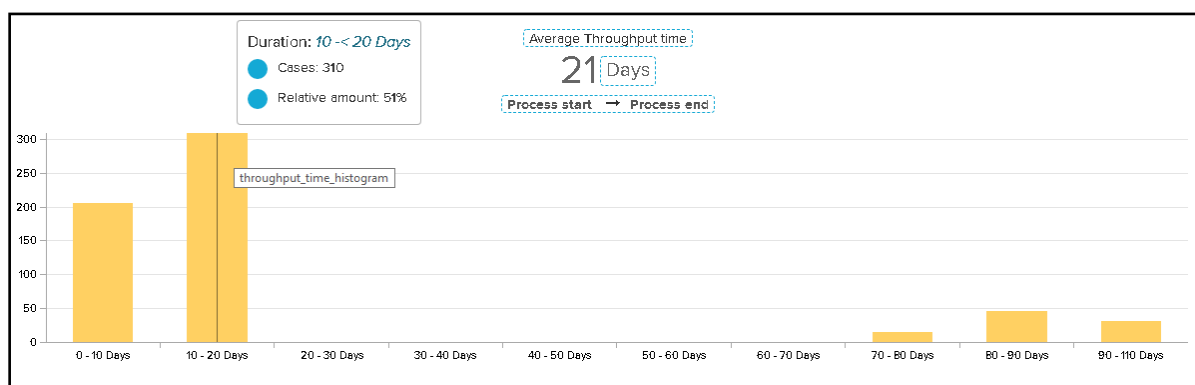
Celkový přehled nabízí například počet případů za den nebo počet událostí za den. Navíc, se zde oproti předchozím dvou nástrojům (ProM a Disco), objevila nová možnost náhledu na „Happy Path“, což v kontextu s modelováním je původní scénář procesu bez výjimečných a chybových cest. Obrázek 26 zobrazuje původní scénář procesu.



Obrázek 26: Happy path v Celonisu

*Zdroj: vlastní zpracování*

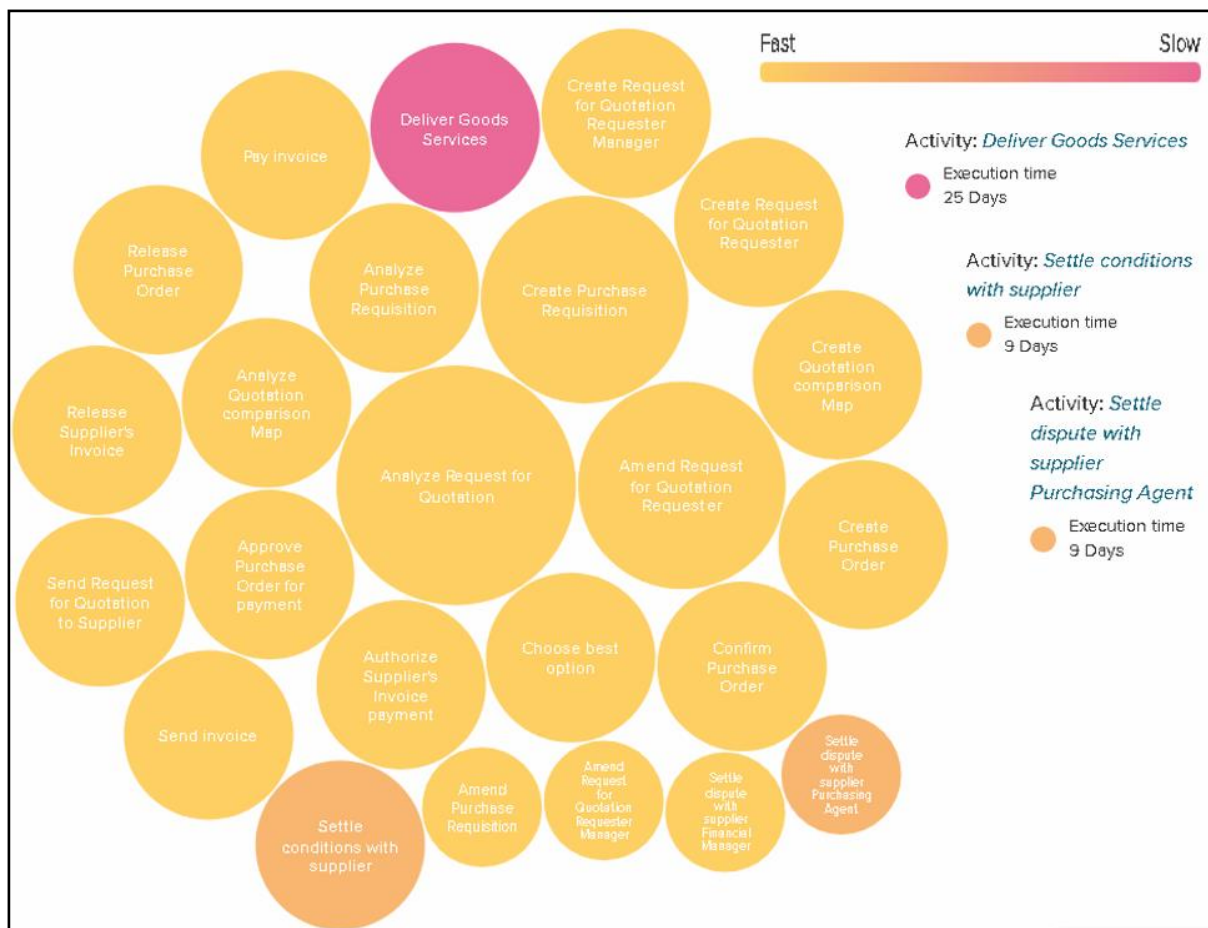
Statistika průchodu případů reálného procesu dává na výběr pouze jeden graf, který vyjadřuje počet případů v závislosti na dobu trvání od začátku do konce procesu. Na obrázku 27 je vidět 310 případů, které trvaly v rozmezí 10-20 dnů, a které zároveň vyjadřují 51 % relativního množství případů z celku. Dále jsou v této části vypsány cesty mezi aktivitami, které značně zpomalují průběh procesu. Například cesta mezi aktivitami „Create Request for Quation“ a „Analyze Request for Quation“ zbrzdí proces o 6 dnů a celkově se toto zpomalení týká 39 % případů.



Obrázek 27: Počet případů ve dnech v Celonisu

*Zdroj: vlastní zpracování*

Statistika aktivit se týká především frekvence a doby trvání aktivit. Obrázek 28 zobrazuje celkovou dobu trvání jednotlivých aktivit. Například aktivita „Deliver Goods Services“ trvala celkově 25 dní a je tak nejpomalejší aktivitou ze všech. Za zmínku stojí také aktivity „Settle conditions with Supplier“ a „Settle dispute with supplier Purchasing agent“, které mají celkovou dobu trvání 9 dní.



Obrázek 28: Celková doba trvání aktivit v Celonisu

*Zdroj: vlastní zpracování*

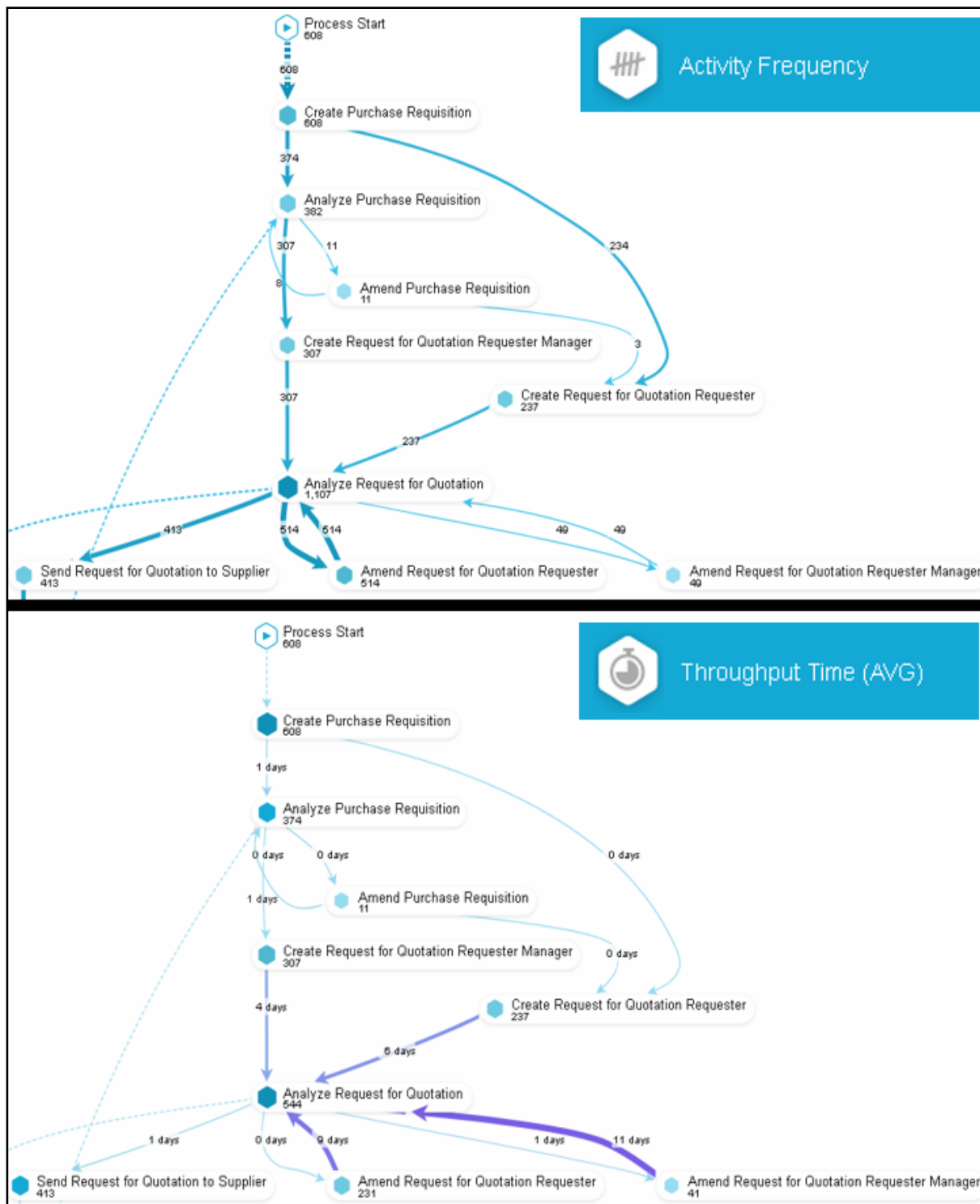
### 5.4.3 Prozkoumávání reálného modelu procesu v Celonisu

Většina komerčních nástrojů nepodporuje různé formáty výstupů jako ProM, ale výsledný reálný model procesu je založený na modifikovaném Fuzzy Mineru. Celonis není výjimkou a stejně jako v Disco, je výstupem reálného modelu procesu Fuzzy model.

Obrázek 30 zobrazuje reálný model procesu vytvořený v Celonisu. Vytvořený model je stejný jako v programu Disco nebo ProM, liší se pouze designem. Podobně jako v Disco, lze v Celonisu nastavit frekvenci aktivit pomocí vertikálního posuvníku, přičemž 100 % aktivit zobrazuje celkový model se všemi aktivitami a naopak. Zároveň je možné posouvat podobným posuvníkem pro zobrazení cest mezi aktivitami. Při posunutí jak jedním, tak

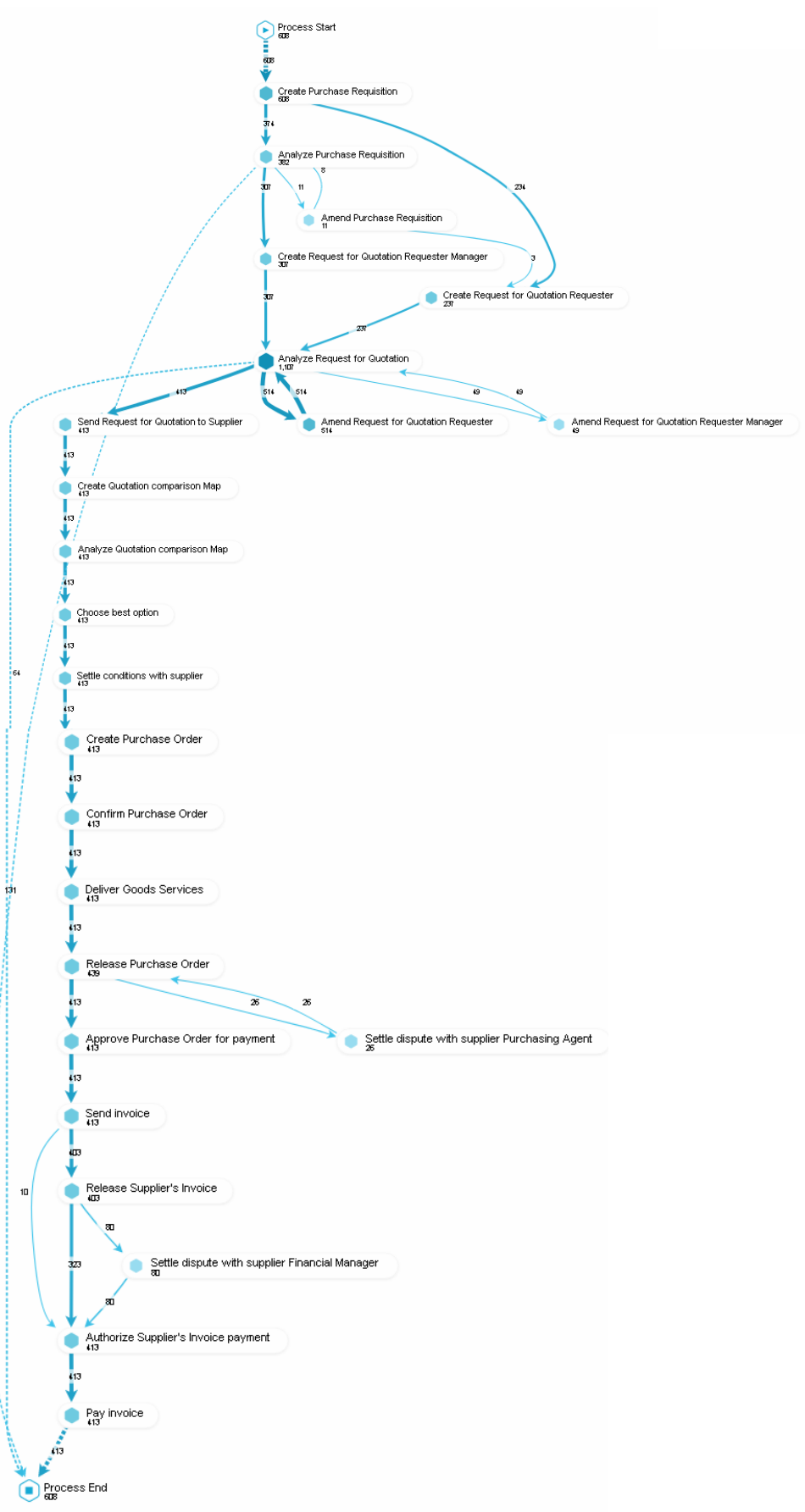
druhým posuvníkem, se po dobu přibližně dvou vteřin zvýrazní nové aktivity nebo cesty. Disco ani ProM tuto funkci nemají k dispozici. Dále je možné zobrazit v modelu frekvenci aktivit nebo případů a celkový průběh případů v procesu v časovém rámci.

Obrázek 29 zobrazuje frekvence aktivit a průměrnou dobu trvání případů.



Obrázek 29: Frekvence aktivit a průměrná doba trvání případů v Celonisu

*Zdroj: vlastní zpracování*

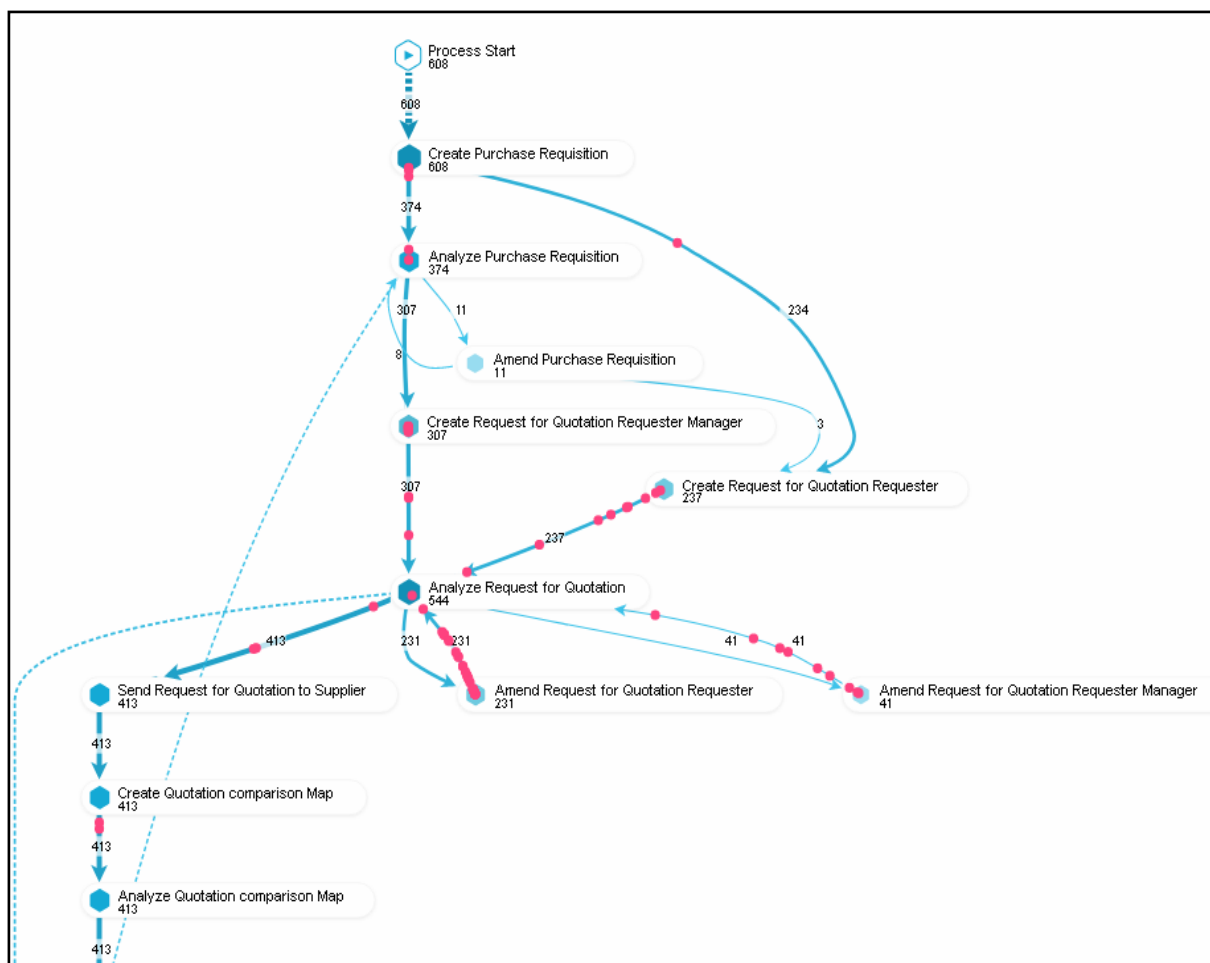


Obrázek 30: Model reálného procesu v Celonis

Zdroj: vlastní zpracování

#### 5.4.4 Animace reálného modelu procesu v Celonisu

Na obrázku 31 je zobrazena část animace v Celonisu, která funguje na stejném principu jako v programu Disco nebo ProM. Fialové tečky představují probíhající případy. Případy se nejvíce hromadí v aktivitě „Analyze Request for Quotation“ a to především díky tomu, že značný počet případů dále putuje do aktivit, které nepatří do původního scénáře a jedná se o slepé smyčky. Po vyřízení těchto slepých smyček putují případy zpět. Díky tomu vzniká přetížené místo a tato aktivita brzdí celý proces.



Obrázek 31: Animace reálného modelu procesu v Celonisu

*Zdroj: vlastní zpracování*

### 5.5 Shrnutí

Tato kapitola shrnuje pátou kapitolu, ve které se realizovala vlastní analýza procesu a byl vytvořen reálný model procesu, jehož výstupem byl Fuzzy model. Pro analýzu byly využity tři nástroje – ProM, Disco a Celonis, kde u každého bylo představeno prostředí softwaru a byly vysvětleny kroky, které vedly k vytvoření reálného modelu procesu. Tyto kroky představují import event logu a statistiku event logu. Po vytvoření reálného modelu procesu



bylo možné využít funkce animace procesu, která vizualizovala průběh jednotlivých případů v procesu.

V případě importu event logu je zřejmé, že se nástroje trochu liší. Například ProM nedokáže importovat soubory ve formátu CSV nebo XLS a je potřeba je před samotnou analýzou transformovat pomocí plug-inů nebo jiným nezávislým softwarem do podporovaných formátů, jimiž jsou XES nebo MXML. Disco i Celonis si dokážou poradit prakticky s většinou dostupných formátů, ale výhodou těchto nástrojů oproti ProM je především možnost importu event logů rovnou z databáze a aplikování analýzy online. Nejeftivnější je online analýza, protože organizace má přehled nad právě probíhajícími případy v procesu a může tak ovlivnit jejich průběh.

Na druhou stranu, u komerčních a nekomerčních nástrojů je dobře vidět rozdíl ve zpracování statistiky. ProM nabízí základní deskriptivní statistiku a několik málo grafů zachycujících frekvenci případů a událostí. Dále nabízí seznam případů, který je možný postupně „rozklikat“ a dostat se tak k požadovaným informacím. Nicméně tento způsob není intuitivní a není na první pohled jednoduché poznat, kde by mohl být potenciální problém. Komerční nástroje jsou více zaměřené na statistiku a poskytují přehledné grafy rozdělené podle atributů. Celonis navíc poskytuje „happy path“, což je původní scénář procesu bez výjimečných a chybových cest. ProM a Disco tuto funkci nemají.

Všechny nástroje vytvořily stejný reálný model procesu na základě Fuzzy mineru, přestože komerční nástroje disponují modifikovaným Fuzzy minerem. Na obrázcích všech reálných modelů si lze všimnout zvýrazněných cest, kde jejich tloušťka charakterizuje frekvenci využití. Například aktivita „Analyze Request for Quotation“ je velmi frekventovaná, a tudíž lze usoudit, že zde bude docházet k přetížení. Na základě frekventovanosti této aktivity je jasné, že průměrná doba vyřízení bude taktéž delší.

Dále stojí za povšimnutí uzavřená dominantní smyčka (aktivita „Amend Request for Quotation Requester“), která není standardním krokem v původním scénáři. Tento krok není důležitý, avšak je nejfrekventovanější aktivitou v celém procesu. Návrh na zlepšení procesu je určité v této aktivitě.

Z procesu lze také vyčíst, že v součtu je nejdéle trvající aktivitou „Deliver Goods Services“, což lze považovat za logické, neboť doručení objednávky bude trvat mnohem déle než vyřízení jiné aktivity několika kliknutími v informačním systému nebo telefonátem. Na druhou stranu aktivity jako „Settle Conditions with Supplier“ a „Settle Dispute with Supplier Purchasing Agent“, které trvaly v součtu 9 dní, by již mohly být předmětem k diskusi.

V případě animací všechny modely vykazovaly stejný trend, kdy přibližně od poloviny března do konce září se aktivita „Analyse Request for Quation“ přetížila a značně brzdila celý průběh procesu.

## ZÁVĚR

Diplomová práce byla rozdělena na několik kapitol, přičemž první kapitola pojednávala o podstatě Process Miningu jako o poměrně nové vědní disciplíně v oblasti Process Managementu, která propojuje Data Science (vědu o datech) a Process Science (vědu o procesech). Process Mining využívá event logy, jež lze vysvětlit jako protokoly událostí zaznamenané informačním systémem, které se pomocí specializovaných algoritmů analyzují a na základě vytvořeného reálného modelu procesu zjišťují trendy, odchylky a různé vzory.

Process Mining se především využívá ve chvíli, kdy je v prováděných činnostech zjištěna neefektivita či slabina, ale nejsme si jisti, kde přesně nežádoucí jev nastává a jaké jsou jeho příčiny. Vedení společnosti kvůli tomu často organizuje různé schůzky a porady manažerů, kde může například docházet k nedorozuměním na základě subjektivních názorů. Za neefektivní činnost při optimalizaci netriviálních procesů se dá považovat i vedení dialogů a investigace se zaměstnanci a manažery napříč celou firmou nebo testování pouhého vzorku položek v procesu.

Cílem Process Miningu je zlepšit efektivitu procesů a lépe porozumět, jak vlastně v praxi fungují na základě faktů, a ne fikce. Organizace tak mohou mít jistotu, že reálné procesy probíhají přesně tak, jak byly původně navrženy.

V první části diplomové práce byly také vysvětleny 3 hlavní kroky Process Miningu:

- prozkoumávání procesu,
- kontrola shody,
- zlepšení modelu.

První krok, prozkoumávání procesu, je technika, která jako vstup do analýzy využívá event logy vygenerované informačním systémem a pomocí algoritmů implementovaných v Process Miningových nástrojích se vygeneruje reálný model procesu bez použití jakékoliv předchozí informace.

Druhý krok, kontrola shody, je technika, která se používá pro zjišťování odchylek, predikci, rozhodování a k systémovému doporučení. Při kontrole shody je reálný model vytvořený z event logů porovnán s původně navrženým modelem procesu.

Třetí krok, zlepšení modelu, je technika, při které se zdokonaluje nebo rozšiřuje původní model pomocí informací získaných z reálného modelu procesu vytvořeného z event logů.

Druhá část diplomové práce se zabývala event logy, a jejich popisem a strukturou. Jak již bylo řečeno, event logy jsou předpokladem pro využití Process Miningu. Event logy se dají nazvat jako protokoly událostí a jsou to tedy údaje, které jsou generované informačním systémem nebo manuálně vkládané pracovníkem do informačního systému.

Každý event log musí obsahovat tři povinné atributy: ID případu, ID události a časovou stopu události. Bez těchto atributů by nešly dát údaje do procesních souvislostí a nebylo by tak možné vytvořit reálný model procesu. Každá událost je charakterizovaná vlastnostmi, které jsou již nepovinné. Vlastnosti mají také povahu atributů a mají svůj datový typ. Nejčastěji se využívají ke komplexnější analýze a lze pomocí nich vytvořit rozsáhlejší modely, které lépe zachycují celý průběh procesu. Příkladem nepovinných atributů jsou zdroje, role nebo náklady.

Přístup k event logům může být různorodý. Některé Process Miningové nástroje podporují propojení s databázovým systémem a je možné analyzovat tato data online, což je nejefektivnější způsob. Je také možné extrahovat data z databázového systému. Většinou se data nenachází v jedné databázi, proto je nutné extrahovat všechna data týkající se zkoumaného procesu, vytrždit je a vložit do jednoho souboru. Nejčastější formáty souborů jsou CSV nebo XLS. Avšak ne všechny Process Miningové nástroje si s tímto formátem dokáží poradit, proto pro tyto účely vznikl standardní formát XES (eXtensible Event Stream), který má řadu výhod.

Ve třetí části této práce byly představeny Process Miningové nástroje a jejich charakteristika. V závěru této kapitoly byly vyhodnoceny tři vybrané nástroje a shrnuty v tabulce dle kritérií. Ještě předtím, než začal být Process Mining populární a dostal se do povědomí firem, existovaly pouze nekomerční open source nástroje. Jedním z nich je ProM, který byl vyvinut na technické univerzitě v Eindhoven profesorem Wilem Aalstem a spol. Tento program patří mezi nejvyužívanější nástroje v Process Miningu nekomerční licence. Jedná se o nezávislý software implementovaný v Javě, a proto ho lze bezplatně získat na oficiálních webových stránkách ProM. Funkce v ProM jsou řešeny pomocí plug-inů, tj. zásuvných modulů, kterých je celkem kolem 1500. Plug-iny mohou být užitečné například pro dolování v event logech, import event logů, export vytvořeného modelu, konverzi modelu do jiného formátu nebo různé analýzy, například vytvoření analýzy sociální sítě na základě event logů.

Vzhledem k tomu, že se za poslední roky rapidně zvýšila poptávka po Process Miningu, firmy již nabízejí desítky komerčních nástrojů. Charakterizovat a porovnávat všechny nemá smysl, proto byly pro tuto práci vybrány dva nejznámější komerční nástroje, Disco a Celonis.

V další části třetí kapitoly se dle kritérií porovnávaly tři Process Miningové nástroje – ProM, Disco a Celonis. Při výběru nástrojů bylo přihlíženo k několika faktorům, aby bylo možné objektivně porovnat komerční a nekomerční nástroje. Disco i Celonis nabízejí akademické licence, díky kterým lze získat přístup k většině funkcí jako je například neomezený počet událostí v event logu, export modelu nebo animace reálného modelu procesu. Tyto funkce byly pro účely této diplomové práce poskytnuty společnostmi a díky akademické licenci bylo možné nástroje objektivně porovnat.

Pro komparaci nástrojů byla využita kritéria jako například podpora formátu event logu při importu, počet událostí v event logu při importu, formát výstupu reálného modelu, možnost filtrování event logu, možnost prozkoumávání procesu nebo kontrola shody, tvorba sociální sítě, animace procesu, podporované platformy, licence anebo shlukování případů.

Ve čtvrté kapitole byly zmíněny tři hlavní algoritmy, které se nejvíce využívají v praxi – Heuristic miner, Inductive miner a Fuzzy miner.

Heuristic miner je de facto první algoritmus, jež bylo možné použít pro reálné event logy, které by ale neměly obsahovat příliš událostí. Podstatou tohoto algoritmu je, že odvozuje logické členy „or“ a „and“ od vztahových závislostí a bere v potaz frekvenci a sekvenci událostí při tvorbě reálného modelu. Výstupem je model ve formátu Heuristické sítě a velkou výhodou je oproti jiným algoritmům je, že tento výstup lze transformovat na Petriho sítě pro hlubší analýzu.

Inductive miner je vylepšený Heuristic miner, který je založen na procesních stromech a netrpí tak problémy spojené s deadlockem, livelockem nebo jinými anomáliemi. Inductive miner je základním algoritmem ve skupině Inductive mining, který dokáže zpracovat rozsáhlé event logy a vytvořit zobecněné modely. Výstupem jsou Petriho síť nebo BPMN.

Fuzzy miner patří mezi nejnovější algoritmy pro zjišťování procesu, který využívá metriky významnosti nebo korelace, aby zjednodušil procesní model na požadovanou úroveň a zdůraznil frekventované aktivity a cesty. Fuzzy miner může také vynechat méně důležité aktivity (nebo je skrýt v klastrech), pokud jich je příliš. Výstupem je Fuzzy model a tento algoritmus najde největší uplatnění ve složitých a nestrukturovaných event logech.

Většina komerčních nástrojů v dnešní době využívá modifikovaný Fuzzy miner.

Poslední kapitola se zabývala realizací analýzy event logů ve vybraných nástrojích. Pro analýzu byl vybraný Fuzzy miner kvůli objektivnosti, protože výstupem u Disco a Celonis je pouze modifikovaný Fuzzy model. Součástí analýzy bylo i seznámení čtenáře s prostředím

nástrojů a výsledky statistik. V závěru kapitoly byly vždy vytvořeny animace reálných modelů procesů v jednotlivých nástrojích.

Přínos práce spočívá především ve shrnutí podstaty Process Miningu a jeho využití. Dále poukázání na existenci nekomerčních a komerčních nástrojů, jejich popis a porovnání na základě vybraných kritérií. Za přínosné lze rovněž považovat uvedení nejvyužívanějších metod v Process Miningu a aplikaci Fuzzy mineru v realizované analýze ve vybraných nástrojích. Tato práce může sloužit jako podklad pro detailnější analýzu event logů a tvorbu reálného modelu procesu.

Cílem této diplomové práce byl popis a charakterizace metod a nástrojů využívaných v Process Miningu. Dále byla vypracována analýza reálných event logů a vytvořen reálný model procesu ve vybraných nástrojích. Konkrétně se jednalo o nekomerční nástroj určení především pro akademické účely – ProM a komerční nástroje Disco a Celonis.

Z výše uvedeného lze usoudit, že cíl byl splněn.

## POUŽITÁ LITERATURA

- [1] AALST, Wil van der. Four quality criteria for Process Discovery. Coursera [online video]. Eindhoven University of Technology [cit. 2018-04-29]. Dostupné z: <https://www.coursera.org/learn/process-mining/lecture/D384Y/3-1-four-quality-criteria-for-process-discovery>
- [2] AALST, Wil van der. Business Process Management: A Comprehensive Survey [online]. Eindhoven: University of Technology [cit. 2018-04-29]. Dostupné z: <http://wwwis.win.tue.nl/~wvdaalst/publications/p712.pdf>
- [3] AALST, Wil van der. Process Discovery and Conformance Checking Using Passages [online]. Eindhoven: University of Technology [cit. 2018-04-29]. Dostupné z: <http://www.pads.rwth-aachen.de/wvdaalst/publications/p802.pdf>
- [4] AALST, Wil van der. Process Mining in the Large: A Tutorial [online]. Eindhoven: University of Technology [cit. 2018-04-29]. Dostupné z: <http://wwwis.win.tue.nl/~wvdaalst/publications/p712.pdf>
- [5] AALST, Wil van der. Process Mining: Data Science in action. 2nd Edition Berlin: Springer, 2016. ISBN 978-3-662-49850-7.
- [6] AALST, Wil van der. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Berlin: Springer, 2011. ISBN 978-3-642-19344-6.
- [7] AALST, Wil van der. ProM framework [online]. [cit. 2018-04-29]. Dostupné z: <http://www.pads.rwth-aachen.de/wvdaalst/publications/p264.pdf>
- [8] AALST, Wil van der. Workflow Mining: Discovering process models from event logs [online]. Eindhoven: University of Technology [cit. 2018-04-29]. Dostupné z: <http://www.pads.rwth-aachen.de/wvdaalst/publications/p245.pdf>
- [9] AALST, Will van der. Event logs and process models. Coursera [online]. Eindhoven: University of Technology, 2017 [cit. 2018-04-29]. Dostupné z: <https://www.coursera.org/learn/process-mining/lecture/uA8rJ/2-1-event-logs-and-process-models>
- [10] About ProM. Process mining [online]. [cit. 2018-04-29]. Dostupné z: <http://www.processmining.org/prom/start>
- [11] Algorithm: Definition [online]. 2013 [cit. 2018-04-29]. Dostupné z: <https://techterms.com/definition/algorithm>

- [12] Automated Business Process Discovery (ABPD). Gartner [online]. [cit. 2018-04-29]. Dostupné z: <https://www.gartner.com/it-glossary/automated-business-process-discovery-abpd>
- [13] CASTROUNIS, Alex. What is Data Science and what does a Data Scientist do. Kdnuggets[online]. InnoArchiTech, 2017 [cit. 2018-04-29]. Dostupné z: <https://www.kdnuggets.com/2017/03/data-science-data-scientist-do.html>
- [14] CoBeFra: Comprehensive Benchmarking Framework for Conformance Checking. Process mining [online]. Belgium: Faculty of Business and Economics [cit. 2018-04-29]. Dostupné z: <http://www.processmining.be/cobefra/>
- [15] Data Science and Data Analytics and Machine Learning. Simplilearn [online]. [cit. 2018-04-29]. Dostupné z: <https://www.simplilearn.com/data-science-vs-data-analytics-vs-machine-learning-article>
- [16] Disco. Fluxicon [online]. [cit. 2018-04-29]. Dostupné z: <https://fluxicon.com/disco/>
- [17] DUMAS, M. Fundamentals of Business Process Management. 2nd Edition. Berlin: Springer, 2018. ISBN 978-3-642-33143-5.
- [18] Event logs. Process mining [online]. Eindhoven: University of Technology, 2016 [cit. 2018-04-29]. Dostupné z: <http://www.processmining.org/logs/start>
- [19] Future of process mining [online]. 2016 [cit. 2018-04-29]. Dostupné z: <https://www.tue.nl/en/university/departments/mathematics-and-computer-science/news/10-11-2016-the-future-of-process-mining-trailblazing-cooperation-between-research-industry/>
- [20] Fuzzy miner [online video]. Eindhoven: University of technology [cit. 2018-04-29]. Dostupné z: <https://www.futurelearn.com/courses/process-mining/0/steps/15645>
- [21] HAU, Irene. What is the future of process mining. Quora [online]. 2018 [cit. 2018-04-29]. Dostupné z: <https://www.quora.com/What-is-the-future-of-process-mining>
- [22] Heuristic miner [online video]. Eindhoven: University of Technology [cit. 2018-04-29]. Dostupné z: <https://www.futurelearn.com/courses/process-mining/0/steps/15639>
- [23] Intro to XES. Fluxicon [online]. [cit. 2018-04-29]. Dostupné z: <https://fluxicon.com/blog/2010/09/intro-to-xes/>
- [24] KPI. Klip folio [online]. [cit. 2018-04-29]. Dostupné z: <https://www.klipfolio.com/resources/kpi-examples>



- [25] Livelock. Techopedia [online]. [cit. 2018-04-29]. Dostupné z: <https://www.techopedia.com/definition/3723/livelock>
- [26] Manifesto: Process mining [online]. [cit. 2018-04-29]. Dostupné z: <http://www.win.tue.nl/ieeetfpm/downloads/Process%20Mining%20Manifesto.pdf>
- [27] Matice odpovědnosti. Management mania [online]. [cit. 2018-04-29]. Dostupné z: <https://managementmania.com/cs/matice-odpovednosti-raci>
- [28] MONNAPPA, Avantika. Data Science vs Big Data vs Data Analytics [online]. 2018 [cit. 2018-04-29]. Dostupné z: <https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>
- [29] PALMER, Nathaniel. What is BPM [online]. 2014 [cit. 2018-04-29]. Dostupné z: <https://bpm.com/what-is-bpm>
- [30] PMLab. Github [online]. 2017 [cit. 2018-04-29]. Dostupné z: <https://github.com/josepcarmona/PMLAB>
- [31] Process Mining: A historical perspective [online]. [cit. 2018-04-29]. Dostupné z: [http://www.processmining.org/\\_media/presentations/2013/8-process-mining-camp2013-history-of-process-mining.pdf](http://www.processmining.org/_media/presentations/2013/8-process-mining-camp2013-history-of-process-mining.pdf)
- [32] Process Science [online]. [cit. 2018-04-29]. Dostupné z: [http://www.pads.rwth-aachen.de/wvdaalst/process\\_science/process\\_science.html](http://www.pads.rwth-aachen.de/wvdaalst/process_science/process_science.html)
- [33] ProM. ProM tools [online]. Eindhoven: University of Technology [cit. 2018-04-29]. Dostupné z: <http://www.promtools.org/doku.php?id=start>
- [34] ProMImport. Process mining [online]. Eindhoven: University of Technology [cit. 2018-04-29]. Dostupné z: <http://www.processmining.org/promimport/start>
- [35] Rapid miner: About us [online]. [cit. 2018-04-29]. Dostupné z: <https://rapidminer.com/us/>
- [36] Research ProM. Process mining [online]. Eindhoven: University of Technology, 2016 [cit. 2018-04-28]. Dostupné z: <http://www.processmining.org/research/start>
- [37] ROUSE, Margaret. Algorithm: Definition. Whatis.techtarget [online]. 2016 [cit. 2018-04-29]. Dostupné z: <https://whatis.techtarget.com/definition/algorithm>
- [38] ROUSE, Margaret. Deadlock: Definition. Whatis.techtarget [online]. 2005 [cit. 2018-04-29]. Dostupné z: <https://whatis.techtarget.com/definition/deadlock>

- [39] ROZINAT, Anne. ProM tips: Which mining algorithm should you use. Fluxicon [online]. 2017 [cit. 2018-04-29]. Dostupné z: <https://fluxicon.com/blog/2010/10/prom-tips-mining-algorithm/>
- [40] ROZINAT, Anne. What is the best software for process mining. Quora [online]. 2017 [cit. 2018-04-29]. Dostupné z: <https://www.quora.com/What-is-the-best-software-for-process-mining>
- [41] ROZINAT, Anne. Disco User's Guide. Fluxicon [online]. [cit. 2018-04-29]. Dostupné z: <https://fluxicon.com/disco/files/Disco-User-Guide.pdf>
- [42] ROZINAT, Anne. ProM tips: Which log filters should you use [online]. 2014 [cit. 2018-04-29]. Dostupné z: <http://fluxicon.com/blog/2010/12/prom-tips-log-filters/>
- [43] Social network analysis in ProM. Futurelearn [online video]. [cit. 2018-04-29]. Dostupné z: <https://www.futurelearn.com/courses/process-mining/0/steps/15652>
- [44] Tag based language. Your dictionary [online]. [cit. 2018-04-29]. Dostupné z: <http://www.yourdictionary.com/tag-based-language>
- [45] Xes. Xes-standard [online]. 2017 [cit. 2018-04-28]. Dostupné z: <http://www.xes-standard.org>
- [46] XES: eXtensible Event Stream [online]. Eindhoven University of Technology: Process Mining Group, 2016 [cit. 2018-04-29]. Dostupné z: <http://www.processmining.org/logs/xes>
- [47] Xesame. Process mining [online]. Eindhoven: University of Technology [cit. 2018-04-29]. Dostupné z: <http://www.processmining.org/xesame/start>