

CS 4072 - Topics in CS

Process Mining

Lecture # 02

February 15, 2022

Spring 2022

FAST - NUCES, CFD Campus

Dr. Rabia Maqsood

rabia.maqsood@nu.edu.pk

Today's Topics

- ▶ Introduction to Process Mining (continued)
 - ▶ Quick recap
 - ▶ Event log data
 - ▶ Process mining tasks

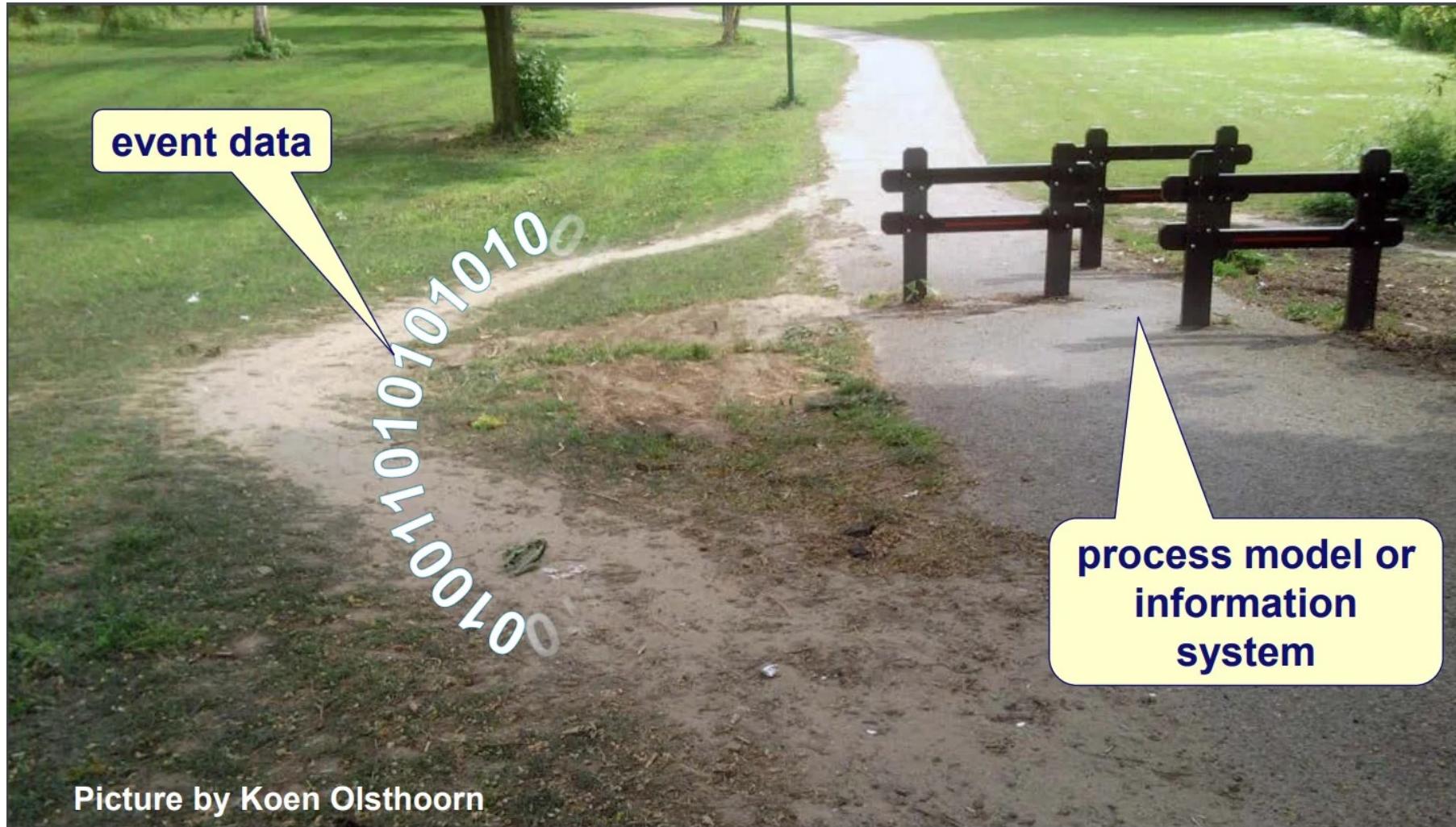
Join the Google Classroom

Class name: PM-S2022
Code: h6mmgbj

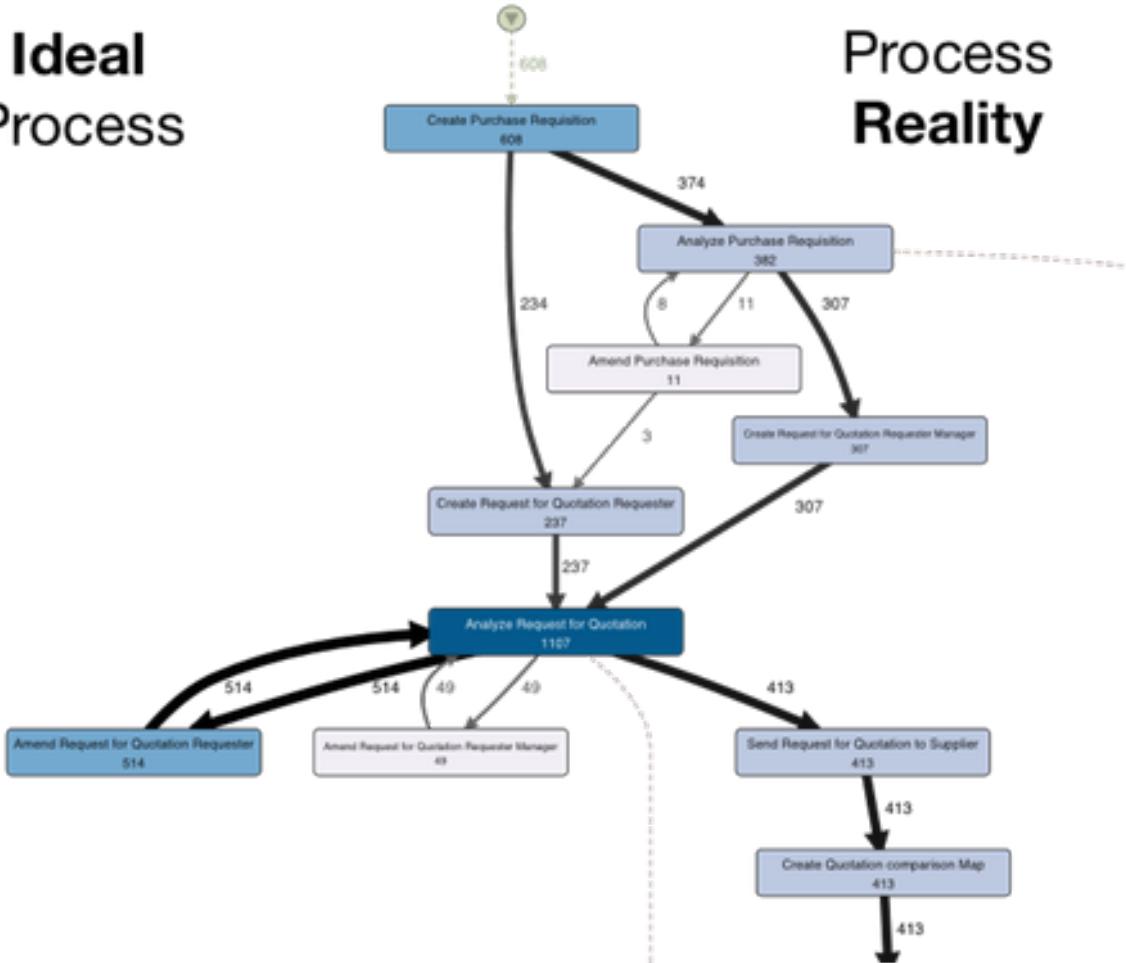
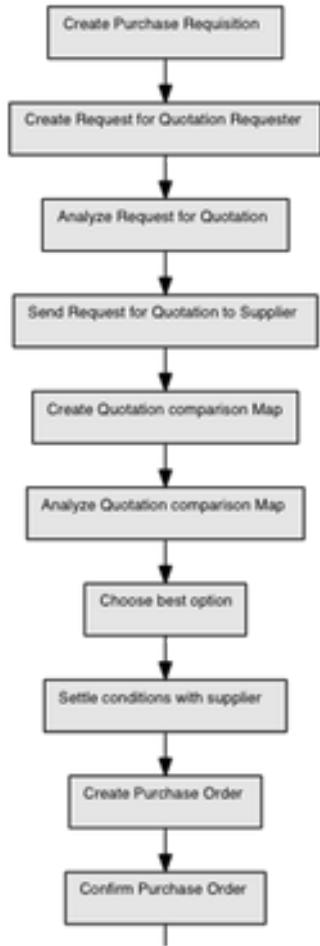
What is Process Mining?

- ▶ Process mining is an emerging discipline providing comprehensive set of tools to provide fact-based insights and to support process improvements.
 - ▶ Existing data mining techniques are too data-centric to provide a comprehensive understanding of the end-to-end processes in an organization.
 - ▶ Business intelligence tools focus on simple dashboards and reporting rather than clear-cut business process insights.
- ▶ Process mining **bridges the gap** between traditional model-based process analysis (e.g., simulation and other business process management techniques) and data-centric analysis techniques such as machine learning and data mining.

Why Process Mining?



Why Process Mining?



People tend to deviate from an “ideal” or expected process (for many reasons, e.g., to handle exceptions).

Nobody knows about the execution of a process in reality.

Thus, process mining takes data as an evidence to give us a “real picture” at the least.

But its not limited to just process discovery.
We will discuss later.

Event Data

Case ID	Activity	Resource	Timestamp	product	prod-price	quantity	address
...
6350	place order	Aiden	2018/02/13 14:29:45.000	APPLE iPhone 6 16 GB	639,00 €	5	NL-7751DG-21
6283	pay	Lily	2018/02/13 14:39:25.000	SAMSUNG Galaxy S6 32 GB	543.99	3	NL-7828AM-11a
6253	prepare delivery	Sophia	2018/02/13 15:01:33.000	APPLE iPhone 6 16 GB	639,00 €	3	NL-7887AC-13
6257	prepare delivery	Aiden	2018/02/13 15:03:43.000	SAMSUNG Galaxy S6 32 GB	543.99	1	NL-9521KJ-34
6185	confirm payment	Emily	2018/02/13 15:05:36.000	SAMSUNG Galaxy S4	329,00 €	1	NL-9521GC-32
6218	confirm payment	Emily	2018/02/13 15:08:11.000	APPLE iPhone 6s Plus 64 GB	969,00 €	2	NL-7948BX-10
6245	make delivery	Michael	2018/02/13 15:14:04.000	APPLE iPhone 6 16 GB	639,00 €	3	NL-7905AX-38
6272	pay	Emily	2018/02/13 15:20:36.000	APPLE iPhone 6 16 GB	639,00 €	1	NL-7821AC-3
6269	pay	Charlotte	2018/02/13 15:25:21.000	SAMSUNG Galaxy S4	329,00 €	1	NL-7907EJ-42
6212	prepare delivery	Sophia	2018/02/13 15:43:39.000	HUAWEI P8 Lite	234,00 €	1	NL-7905AX-38
6323	send invoice	Alexander	2018/02/13 15:46:08.000	APPLE iPhone 6 16 GB	639,00 €	1	NL-7833HT-15
6246	confirm payment	Jack	2018/02/13 15:56:03.000	SAMSUNG Galaxy S4	329,00 €	3	NL-7833HT-15
6347	send invoice	Jack	2018/02/13 15:57:42.000	SAMSUNG Galaxy S4	329,00 €	3	NL-7905AX-38
6351	place order	Zoe	2018/02/13 16:17:37.000	APPLE iPhone 5s 16 GB	449,00 €	3	NL-9521GC-32
6204	prepare delivery	Sophia	2018/02/13 16:31:28.000	SAMSUNG Core Prime G361	135,00 €	1	NL-7828AM-11a
6204	make delivery	Kaylee	2018/02/13 16:51:54.000	SAMSUNG Core Prime G361	135,00 €	1	NL-7828AM-11a
6265	confirm payment	Lily	2018/02/13 16:55:55.000	SAMSUNG Galaxy S4	329,00 €	4	NL-9521GC-32
6250	confirm payment	Jack	2018/02/13 17:03:26.000	MOTOROLA Moto G	199,00 €	4	NL-7942GT-2
6328	send invoice	Lily	2018/02/13 17:30:16.000	APPLE iPhone 6s 64 GB	858,00 €	4	NL-9514BV-16
6352	place order	Aiden	2018/02/13 17:53:22.000	APPLE iPhone 6 16 GB	639,00 €	2	NL-9514BV-16
6317	send invoice	Jack	2018/02/13 18:45:30.000	APPLE iPhone 6s 64 GB	858,00 €	5	NL-7907EJ-42
6353	place order	Sophia	2018/02/13 20:16:20.000	APPLE iPhone 5s 16 GB	449,00 €	4	NL-7751AR-19
...

event

71,043 events
12,666 cases
7 activities

Event Data

Case ID	Activity	Resource	Timestamp	product	prod-price	quantity	address
...
6350	place order	Aiden	2018/02/13 14:29:45.000	APPLE iPhone 6 16 GB	639,00 €	5	NL-7751DG-21
6283	pay	Lily	2018/02/13 14:39:25.000	SAMSUNG Galaxy S6 32 GB	543,99	3	NL-7828AM-11a
6253	prepare delivery	Sophia	2018/02/13 15:01:33.000	APPLE iPhone 6 16 GB	639,00 €	3	NL-7887AC-13
6257	prepare delivery	Aiden	2018/02/13 15:03:43.000	SAMSUNG Galaxy S6 32 GB	543,99	1	NL-9521KJ-34
6185	confirm payment	Emily	2018/02/13 15:05:36.000	SAMSUNG Galaxy S4	329,00 €	1	NL-9521GC-32
6218	confirm payment	Emily	2018/02/13 15:08:11.000	APPLE iPhone 6s Plus 64 GB	969,00 €	2	NL-7948BX-10
6245	make delivery	Michael	2018/02/13 15:14:04.000	APPLE iPhone 6 16 GB	639,00 €	3	NL-7905AX-38
6272	pay	Emily	2018/02/13 15:20:36.000	APPLE iPhone 6S 32 GB	639,00 €	1	NL-7821AC-3
6269	pay	Charlotte	2018/02/13 15:25:21.000	SAMSUNG Galaxy S4	329,00 €	1	NL-7907EJ-42
6212	prepare delivery	Sophia	2018/02/13 15:43:39.000	HUAWEI P8 Lite	234,00 €	1	NL-7905AX-38
6323	send invoice	Alexander	2018/02/13 15:46:08.000	APPLE iPhone 5S 16 GB	329,00 €	1	NL-7833HT-15
6246	confirm payment	Jack	2018/02/13 15:56:03.000	SAMSUNG Galaxy S4	329,00 €	3	NL-7833HT-15
6347	send invoice	Jack	2018/02/13 15:57:42.000	SAMSUNG Galaxy S4	329,00 €	3	NL-7905AX-38
6351	place order	Zoe	2018/02/13 16:17:37.000	APPLE iPhone 5s 16 GB	449,00 €	3	NL-9521GC-32
6204	prepare delivery	Sophia	2018/02/13 16:31:28.000	SAMSUNG Galaxy S6 32 GB	543,99	1	NL-7828AM-11a
6204	make delivery	Kaylee	2018/02/13 16:51:54.000	SAMSUNG Galaxy Note 5 64 GB	858,00	1	NL-7828AM-11a
6265	confirm payment	Lily	2018/02/13 16:55:55.000	SAMSUNG Galaxy S4	329,00 €	4	NL-9521GC-32
6250	confirm payment	Jack	2018/02/13 17:03:26.000	MOTOROLA Moto G	199,00 €	4	NL-7942GT-2
6328	send invoice	Lily	2018/02/13 17:30:16.000	APPLE iPhone 6s 64 GB	858,00 €	4	NL-9514BV-16
6352	place order	Aiden	2018/02/13 17:53:22.000	APPLE iPhone 6 16 GB	639,00 €	2	NL-9514BV-16
6317	send invoice	Jack	2018/02/13 18:45:30.000	APPLE iPhone 6s 64 GB	858,00 €	5	NL-7907EJ-42
6353	place order	Sophia	2018/02/13 20:16:20.000	APPLE iPhone 5s 16 GB	449,00 €	4	NL-7751AR-19
...

event =
case +
activity +
timestamp +

Let's look at orders 6350, 6351 and 6352

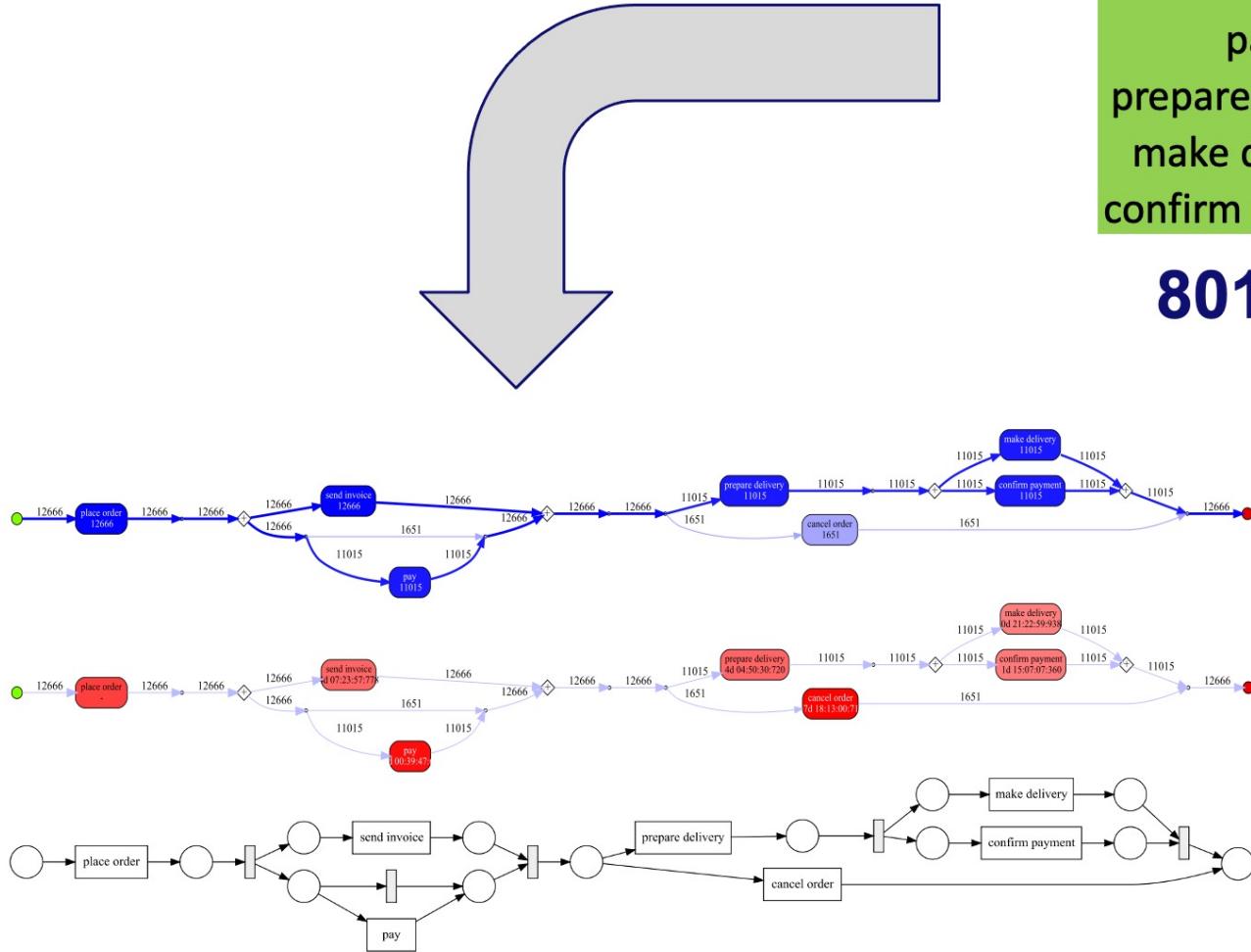
Case ID	Activity	Timestamp
6350	place order	2018/02/13 14:29:45.000
6351	place order	2018/02/13 16:17:37.000
6352	place order	2018/02/13 17:53:22.000
6352	send invoice	2018/02/19 09:20:28.000
6351	send invoice	2018/02/19 16:08:07.000
6350	send invoice	2018/02/21 09:38:16.000
6350	pay	2018/03/02 12:39:37.000
6352	pay	2018/03/05 15:46:47.000
6351	cancel order	2018/03/06 10:17:01.000
6350	prepare delivery	2018/03/07 13:50:35.000
6350	make delivery	2018/03/07 16:41:01.000
6350	confirm payment	2018/03/07 16:53:00.000
6352	prepare delivery	2018/03/07 17:05:59.000
6352	confirm payment	2018/03/07 17:59:55.000
6352	make delivery	2018/03/08 09:54:36.000

place order
send invoice
pay
prepare delivery
make delivery
confirm payment

place order
send invoice
cancel order

place order
send invoice
pay
prepare delivery
confirm payment
make delivery

Using the whole event log



place order
send invoice
pay
prepare delivery
make delivery
confirm payment

8016 x

place order
send invoice
cancel order

1651 x

place order
send invoice
pay
prepare delivery
confirm payment
make delivery

2962 x

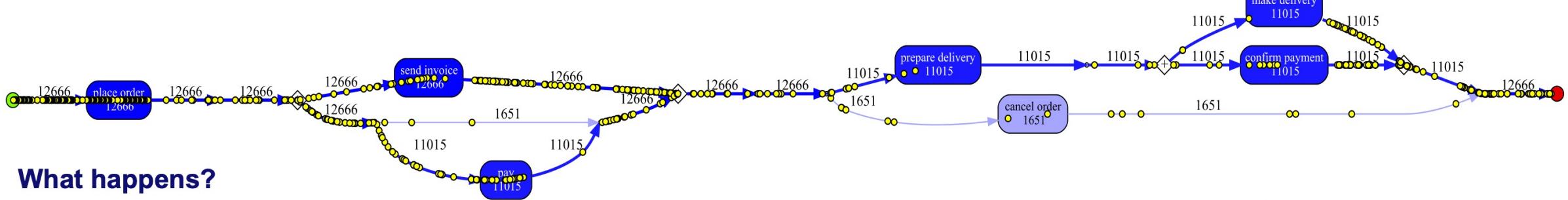
place order
pay
send invoice
prepare delivery
make delivery
confirm payment

30 x

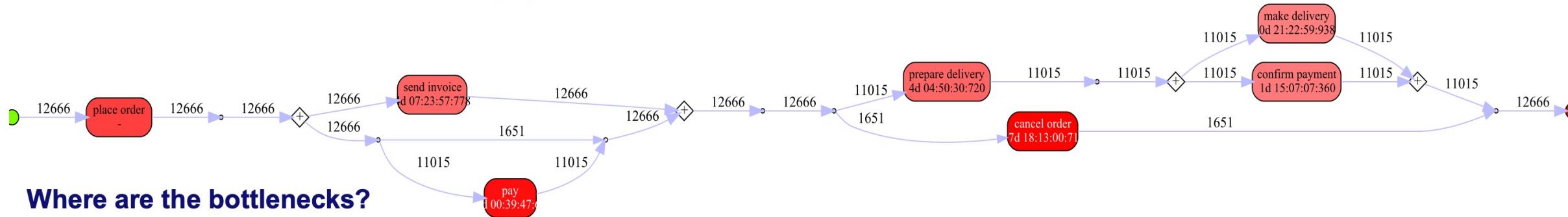
place order
pay
send invoice
prepare delivery
confirm payment
make delivery

7 x

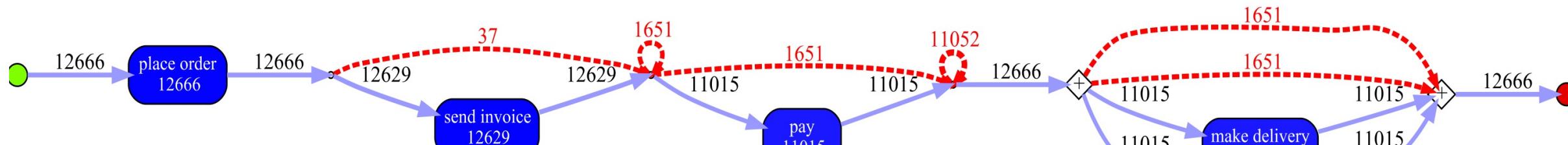
Performance and Compliance



What happens?

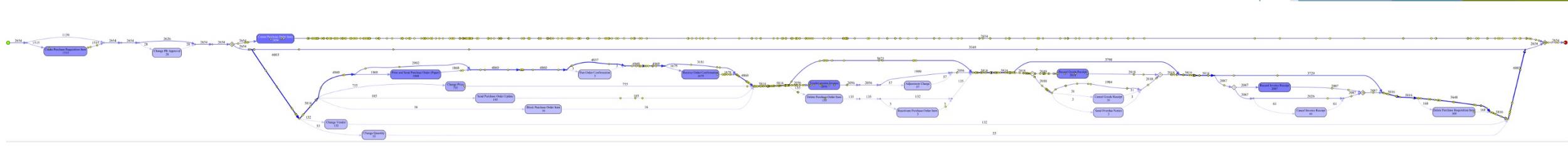


Where are the bottlenecks?

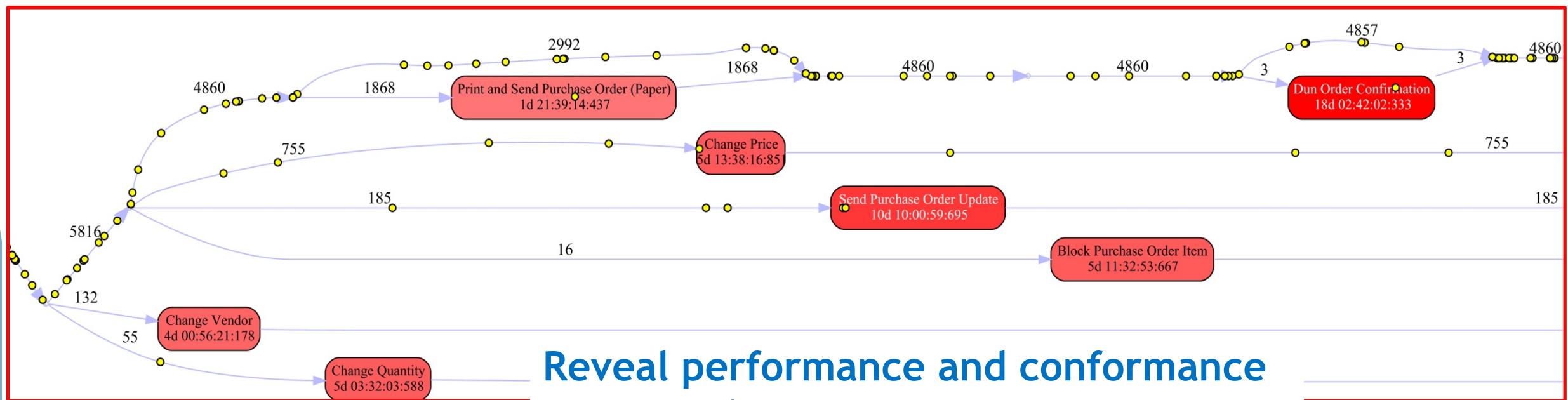
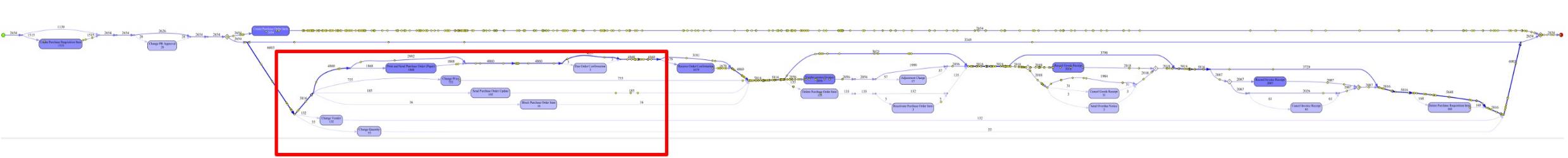


Where do we deviate from the happy path?

Process mining helps organizations to address these problems



Process mining helps organizations to address these problems



Reveal performance and conformance issues and suggest actions.

Event log

- ▶ We assume the existence of an **event log** where each event refers to a **case**, an **activity**, and a point in **time**.
- ▶ An **event log** can be seen as a **collection of cases**.
- ▶ A **case** can be seen as a **trace/sequence of events**.

An example log

student name	course name	exam date	mark
Peter Jones	Business Information systems	16-1-2014	8
Sandy Scott	Business Information systems	16-1-2014	5
Bridget White	Business Information systems	16-1-2014	9
John Anderson	Business Information systems	16-1-2014	8
Sandy Scott	BPM Systems	17-1-2014	7
Bridget White	BPM Systems	17-1-2014	8
Sandy Scott	Process Mining	20-1-2014	5
Bridget White	Process Mining	20-1-2014	9
John Anderson	Process Mining	20-1-2014	8

case id

activity name

timestamp

other data

Another event log: order handling

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

Another event log: patient treatment

patient	activity	timestamp	doctor	age	cost
5781	make X-ray	23-1-2014@10.30	Dr. Jones	45	70.00
5541	blood test	23-1-2014@10.18	Dr. Scott	61	40.00
5833	blood test	23-1-2014@10.27	Dr. Scott	24	40.00
5781	blood test	23-1-2014@10.49	Dr. Scott	45	40.00
5781	CT scan	23-1-2014@11.10	Dr. Fox	45	1200.00
5833	surgery	23-1-2014@12.34	Dr. Scott	24	2300.00
5781	handle payment	23-1-2014@12.41	Carol Hope	45	0.00
5541	radiation therapy	23-1-2014@13.57	Dr. Jones	61	140.00
5541	radiation therapy	23-1-2014@13.08	Dr. Jones	61	140.00
...

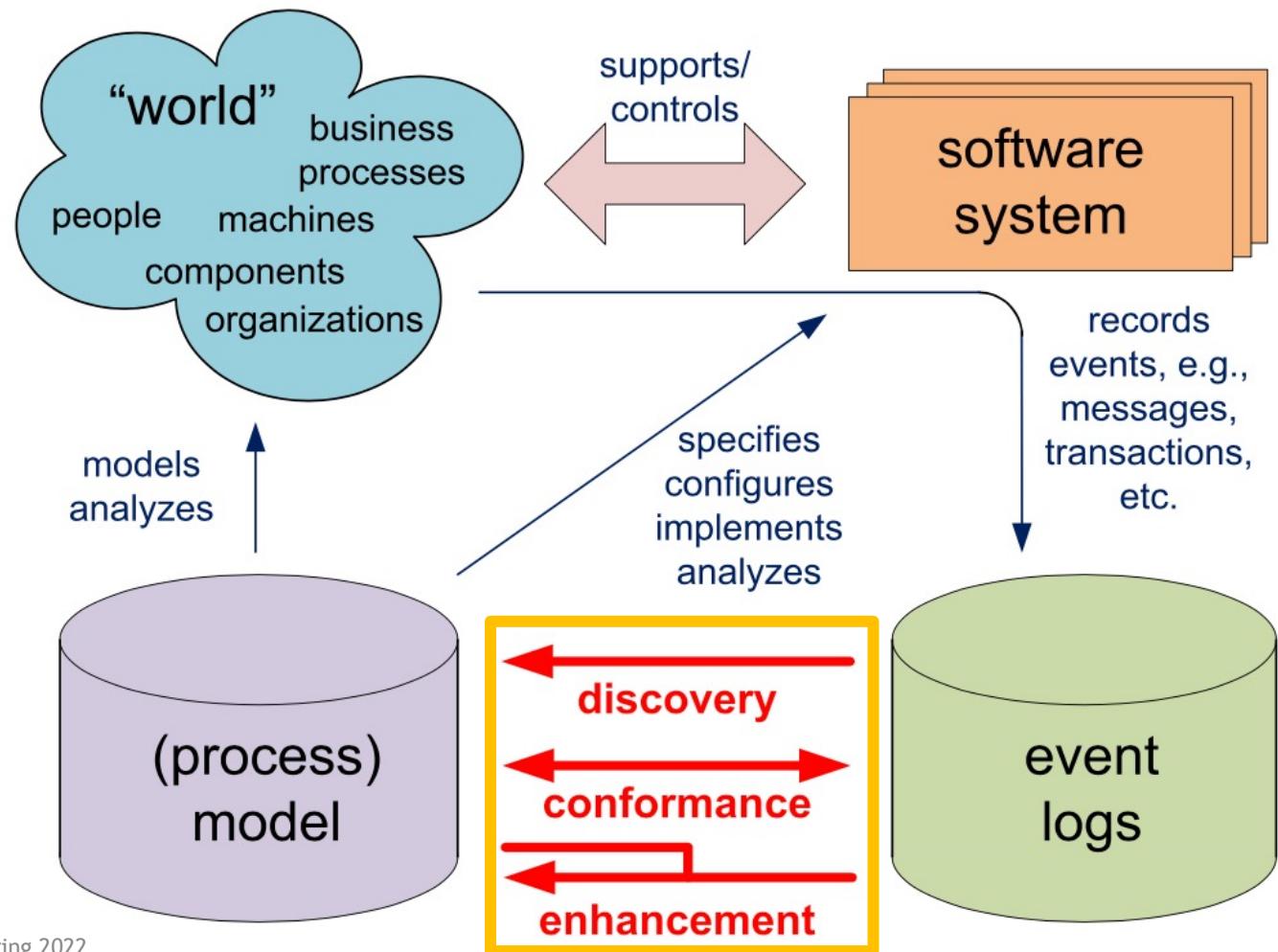
case id **activity name** **timestamp** **resource** **other data**

Event log example

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 casually	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 request	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	...	

Process Mining Tasks

Process Mining Tasks



Model Discovery

Given an event log as input, process mining techniques can discover a model.

Table 2.2 A more compact representation of log shown in Table 2.1: $a = \text{register request}$, $b = \text{examine thoroughly}$, $c = \text{examine casually}$, $d = \text{check ticket}$, $e = \text{decide}$, $f = \text{reinitiate request}$, $g = \text{pay compensation}$, and $h = \text{reject request}$

Case id	Trace
1	$\langle a, b, d, e, h \rangle$
2	$\langle a, d, c, e, g \rangle$
3	$\langle a, c, d, e, f, b, d, e, g \rangle$
4	$\langle a, d, b, e, h \rangle$
5	$\langle a, c, d, e, f, d, c, e, f, c, d, e, h \rangle$
6	$\langle a, c, d, e, g \rangle$
...	...

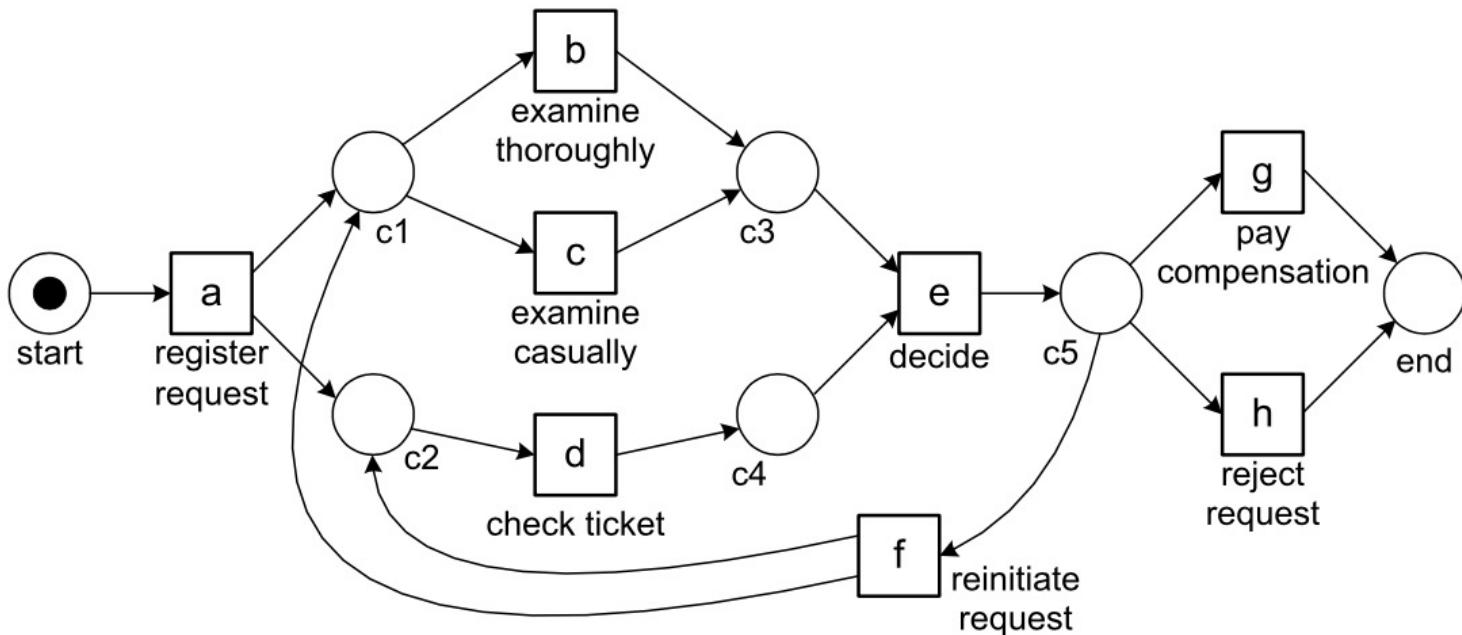


Fig. 2.6 The process model discovered by the α -algorithm [157] based on the set of traces $\{\langle a, b, d, e, h \rangle, \langle a, d, c, e, g \rangle, \langle a, c, d, e, f, b, d, e, g \rangle, \langle a, d, b, e, h \rangle, \langle a, c, d, e, f, d, c, e, f, c, d, e, h \rangle, \langle a, c, d, e, g \rangle\}$

Model Discovery

Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E

Case 1

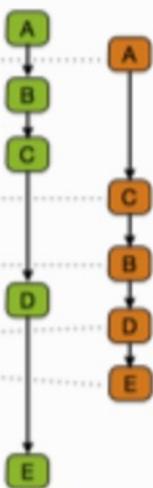


Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E

Case 1



Case 2



Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E

Case 1



Case 2



Case 3



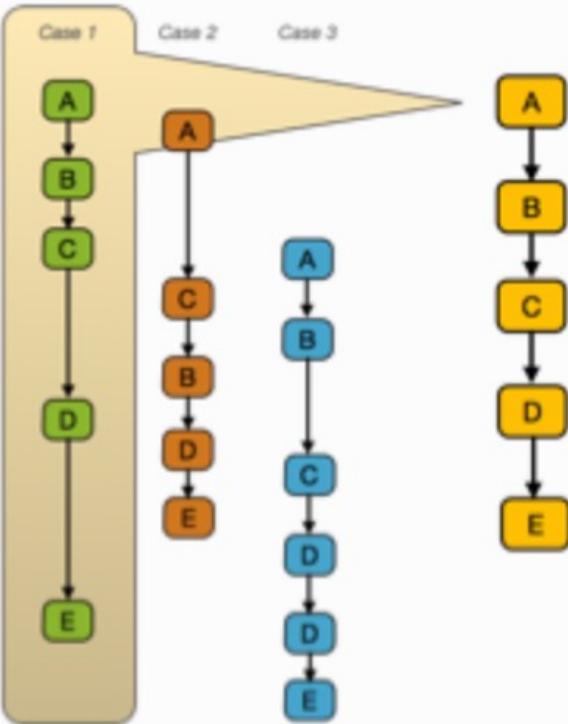
Step (1)

Step (2)

Step (3)

Model Discovery

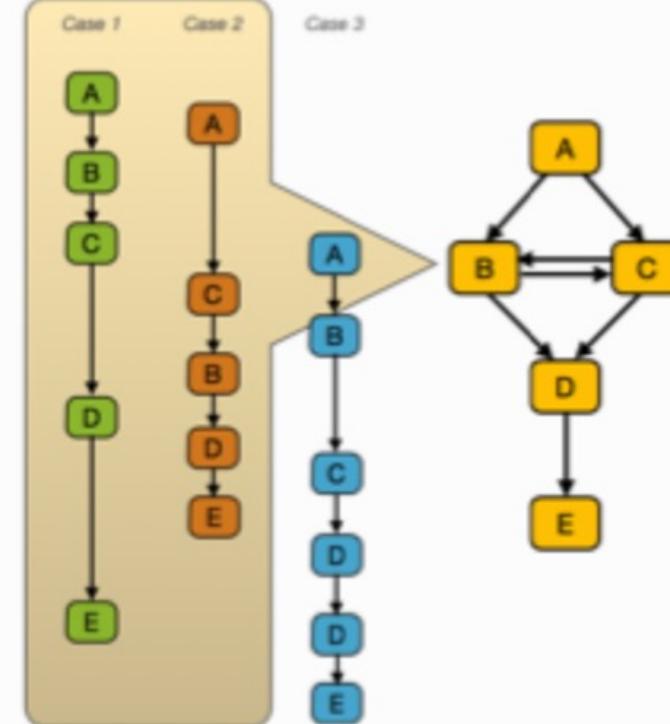
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E



Step (4)

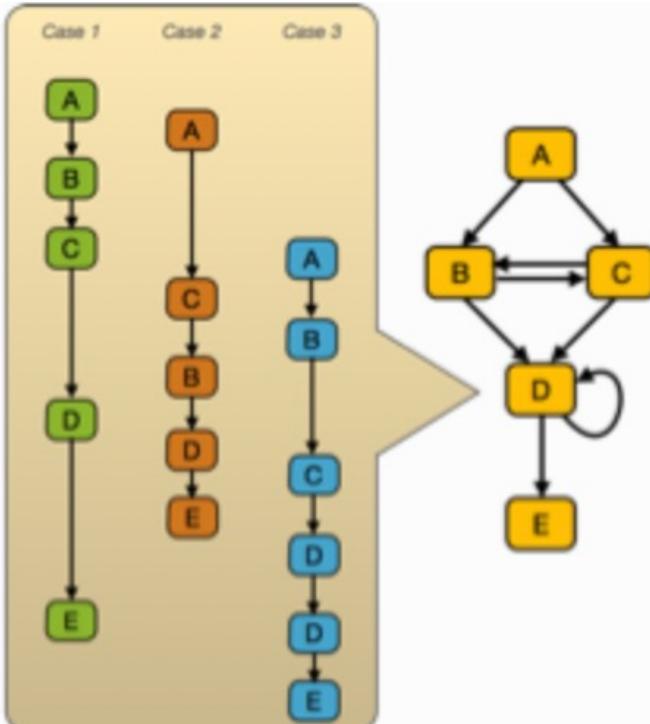
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E

Step (5)



Model Discovery

Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E



Step (6)

But reality is certainly not that simple.

We need sophisticated algorithms which can extract/discover models for event logs with large sizes and complex nature.

Model Discovery

What we should be careful about while model discovery?

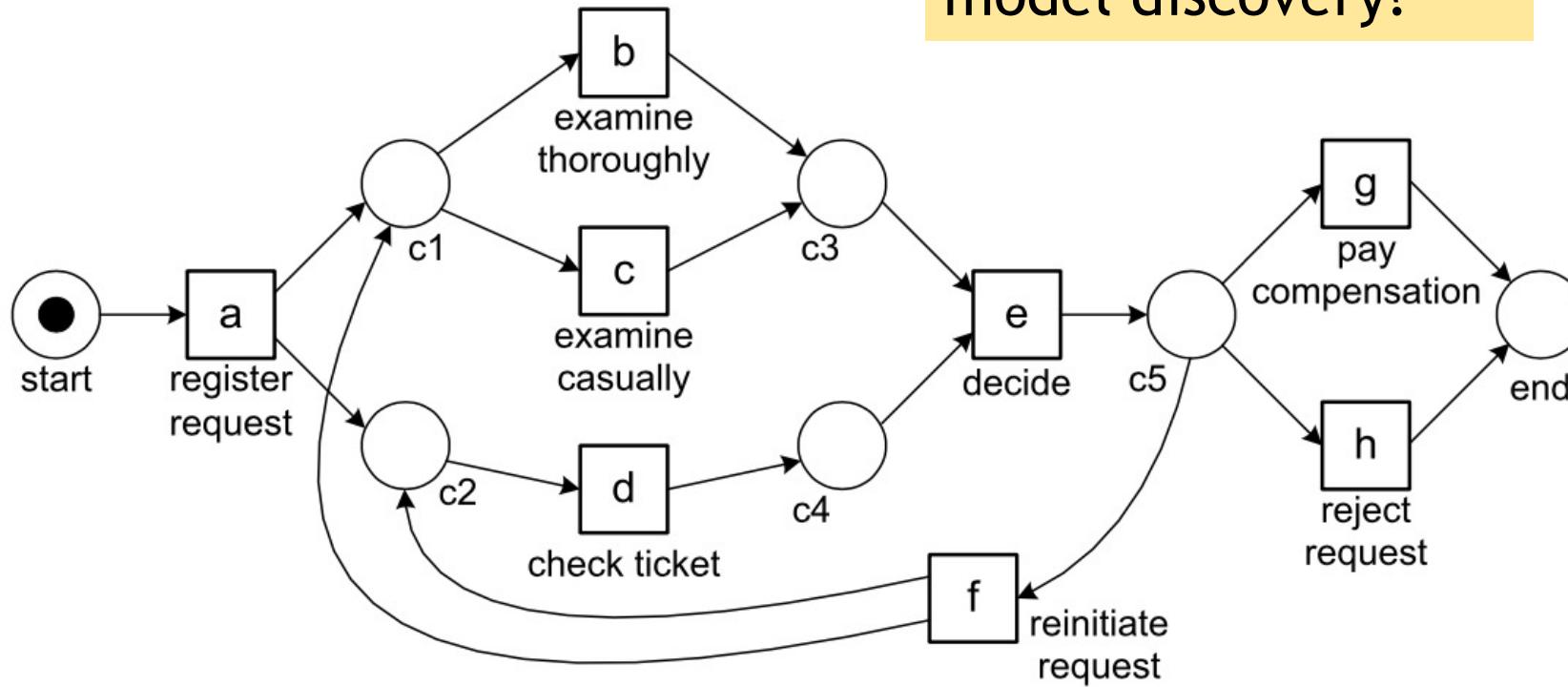


Fig. 2.6 The process model discovered by the α -algorithm [157] based on the set of traces $\{\langle a, b, d, e, h \rangle, \langle a, d, c, e, g \rangle, \langle a, c, d, e, f, b, d, e, g \rangle, \langle a, d, b, e, h \rangle, \langle a, c, d, e, f, d, c, e, f, c, d, e, h \rangle, \langle a, c, d, e, g \rangle\}$

Model Discovery: our concerns

- ▶ We need **generalized models** (i.e., allow probable behaviors that are not present in the input event log).
- ▶ Challenge is to balance between **overfitting** and **underfitting**.
 - ▶ Overfitting: model is too specific and only allows for the “accidental behavior” observed
 - ▶ Underfitting: model is too general and allows for behavior unrelated to the behavior observed

Traces $\langle a, d, c, e, f, b, d, e, g \rangle$ and $\langle a, c, d, e, f, c, d, e, f, c, d, e, f, c, d, e, f, b, d, e, g \rangle$ are also possible.

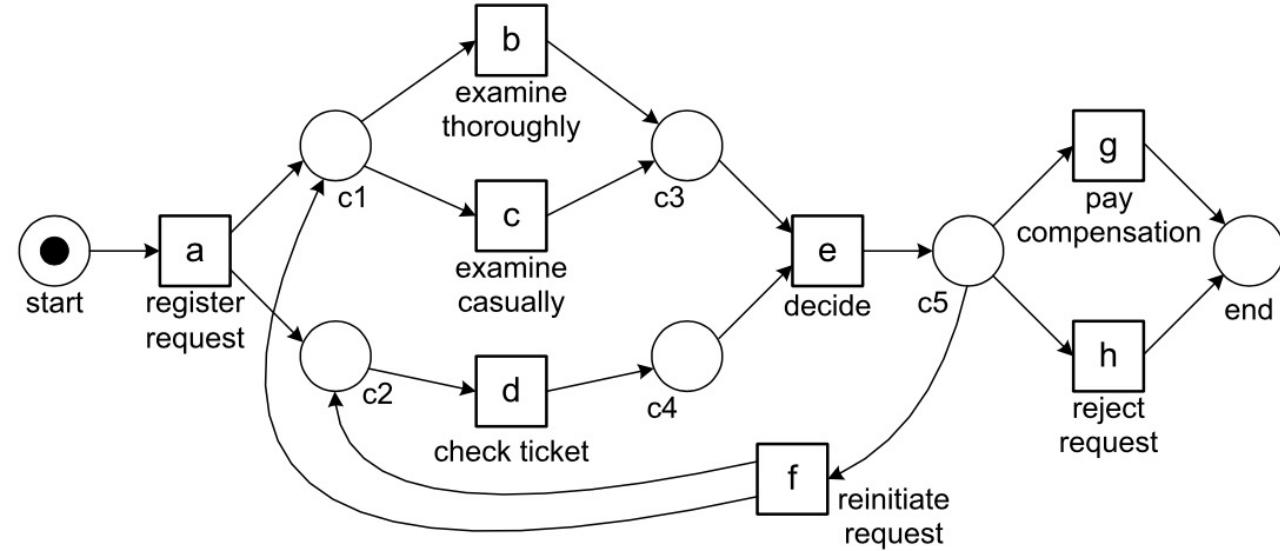


Fig. 2.6 The process model discovered by the α -algorithm [157] based on the set of traces $\{\langle a, b, d, e, h \rangle, \langle a, d, c, e, g \rangle, \langle a, c, d, e, f, b, d, e, g \rangle, \langle a, d, b, e, h \rangle, \langle a, c, d, e, f, d, c, e, f, c, d, e, h \rangle, \langle a, c, d, e, g \rangle\}$

Model Discovery

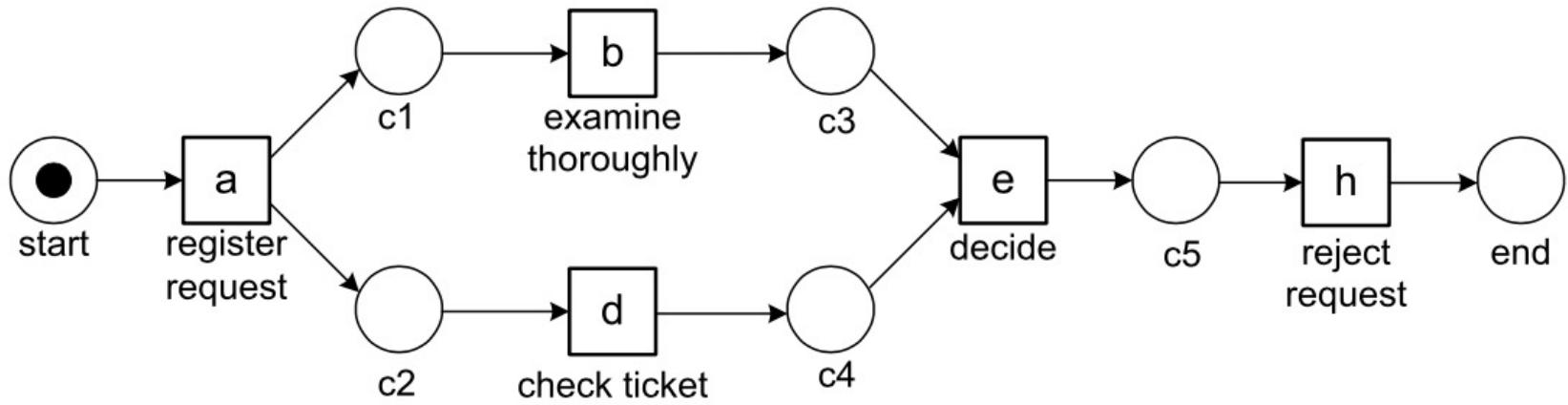


Fig. 2.7 The process model discovered by the α -algorithm based on cases 1 and 4, i.e., the set of traces $\{\langle a, b, d, e, h \rangle, \langle a, d, b, e, h \rangle\}$

Not a generalized model

Conformance Checking

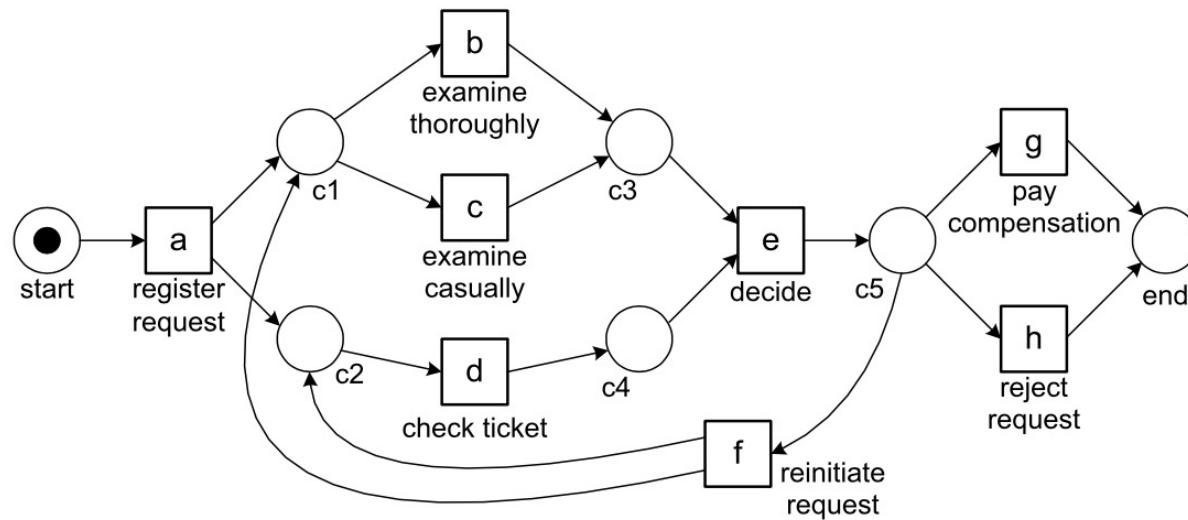
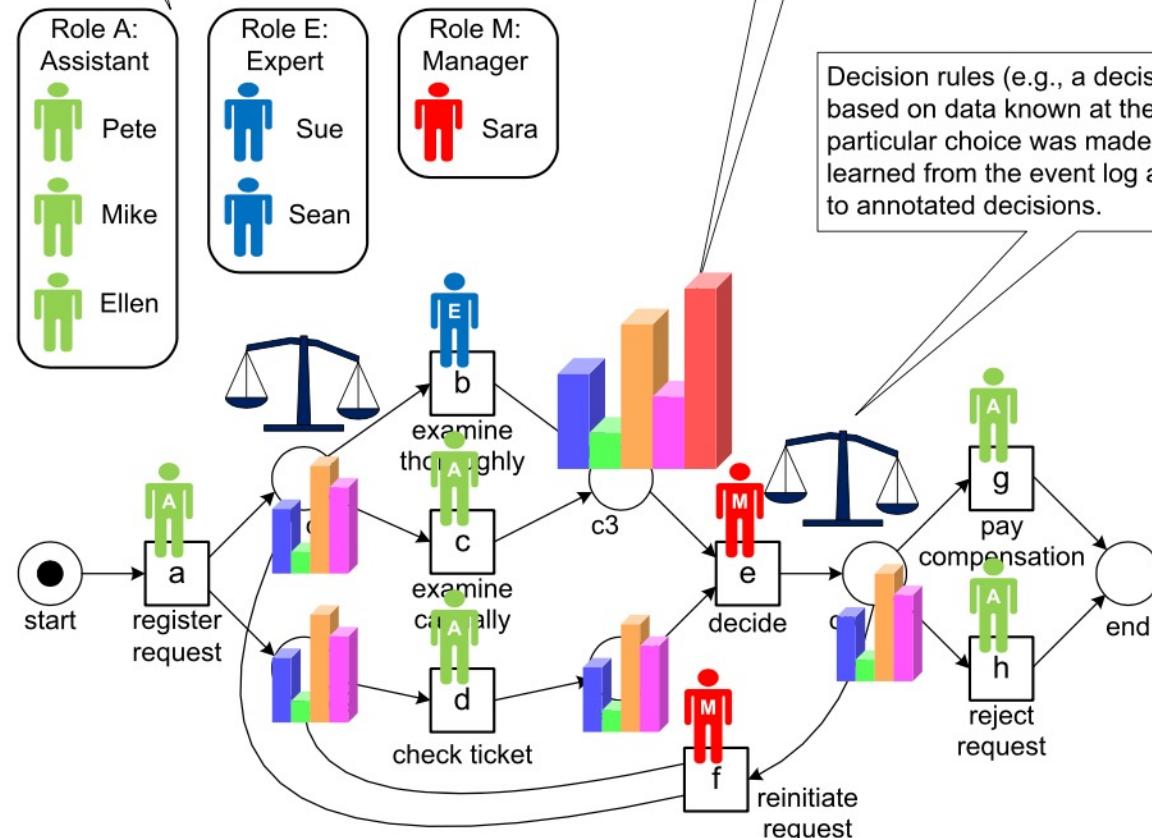


Fig. 2.6 The process model discovered by the α -algorithm [157] based on the set of traces $\{\langle a, b, d, e, h \rangle, \langle a, d, c, e, g \rangle, \langle a, c, d, e, f, b, d, e, g \rangle, \langle a, d, b, e, h \rangle, \langle a, c, d, e, f, d, c, e, f, c, d, e, h \rangle, \langle a, c, d, e, g \rangle\}$

Case id	Trace
1	$\langle a, b, d, e, h \rangle$
2	$\langle a, d, c, e, g \rangle$
3	$\langle a, c, d, e, f, b, d, e, g \rangle$
4	$\langle a, d, b, e, h \rangle$
5	$\langle a, c, d, e, f, d, c, e, f, c, d, e, h \rangle$
6	$\langle a, c, d, e, g \rangle$
7	$\langle \mathbf{a}, \mathbf{b}, \mathbf{e}, \mathbf{g} \rangle$
8	$\langle \mathbf{a}, \mathbf{b}, \mathbf{d}, \mathbf{e} \rangle$
9	$\langle a, d, c, e, f, d, c, e, f, b, d, e, h \rangle$
10	$\langle \mathbf{a}, \mathbf{c}, \mathbf{d}, \mathbf{e}, \mathbf{f}, \mathbf{b}, \mathbf{d}, \mathbf{g} \rangle$

Process Enhancement

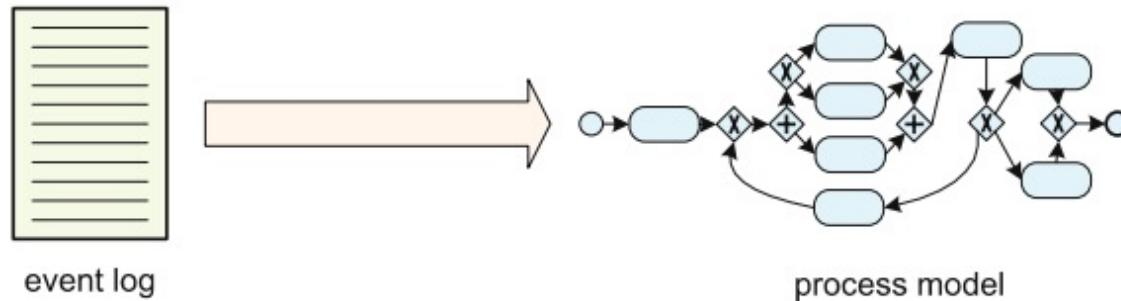


The event log can be used to discover roles in the organization (e.g., groups of people with similar work patterns). These roles can be used to relate individuals and activities.

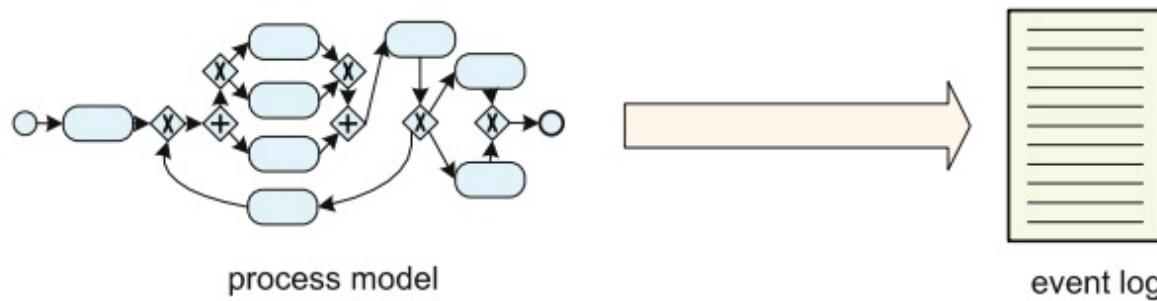
Performance information (e.g., the average time between two subsequent activities) can be extracted from the event log and visualized on top of the model.

Decision rules (e.g., a decision tree based on data known at the time a particular choice was made) can be learned from the event log and used to annotated decisions.

Play-In



Play-Out



Replay

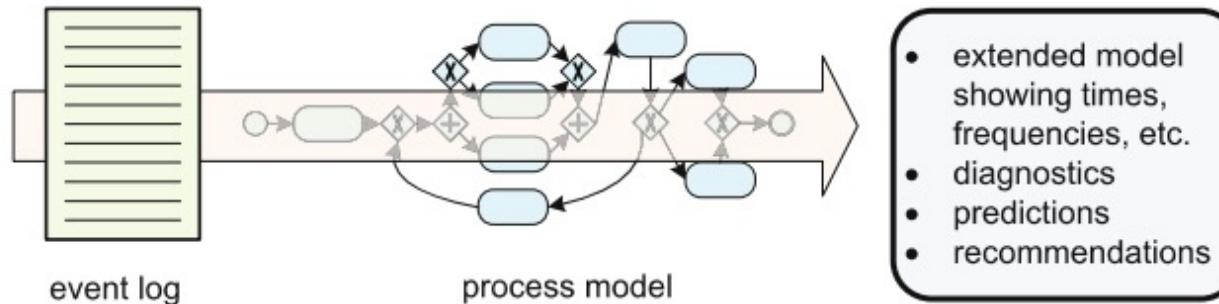


Fig. 1.8 Three ways of relating event logs (or other sources of information containing example behavior) and process models: *Play-in*, *Play-out*, and *Replay*

Reading Material

- ▶ Chapter 2: Aalst