



# Sok-sok adat "Az adat az új olaj!"

Az elmúlt 10 percben több adatot generáltunk mint 2003 előtt összesen.

Adatokat generálunk:

- jegyet vásárolunk egy repülőjára
- veszünk egy kávét
- feltöltjük üzemanyaggal az autónkat
- kezdeményezünk egy telefonhívást
- küldünk egy emailt
- megnézünk egy klippet a youtube-on
- stb.



# Honnan?

Az iPhone-on pl. nagyon sok féle szenzor található:

- camera (elől, hátul),
- mágnesmérő (magnetometer),
- érintés érzékelő,
- gyorsulásmérő (accelerometer),
- Bluetooth,
- GPS,
- ujjlenyomat vizsgáló,
- környezeti fényerősség érzékelő (ambient light sensor)



# Vizsgálható kérdések 1

Vegyük példaként az egészségügyi ellátást:

- Minden orvos követi a protokollt?
- Miért kell a betegnek olyan sokáig várnia?
- Mi okozza a késedelmet?
- Meg tudjuk jósolni a páciensek várakozási idejét?
- Meg tudjuk előre határozni, hogy a következő napon az adott osztályon mekkora személyzetre lesz szükség?
- Hogyan lehet csökkenteni a költségeket a minőségi ellátás megtartása mellett?



## Vizsgálható kérdések 2

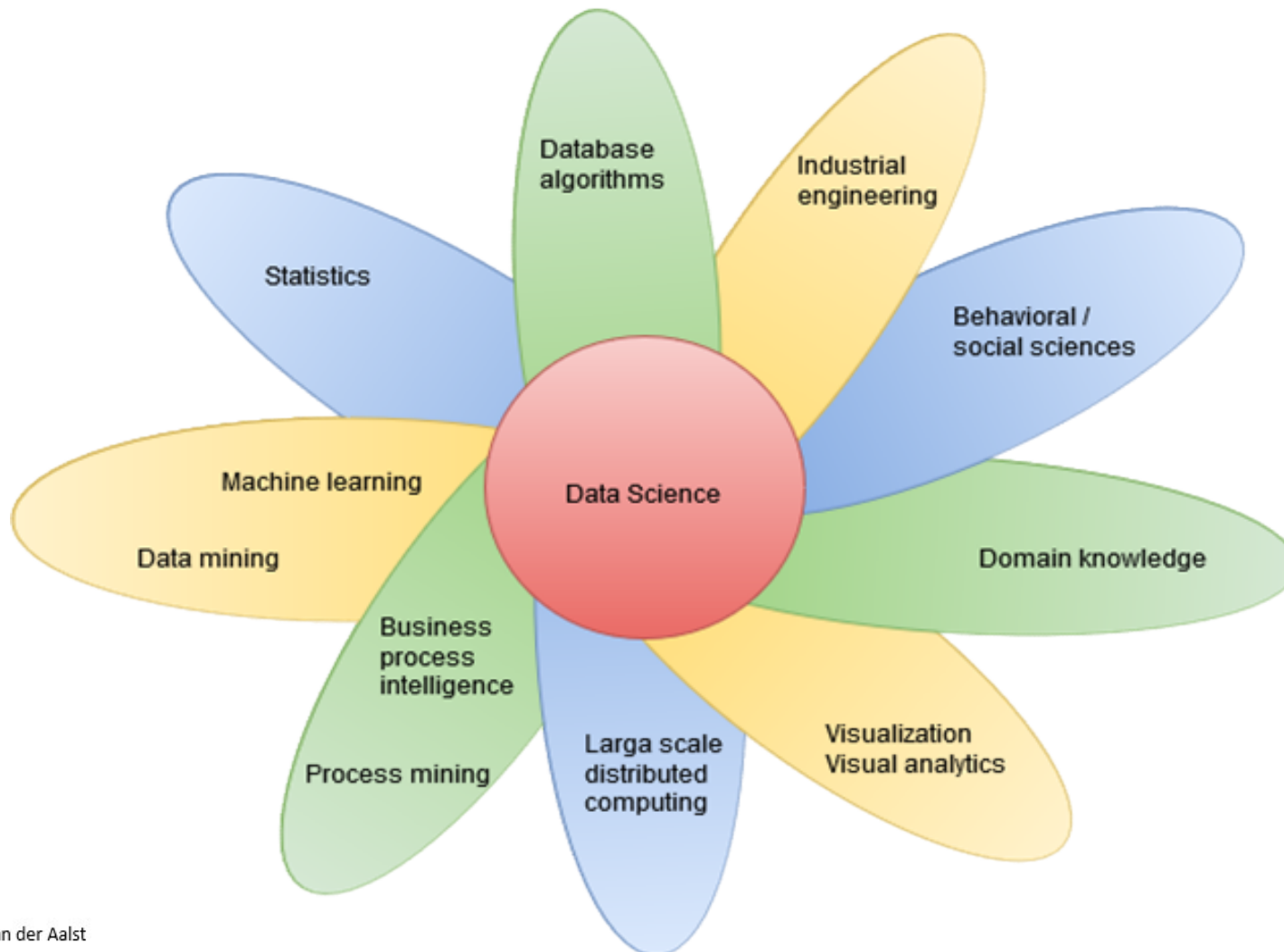
Vegyünk példaként egy gyártási folyamatot:

- Melyik komponens törött el, hibásodott meg?
- Mikor tört el/hibásodott meg és miért?
- Melyik komponenst kell kicserélni?
- Meg tudjuk jósolni, hogy melyik komponenssel fog történni valami?
- Tudunk-e tanulni a létező problémákból, hogy mely részeket kell javítani?
- Mi okozza a késedelmet a folyamatban?
- Meg tudjuk előre határozni, hogy a következő napon az adott osztályon mekkora személyzetre lesz szükség?
- Hogyan lehet csökkenteni a költségeket a minőségi ellátás megtartása mellett?



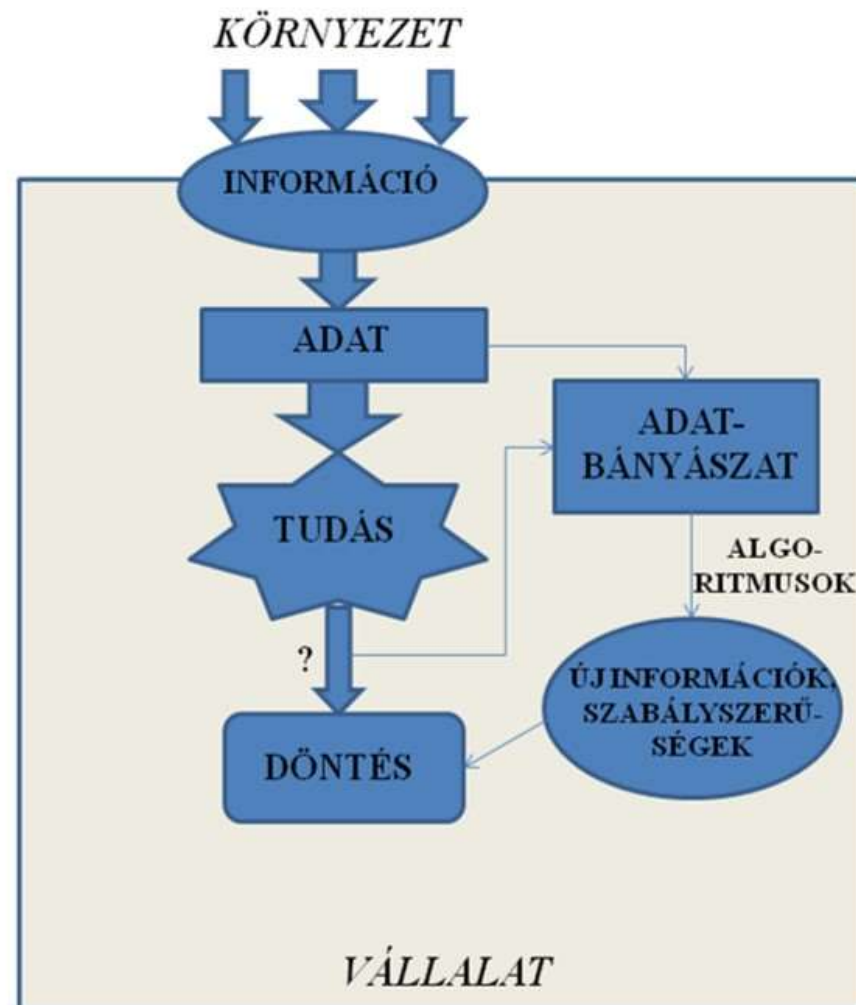
# Kérdések megválaszolása

Számos tudományterületre van szükség:



# Adatok felhasználása

A folyamatbányászat az adatbányászat egy speciális területének is tekinthető



# Folyamatbányászat vs. adatbányászat

- mindkettő adatokkal dolgozik
- az adatbányászati technikák nem folyamat centrikusak
- a tradicionális adatbányászati technikákkal a folyamat felfedezés, összehasonlítás, torlódás elemzés nem elvégezhető

A két módszer együttes használatával **komplex kérdésekre adhatunk választ.**



# Mi is a folyamatbányászat?

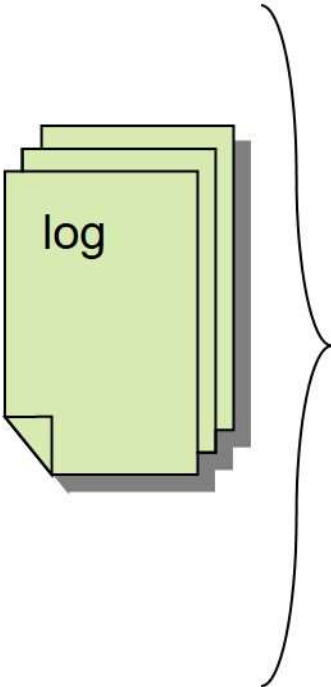
- A folyamatbányászat célja: **a folyamati adatokból valódi hasznosítható tudásanyag kinyerése.**
- Felderíthetők és megjeleníthetők azok az összefüggések (pl. az ügyfélcsoportok, régiók, a termékek stb. között), amelyek
  - a hosszú átfutási időért,
  - a magas költségekért,
  - a rossz minőségért felelősek.
- Felderíthetőek a gyártási folyamatban megjelenő hibák, amelyek
  - a leállásokért,
  - a hosszabb gyártási időért felelősek.

# Kiemelt területek

- Folyamat felfedezése , feltérképezése (Process discovery): **mi történik valójában?**
- Megfelelőség elemzés (Conformance checking): **az történik amit elvárunk, amit előre meghatároztunk?**
- Teljesítményelemzés (Performance analysis): **hol lehet javítani, változtatni?**
- Folyamat előrejelzés (Process prediction): **vajon egy eset be fog következni?**
- Folyamat fejlesztés, tökéletesítés (Process improvement): **hogyan alakítsuk át a folyamatot?**
- Stb.

# A folyamatbányászat perspektívái

Minden folyamatot megvizsgálhatunk más-más szempont szerint is, ezeket perspektíváknak nevezzük.



Tartalma	Perspektíva
A feladatok végrehajtásának körülményei és azok kapcsolatai	Ellenőrzött folyam (control flow) perspektíva
A végrehajtó személyek / rendszerek	Szervezeti (organizational) perspektíva
További információk (adatmezők értékei) a feladatokról	Eset szerinti (case) perspektíva

# Hol alkalmazhatjuk a folyamatbányászatot?

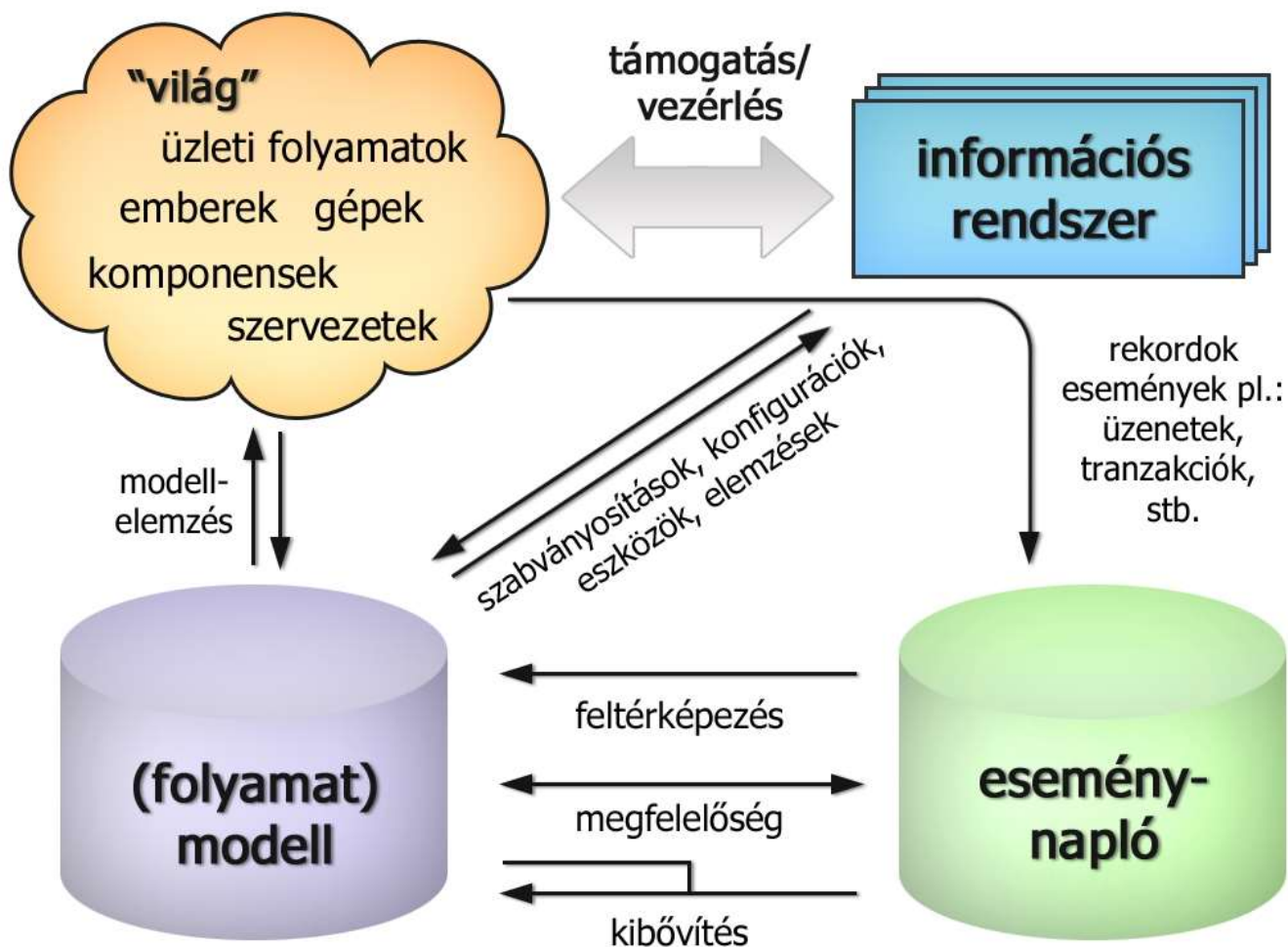
- Önkormányzatoknál (pl.: Alkmaar, Heusden, Harderwijk)
- Kormányzati ügynökségeknél (pl.: Rijkswaterstaat, Centraal Justitiele Incasso Bureau, Justice department)
- Biztosításokhoz kapcsolódó társaságoknál (pl.: UWV)
- Bankoknál (pl.: ING Bank)
- Kórházakban (pl.: AMC kórház, Catharina kórház)
- Multi cégeknél (pl.: DSM, Deloitte)
- Média vállalatoknál (pl.: Winkwaves)
- High-tech rendszerek gyártóinál és fogyasztóinál (pl.: Philips Healthcare, ASML, Thales)
- stb.

## Kiindulási pont: esemény adatok

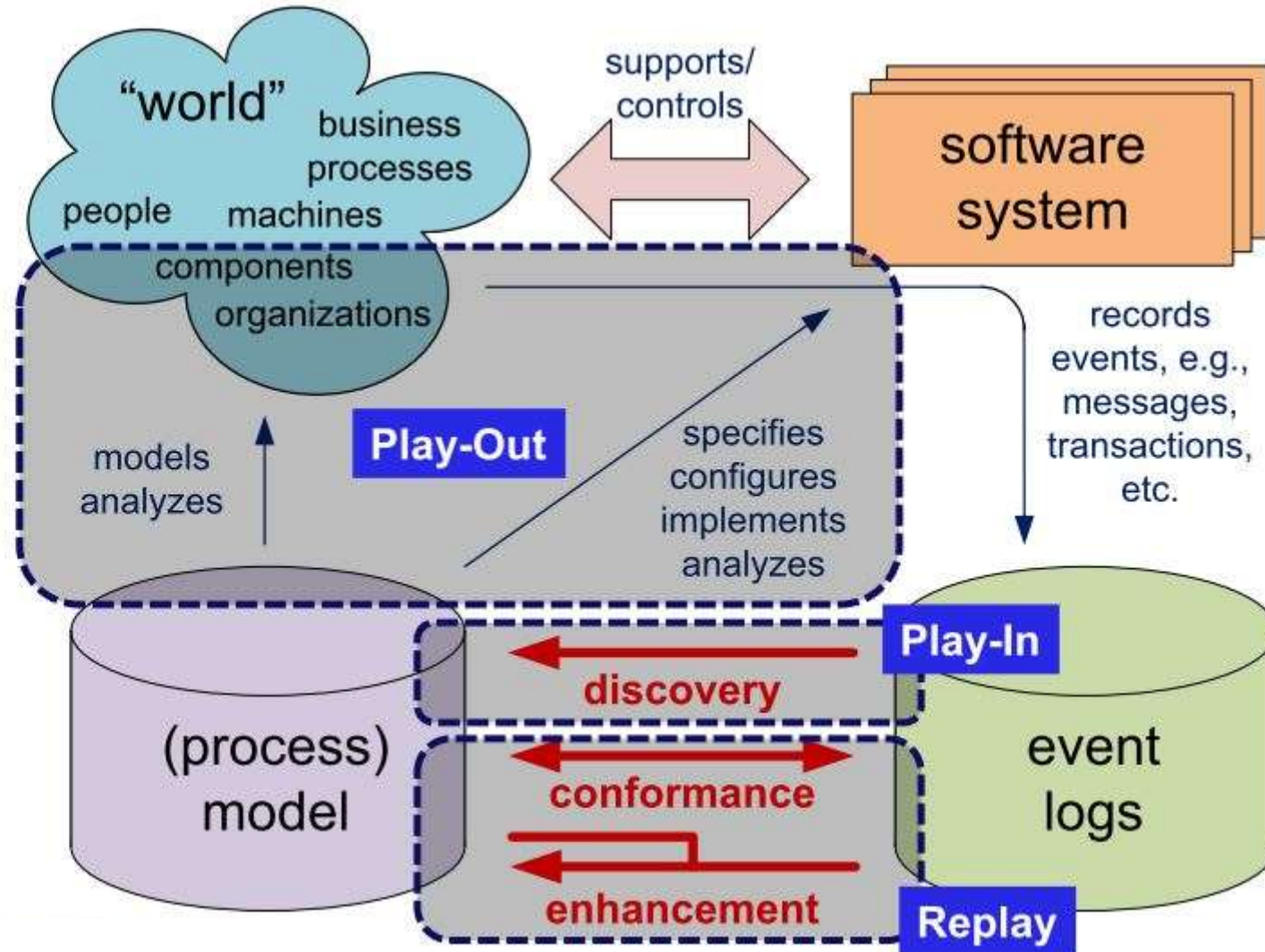
Minden sor egy eseményt azonosít (eset azonosító, esemény neve, időbélyeg, egyéb).

Hallgató neve	Kurzus	Vizsga időpontja	Jegy
Kovács Péter	Programozás I.	2016.01.15	4
Molnár Hajnalka	Programozás I.	2016.01.15	5
Fehér Klára	Programozás I.	2016.01.15	2
Fekete Péter	Programozás I.	2016.01.15	3
Kovács Péter	Programozás II.	2016.01.18	4
Molnár Hajnalka	Programozás II.	2016.01.18	5
Fehér Klára	Mesterséges intelligencia	2016.01.22	3
Kovács Péter	Mesterséges intelligencia	2016.01.22	5
Molnár Hajnalka	Mesterséges intelligencia	2016.01.22	4

# A folyamatbányászat magas szintű modell diagramja



# Folyamatbányászati feladatkörök



## Kapcsolatok: folyamat modellek - esemény logok

- **Play-out** (szimuláció, munkafolyam automatizálás, modell ellenőrzés stb.): A folyamatból kell lefutásokat generálni. Ezeknek a lefutásoknak a száma ugrásszerűen növekedik a folyamat bonyolultságától függően. Amennyiben a folyamat tartalmaz visszacsatolást, akkor a lehetséges lefutások halmaza akár végtelen nagy is lehet.
- **Play-in** (logokból modell felépítése (folyamat felfedezés)): Minél nagyobb a lefutások halmaza, annál jobb folyamat generálható hozzá. Korlátot szab a folyamat bonyolultsága, minél összefüggőbb, annál nehezebb ütemezett folyamatot pontosan visszaadni, hiszen csak a valós lefutások álnak rendelkezésre inputnak.
- **Replay** (adott a modell és a viselkedés, megfelelés vizsgálat, teljesítmény elemzés, előrejelzés, torlódás vizsgálat stb.): Egy meglévő folyamathoz kell a valós lefutásokat viszonyítani



# Példa1: A folyamat felfedezés hasonlít a nyelvtanuláshoz

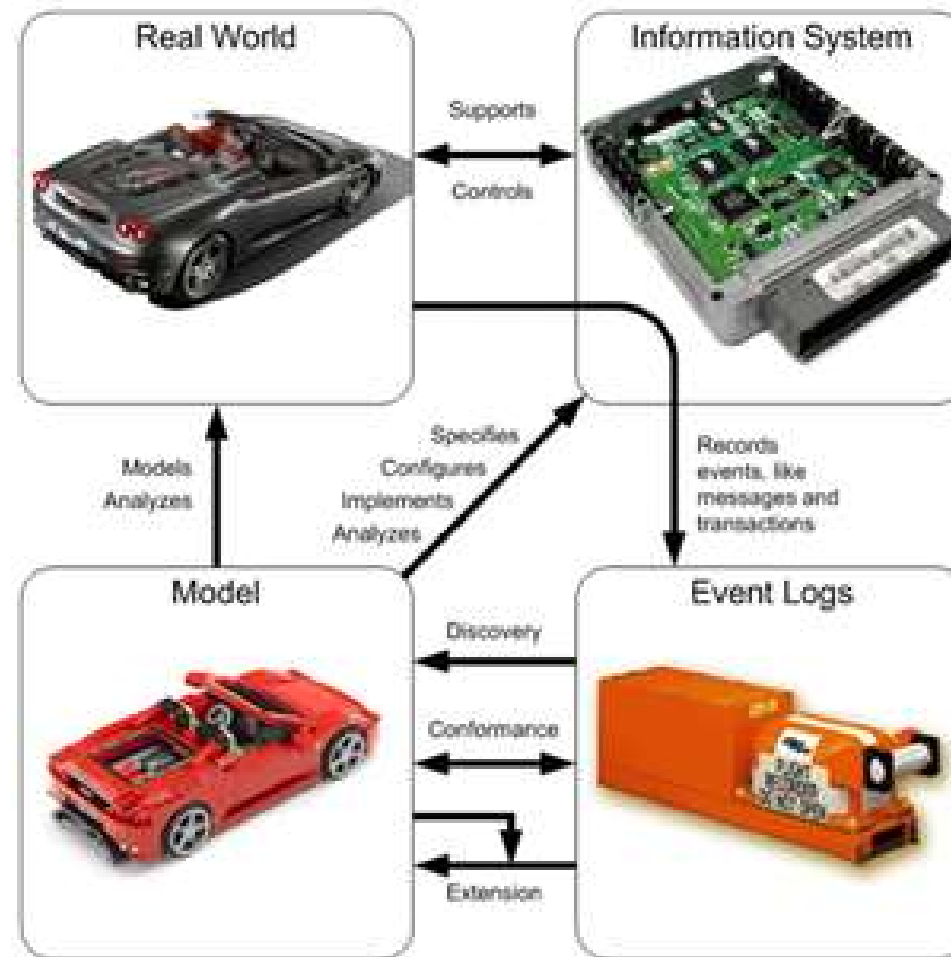
A mondatok olyanok mint a trace-ek az esemény logban, a nyelv pedig a folyamat modell.



Diagram by W. van der Aalst

Az összehasonlítás elemzés hasonlít a helyesírás ellenőrzéshez: a nyelv modellje a helyesírás ellenőrző.

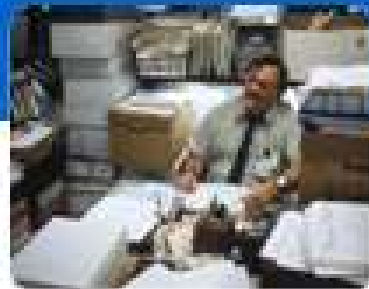
# Példa2: járművek fejlesztése



# Honnan érkeznek az esemény adatok?

- adatbázis rendszerekből (pl. páciensek adatai egy kórházban)
- kórházi információs rendszerekből (pl. Chipsoft)
- tranzakciós logokból (pl. kereskedelmi rendszer)
- üzleti rendszerekből, vállalati erőforrás tervező rendszerekből (pl. SAP, Oracle)
- üzenet logokból (pl. IBM middleware)
- folyamatirányító rendszerekből (pl. WebSphere software, Staffware, Flower, Eastman software, People Soft, Adept)
- web szerverek (pl. Apache)
- Szoftver Konfiguráció Menedzsment rendszerekből (pl. Subversion)
- ügyfél kapcsolatokat menedzselő rendszerekből (pl.: Microsoft Dynamics CRM)
- vesszővel elválasztott értékek fájljaiból (CSV) vagy táblázatokból
- stb.

# Példák:



## Egy másik példa esemény log-ra, páciensek adatai

páciens ID	művelet	időpont	orvos	kor	költség
23456	CT scan	2016.01.17 17:10	dr. Kovács	23	40.000
25673	MR	2016.01.17 12:00	dr. Molnár	36	60.000
27654	MR	2016.01.18 11:20	dr. Molnár	45	60.000
21117	radiológiai terápia	2016.01.18 11:30	dr. Kiss	65	10.000
24567	vérvizsgálat	2016.01.18 11:30	Mérő Henrietta	34	25.000
27673	CT scan	2016.01.18 12:30	dr. Kovács	37	40.000
26765	CT scan	2016.01.18 12:30	dr. Kovács	51	40.000
23654	vérvizsgálat	2016.01.18 13:00	Mérő Henrietta	21	25.000
27134	vérvizsgálat	2016.01.18 13:10	Mérő Henrietta	70	25.000

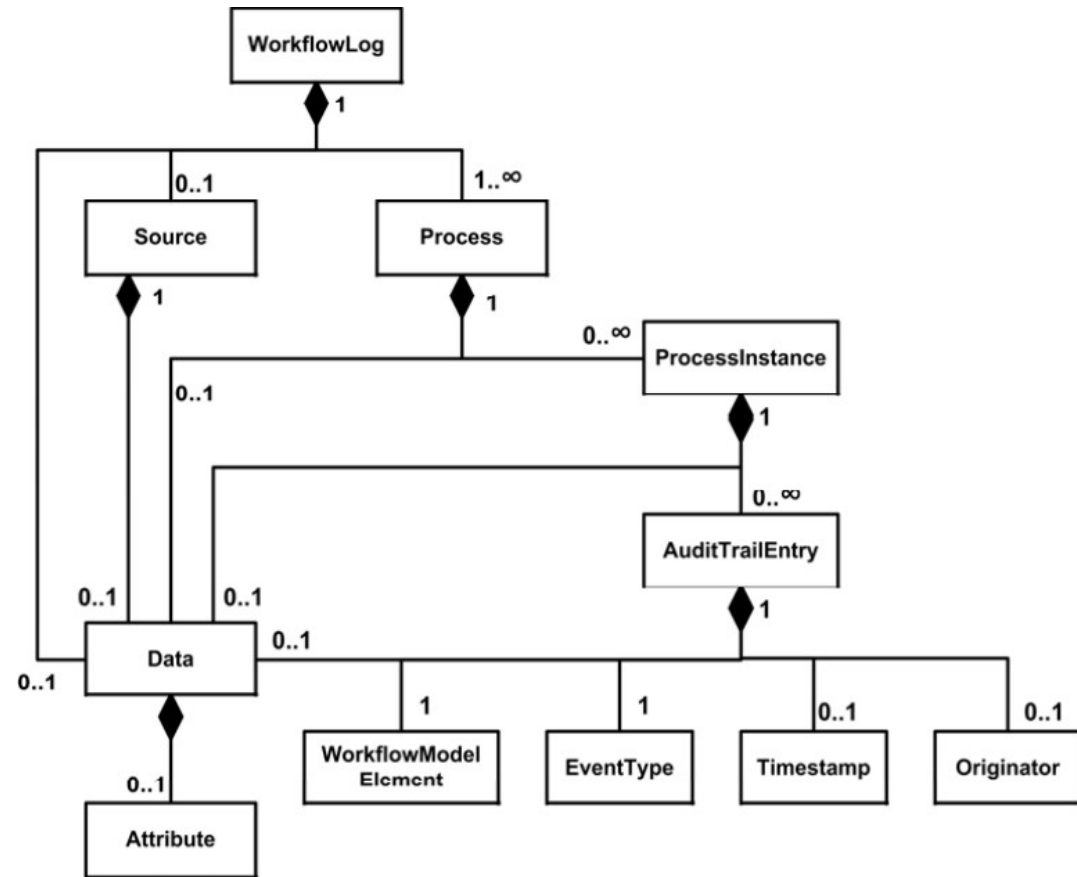
Minden sor egy eseményt azonosít (eset azonosító, esemény neve, időbélyeg, erőforrás, költség).

Az esemény adatok ilyen és hasonló listáit tudjuk elemezni.

# Kiindulási pont: esemény napló



esemény logok, ellenőrzési naplók, adatbázisok, üzenet logok stb.



egyértelmű esemény log  
(MXML formátum)

# A Staffware rendszer saját formátumú naplózásának MXML formátumra való fordításából kapott eseménynapló egy részlete:

## Esemény log:

- folyamatok
  - folyamat előfordulások

## Egy esemény:

- művelet neve
- (esemény típusa)
- (végrehajtó)
- (időbélyeg)
- (adat)

```
<Source program="Staffware">
  <Data>
    <Attribute name="version">7.0</Attribute>
  </Data>
</Source>
<Process id="main_process">
  <Data>
    <Attribute name="description">complaints handling</Attribute>
  </Data>
<ProcessInstance id="Case 1">
  <AuditTrailEntry>
    <WorkflowModelElement>Case start</WorkflowModelElement>
    <EventType unknowntype="case_event">unkown</EventType>
    <Timestamp>2011-04-16T11:06:00:000+001:00</Timestamp>
  </AuditTrailEntry>
  <AuditTrailEntry>
    <WorkflowModelElement>Register
    complaint</WorkflowModelElement>
    <EventType>schedule</EventType>
    <Timestamp>2011-04-16T11:16:00:000+001:00</Timestamp>
    <Originator>jvluin@staffw</Originator>
  </AuditTrailEntry>
```

Case	Activity	Timestamp	Resource
432	register travel request (a)	18-3-2014:8.16	John
432	get support from local manager (b)	18-3-2014:8.26	Mary
432	check budget by finance (d)	18-3-2014:8.66	John
432	decide (e)	18-3-2014:8.98	Sue
432	accept request (g)	18-3-2014:8.48	Mary

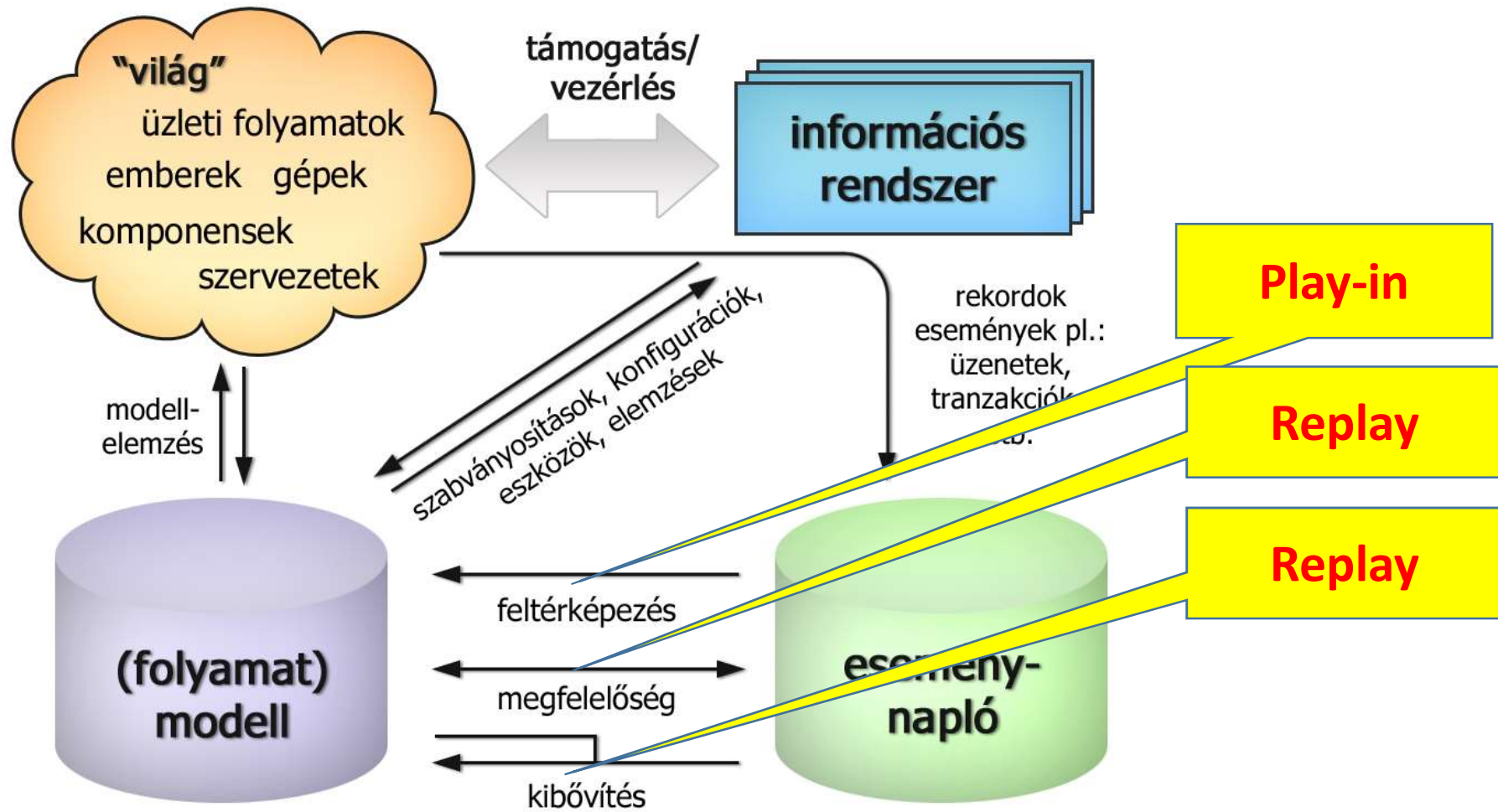


# Valós lefutások modelljei és esemény logok

Diagram by W. van der Aalst

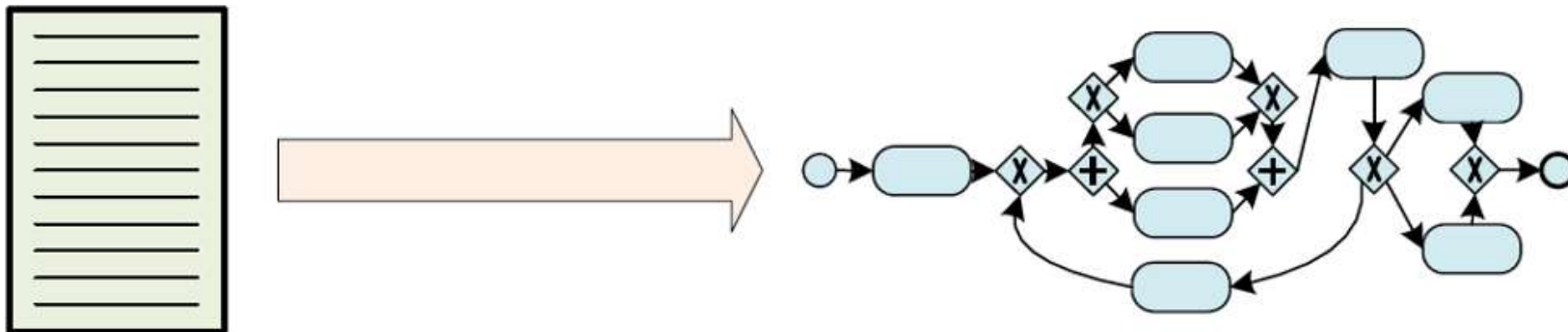


# A folyamatbányászat típusai




# Play-In

Folyamat felfedezés: a megfigyelt viselkedés alapján a folyamati modell előállítás, felépítése

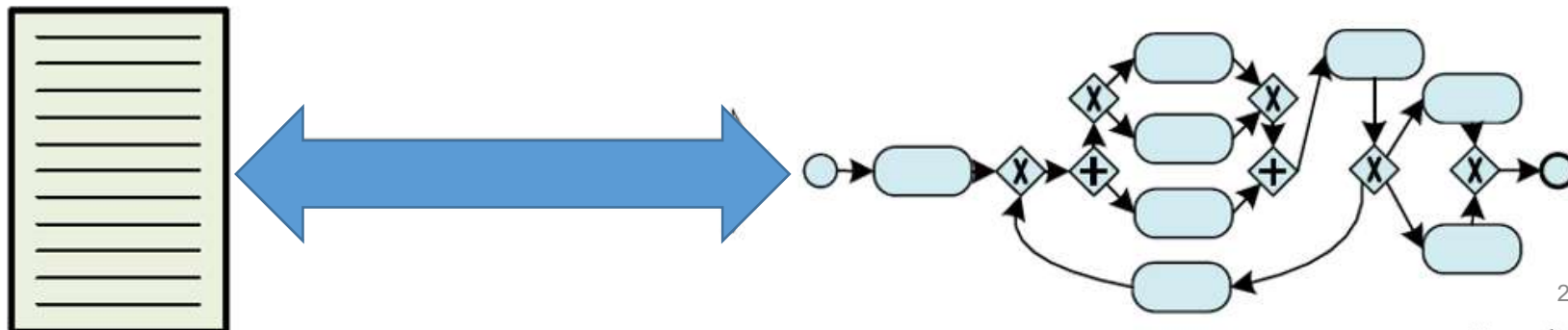


# Replay

A folyamatbányászat legfontosabb formája, amihez tekintjük a modellt és a viselkedést (valós lefutásokat)  szembeállítjuk a modellt a valósággal

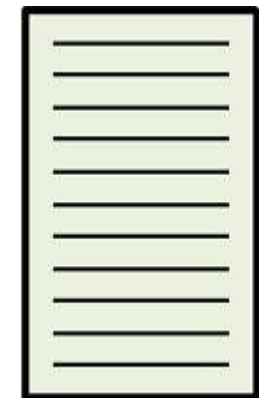
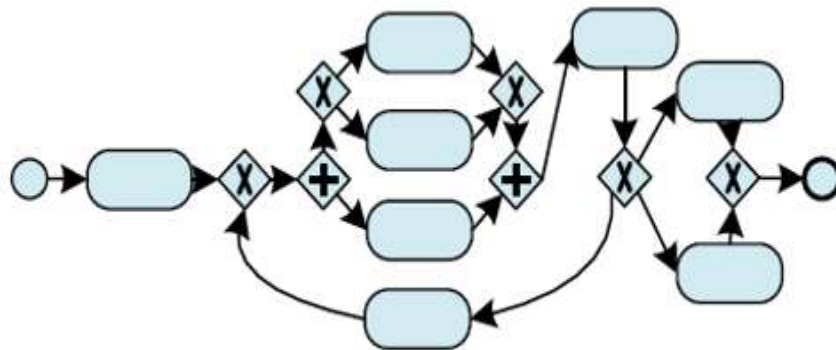
Használható:

- megfelelőség ellenőrzéshez (conformance checking),
- előrejelzéshez (prediction),
- torlódás elemzéshez (bottleneck analysis)
- diagnosztikához (diagnostics)
- stb.



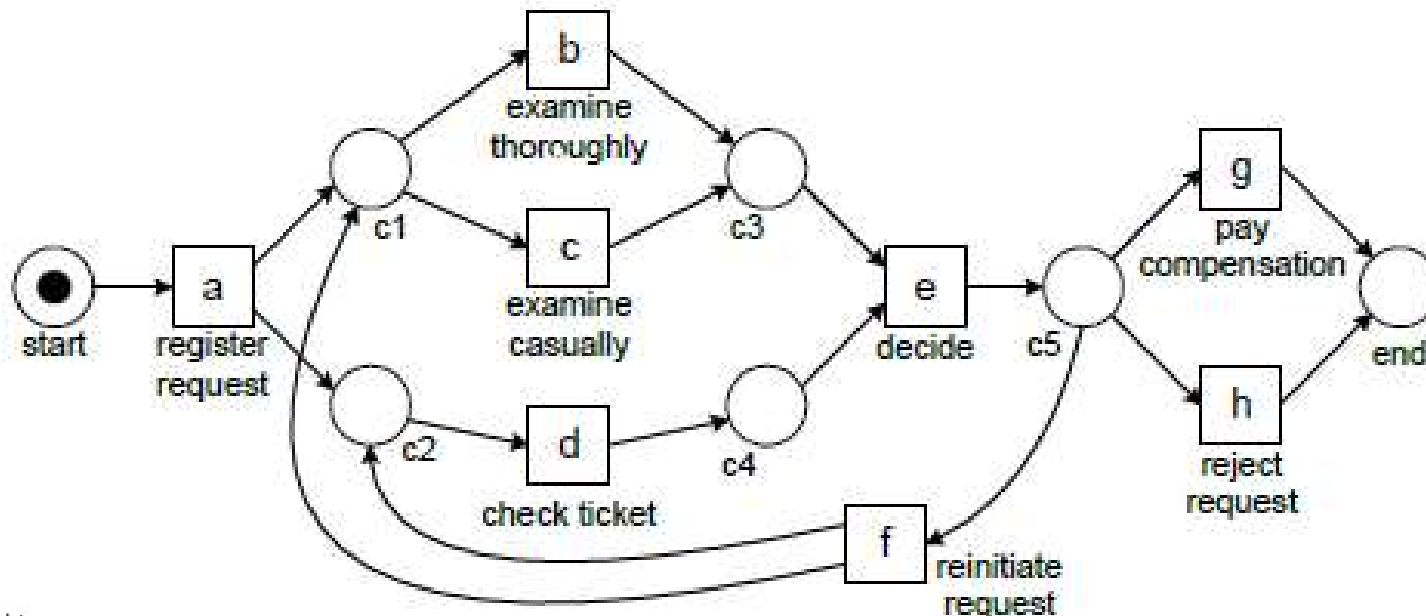
# Play-Out

- Szimuláció
- Folyamat automatizálás
- Menedzsment játékok
- Modell ellenőrzés

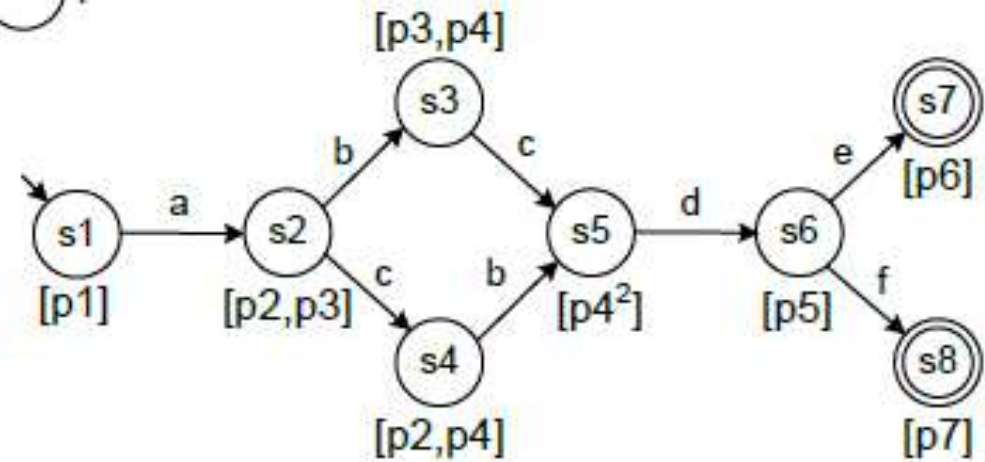
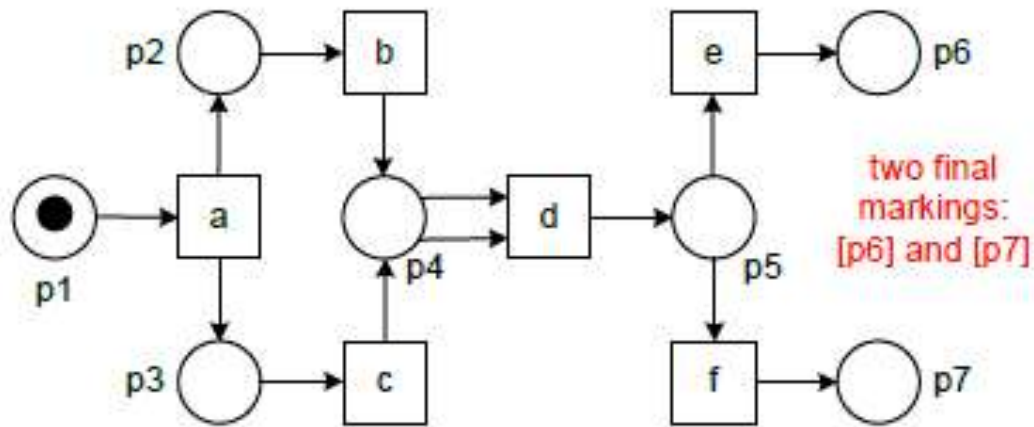


# A „jó” reprezentáció kiválasztása! Vezérlési folyamat reprezentálása Petri háló segítségével

- Az eset kezdődik az **a** művelettel és befejeződik a **g** vagy **h** művelettel.
- A **d** művelet konkurens a **b** vagy **c** művelettel.
- Az **e** művelet várakozik mindaddig, amíg (**d** és **b**) vagy (**d** és **c**) be nem fejeződik.
- Három féle döntési lehetőségünk van: **f**, **g** vagy **h**.

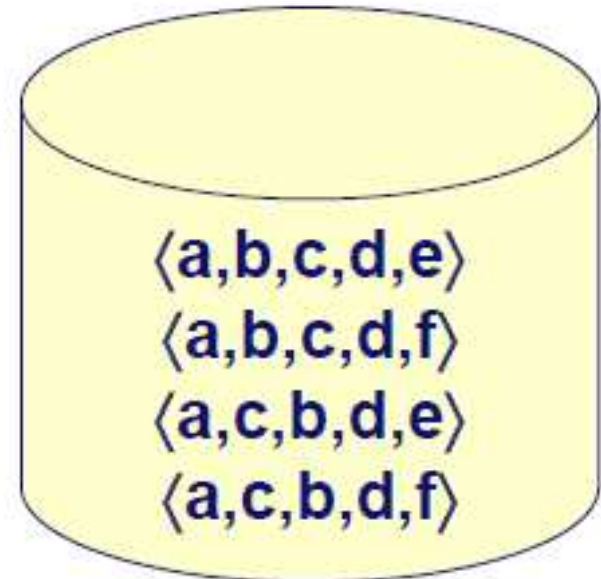
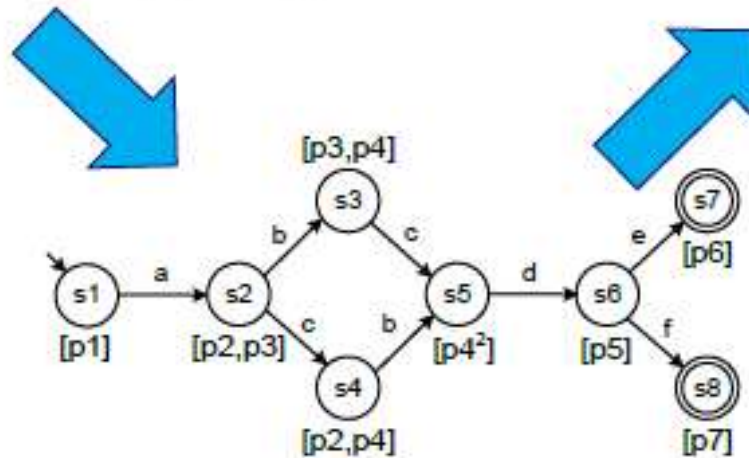
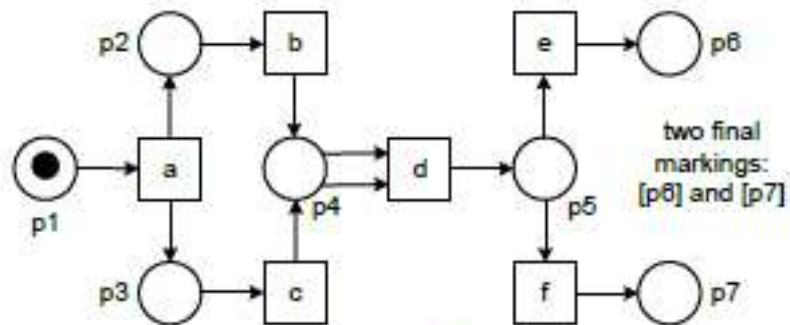


# A Petri háló - Átmenet rendszer

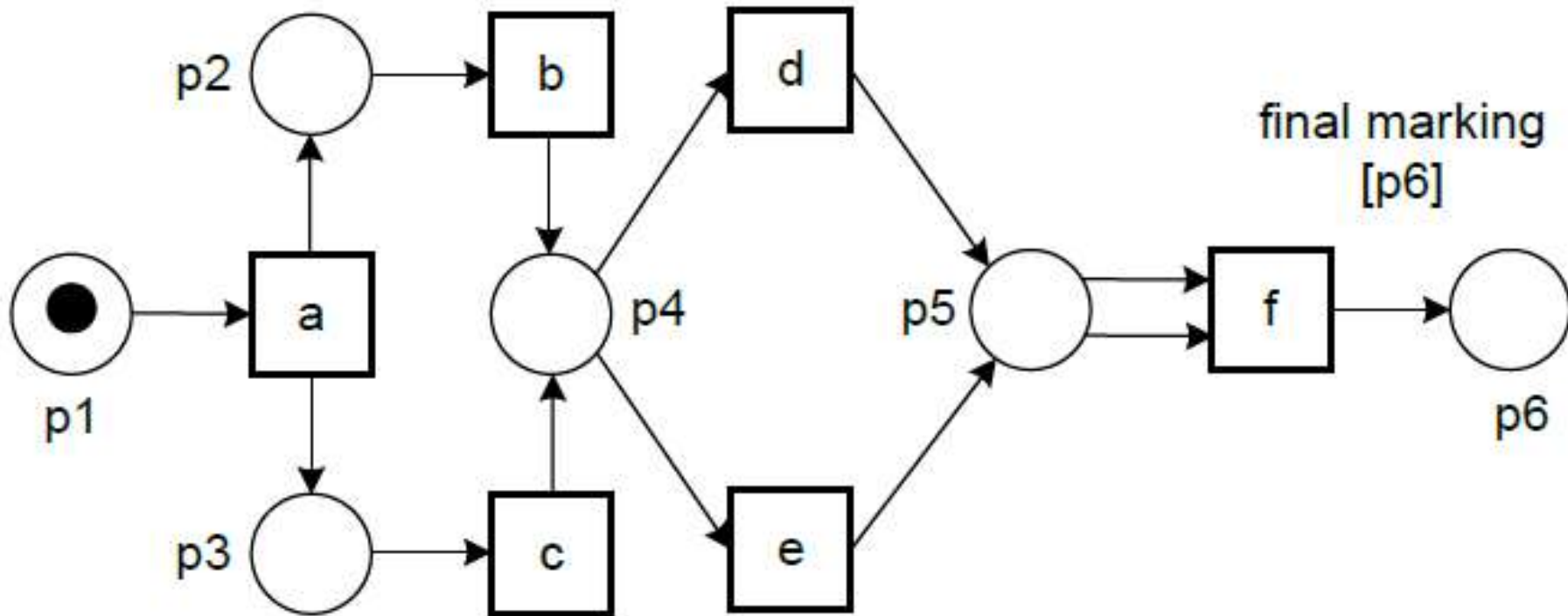


© 2004 W. van der Aalst, RWTH Aachen University

# Play-Out

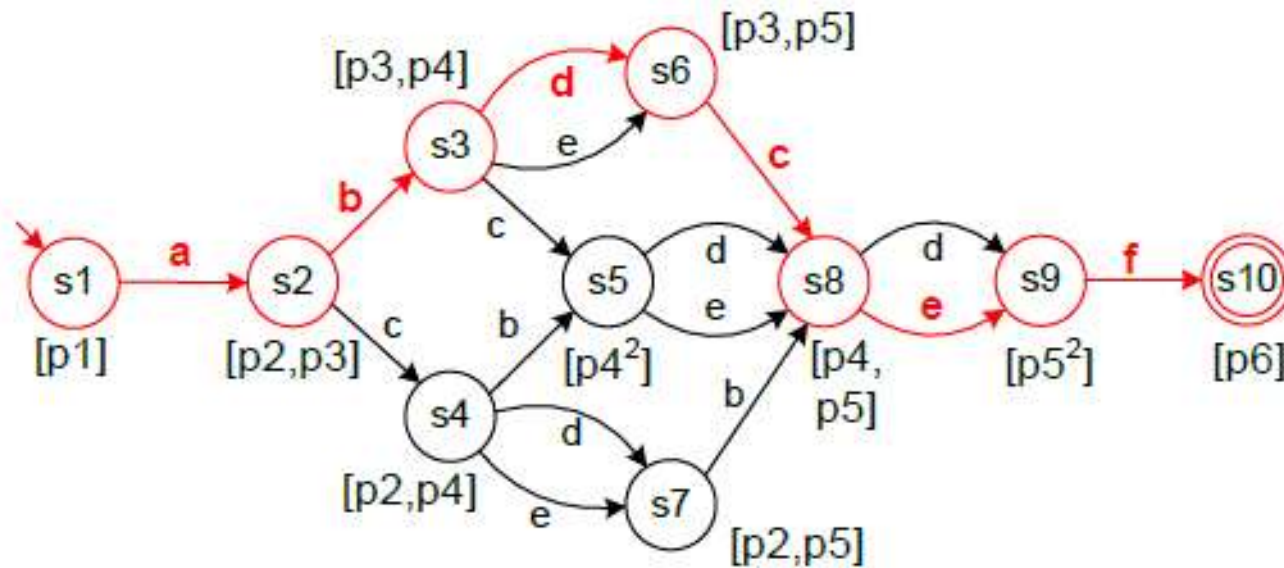


Play-out modell: határozza meg az átmenet rendszert és minden lehetséges teljes trace-t





# Példa egy útvonalra



- <a,b,c,d,d,f>
- <a,b,c,d,e,f>
- <a,b,c,e,d,f>
- <a,b,c,e,e,f>
- <a,c,b,d,d,f>
- <a,c,b,d,e,f>
- <a,c,b,e,d,f>
- <a,c,b,e,e,f>
- <a,b,d,c,d,f>
- <a,b,d,c,e,f>**
- <a,b,e,c,d,f>
- <a,b,e,c,e,f>
- <a,c,d,b,d,f>
- <a,c,d,b,e,f>
- <a,c,e,b,d,f>
- <a,c,e,b,e,f>