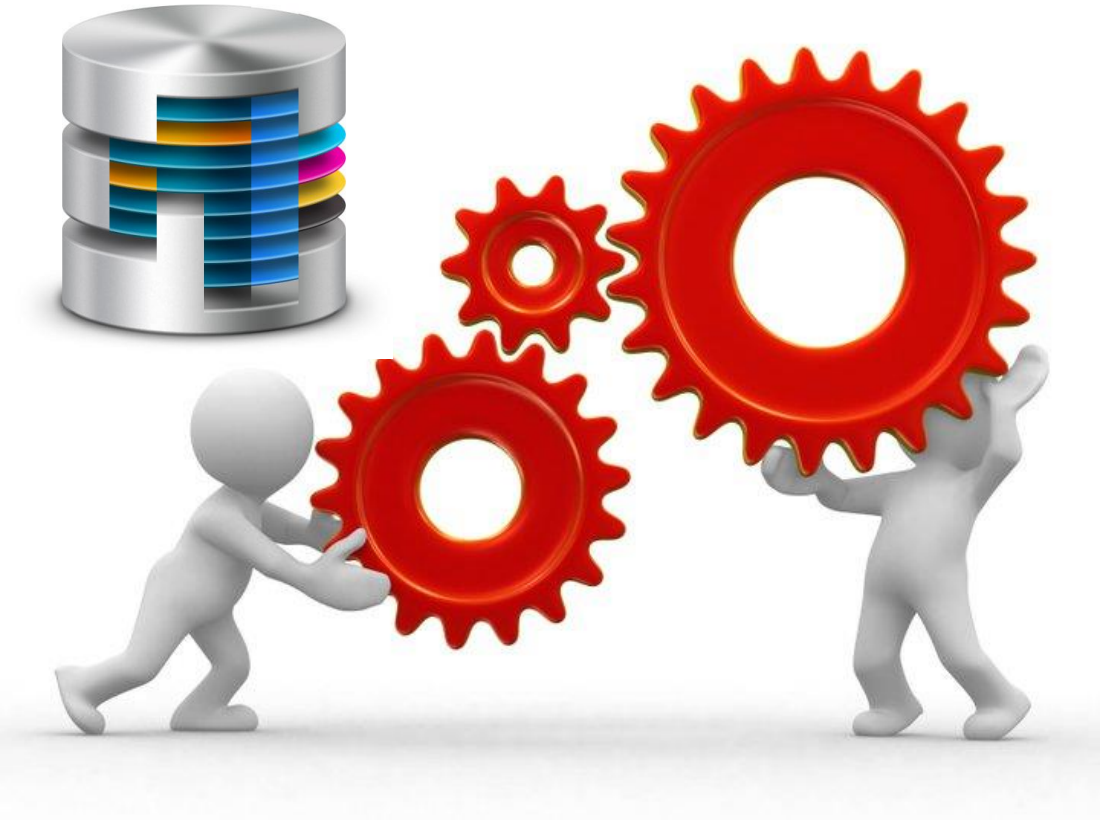
Accelerating Process Mining using Relational Databases

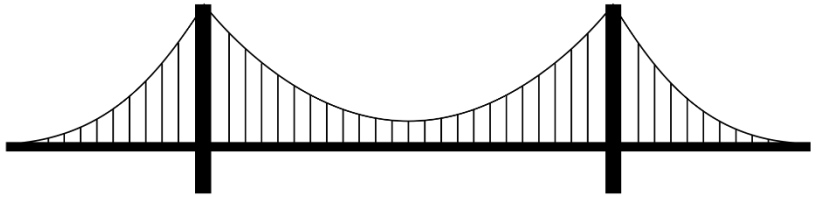


Alifah Syamsiyah

Eindhoven University of Technology

Process Mining

Data Science

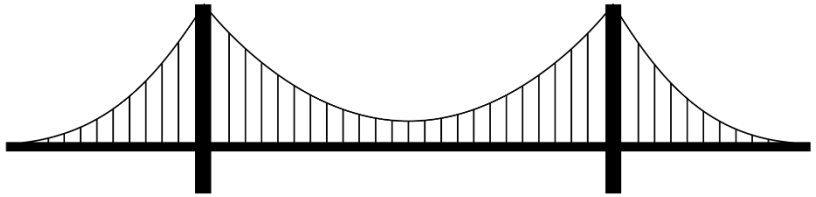


Process Science

Process Mining

Data Science

Data Mining
Machine Learning
Distributed Systems



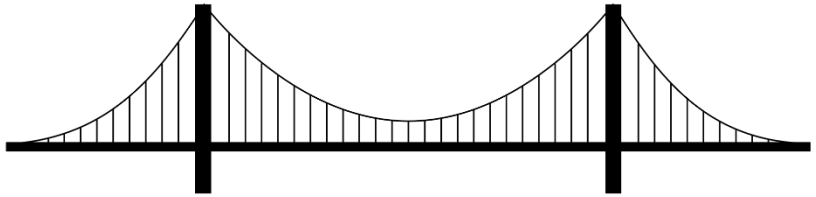
Process Science

ProM
process mining workbench

Process Mining

Data Science

Data Mining
Machine Learning
Distributed Systems



Process Science

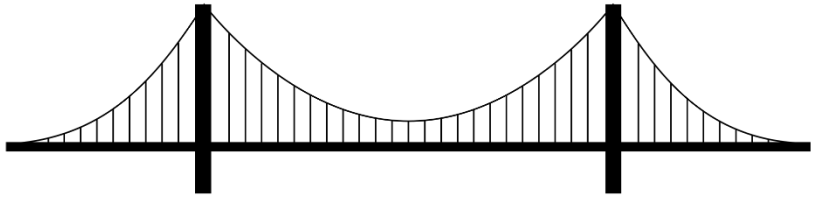
ProM
process mining workbench

Process Mining

“process agnostic”

Data Science

Data Mining
Machine Learning
Distributed Systems



Process Science

Business Process Management
Workflow Management
Optimization

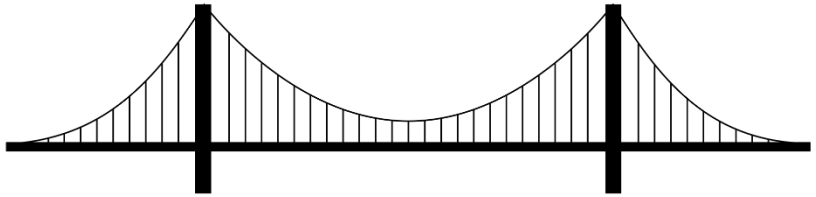
Process Mining

"process agnostic"

Data Science

Data Mining
Machine Learning
Distributed Systems

“process agnostic”



Process Mining

Process Science

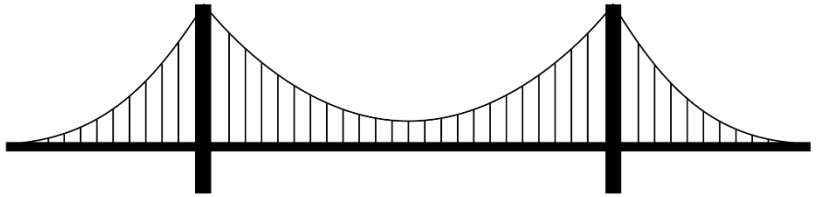
Business Process Management
Workflow Management
Optimization

“model-driven”

Data Science

Data Mining
Machine Learning
Distributed Systems

"process agnostic"



Process Science

Business Process Management
Workflow Management
Optimization

"model-driven"

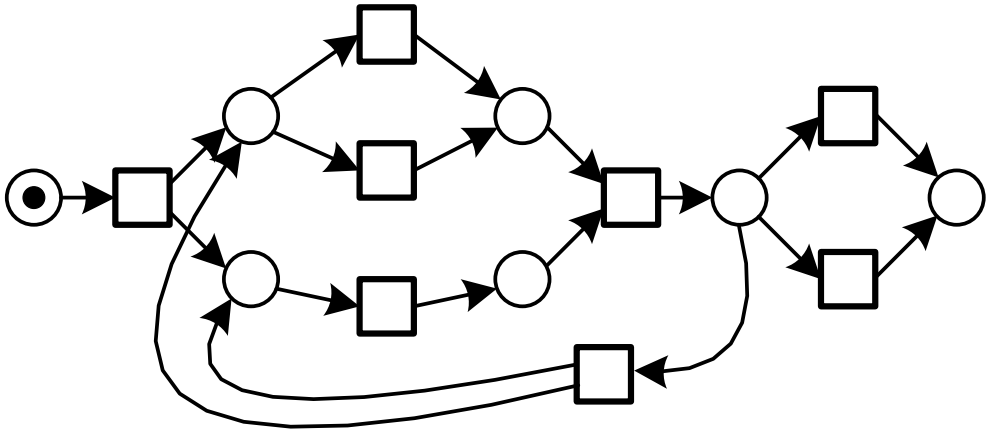
Process Mining

"to turn *event data* into insights and actions
in order to improve *processes*"

ProM
process mining workbench

Event Data

Event Data



Process Model

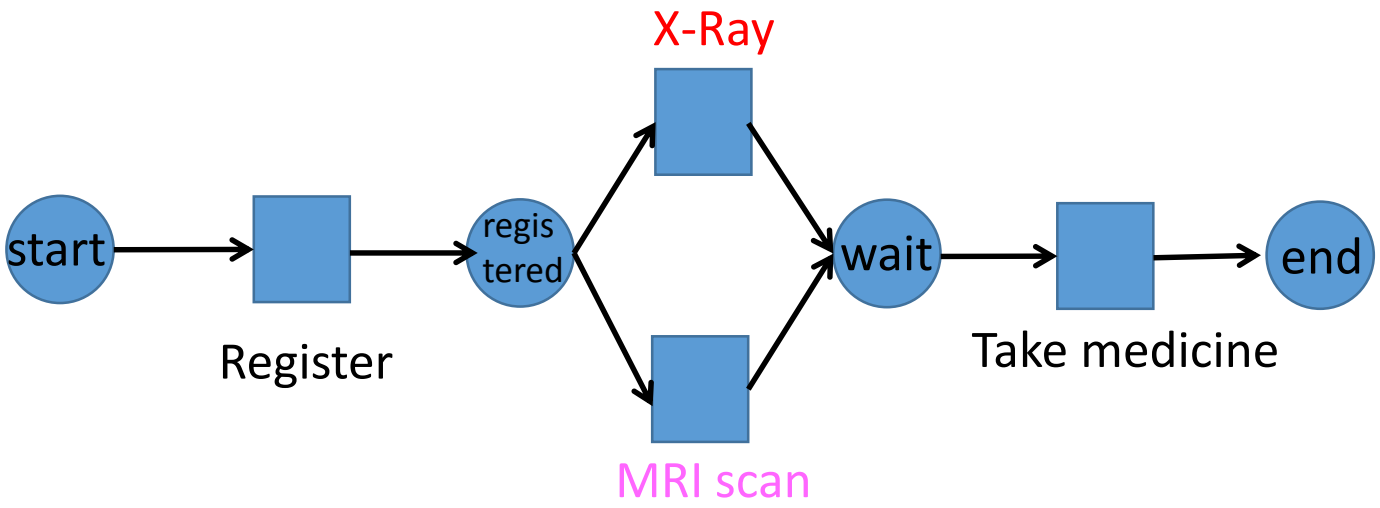
Event Data

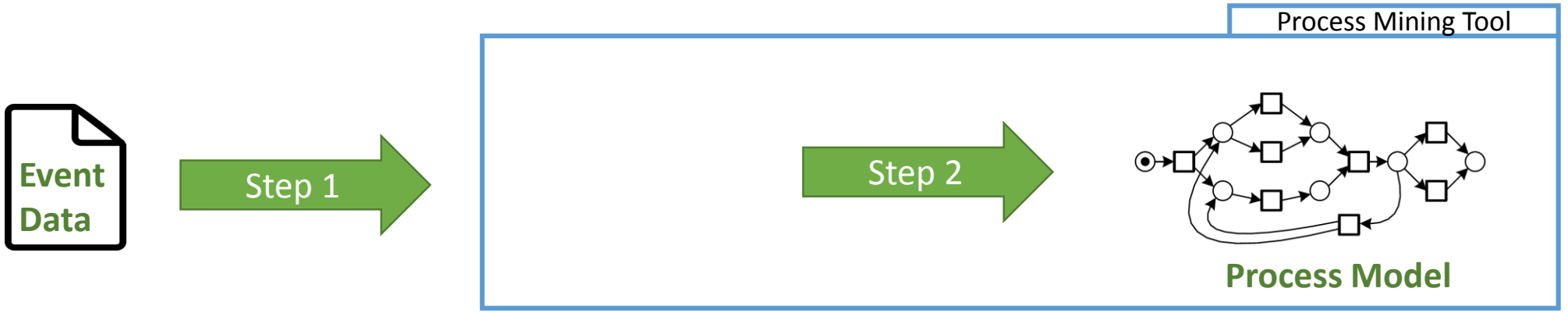
< Register, X-Ray, Take medicine >,
< Register, MRI scan, Take medicine >,
…

Event Data

< Register, X-Ray, Take medicine >,
< Register, MRI scan, Take medicine >,
…

Event Data




Process Mining Tool
Event Data
Step 1
Step 2
Process Model

Process Mining Tool
Event Data
Step 1
Intermediate Structure
Step 2
Process Model

Directly Follows Relation (DFR)

Directly Follows Relation (DFR)

< Register, X-Ray, Take medicine >

Directly Follows Relation (DFR)

< Register, X-Ray, Take medicine >

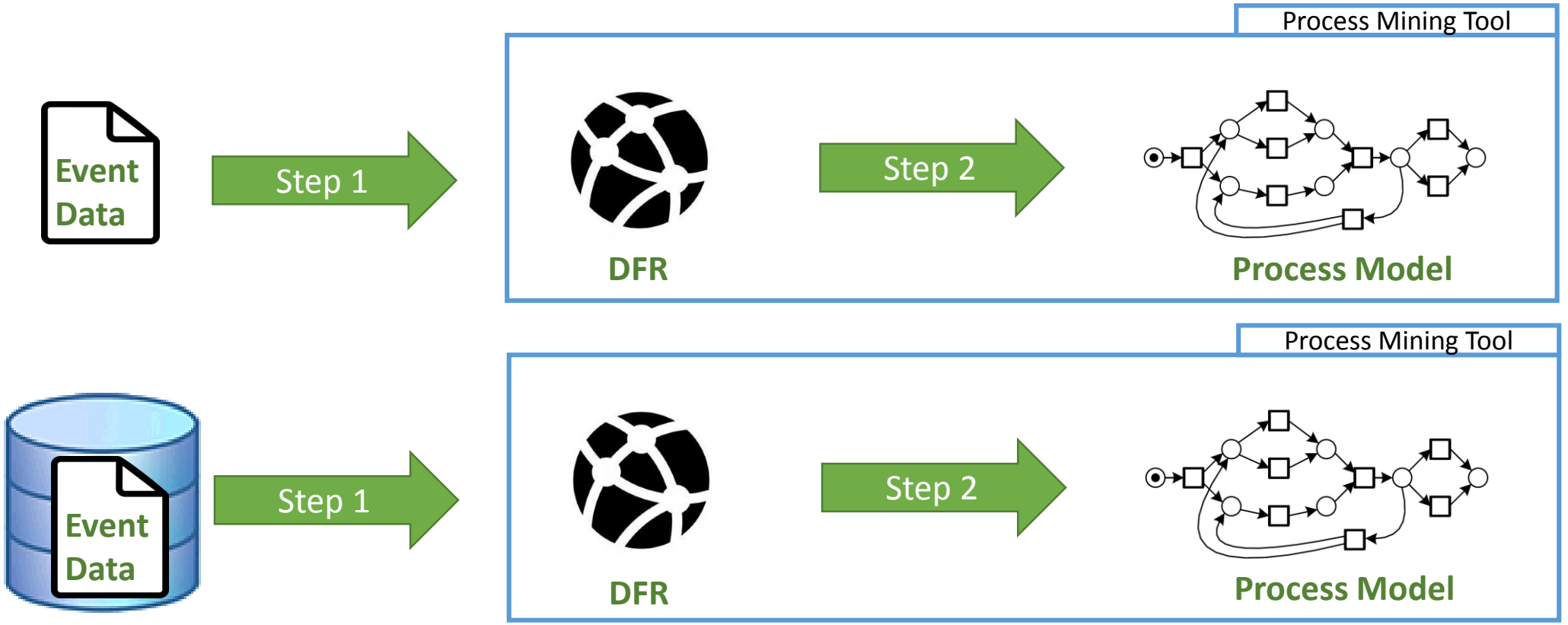
< Register, X-Ray, 1 >

< Register, X-Ray, Take medicine >

Directly Follows Relation (DFR)

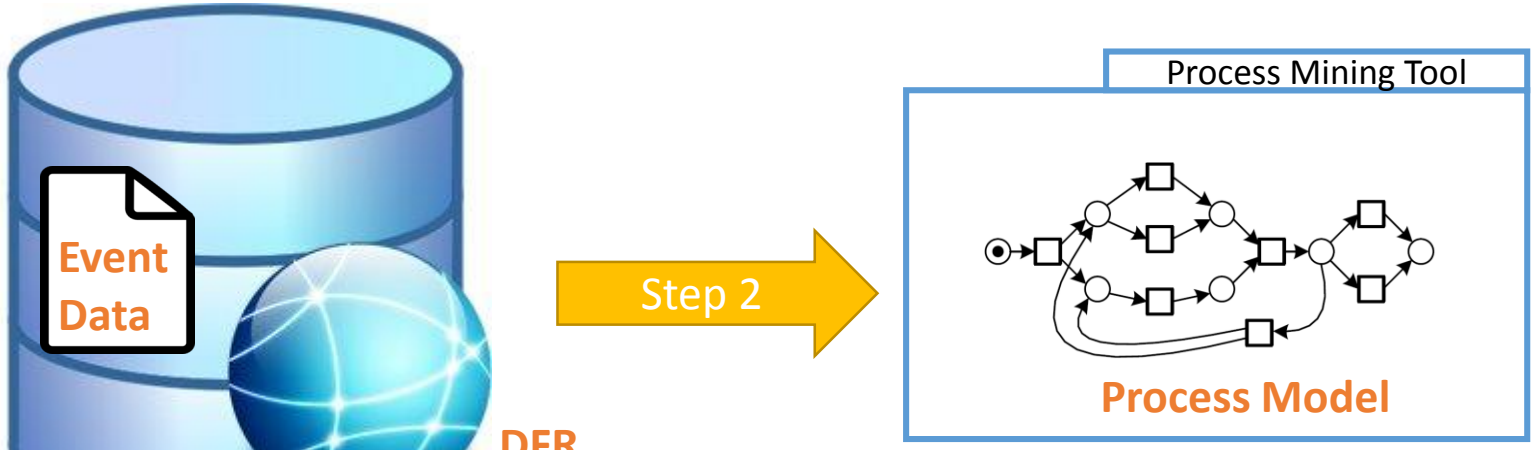
< Register, X-Ray, 1 >
< X-Ray, Take medicine, 1 >

Process Mining Tool
Event Data
Step 1
DFR
Step 2
Process Model


Process Mining Tool
Event Data
Step 1
DFR
Step 2
Process Model
Process Mining Tool
Event Data
Step 1
DFR
Step 2
Process Model

Process Mining Tool
Event Data
Step 1
DFR
Step 2
Process Model

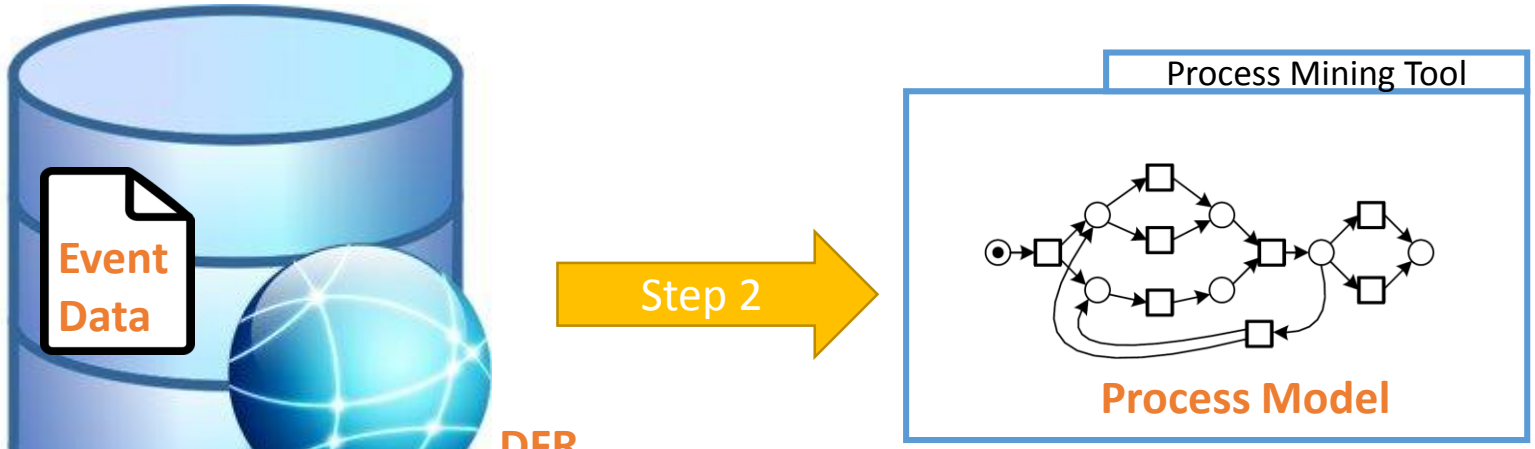
Process Mining Tool
Event Data
Step 1
DFR
Step 2
Process Model


Event Data
DFR
Step 2
Process Mining Tool
Process Model

Log

Event Data
Step 1
Process Mining Tool
DFR
Step 2
Process Model

Event Data
Step 1
Process Mining Tool
DFR
Step 2
Process Model


Event Data
DFR
Step 2
Process Mining Tool
Process Model

Log

Step 1

Process Mining Tool

DFR

Step 2

Process Model

Database

Step 1

Process Mining Tool

DFR

Step 2

Process Model

DFR

Step 2

Process Mining Tool

Process Model

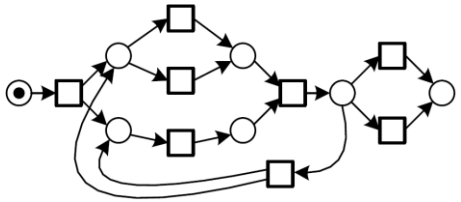
Log

Step 1

Process Mining Tool

DFR

Step 2



Process Model

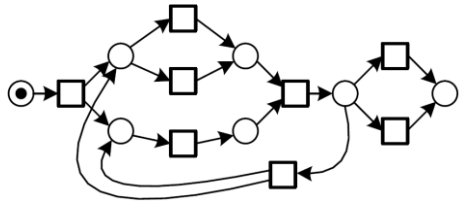
Database

Step 1

Process Mining Tool

DFR

Step 2



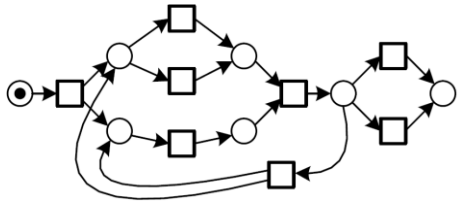
Process Model

Database with DFR

DFR

Step 2

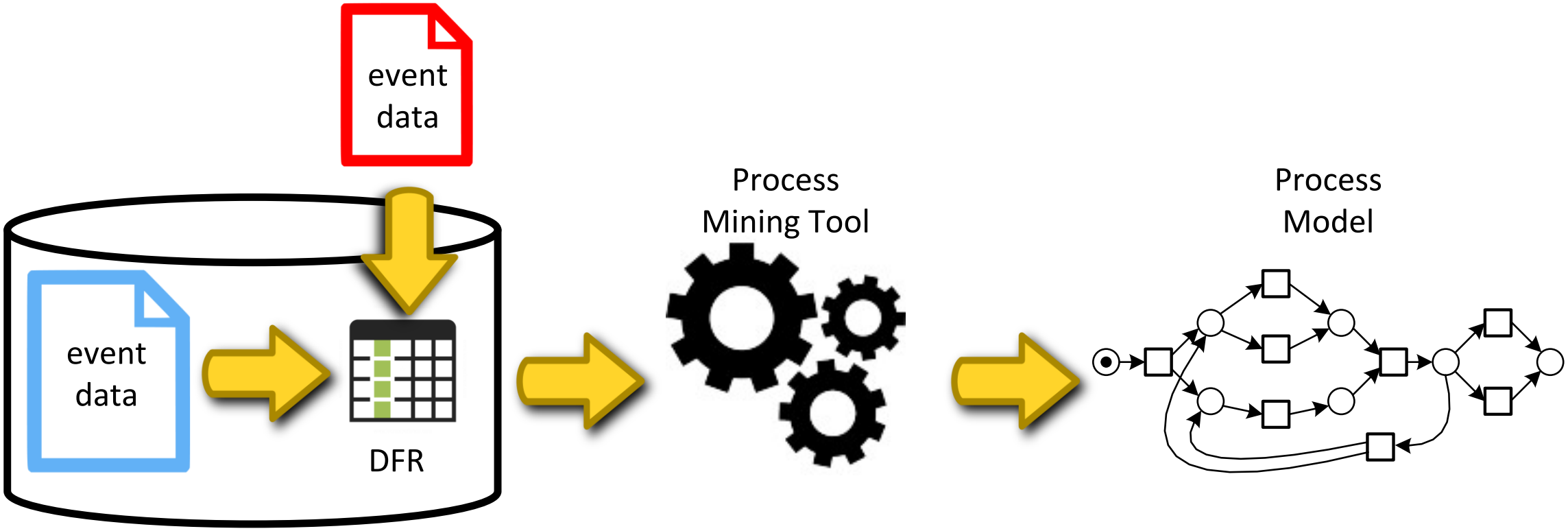
Process Mining Tool



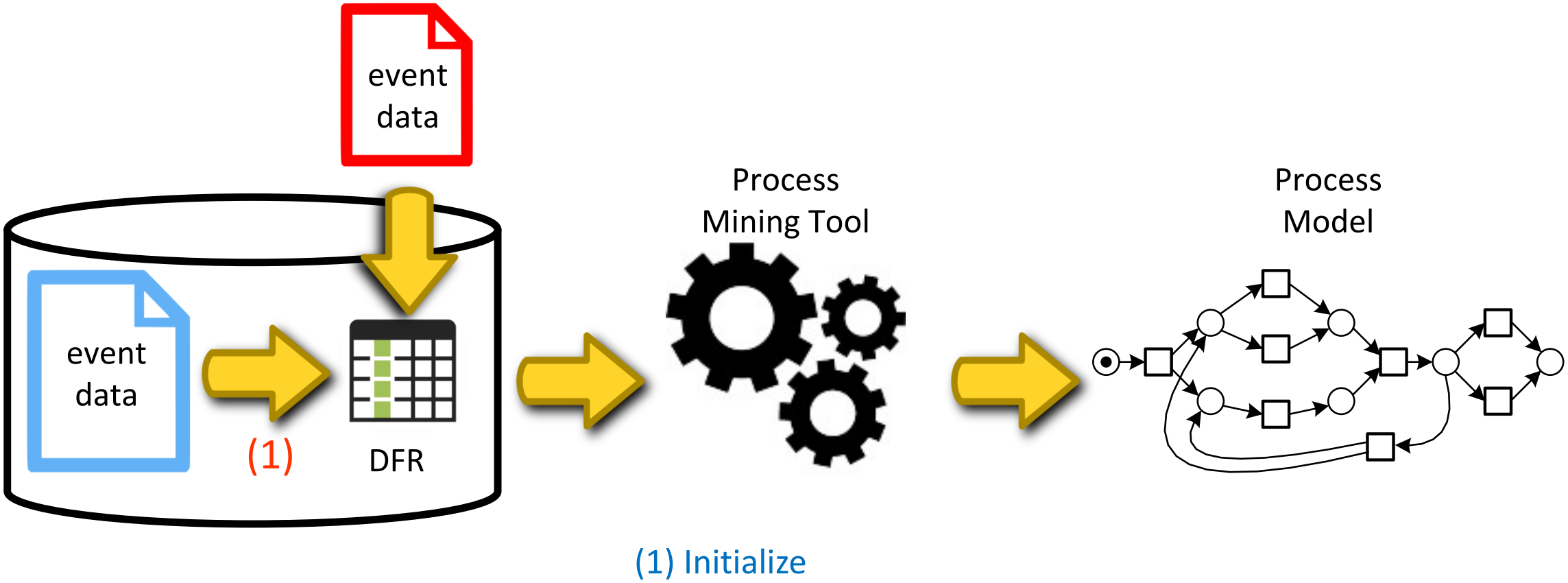
Process Model

“The use of
Relational Databases
to *pre-process event data*
can *accelerate process mining*
while maintaining its scalability”

Database with DFR

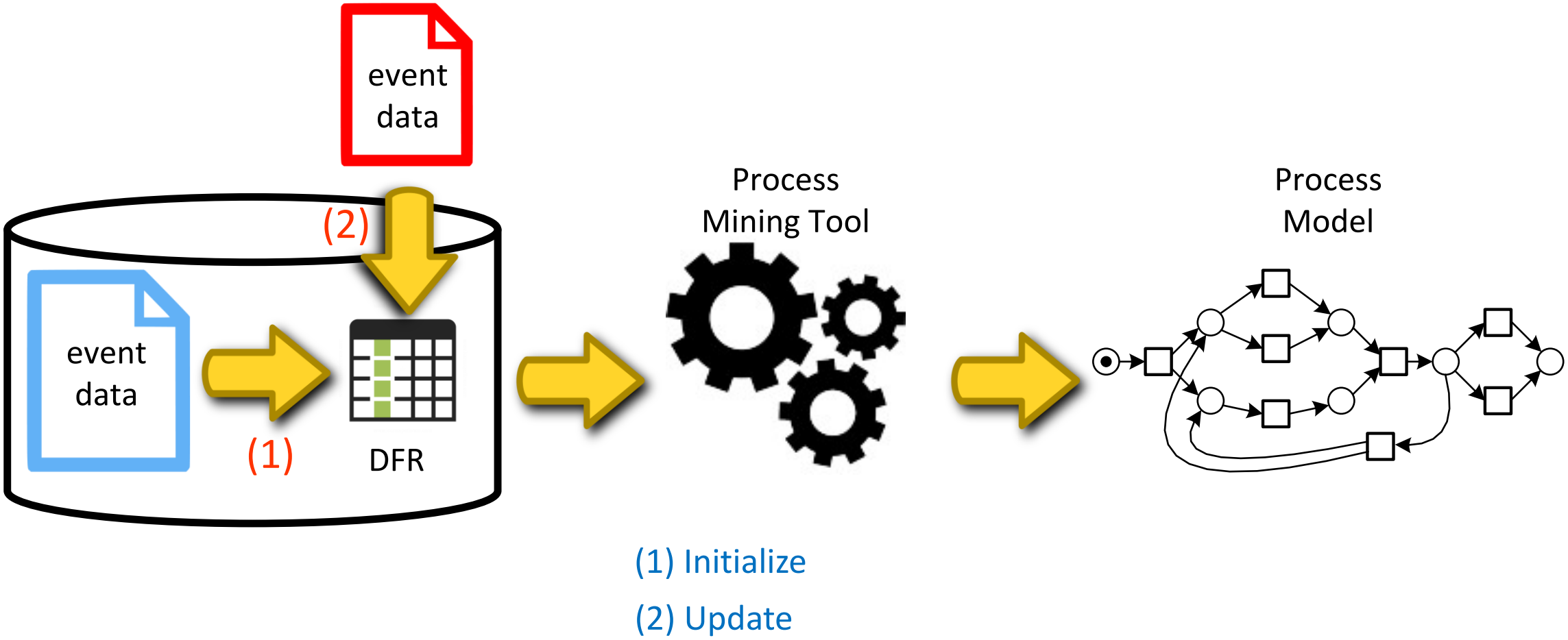


Database with DFR



(1) Initialize

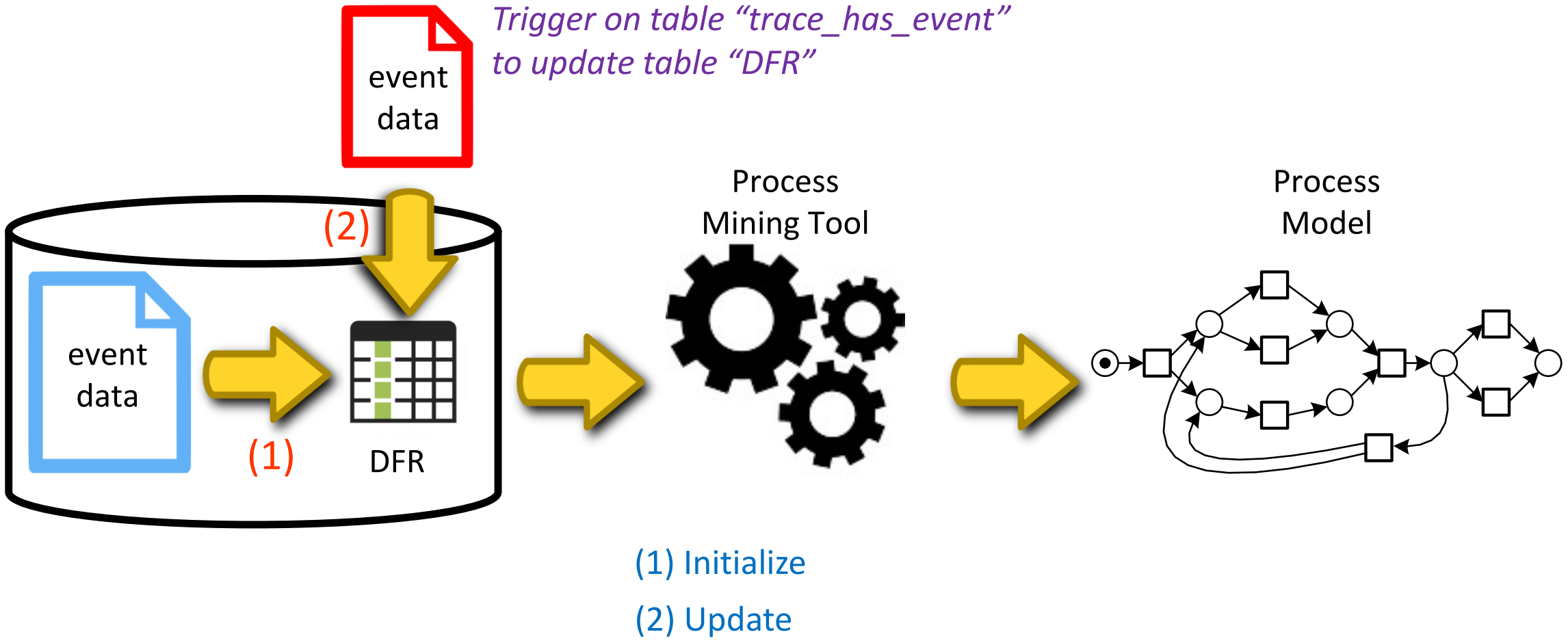
Database with DFR



(1) Initialize

(2) Update

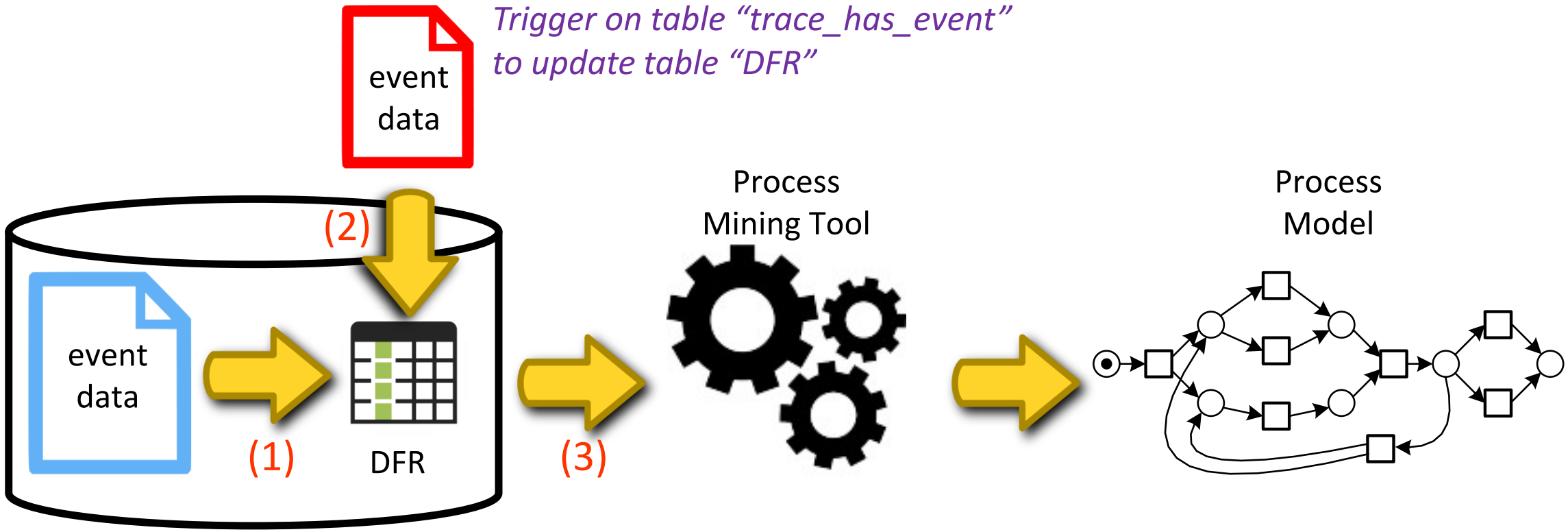
Database with DFR



(1) Initialize

(2) Update

Database with DFR

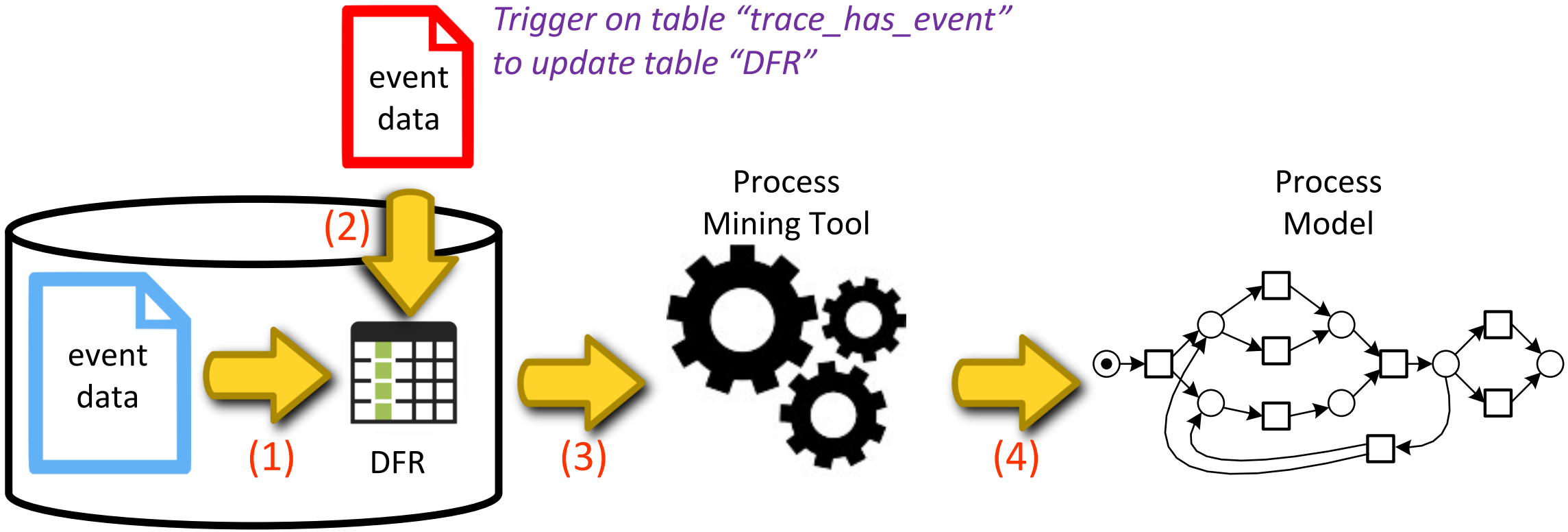


(1) Initialize

(2) Update

(3) Get Directly Follows Relation (DFR)

Database with DFR



(1) Initialize

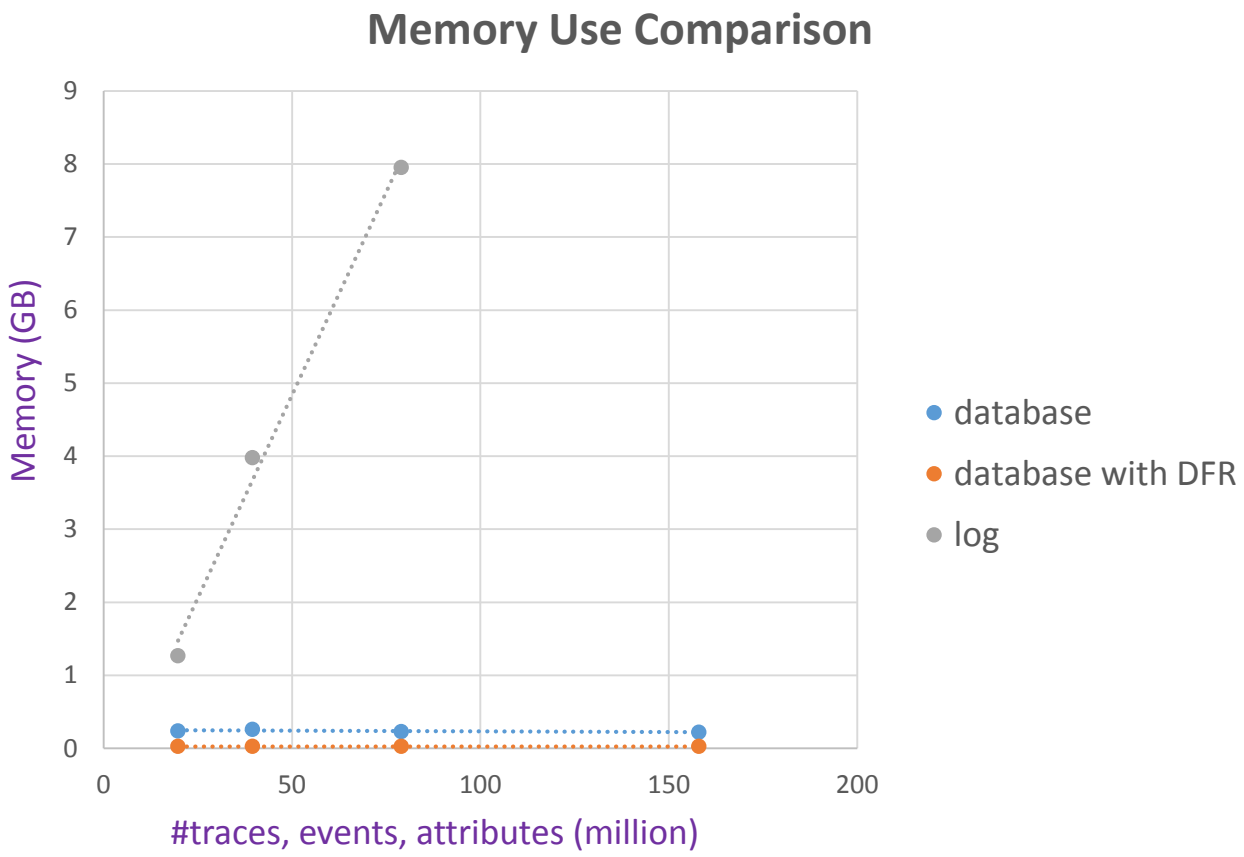
(2) Update

(3) Get Directly Follows Relation (DFR)

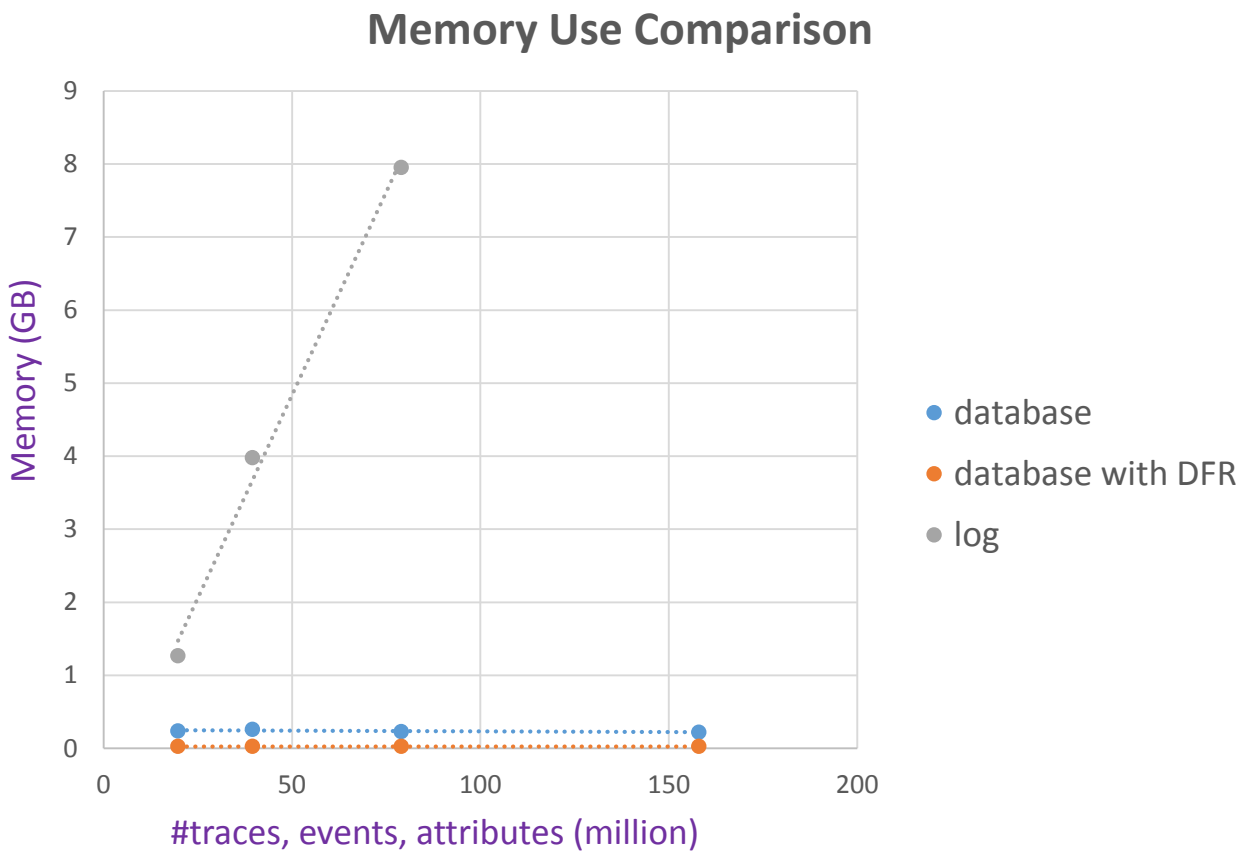
(4) Get process model

Experiment

Experiment

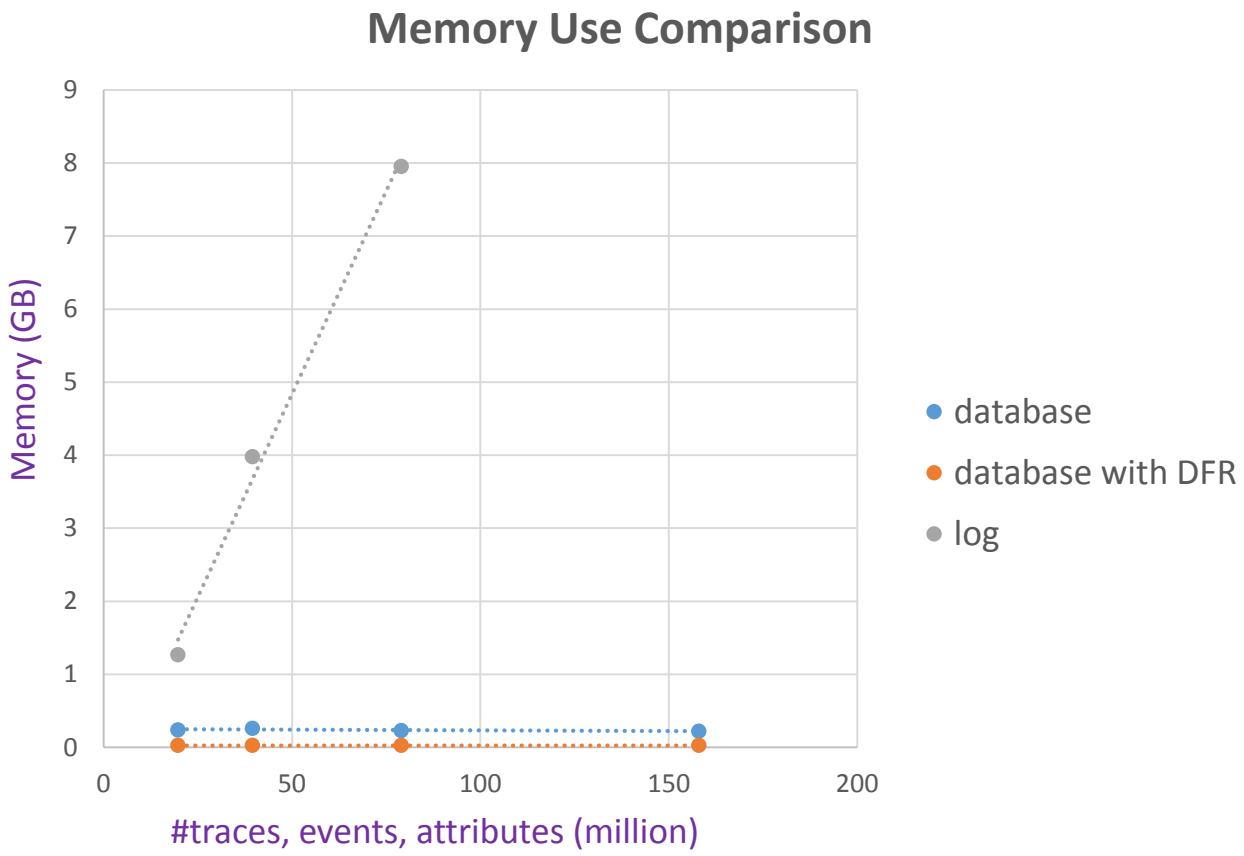

Memory Use Comparison
Memory (GB)
9
8
7
6
5
4
3
2
1
0
0
50
100
150
200
#traces, events, attributes (million)
database
database with DFR
log

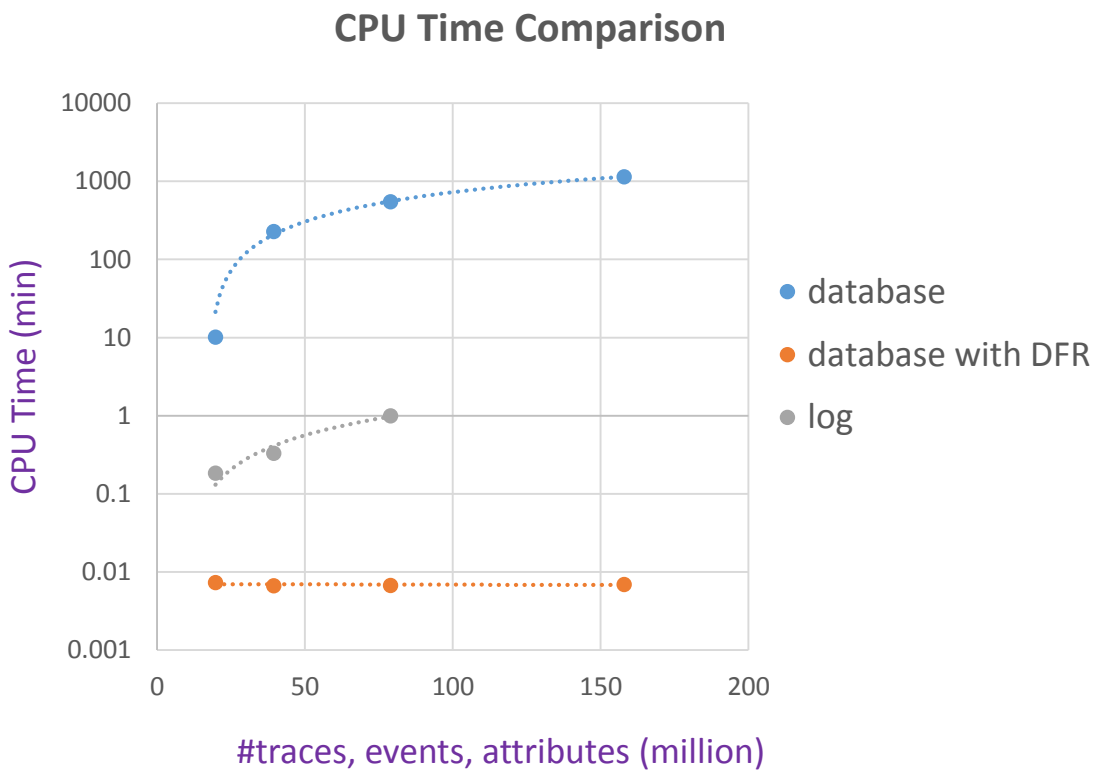
Experiment



"Database approach with DFR uses less memory compared to others"

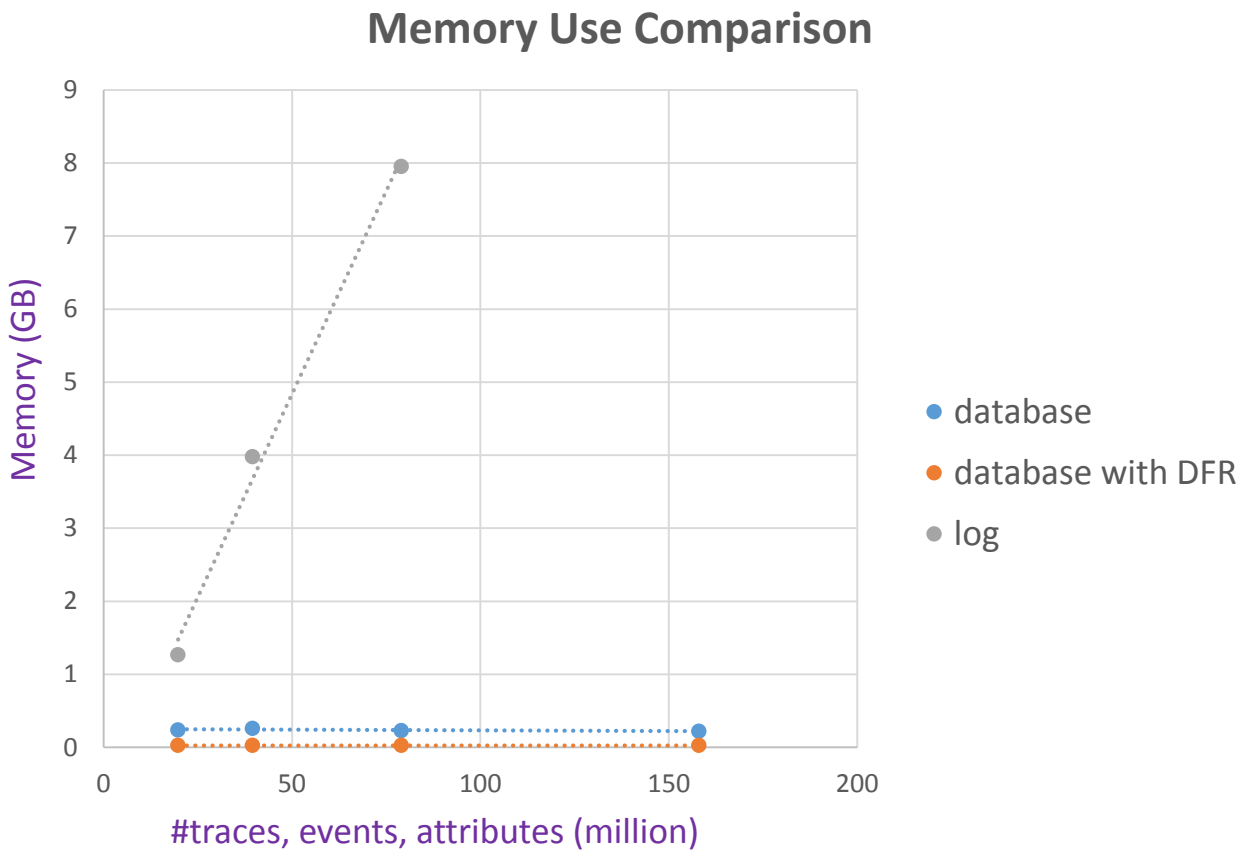
Experiment

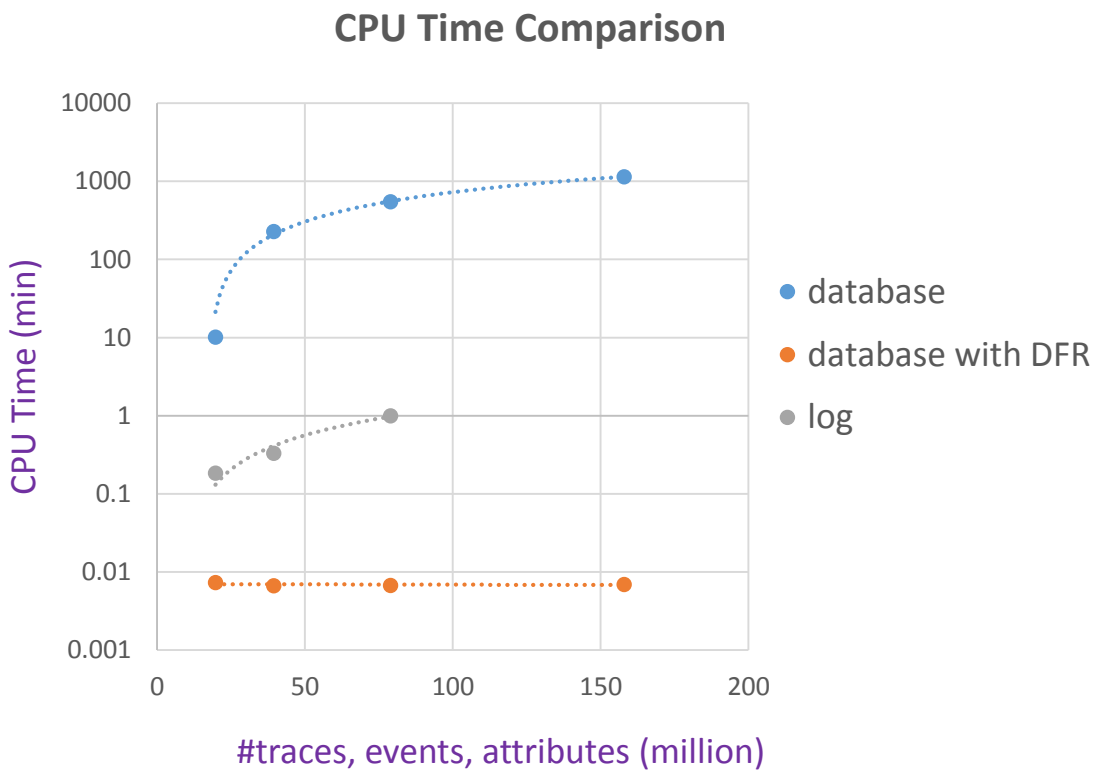




"Database approach with DFR uses less memory compared to others"

Experiment





"Database approach with DFR uses less memory compared to others"

"Database approach with DFR needs less CPU time compared to others"

“The use of
Relational Databases
to *pre-process event data*
can *accelerate process mining*
while maintaining its scalability”

Thank you

Alifah Syamsiyah
Email: a.syamsiyah@tue.nl
Project website: https://www.win.tue.nl/ais/doku.php?id=research:projects:delibida

"The use of
Relational Databases
to *pre-process event data*
can *accelerate process mining*
while maintaining its scalability"