

Mining Configurable Process Models from Collections of Event Logs

Joos Buijs

Boudewijn van Dongen

Wil van der Aalst

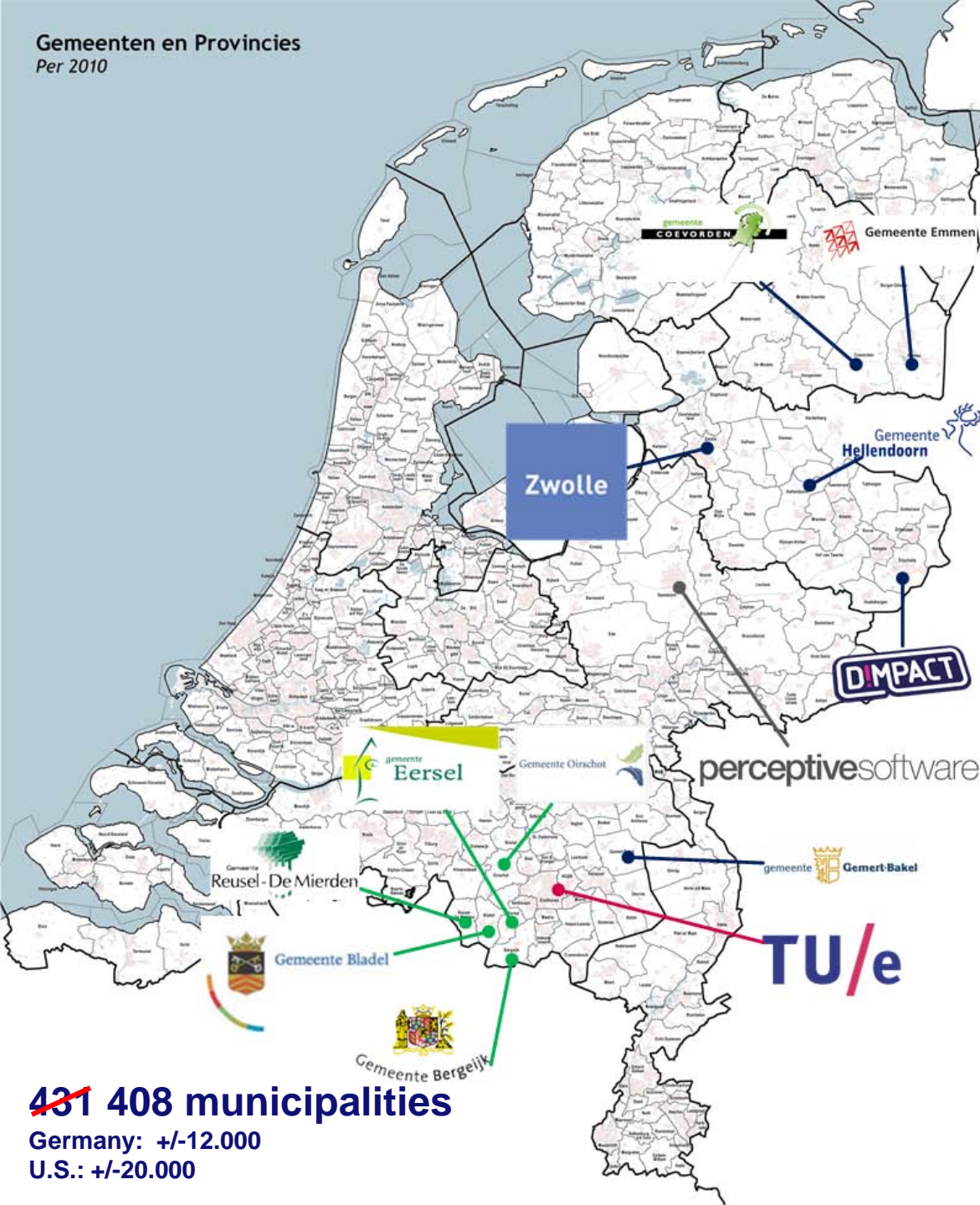
Conditionally accepted for BPM 2013



TU/e

Technische Universiteit
Eindhoven
University of Technology

Where innovation starts



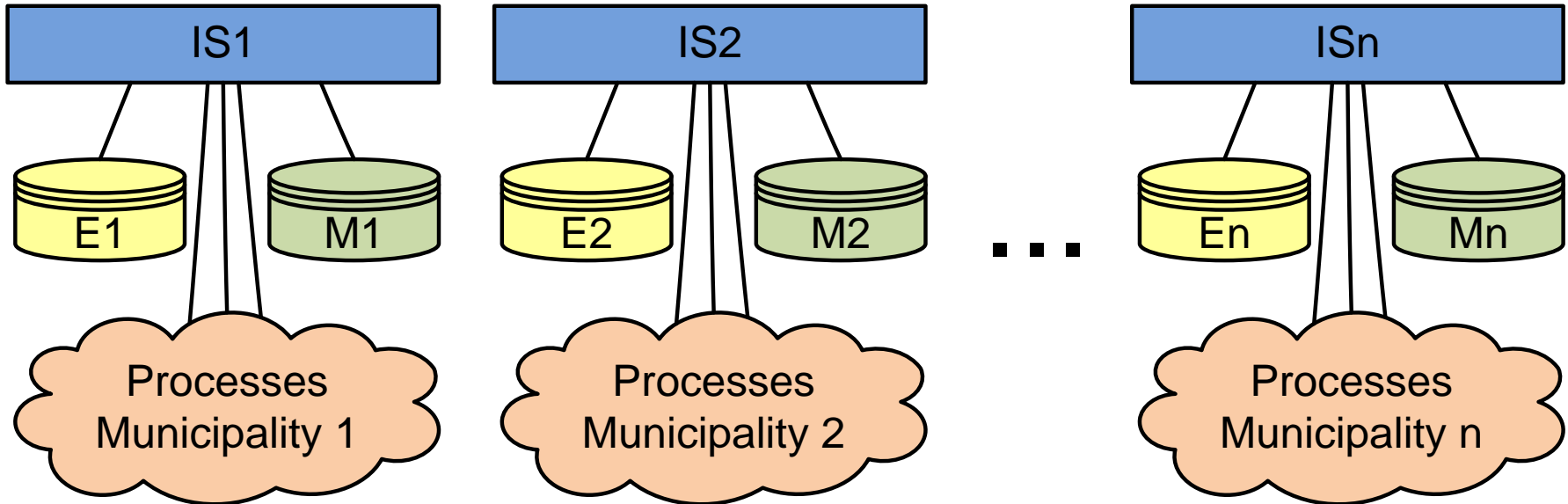
~~431~~ 408 municipalities

Germany: +/-12.000
U.S.: +/-20.000



Software
as a
Service
(SaaS)

Traditional Situation

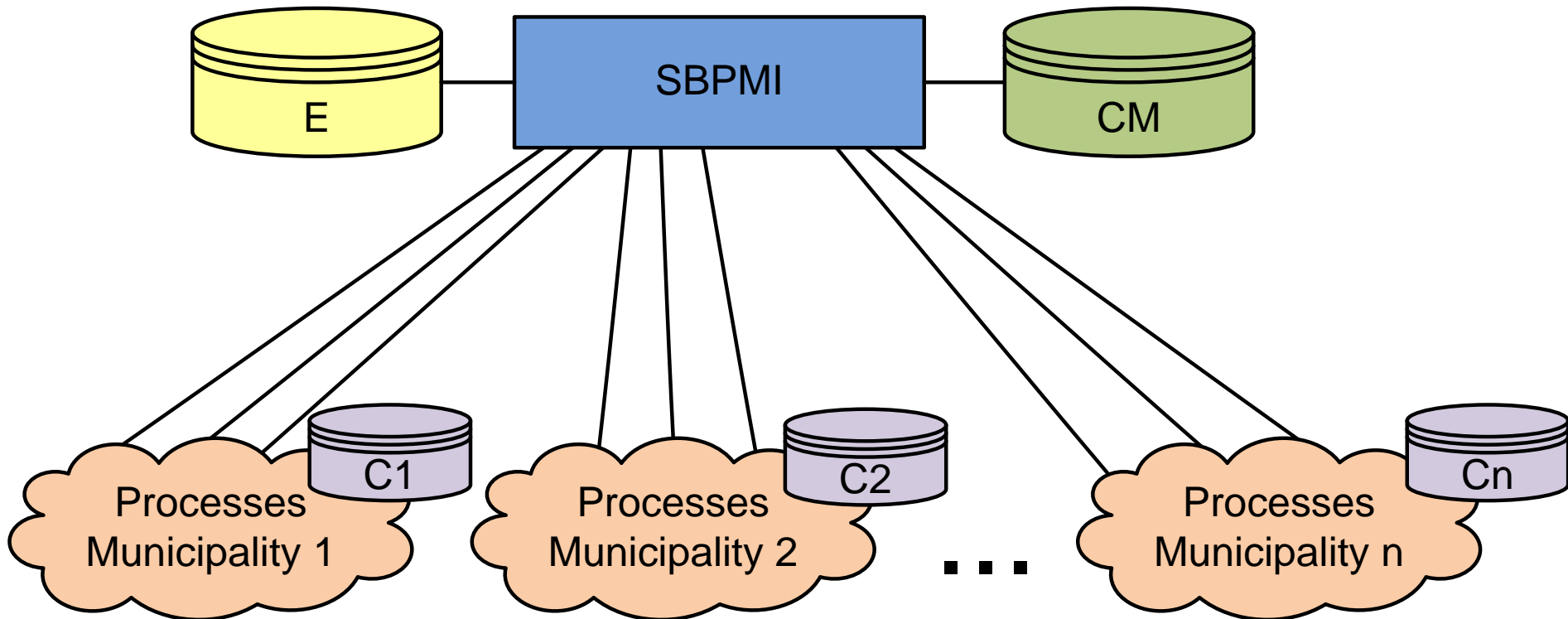


IS = Information System

E = Event log

M = Models

Our Vision



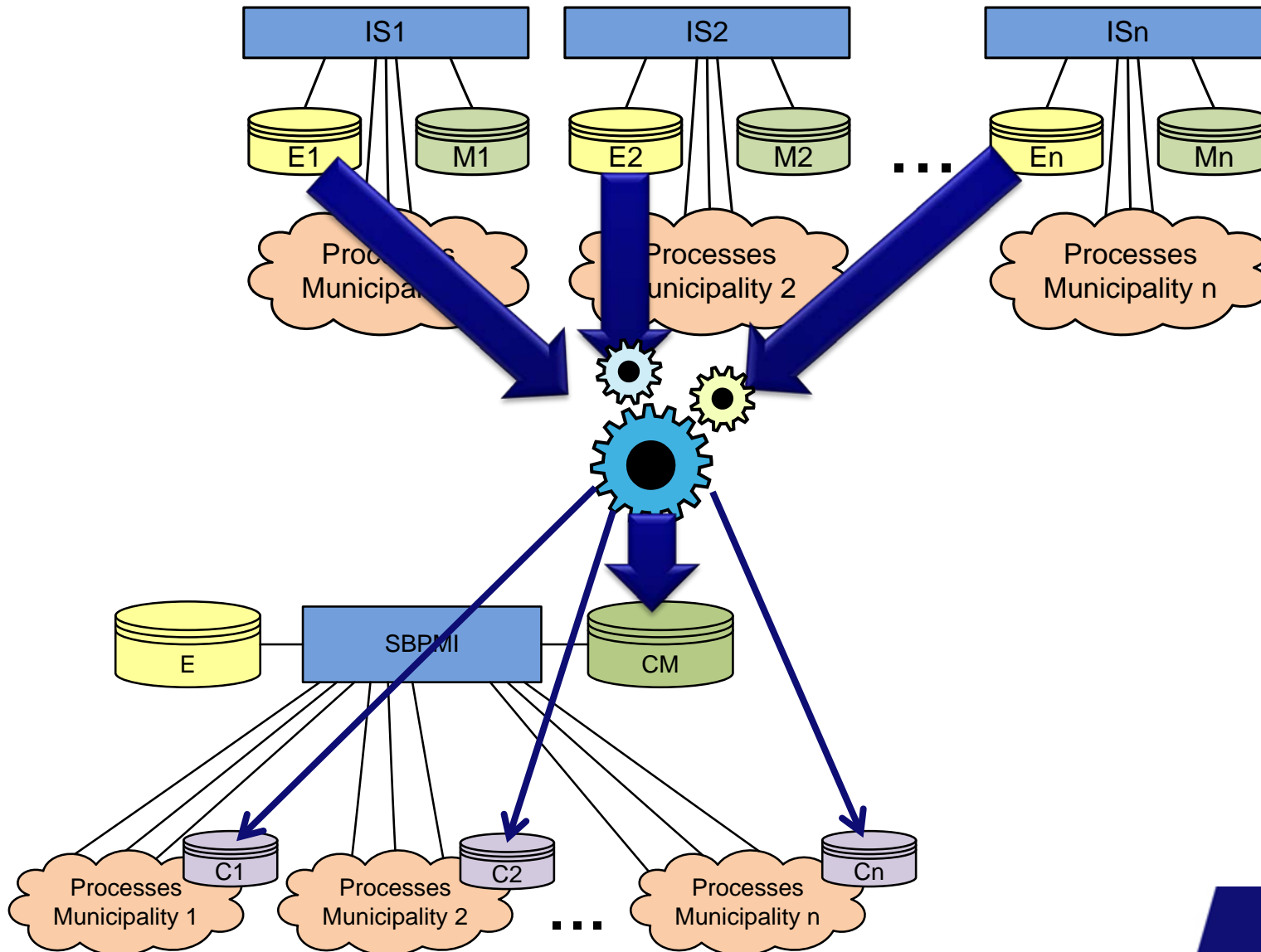
SBPMI = Shared Business Process Management Infrastructure

E = Event log

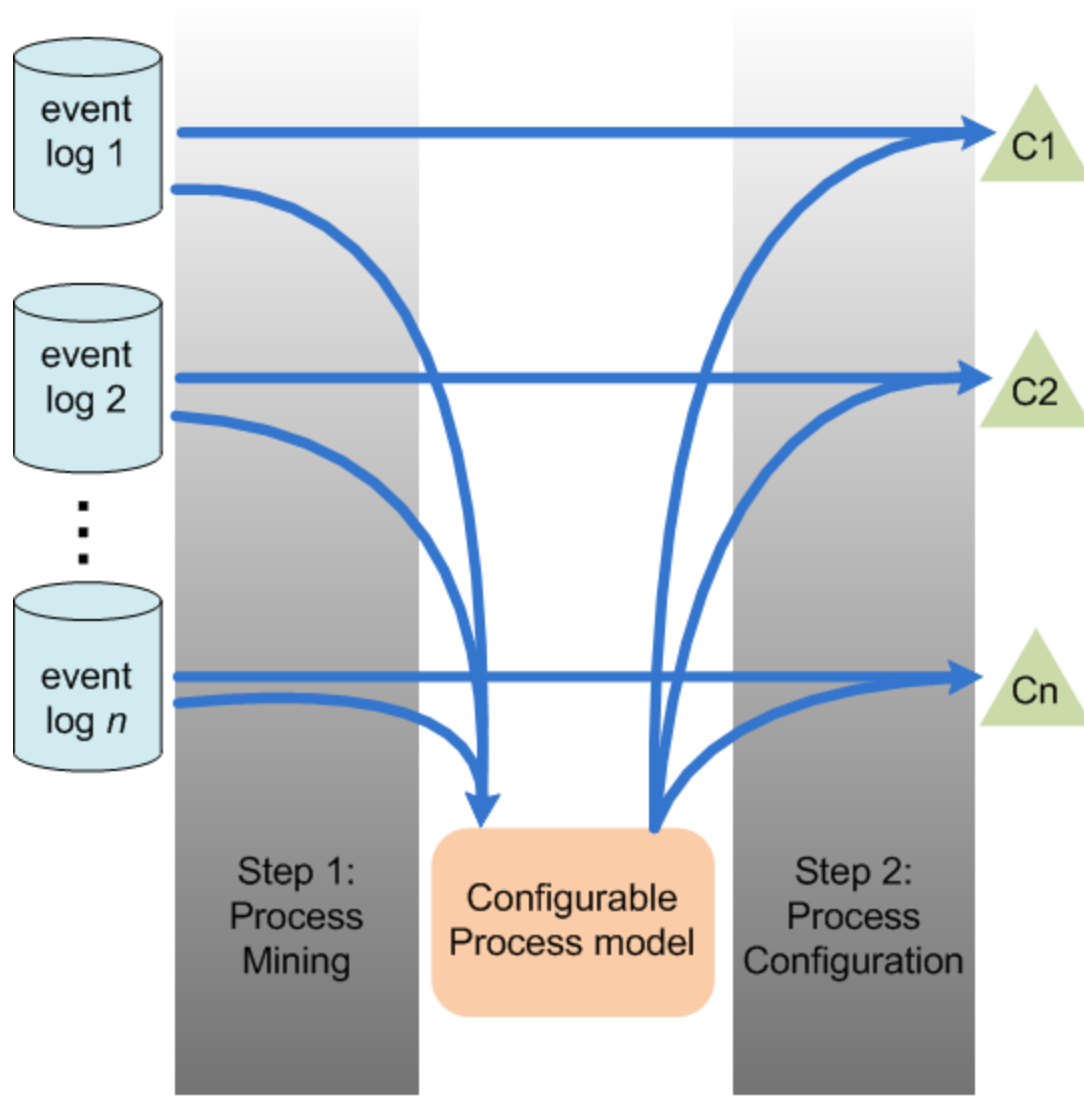
CM = Configurable Models

C = Configuration

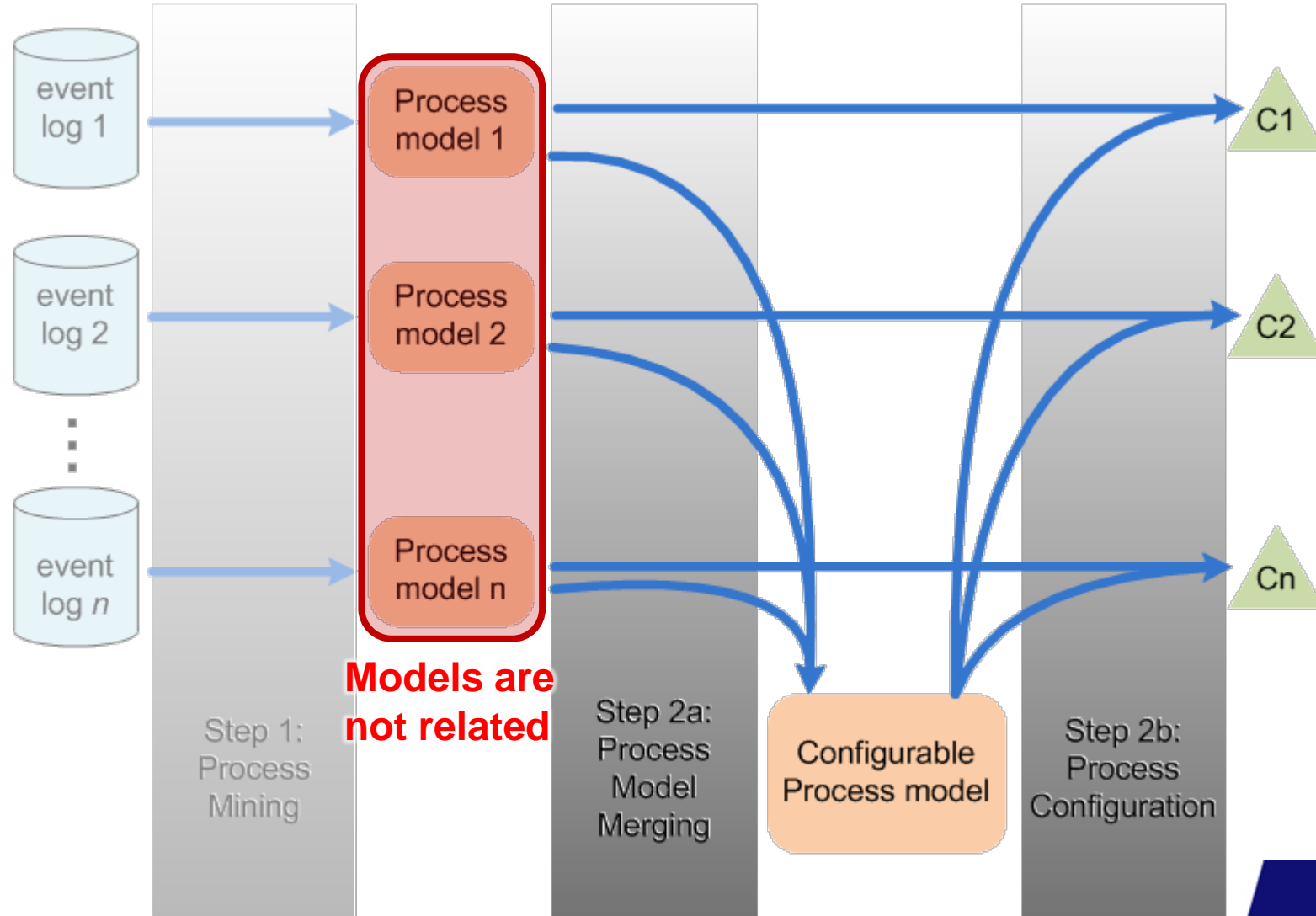
From Old to New



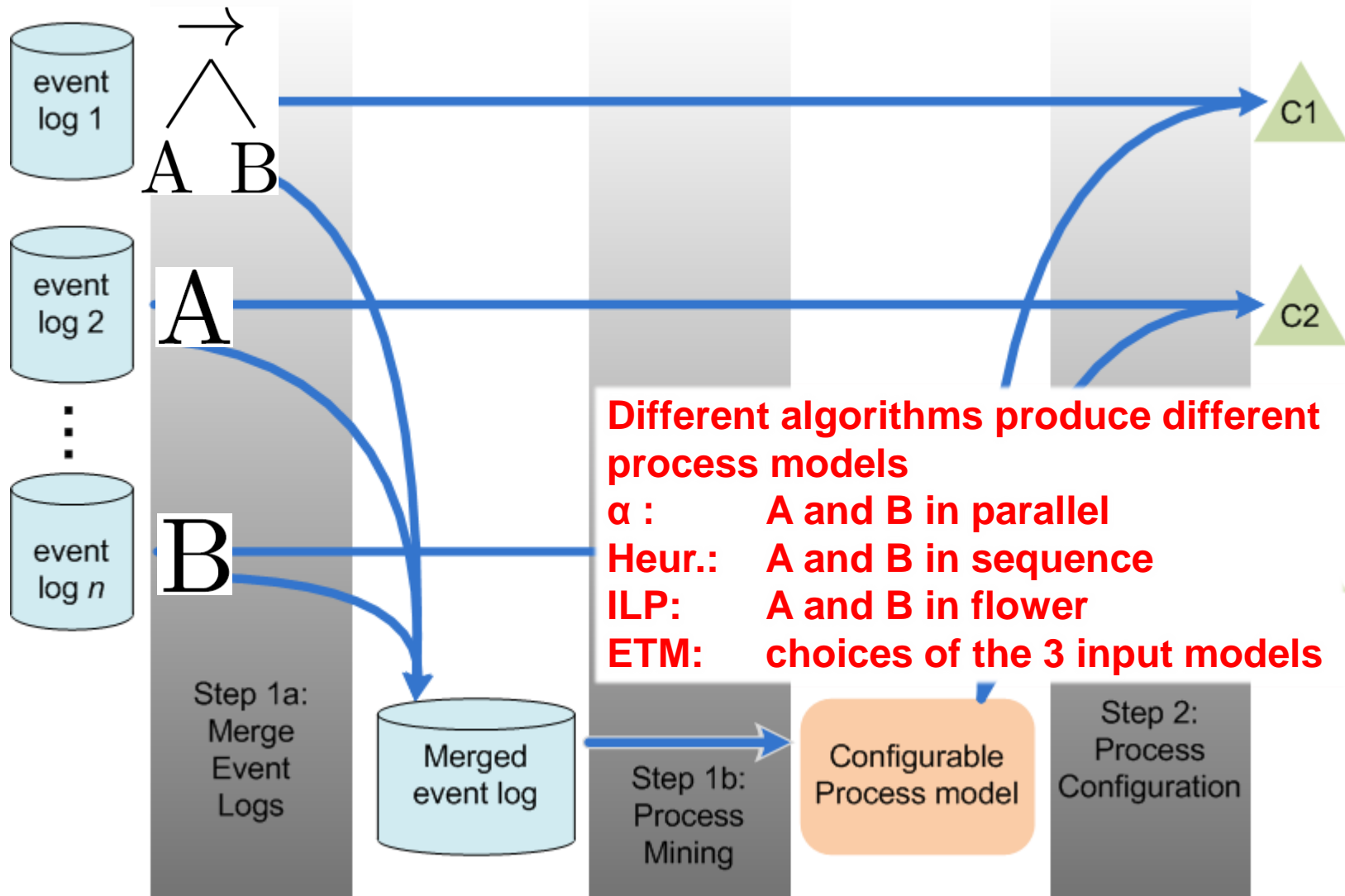
General Idea



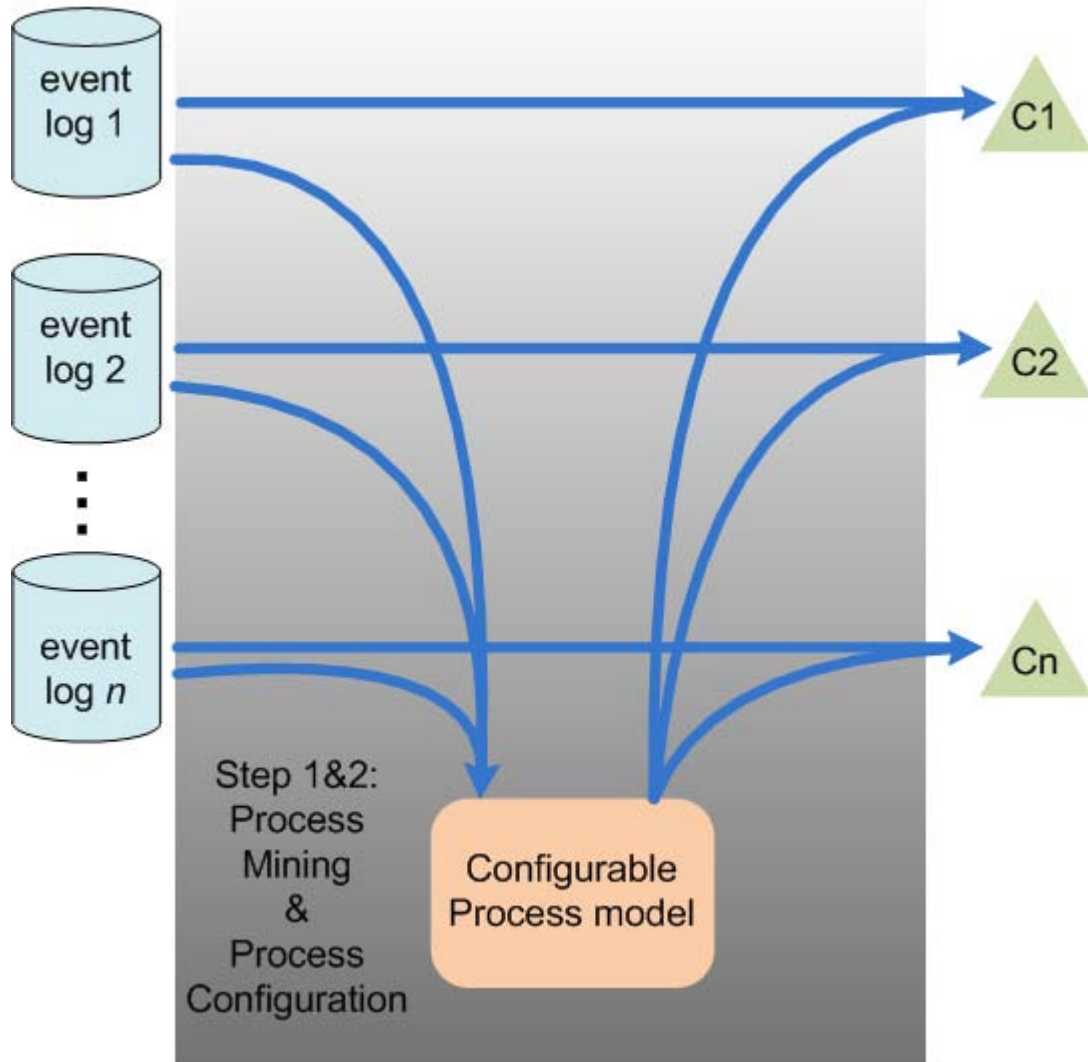
Issue with Model Merge Approach



Issue with Log Merge Approach

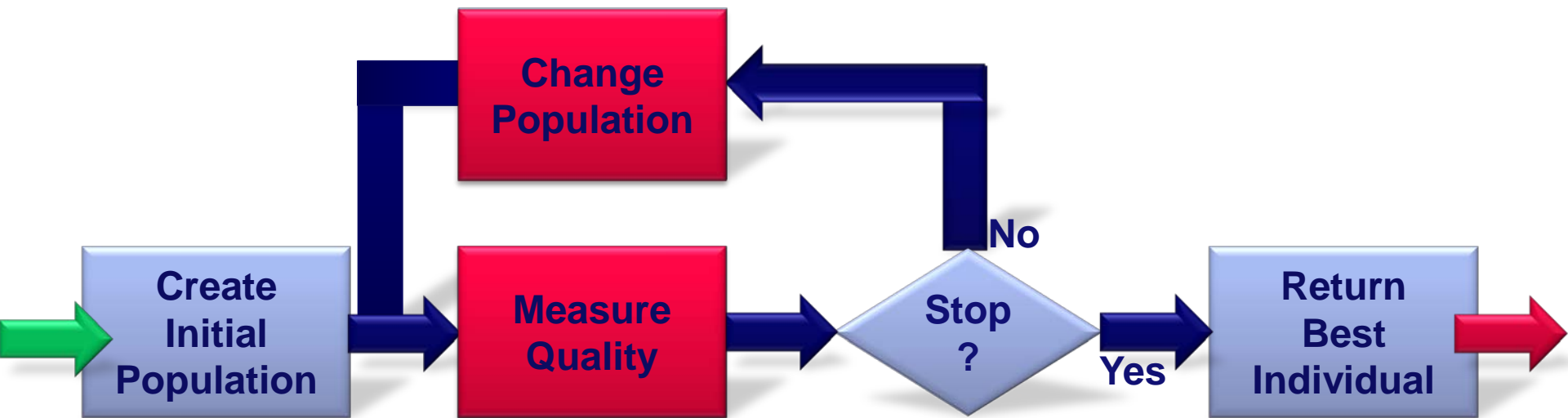


Our Proposal: Integrate Configuration Aspect in Discovery

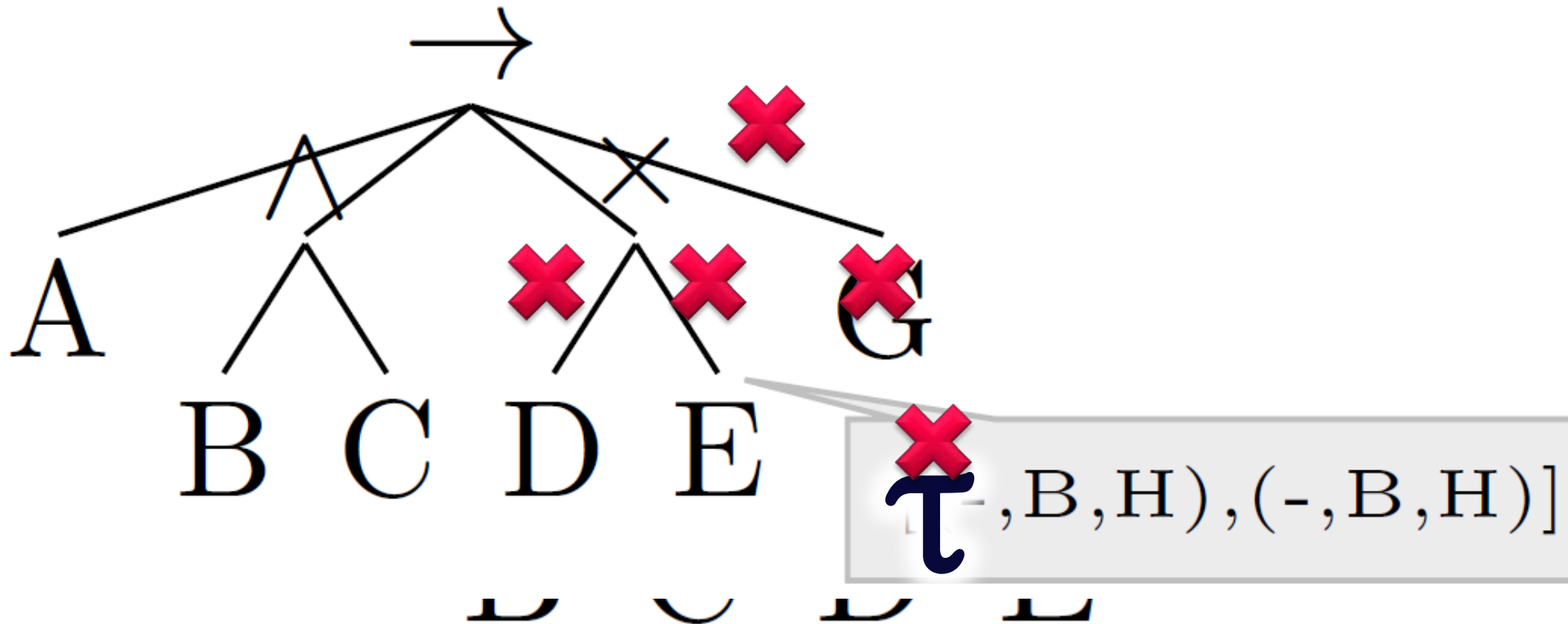


Extending the ETM Algorithm

- **Configurable Process Trees**



Configurable Process Trees

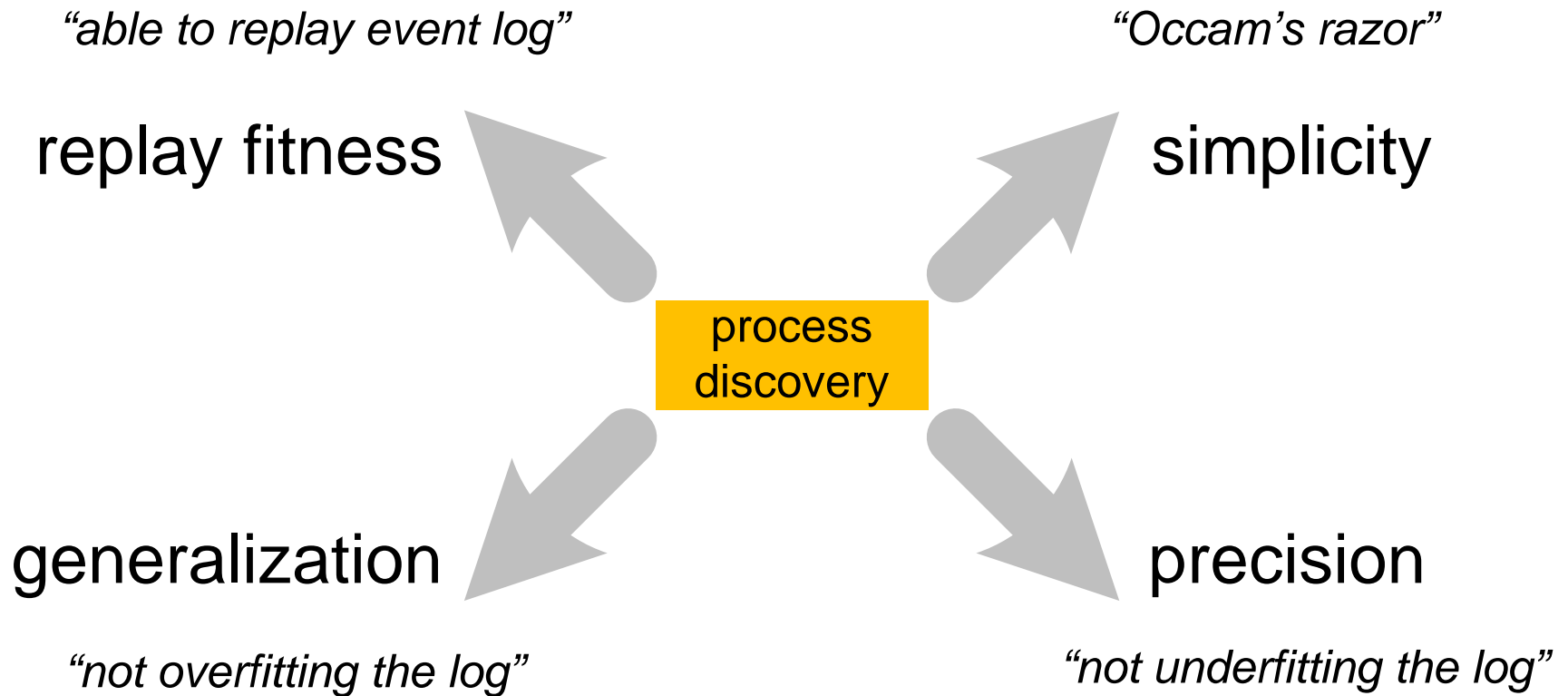


- Configuration points with options that can be set
- Hiding
- Blocking
- Operator Downgrade

Mutation of Configurations

- **All nodes in the tree are potentially configurable with all options enabled but not set**
- **(Randomly) set a configuration option for a specific node and configuration**

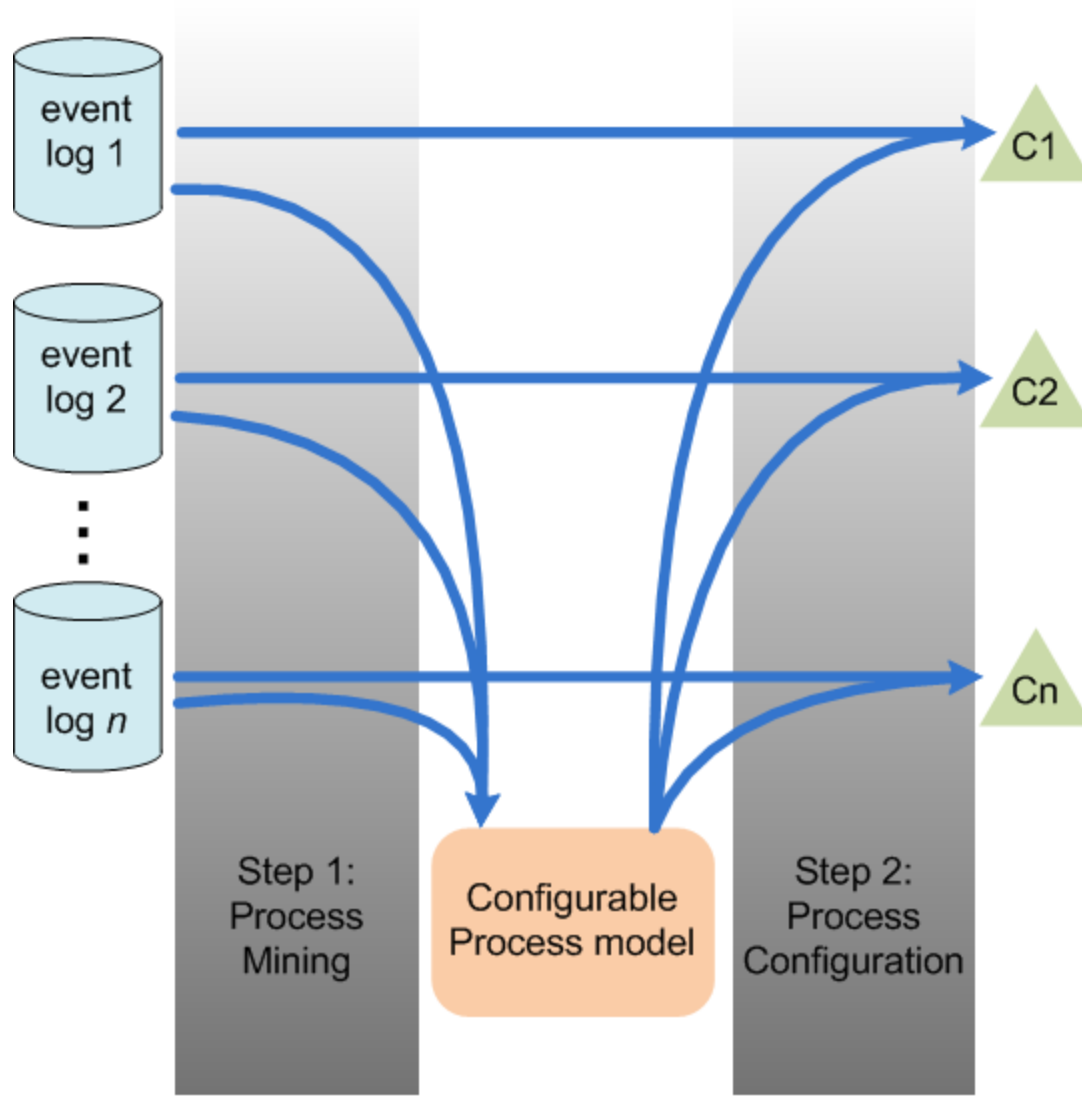
Process Discovery Quality



Evaluation of Configurations

- 1. Apply configuration settings for each individual log to obtain the individual process tree**
- 2. Calculate the overall fitness per log-process tree pair**
- 3. Average these fitness values, weighing them using the number of traces in the log**
- 4. Count the number of used configuration points (less is better)**

General Idea



Running Example: 4 Variants

1

#	Trace
38	A B C D F G
26	A B D C F G
12	A B D C E G
8	A B C F G
6	A B C D E G
4	A D C B F G
2	A C D B F G
1	A D B C F G
1	A D B C E G
1	A C B E G
1	A C B F G

B split in 2
Other D variant
Sequential

No D and G

B,C,D in parallel
D can be skipped

B split in 2
Both D variants
No G

2

#	Trace
20	A B1 B2 C D2 E G
50	A B1 B2 C D2 F G

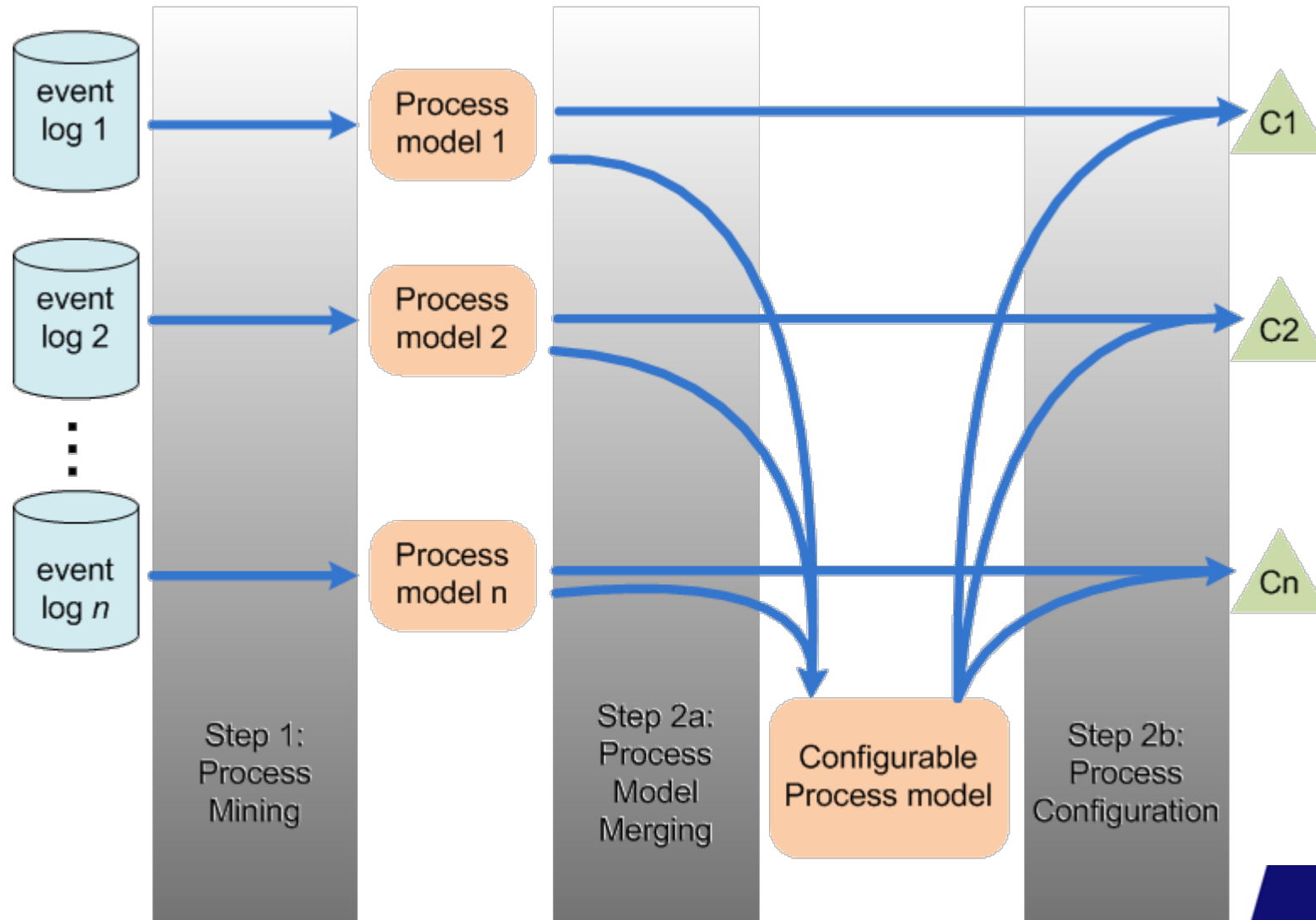
3

#	Trace
120	A C B E
80	A C B F

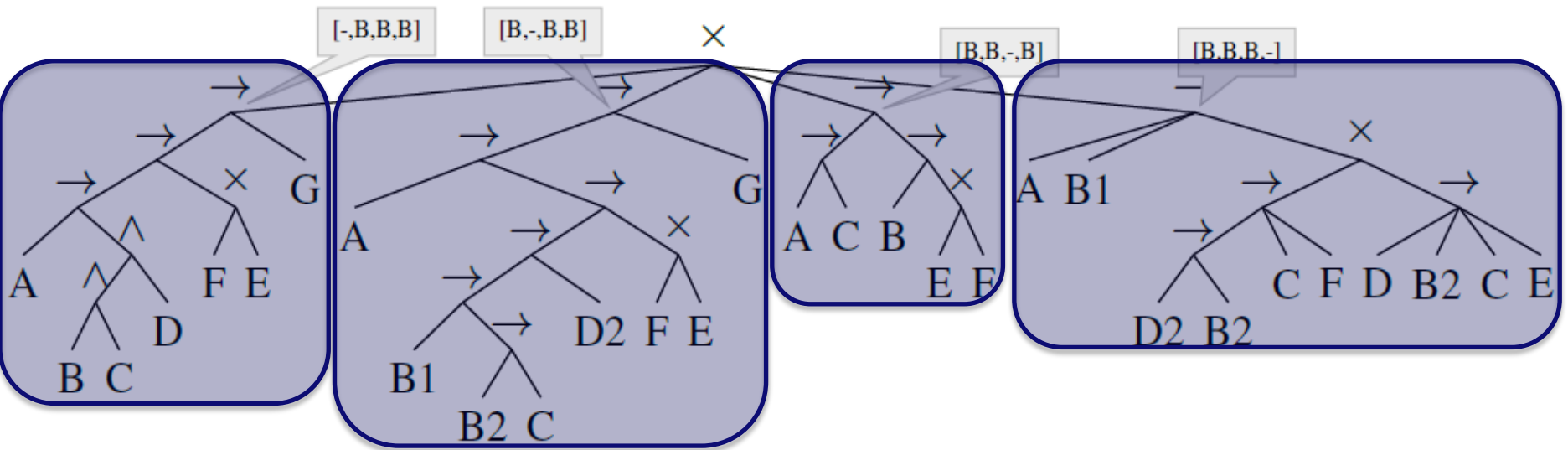
4

#	Trace
20	A B1 D B2 C E
50	A B1 D2 B2 C F

Approach 1 (Gottschalk et. al.)

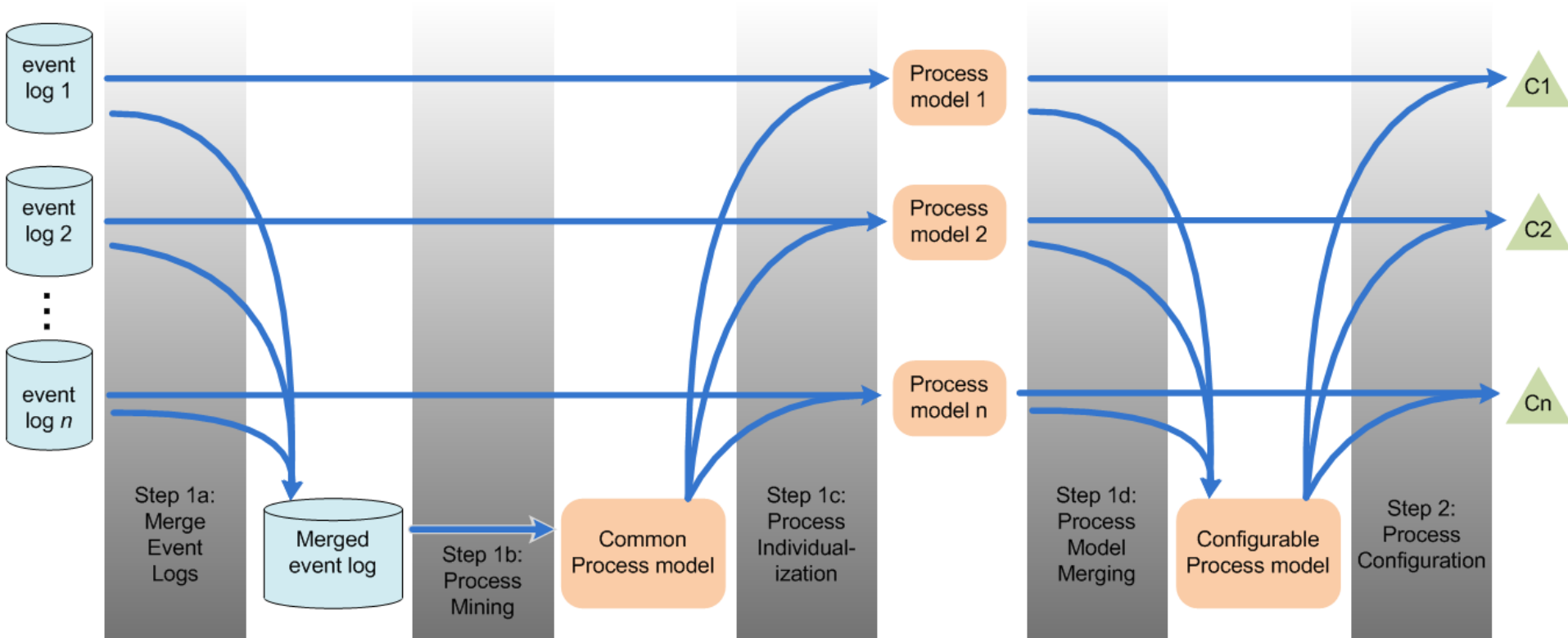


Result Approach 1

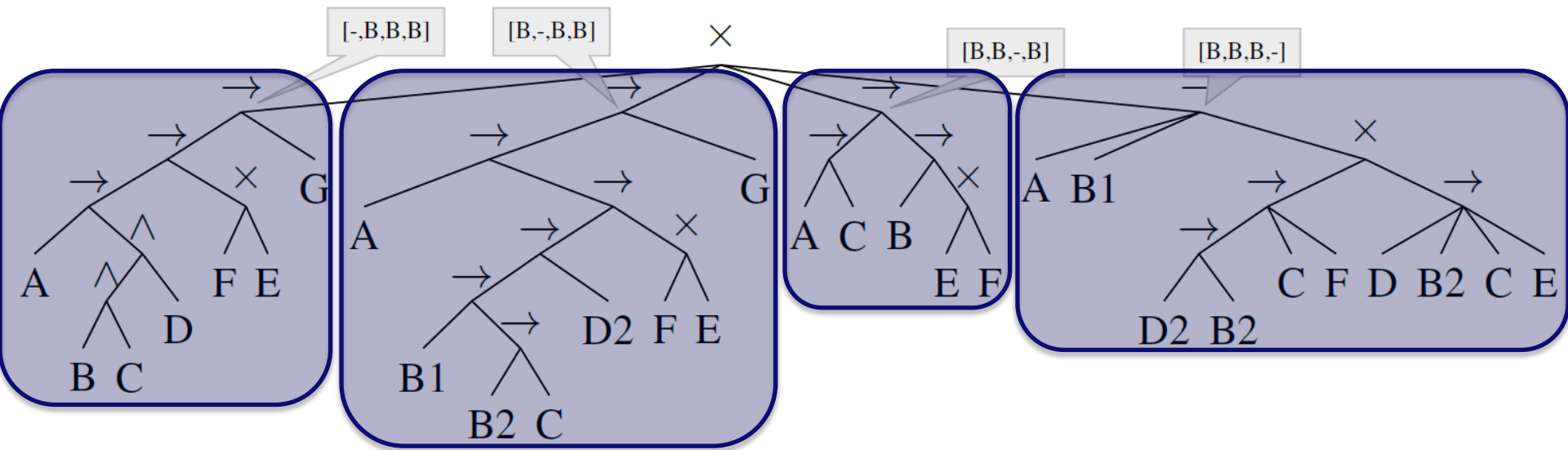


**This is not the idea of a configurable process model:
Common parts should be shared**

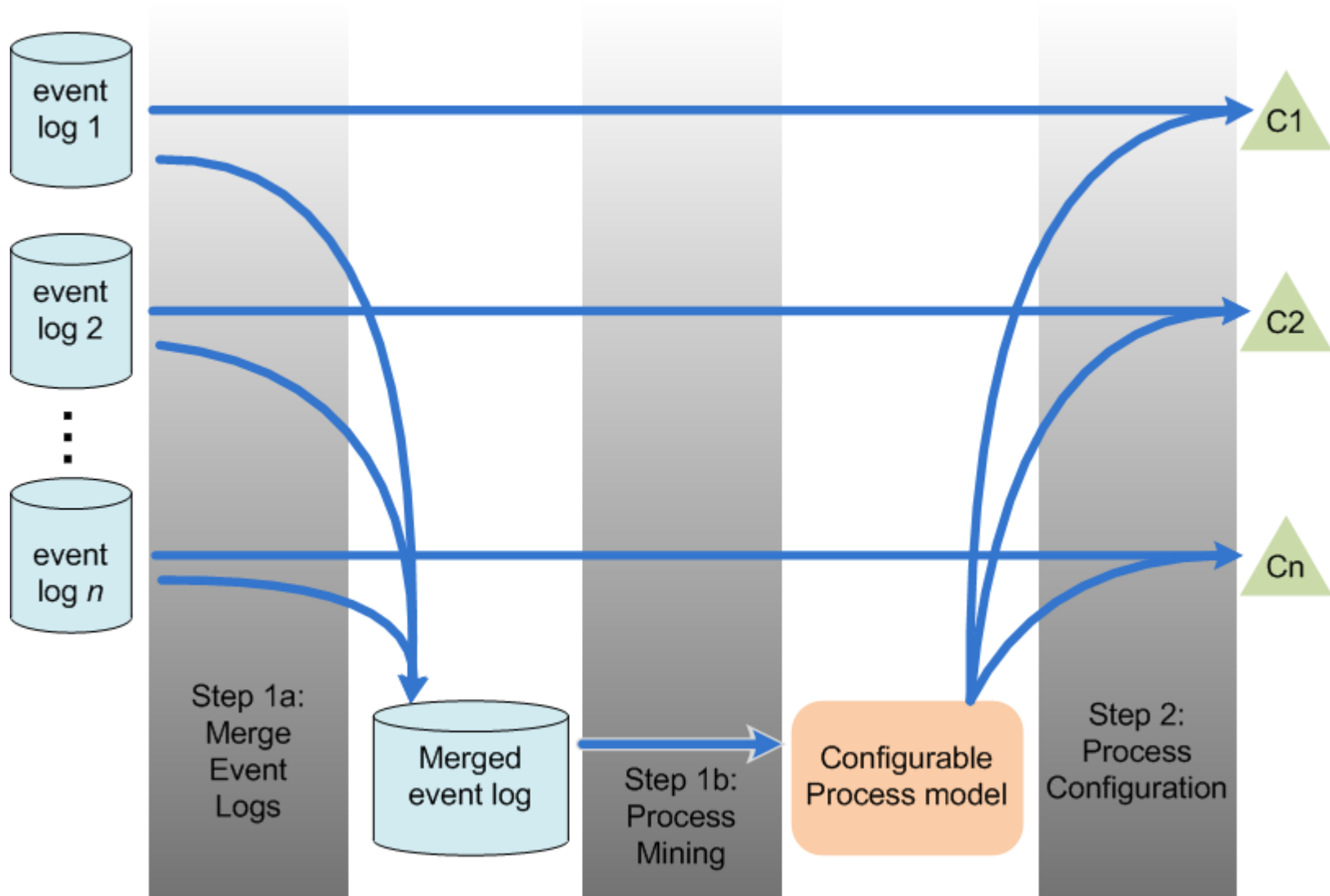
Approach 2



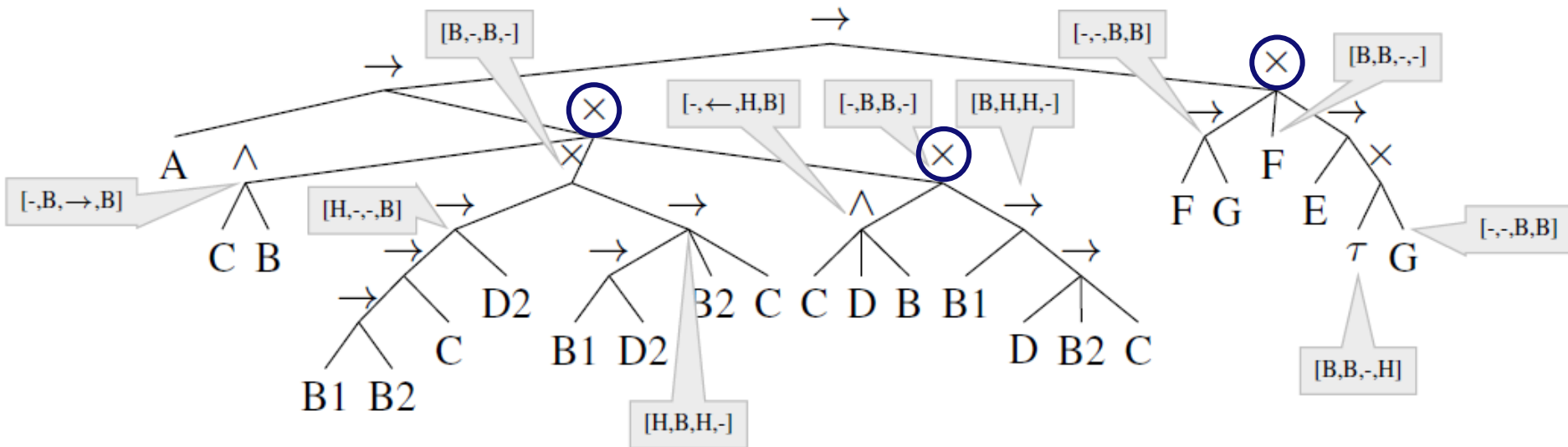
Result Approach 2



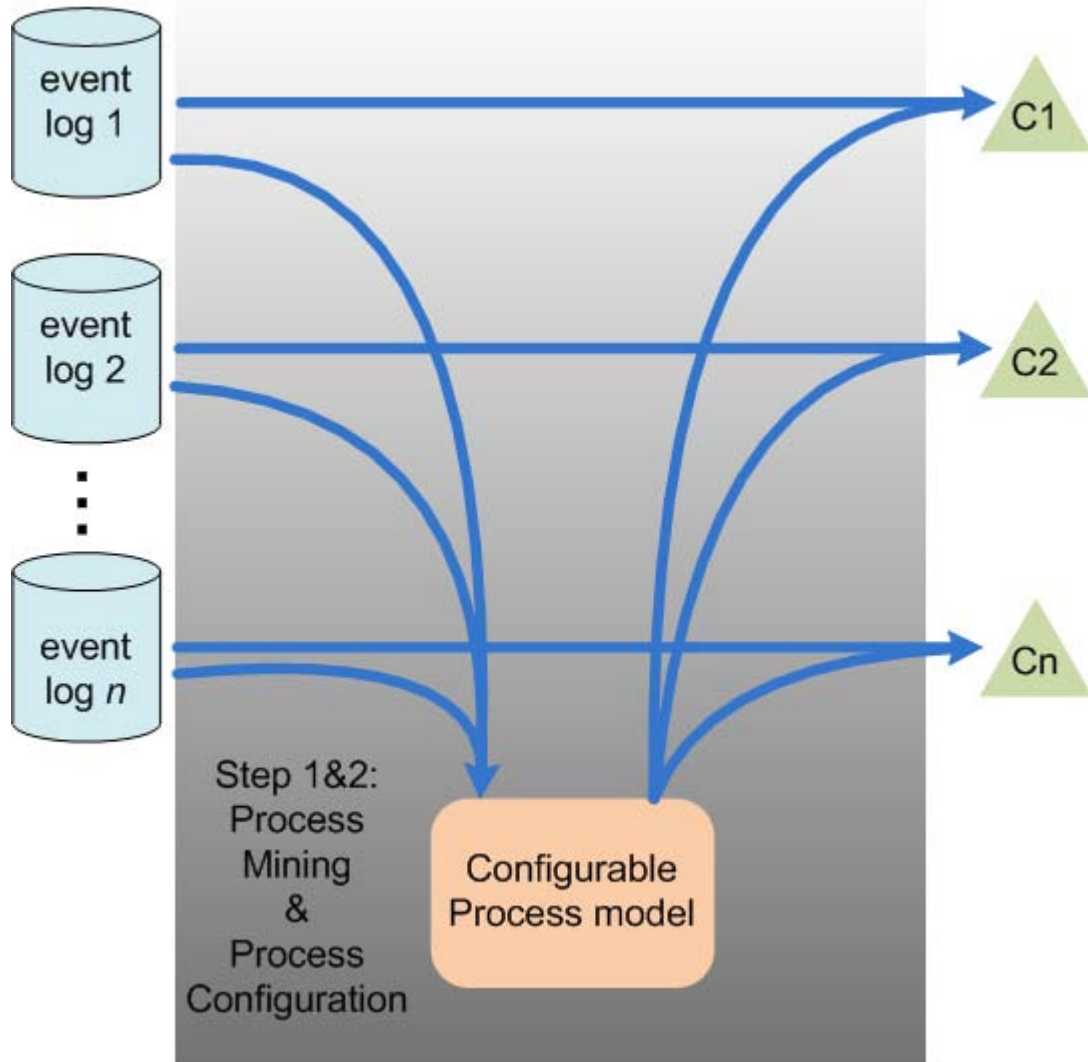
Approach 3 (Gottschalk et. al.)



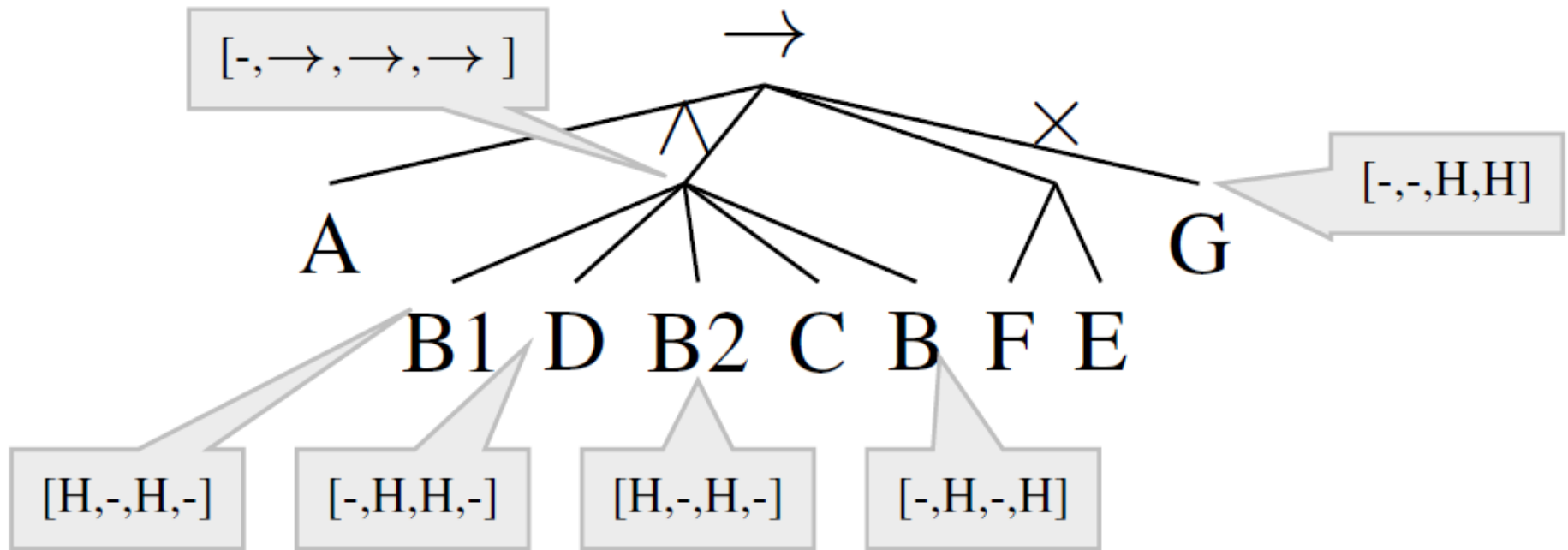
Approach 3: Result



Approach 4



Approach 4: Result



Conclusion

