# Accelerating Process Mining using Relational Databases

Alifah Syamsiyah, Boudewijn F. van Dongen, Wil M.P. van der Aalst

Eindhoven University of Technology
A.Syamsiyah@tue.nl, B.F.v.Dongen@tue.nl, W.M.P.v.d.Aalst@tue.nl

The spectacular growth of data creates a new challenge to turn the data into meaningful value. Process mining is a research discipline that bridges data science and process science whose the goal is to turn event data into insights and actions in order to improve processes. However, once event data files get bigger, scalability quickly drops since the speed required to access the data becomes a limiting factor. This work proposes a new technique based on relational databases to accelerate process mining analysis while maintain its scalability. To this end, we introduce DB-XES as a database schema which resembles the XES structure for storing and pre-processing the event data. We provide a transparent way to access event data stored in DB-XES and we show how this greatly improves on the memory requirements of a state-of-the-art process discovery technique. Moreover, we show how to move the computation of intermediate data structures, such as the directly follows relation, to the database engine, to reduce the time required during process mining. The work is implemented in ProM, and a range of experiments demonstrates the feasibility of our approach.