MSc thesis internship – Resource Usage Computing Platform

Topic description
Anomaly detection and optimal planning of resources for computing platform

The extreme requirements on quality of service (process times, correctness, etc.), combined the large number of transactions that ING handles puts a unique requirement on its infrastructure. In order to handle these loads, ING uses specialized hardware. This hardware is rented from a provider and is paid for based on monthly usage. Usage is expressed in terms of the highest four-hour average CPU time usage. This puts the responsibility for spreading load into the users' hands. On top of this, various types of jobs are run on this system. Job runtimes may vary from microseconds up to multiple hours.

There are multiple directions for research, which are described in detail below:
1. Detecting anomalous jobs and runs on data centre
2. Finding/Learning an optimal job cancellation policy
3. Finding/Learning an optimal schedule for jobs

Detecting anomalous jobs and runs on data centre
Finding which jobs are behaving differently from ‘usual’ could give insight on jobs that require some kind of maintenance. Jobs that take significantly more time than other jobs are suspect to programming mistakes and similarly so for runs of a job when compared to previous runs of this job. This problem can possibly solved using known statistical analyses and anomaly detection methods thus this formulation of the project is fits a Bachelor’s thesis or small paper.

Finding/Learning an optimal job cancellation policy
Some jobs are cancelled by an operator due to running too long. This problem is to find an optimal policies for cancelling long-running jobs. The scenario is such that any unfairly cancelled job (e.g. that would have successfully finished) should be re-run, resulting in CPU time waste. This can possibly be formulated as a multi-armed bandit problem and solved using simulation. An application can be found in a system that cancels jobs or notifies operators (including the job author) and thus reduces costs.

Finding/Learning an optimal schedule for jobs
Runs of a jobs have varying runtimes that may be only partially predictably. Furthermore, novel jobs may be introduced or jobs may be altered so that runtimes can vary over times. However, considering the way ING is billed, an optimal schedule may be approximated using Reinforcement Learning. This problem consists of both predicting job runtimes and handling uncertainly in planning them. On top of this, we could think of additional requirements to runtime e.g. some jobs may have higher priority, may need to be run before other jobs or there may be a predefined ordering of jobs.
About ING
ING is transforming rapidly towards a top notch IT company with a Top Engineering culture, incorporating FinTech start-up mentality and financial innovation. We are end-to-end agile, work in multidisciplinary teams and use cutting edge technology. This enables us to offer our customers the best possible service and respond to rapidly changing customer demands.

You will be working in the ING’s Core Banking University, this is where science and business meet. Within this department top experts from both areas are collaborating on the next-generation digital bank.

Requirements
We are looking for a Master student that can meet the following requirements:

- Bachelor degree in a data science-related field, with great results. Affinity with topics stochastic modeling and optimization, scheduling, statistical models and machine learning.
- The ability to learn quickly and explore new topics
- Affinity with a cooperate research environment
- Data science experience (e.g. projects or working experience besides your study)
- Fluent in English

Interested?
Please get in contact with us: send your CV and grade lists (BSc and MSc so far). If you have any other topic ideas, you can send them as well. Please be aware that a Certificate of Good Conduct and positive Pre-Employment screening are required.

Joost Bosman, IT Manager, ING: joost.bosman@ing.nl