
Time Cost Tradeoff of Pipelined Dataflow Applications

 rik Saule*¹

¹Department of Computer Science [University of North Carolina at Charlotte] (UNCC) – United States

Abstract

In the era of cloud computing and the exascale, high performance computing is no longer about finishing an application as fast as possible. Time is certainly of the essence, but other objectives such as the cost of a computation, the energy it uses or the maximum amount of power it can draw are just as much of a concern. The pipelined data flow programming model has proved to be popular for analyzing large or streaming datasets. We argue that the pipelined data flow model is suited to adapt to different objectives. We use one particular histopathology application as a case study. The simple model of pipelined applications assumes a predictable work volume and a smooth execution of that work. We show that in our application it is easy to predict the amount of work, but its execution is not smooth, as work is released from one stage to the next in bursts. However, simple changes in the application allow us to smooth the release of work over time which makes the execution faster and easier to model. With an accurate model of the application’s execution for a particular configuration, we are able to express the tradeoff between the execution cost of our analysis on a cloud platform and the time it takes to perform the analysis.

*Speaker