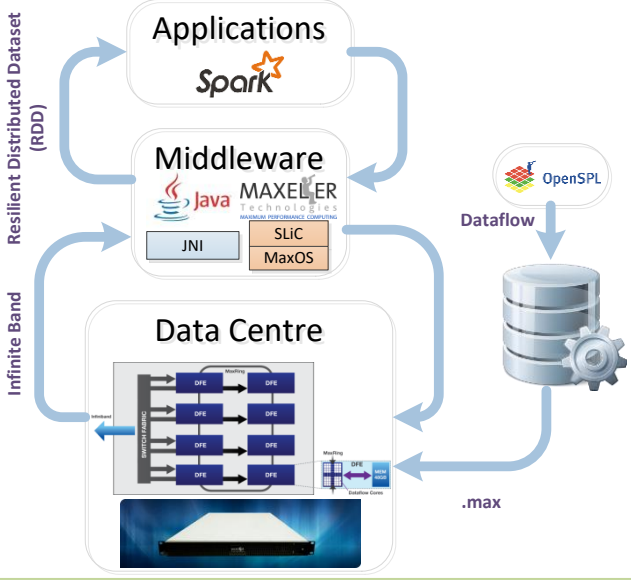
Queen's University Belfast

Accelerating Data Centre Applications with Reconfigurable Dataflow Engines

Introductions

Energy-efficient, programmable accelerators are an important value proposition for pushing the limits on the computation capacity and density of future data centers. Integrating the accelerators in the software stacks of cloud-based data analytics frameworks is, however, an open issue. The goal of this work is to achieve seamless integration of reconfigurable accelerators, Dataflow Engine (DFE), in data analytics programming frameworks.

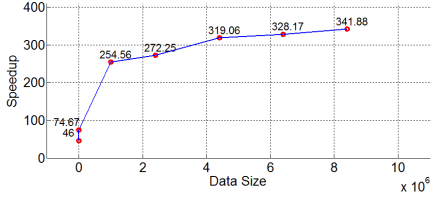
Framework on DFE

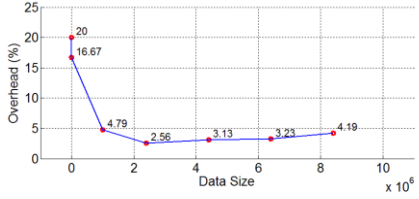


Experiment

❖ Linear Regression:

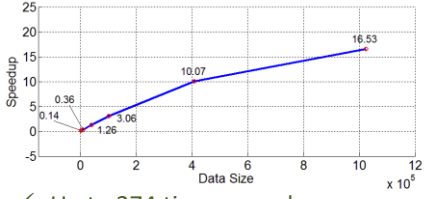
$$a = \bar{y} - b \cdot \bar{x} \quad and \quad b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

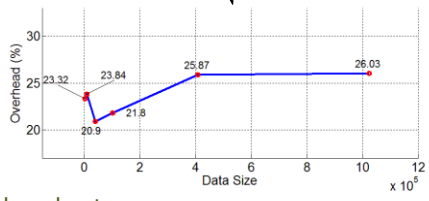




❖ Classification:

$$d(p,q) = d(q,p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$$





✓ Up to 374 times speedup compared Spark on host

✓ Down to 3% middleware overhead by executing kernel on DFE

Conclusion

In this work, Spark is seamlessly integrated with Dataflow Engines on a reconfigurable device. As the data size scales up, the speedup is considerable compared to executing the kernel in Spark on the host CPU. Integrating Spark with DFEs induces overhead compared to the C-based data streaming API and requires further optimization.

Sakil Barbhuiya, Yun Wu, Karen Murphy, Hans Vandierendonck, George Karakonstantis, Dimitrios S. Nikolopoulos

{sbarbhuiya01, yun.wu, k.l.murphy, h.vandierendonck, g.karakonstantis, d.nikolopoulos}@qub.ac.uk

VINEYARD