SAND2016-6736C

The Center for Cyber Defenders

Expanding computer security knowledge

Scalable, Portable, and Productive?

Using K-Means as an exemplar to examine Chapel's relevance to machine learning

Jacob Caswell, St. Olaf College

Project Mentors: Richard Barrett, Org. 5638; Tom Brounstein & Matt Stirling, Org. 5635

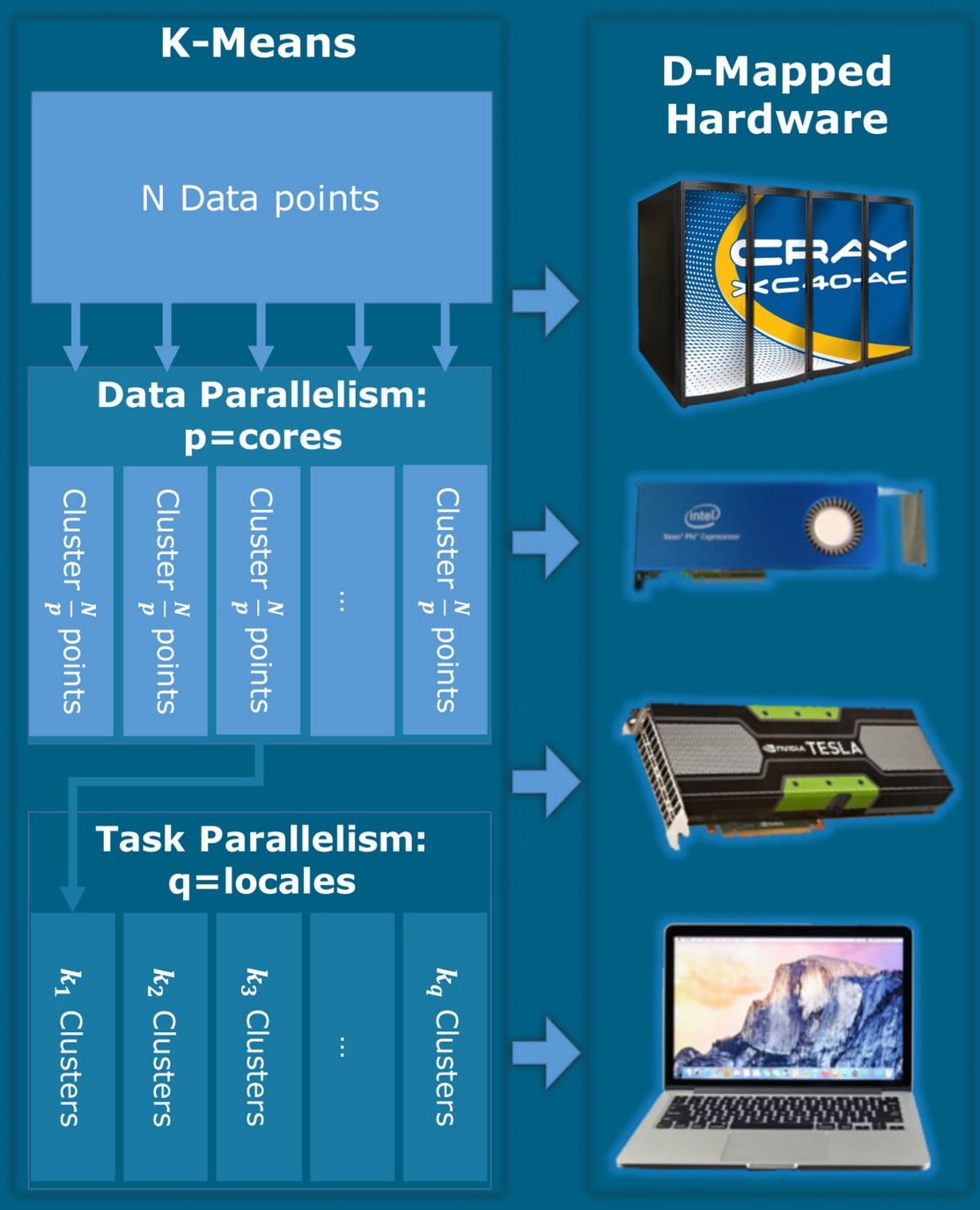
Research Statement:

Chapel is an open source parallel programming language initiated by DARPA to foster the productive use of high performance computing (HPC) resources. Our project investigates Chapel's design principles and resulting performance potential within the context of machine learning and text analysis. Specifically, we propose a k-means implementation designed to test Chapel's productivity and address the needs of the Analytics and Cryptography group at Sandia.

Objectives:

- Outline and prototype a platform for scalable computational social analytics and machine learning using Chapel
- Explore Chapel's productivity, which we define as a function of its expressability, performance, portability, and robustness

Approach:



Results:

P = number of points
C = number of centroids
K = number of clusters
I = number of different initializations

Clustering Algorithm

```
forall points in 1..P {
   for centroids in 2..C {
    // calculate distance
    // update clusters
    // update centroids
   }
}
```

K-Means Application

```
coforall k in 2..K {
   coforall i in 1..I {
    // clustering algorithm
   }
   // Reduce to find best initializations
}
// Reduce to find best clusterings
```

Visualizing k-means clustering on random data*

*Executed on an Intel Xeon E5-2670 processor

Impact and Benefits:

- Enable deeper analysis and unsupervised learning of substantially larger data sets through increased computational capabilities
- Enable rapid prototyping of ideas
- Enable rapid use of new architectures
- Contribute to programming model understanding of abstractions, semantics, and syntax.
- Contribute to the Chapel project using Sandia mission use cases.

Acknowledgments:

This work benefited greatly from discussions with George Stelle of UNM, Greg Titus of Cray, the Analytics and Cryptography group at Sandia, and several of my fellow CCD interns.