-

**CSI 5640- COMPUTER ARCHITECTURE**

**Project on**

---Title---

**By**

Andersen Ball

Cory Ness

Kokila Subramanian

Kyle Zunich

**Under the Guidance of**

Debatosh Debnath

Fall 2022

**Abstract**

Most of the business Clustering is a potential Machine Learning tool to perform determination of intrinsic grouping of unlabeled datasets without any standard criteria. Based on the defined properties and assumptions the datasets can be obtained from the homogenous group using supervised or unsupervised techniques. K-means, an unsupervised clustering algorithm is a useful benchmark to identify the performance characteristcs of the parallel computer architectures. It is a generic tool used for large sale data mining, image segmentation, pattern classification and document retrieval. OpenACC is a parallel programming method used to identify the optimal parallelization patterns for the given problem. The creation of the specific parallel codes on any given architecture is handled by the compiler in the OpenACC that provides the benefit of high level abstraction over the complex underlying heterogeneous architecture of general purpose processors such as NIVIDIA, CUDA and AMD GCN..

In our project we analyze the performance study of parallelization of k-means algorithm to --------- in OpenACC platform. It is employed to explore the performance of the unsupervised clustering method in terms of parallel computing via minimal architecture aware optimization. The OpenACC provides the opportunity to divulge the potential performance capabilities of the CPUs and GPUs. The directives of OpenACC are employed to achieve the parallelism and manage data. The parallelization is studied depending on the complexity of sequential K-means algorithm in terms of data, dimensions, cluster and iterations.