

Images: Scikits Learn,. Comparing Different Clustering Algorithms On Toy Datasets. 2016. Web. 2 Aug. 2016.

**Reach Convergence** 

## Software-Engineered Library Development to Support a High Performance Machine Learning Visualization System

## Operations



## How It Works

• User uploads their data to the NICE server and selects model

DRAM

- Using the CPU and GPU operations, the server clusters the data in the way that the user selected
- Front end uses clustering data to create a visual environment for the client



Leslie Chase, Jason Booth, Anne Demosthene, Tory Leo, Collin Purcell, Selean Ridley, Andrew Tu, Xiangyu Li, Shi Dong, David Kaeli



DRAM





GitHub

Website

- Develop an open source clustering engine
- Use machine learning algorithms
- Accelerate on CPU using Eigen and KMlocal libraries
- Accelerate parallelizable code on GPUs
- Provide visual data representation to user
- Leads to insights into underlying correlations of the data



## Future Work

- Further accelerating algorithms with GPU implementations
- More machine learning algorithms based on beta testing
- Implement a recommendation engine based on users' interests in the different forms of clustering
- May use a utility matrix that holds the views for each model

	K-Means Clustering	Spectral Clustering	Alt. Spectral Clustering	Logistic Regression	Linear Regression	Average Views
User 1	25	32	30	2	3	18.4
User 2			1	14	12	9
User 3		20			1	10.5
User	1 & User 3 =	$\vec{a}\cdot\vec{b}$	= 0.696		(0,1)	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
User 2	2 & User 3 = 1 & User 2 =	$\ \vec{a}\ \ \vec{b}\ $	= -0.214		$\frac{\frac{1}{2}}{90^{\circ}}$	AT 6 2 12
USEL		-	0.737	$(-1,0) \int_{\pi 180^{\circ}}^{75}$		

# Acknowledgements

This research has been generously supported by the following grants: Grant Award ACI – 1559894 from the National Science Foundation Grant Award Number P42ES017198 from the National Institute of **Environmental Health Sciences** NSF IIS-1546428 BIGDATA: IA: Exploring Analysis of Environment and Health Through Multiple Alternative Clustering



### Conclusion

K-Means Clustering

Data will be normalized to find similarity between the users