



# Automated Deployment of OpenStreetMap Data in an Apache Spark Cluster

Rebecca Appelbaum, Colin McKibben, Thinh Vu

# Objectives

- Be able to quickly parse and analyze OpenStreetMap data which typically has several terabytes of data
- Create a process that is automated and easily repeatable



# Technologies

- OpenStack
- Apache Spark
- Chef cookbooks
- OpenStreetMap



# Implementation

- Created a cm command that will create the number of clusters specified
- Examined and downloaded relevant Chef cookbooks
- Created command to pull data from OSM



# Next Steps

- Utilize the chef cookbooks to install Apache Spark on the clusters
- Have data automatically put onto clusters
- Create a test case to prove usability

