
where(ii)

Progress Report

By Priya Gupta, Lisa Barcelo, Melanie Costello

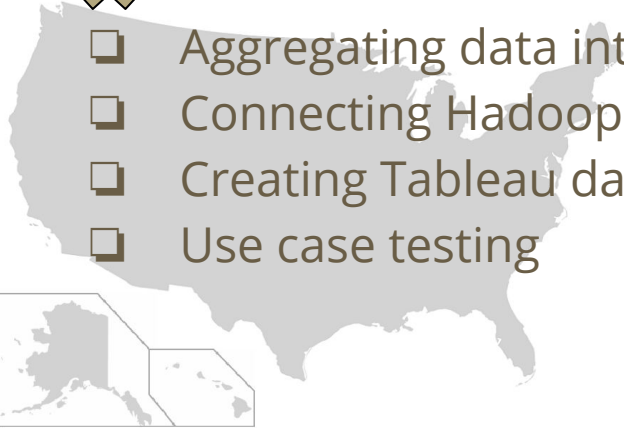
Overview of Activities

- Aggregating income, housing, and traffic data by US Metro Regions
- Combining raw data sources and analyzed to determine most profitable place to live, based on occupation
- Organizing data into RDBMS
- Structuring to optimize analysis and choose best place to live given income and housing parameters



Check List

- ✘ Finalized Data Sources: Zillow, Census, Texas A&M University Dept of Transportation
- ✘ Created ERD
- ✘ Created DDL scripts to clean/load data into hadoop
- ✘ Update ERD to connect tables and create structure of new tables
- ✘ Finalized ETL to transform data into appropriate tables (partitioning)
- ☐ Aggregating data into final display (formulas, time series, etc)
- ☐ Connecting Hadoop and Tableau
- ☐ Creating Tableau dashboards (open up server to make viewable to others)
- ☐ Use case testing



Discussion Points

- Splitting task by data sets -> Zillow, Housing, Census, Traffic
- Multiple data sources together: Metro cities, Region Ids
- Hosting our program/app on a live Tableau dashboard gives us the flexibility to create a simple and attractive user interface while harnessing powerful database connection tools
- Transforming string of dates into data in hadoop? "YYYYMM"

