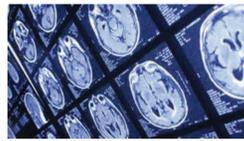




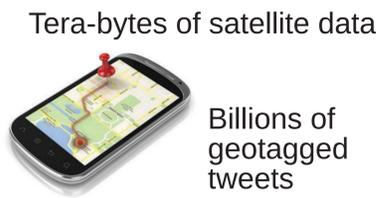
A Demonstration of HadoopViz: An Extensible MapReduce System for Visualizing Big Spatial Data

Ahmed Eldawy Mohamed F. Mokbel Christopher Jonathan
Department of Computer Science and Engineering University of Minnesota

Motivation



Peta-bytes of medical images



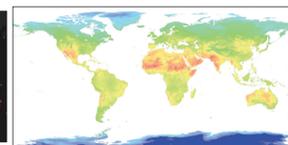
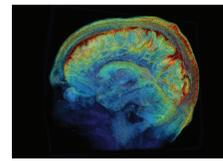
Billions of geotagged tweets



Millions of space images

- Visualization is an international language for communication
- Allows scientists to explore new datasets
- Allows users to easily spot interesting patterns

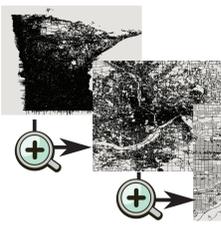
Challenges



Different visualizations for different datasets and applications



Spatial data usually needs to be smoothed



Giga-pixel images to catch all the details

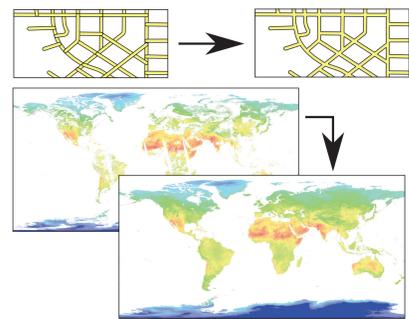


Limited resources of a single machine

Visualization Abstraction

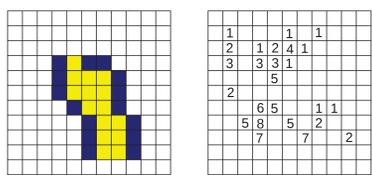
1. smooth

Fuses nearby records together before drawing



2. create-raster

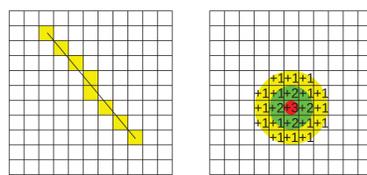
Initializes a canvas on which records are plotted



In-memory image for road networks
2D histogram for frequency heatmaps

3. rasterize

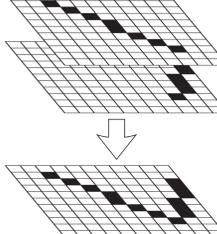
Visualizes a record on a raster layer created by the create-raster function



Bresenham mid-point algorithm
Update histogram with Gaussian Kernel

4. merge

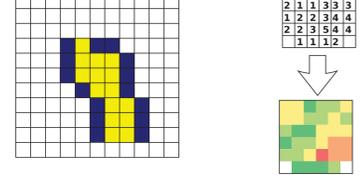
Merges two partial raster layers



Plot images on each other
Add up entries

5. write

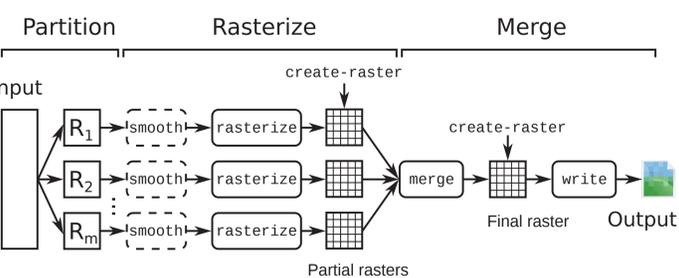
Writes the final raster layer to the output as an image



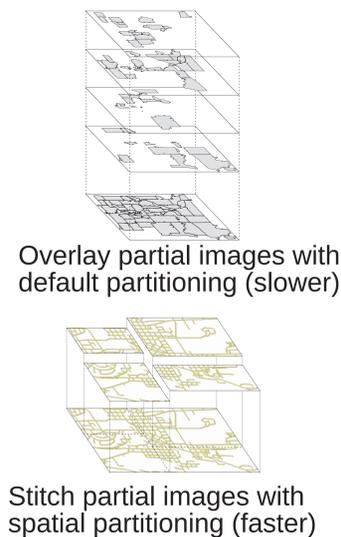
Encode an in-memory image as PNG
Colorize pixels
Encode as PNG

Single Level Plot

Implemented in MapReduce using the five abstract functions

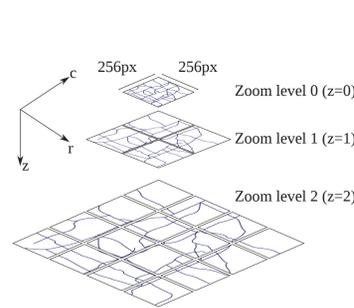


Default Hadoop partitioning is used for small images that do not require smoothing
Otherwise, spatial partitioning is used



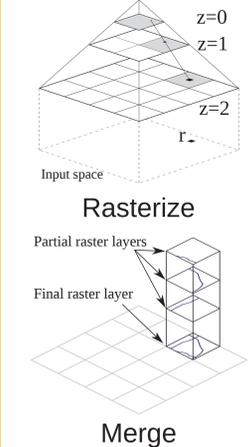
Multilevel Plot

Multilevel Image

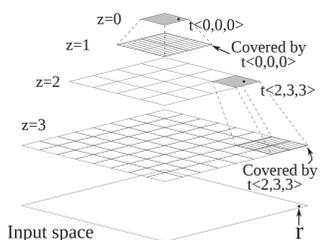


- Millions of tiles
- Supports zoom in/out
- Giga-pixels resolution

Flat Partitioning

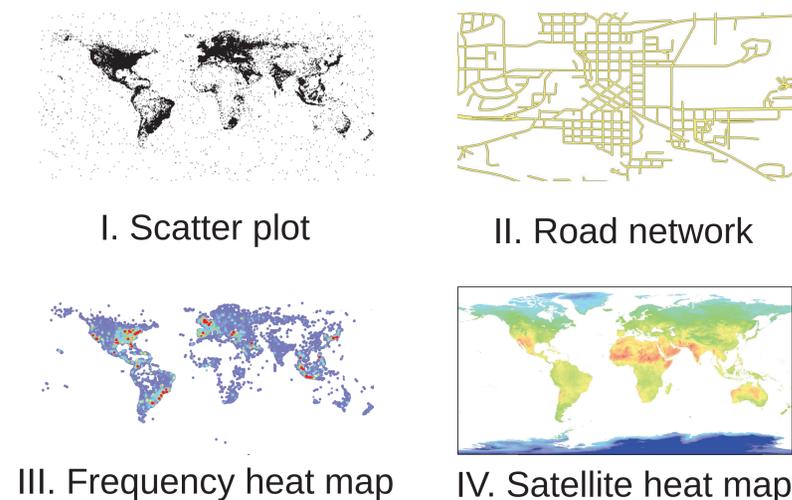


Pyramid Partitioning

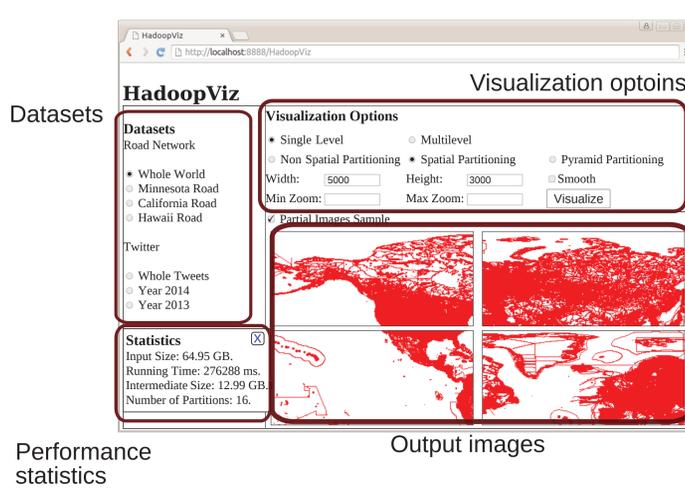


- Better for deeper pyramid levels
- Partition to tiles in levels that are multiple of k
- Each partition generates all tiles in up-to k levels
- No merging step is required

Case Studies



Interface



Performance

