GEOSPATIAL VISUAL ANALYTICS BELONGS TO DATABASE SYSTEMS

Jia Yu
Advisor: Mohamed Sarwat

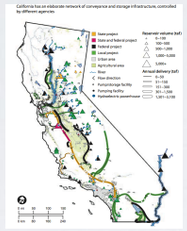
ARIZONA STATE UNIVERSITY

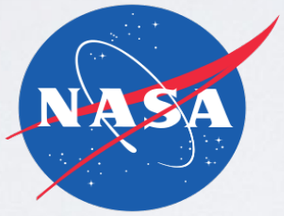
BIG SPATIAL DATA ERA

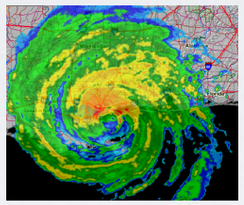
Social Media

1.19 billion monthly active users as of September 30, 2013

Scientific Data



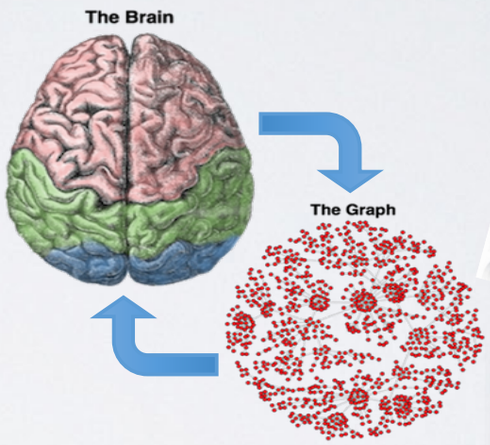




Urban Data

400GB of Road Network Data

Medical Data



85 Billion neurons in the brain nervous system

Mobile Devices

NYC TAXI TRIPS

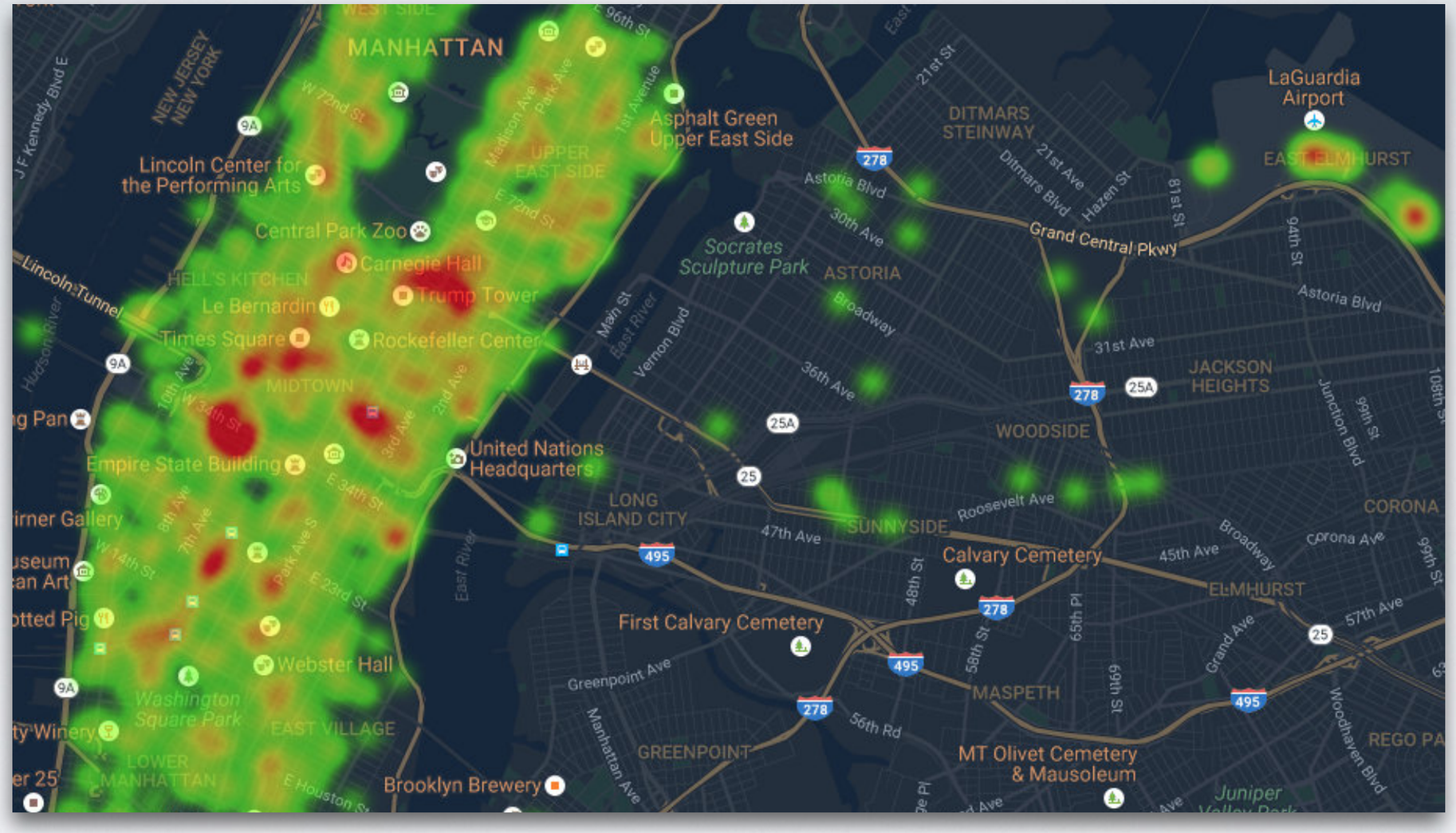
Over a billion taxi trips between 2009 and 2015 released by NYC taxi and limousine services

Pick Up Time	Pick up location	Drop Off Location	Fare amount	…
1	Laguardia	Manhattan	$40	…
2	JFK	Empire State	$30	…
…	…	…	…	…
…	…	…	…	…

Run a sequence of spatial database queries:

Return all the NYC taxi trips for which the pick up location is within the Laguardia airport region

OR



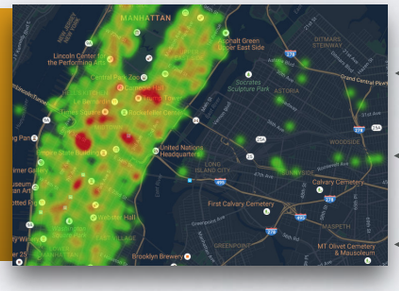
HEAT MAP OF NYC TAXI TRIPS

GEOVIZ

Not Scalable

Google Maps

GeoSpatial Map Visualization Tool

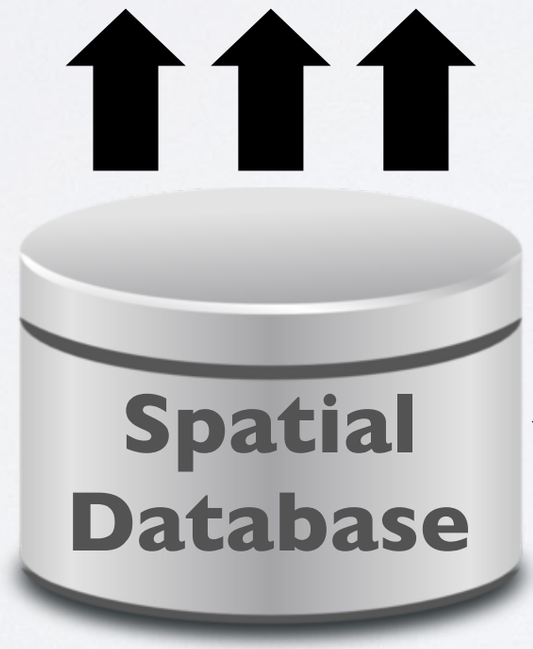


Zoom-In

Pan

Zoom-Out

Not Interactive



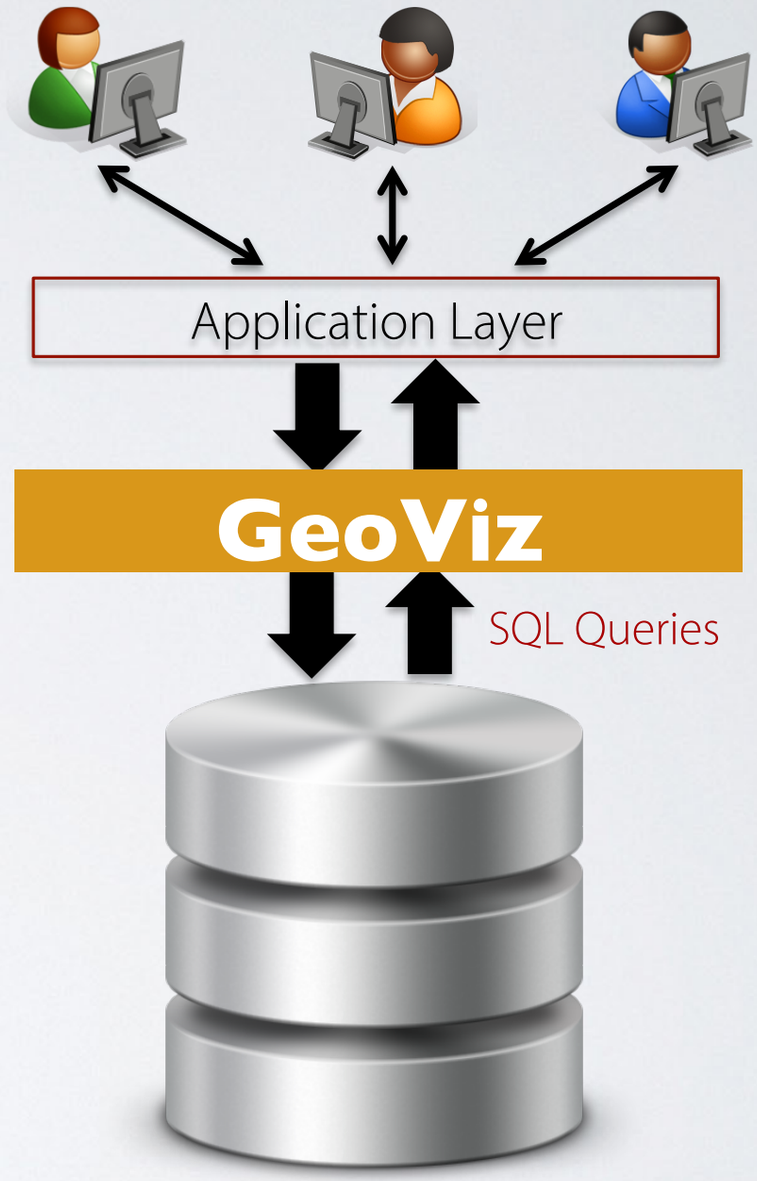
GeoSpatial Data

ON-TOP DB APPROACH

(1) Load / Parse data: HDFS, S3, PostgreSQL

(2) Spatial Query: PostgreSQL, Hadoop, Spark (GeoSpark)

(3) Then do data visualization using Google Map, MapBox, ArcGIS, MapD



Painful process of jumping between query processing and data visualization especially with Big Data

Viz Phase

Data Prep Phase

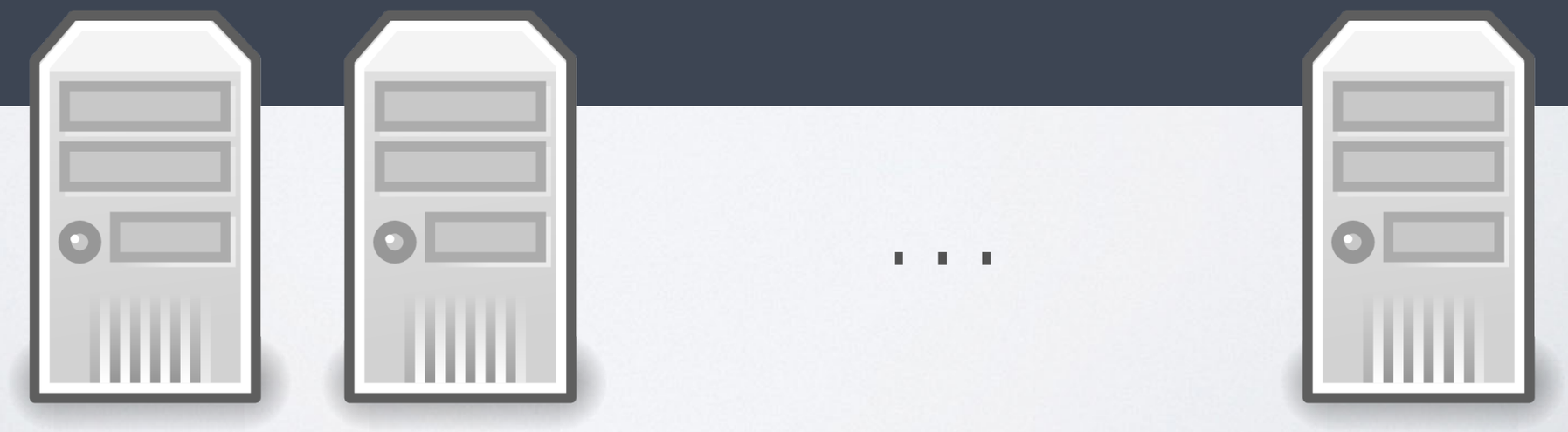
Map Visualization

Scatter Plot | Heat Map | Choropleth Map | . . .

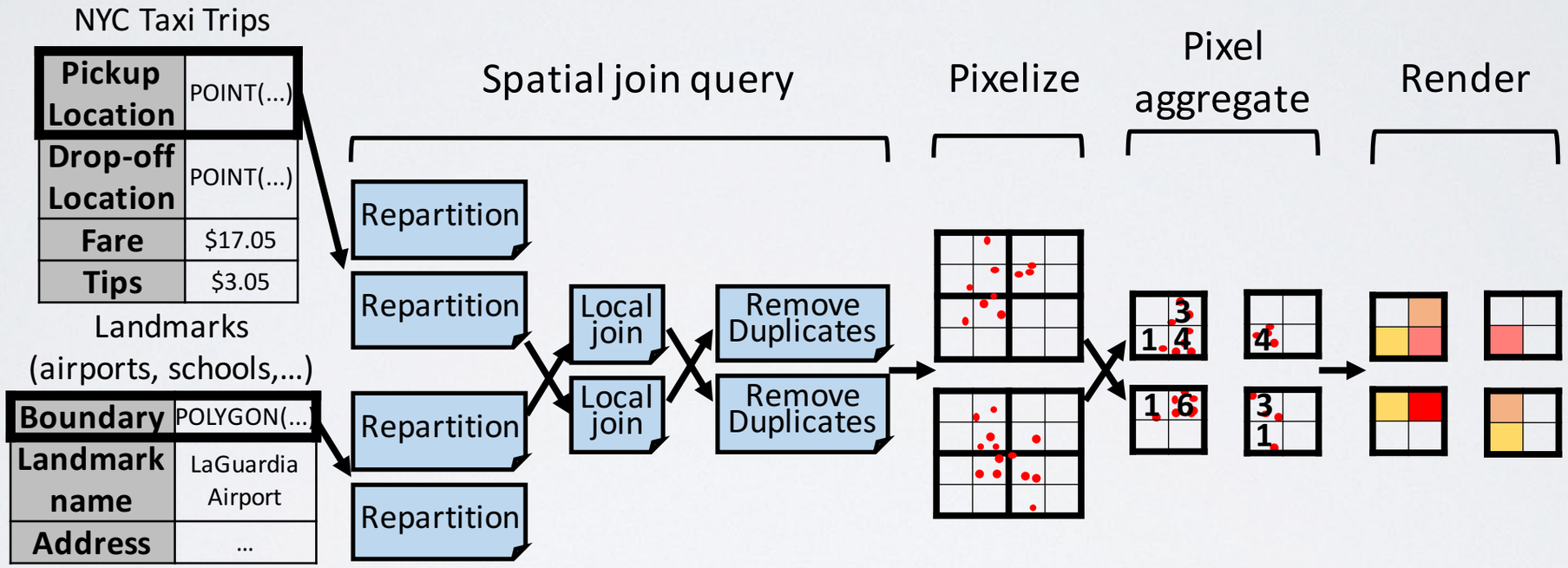
Spatial Query Processing

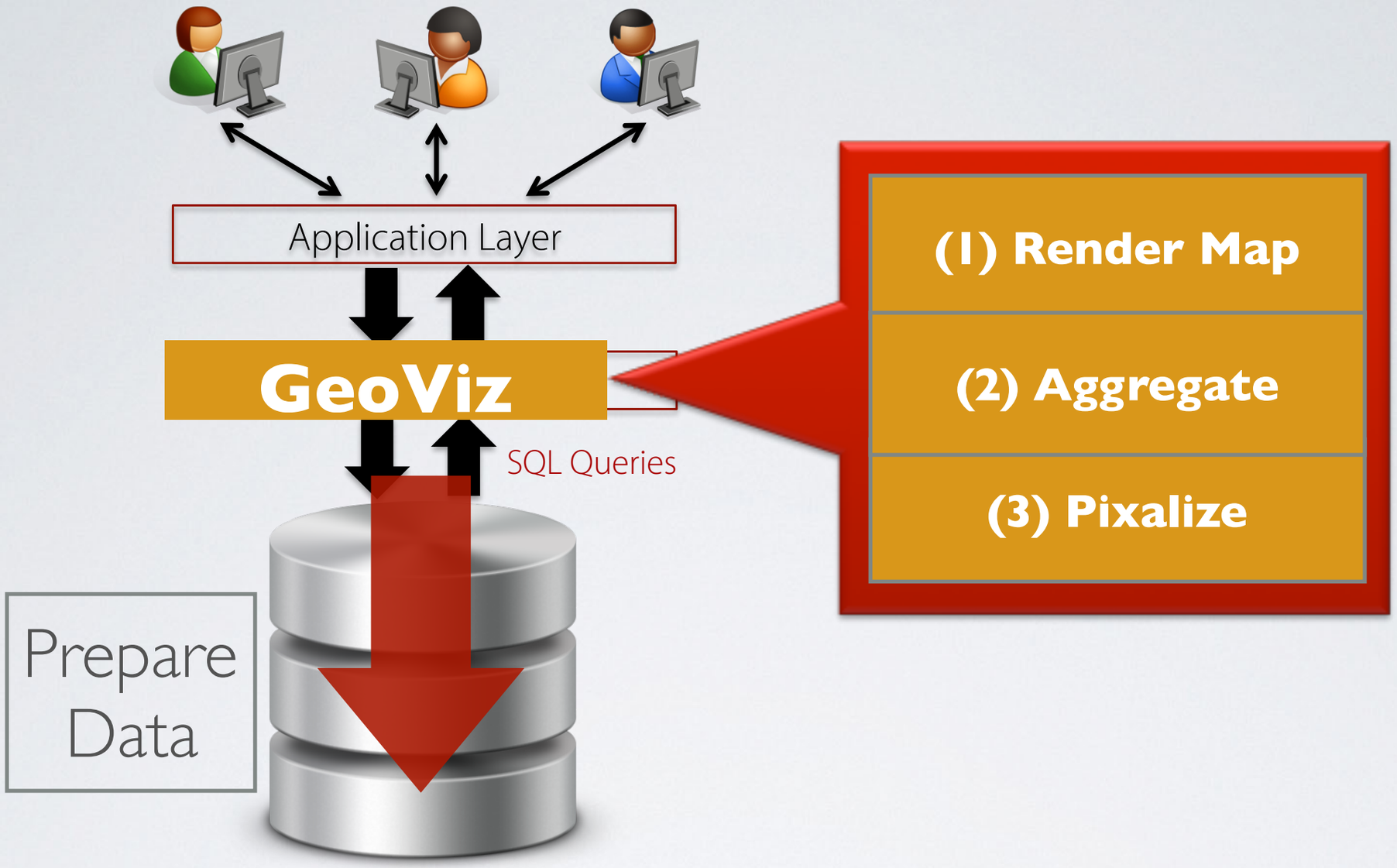
Spatial Datasets

Point Data | Polygon Data | Rectangle Data | . . .



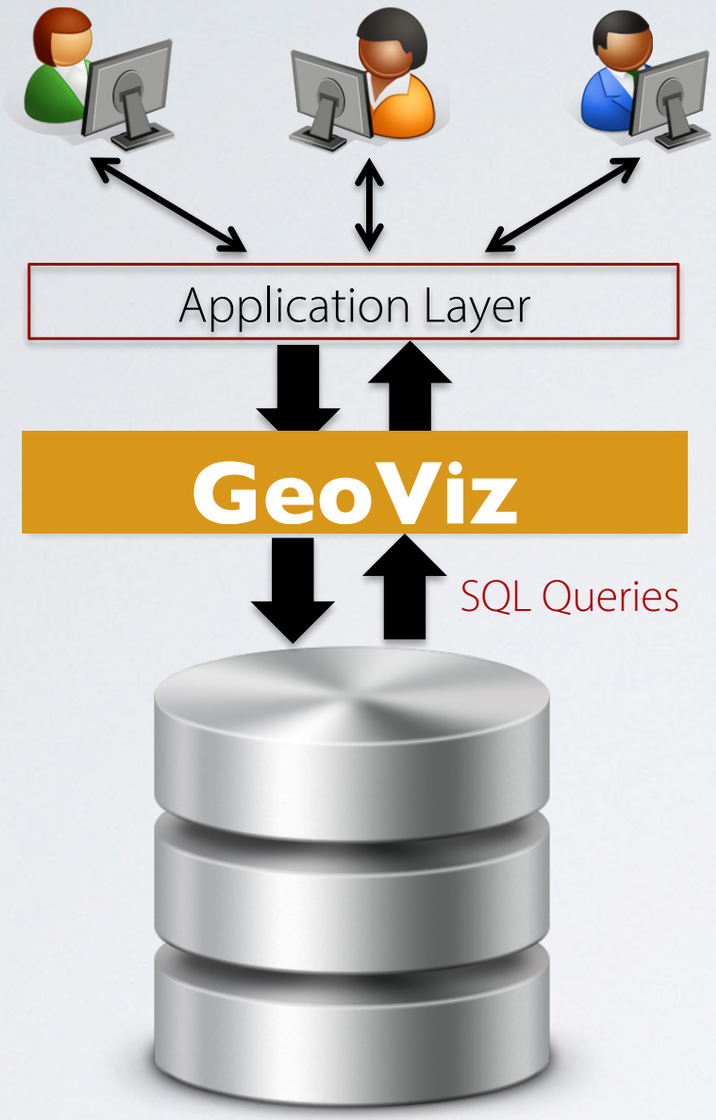
REGULAR EXECUTION PLAN (DISTRIBUTED)



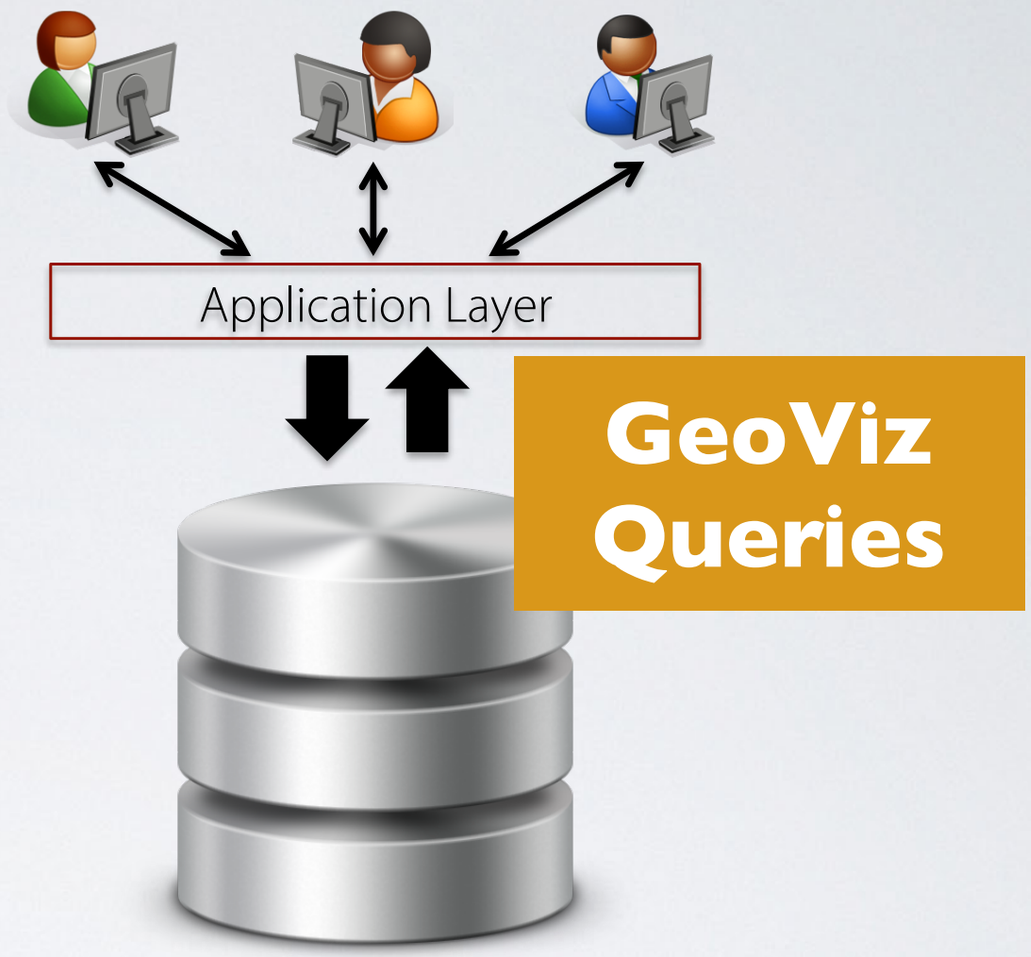


GEOVIZ & DBMS

Implement the Geospatial Visual Analytics functionality on-top of the database system



On-Top DB



GeoVizDB

VISUAL ANALYTICS IN THE DATABASE

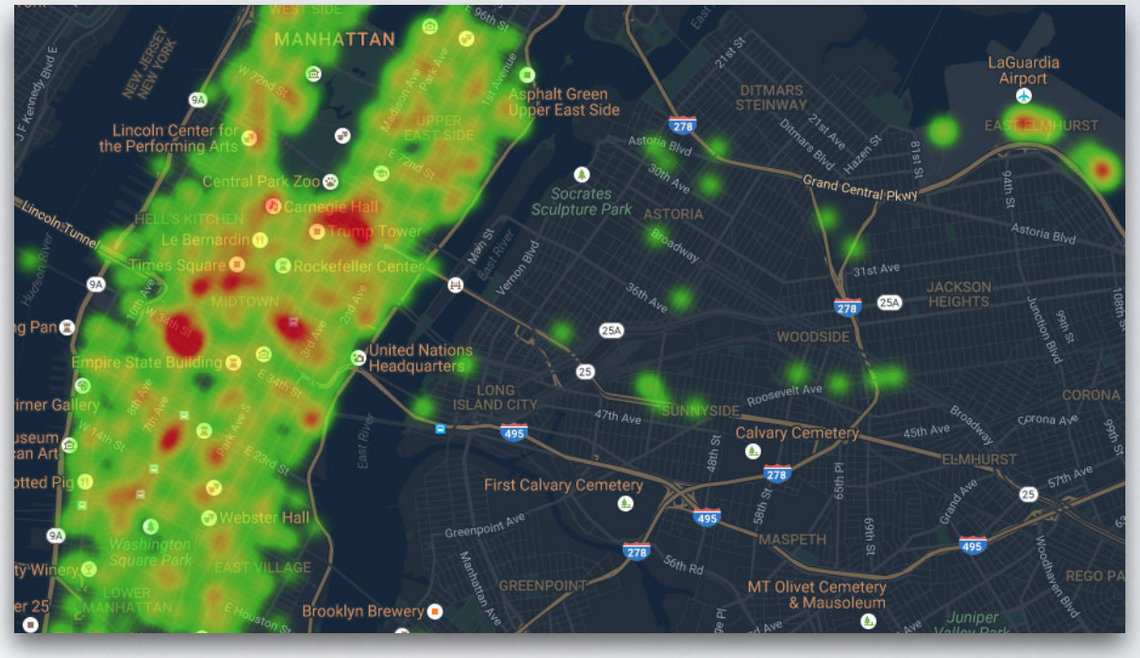
Reduce the overhead of loading the data to a Map Visualization tool

BABYLON

AN END-TO-END VISUAL ANALYTICS SYSTEMS FOR MASSIVE-SCALE GEOSPATIAL DATA

```
SELECT ScatterPlot_OSM_L6 (taxi.pickup)
FROM NYCtaxi taxi
WHERE ST_WITHIN(taxi.pickup,ManhattanBound)
```

Pick Up Time	Pick up location	Drop Off Location	Fare amount	…
1	Laguardia	Manhattan	$40	…
2	JFK	Empire State	$30	…
…	…	…	…	…



GEOVIZ QUERY

Use SQL for both Map Visualization and Data Preparation

BABYLON

```
SELECT ScatterPlot_OSM_L6 (taxi.pickup)
FROM NYCtaxi taxi
WHERE ST_WITHIN(taxi.pickup,ManhattanBound)
```

GeoViz Task Optimizer

Map Visualizaiton (MapViz) Operators

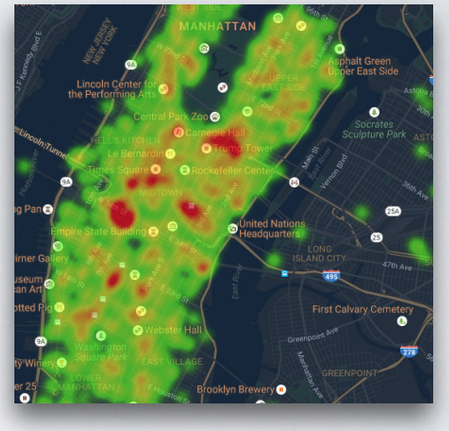
Pixelize	Pixel Aggregate	Render	Overlay

Spatial Query Operators

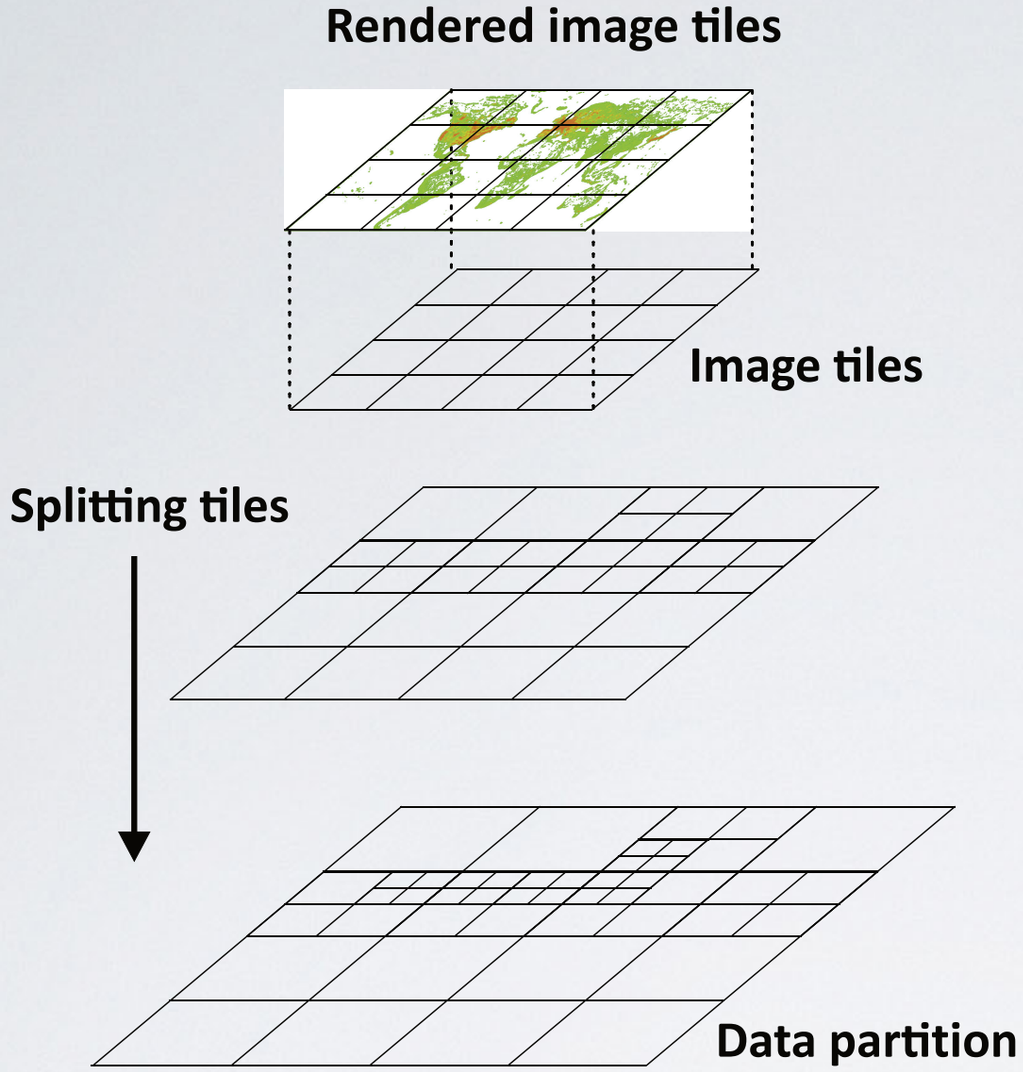
GeoViz Query Execution Plans

GeoViz_Range	GeoViz_Join	…

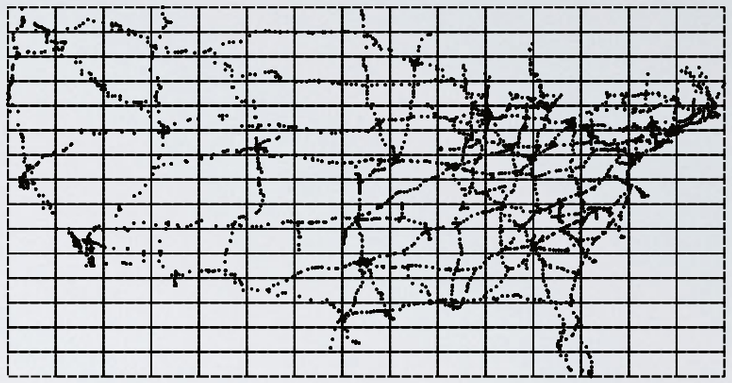
GeoViz-Aware Spatial Partitioner



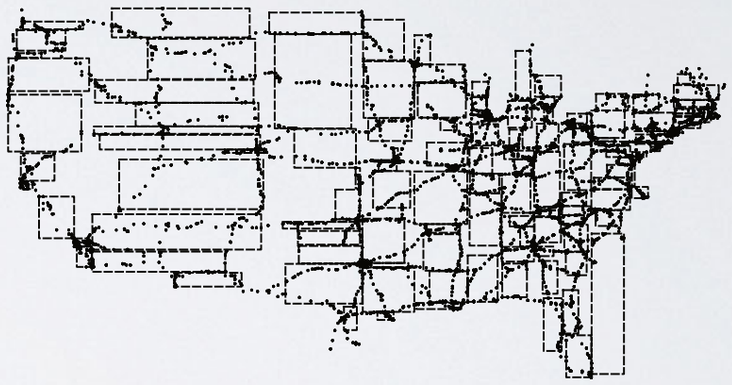
Pick Up Time	Pick up location	Drop Off Location	Fare amount	…
1	Laguardia	Manhattan	$40	…
2	JFK	Empire State	$30	…
…	…	…	…	…



(a) GeoViz-aware partitioner (16 tiles, 50 partitions, maximum tile splitting level = 2)



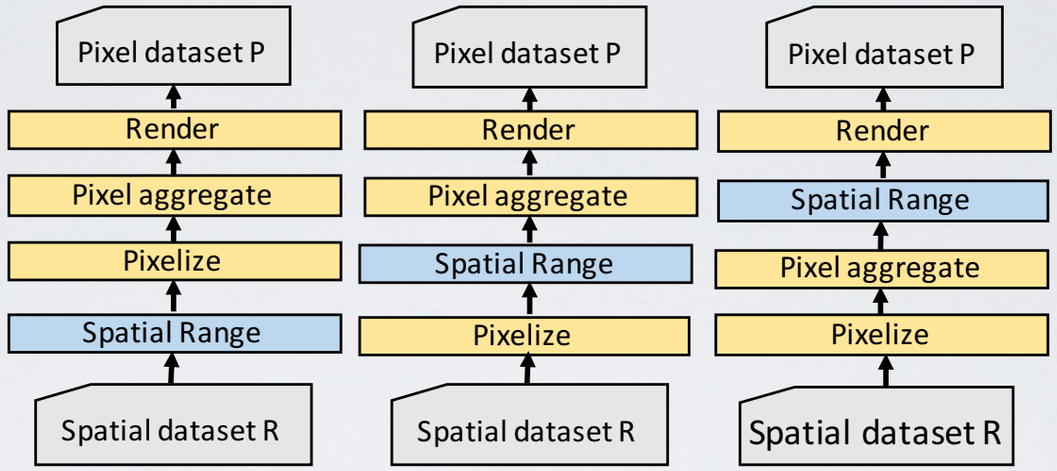
(b) Map tile partitioning

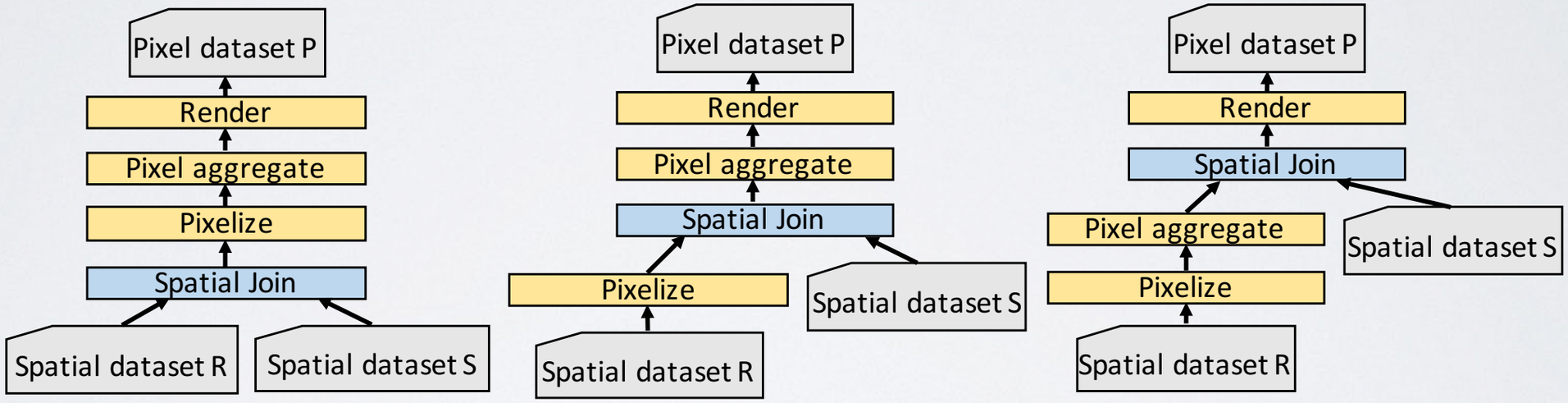


(c) R-Tree partitioning

GEOVIZ PARTITIONING

The partitioning is balanced based upon both the Spatial proximity and the visualization constraints





INTEGRATE OPERATORS

By adopting the GeoVizDB approach, Babylon can achieve more than an order magnitude faster Data to Visualization time than state-of-the-art GeoViz systems

OPTIMIZATION OPPORTUNITIES

Alternate Pixelization and Pixel Aggregation and Spatial Query operators

OPPORTUNITIES

- Leverage existing database ideas, e.g., materialized view maintenance to support dynamic GeoViz

- Leverage GPU-accelerated databases (e.g., MapD) for massive parallelization of data aggregation and map rendering

- More representative and accurate GeoViz-Aware Sampling Techniques

- Deduce Human-Map-Interaction prediction models

Babylon is open-sourced on GitHub

https://github.com/DataSystemsLab/Babylon

DataSys

THE DATA SYSTEMS LAB

https://www.datasyslab.net

@DataSysLab