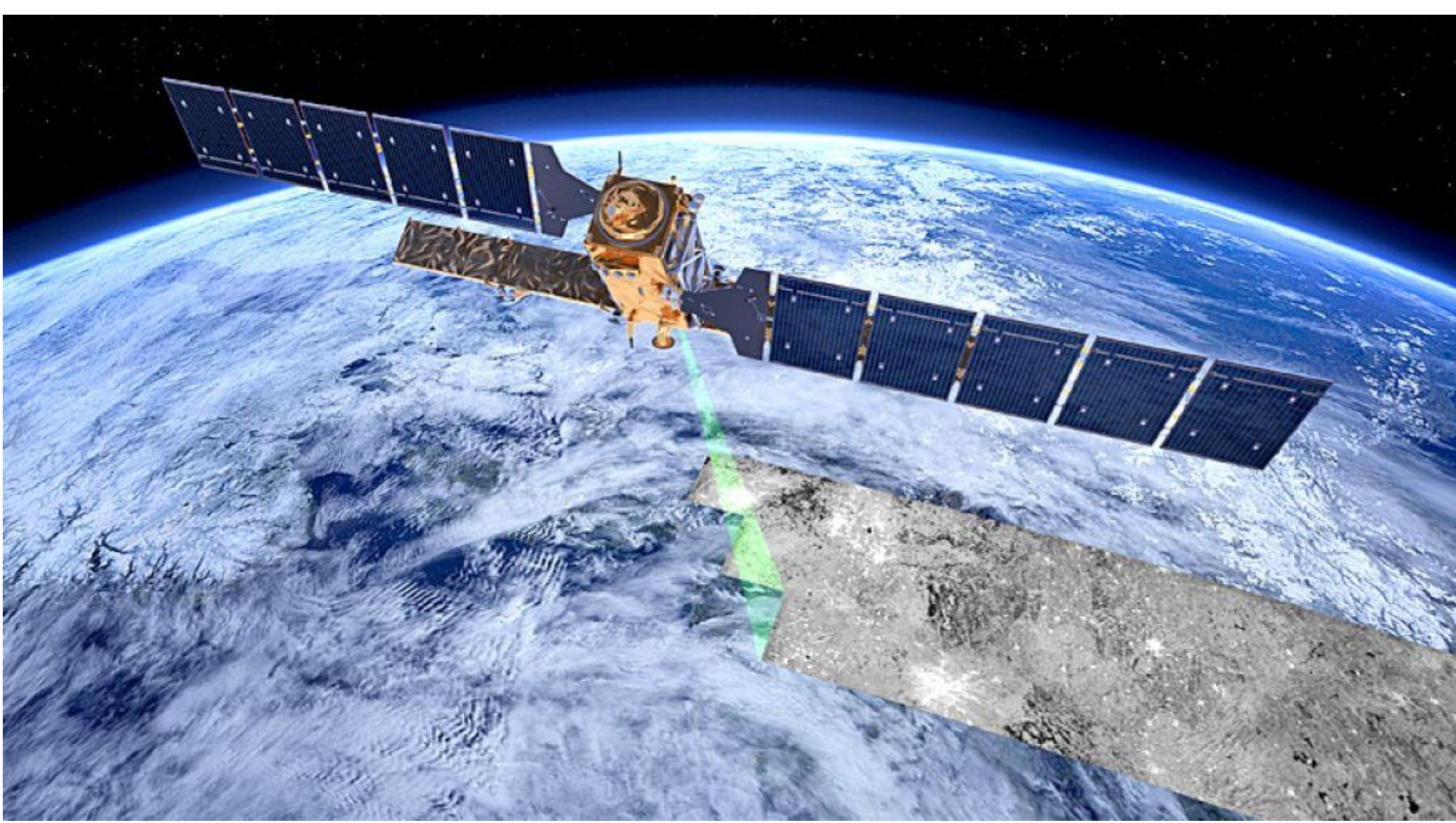


Big Spatial Data

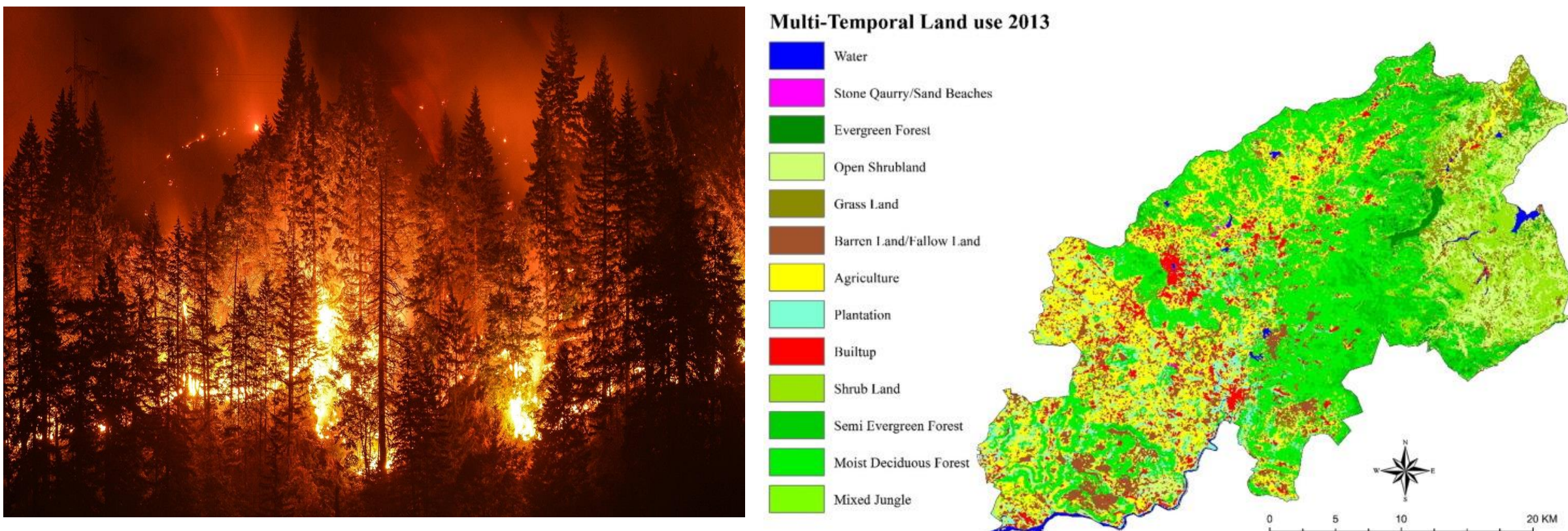


- Petabytes of satellite imagery
- Available in raster format
- Represented using multidimensional arrays



- Highly accurate geographical features such as roads, lakes, etc.
- Available in vector format
- Represented as a collection of coordinates

Applications: Raster + Vector



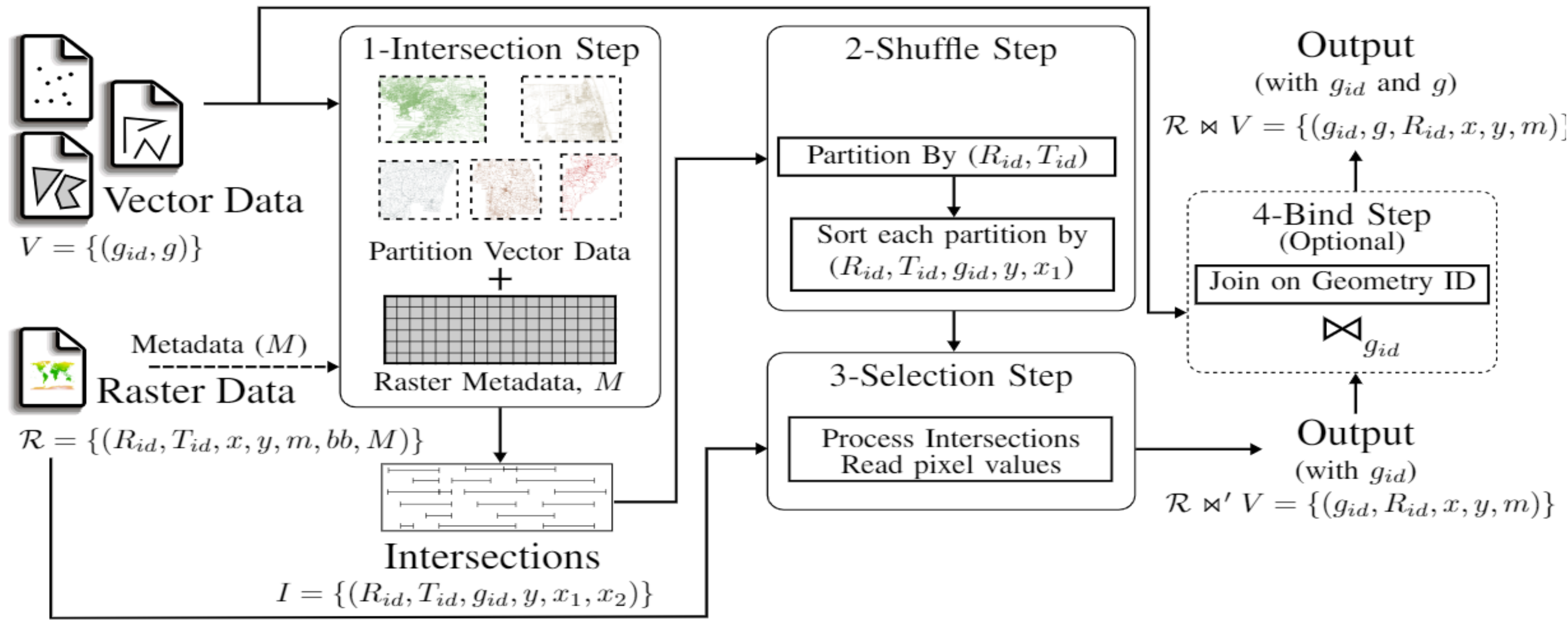
- Need to combine raster and vector data for various applications, such as,
- Crop Yield Mapping
- Modelling the spread of wildfires
- Areal Interpolation
- Studying land use and land cover classification

Existing Big Data Systems



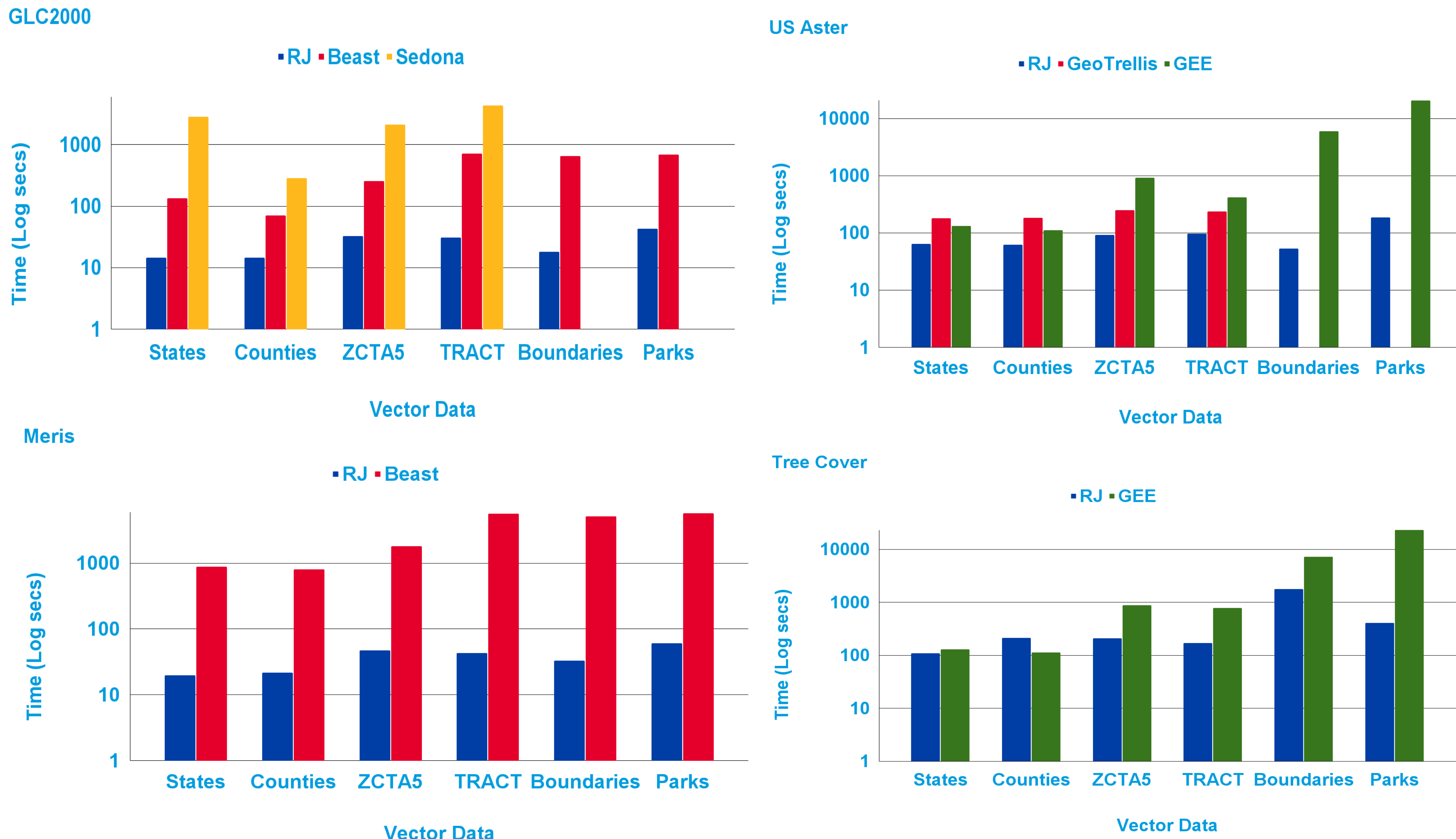
- Focus on either raster or vector data
- Work with different data and query processing models
- Require conversion to work with other forms
- Conversion is computationally expensive and increases the size of data
- Impractical for big data
- Raster-based systems may also require an expensive data loading phase and use linear algebra which leads to complex queries

Proposed System



- Propose a novel approach for concurrent processing of raster + vector data, Raptor Join
- Implemented using an **in-situ approach** in Spark
- Can **directly process raster and vector data** in their native representations
- Modeled as a **relational operator**, easier to use in SQL queries in combination with other relational operators such as selection, join, and aggregation
- Computes *Intersections/Iterator* which minimizes disk access by reading only the required pixels

Results



Vector Datasets

Dataset	Polygons	Segments	#segments #polygons	File Size
Counties	3,108	51,638	17	978 KB
States	49	165,186	3,370	2.6 MB
Boundaries	284	3,817,412	13,440	60 MB
TRACT	74,133	38,467,094	519	632 MB
ZCTA5	33,144	52,894,188	1596	851 MB
Parks	9,747,150	336,145,640	34.5	8.5 GB

Raster Datasets

Dataset	Resolution	File Size
GLC200	40,320x16,353	629 MB
MERIS	129,600x64,800	7.8 GB
US Aster	208,136x89,662	35 GB
Tree Cover	1,296,036x648,018	782 GB

- The proposed Raptor Join (RJ) provides up-to two orders of magnitudes performance gain over baselines.