

地理空间大数据

杨书轩

2020年5月19日

目 录

CONTENTS

- 1 背景
- 2 知识
- 3 资源



Part One

背景

背景

1. 井喷的位置感知数据

- 数十亿的手持设备和物联网设备
- 数千个机载和卫星遥感平台
- 每天产生 2^{60} 字节

2. 蓬勃发展的大数据和机器学习

应用前景

行人流量分析

- 零售商
 - 确定开新店的最佳位置
- 政府机构
 - 公共部门改善城市规划

挑战

数据格式不统一

- GeoJSON, KML, Shapefile, WKT, ESRI Grid,
GeoTIFF,NITF WCS, WFS, WMS,WMITS.....

数据量大

- 单节点难以处理和存储

解决方案

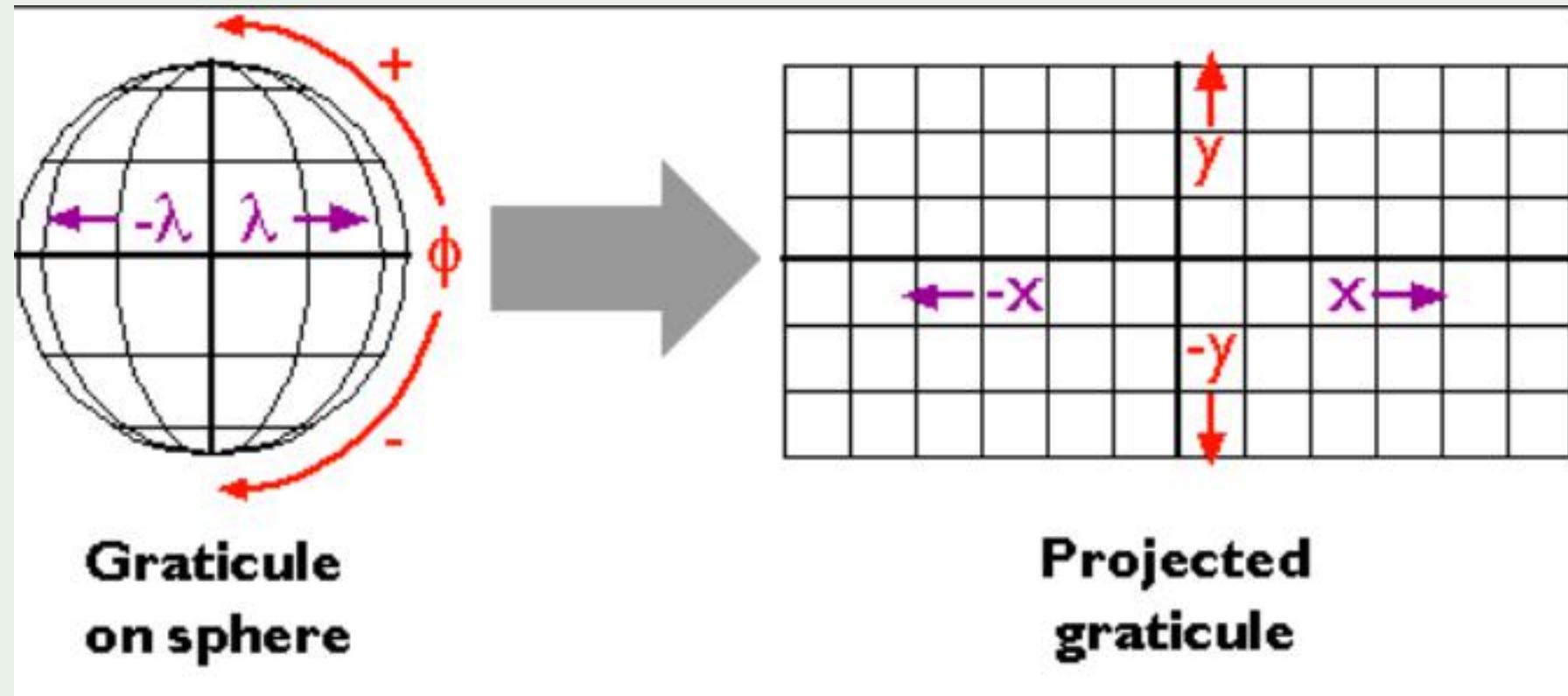
1. 基于 Apache Spark 扩展
 - GeoSpark, GeoMesa, GeoTrellis, Rasterframes
2. UDF(user defined functions)
 - GeoPandas, GDAL, Java Topology Service (JTS)
3. Index
 - 针对地理空间操作进行索引
 - S2, GeoHex, H3



Part Two

知识

坐标参考系统 Coordinate Reference Systems (CRS)



常用的CRS

1. WGS 84 (4326)

- 经纬度
- 121.483333,31.166667

2. Web Mercator (3857)

- 13523462.769429687,3654412.994477549

3. Albers EA (5070)

- -9007089.483105691, 9287044.47565529

地理空间类型

1. 点

➤ (121,31)

2. 线

➤ [(121,31),(122,31)]

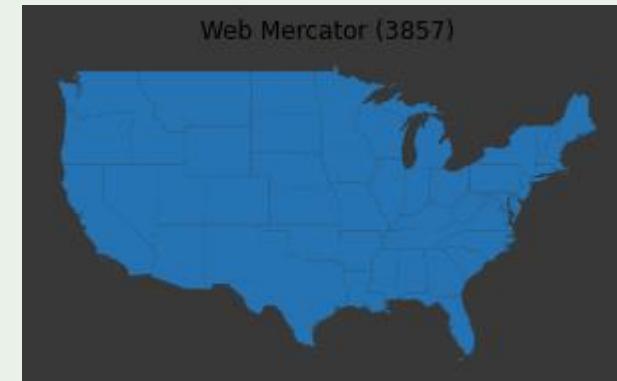
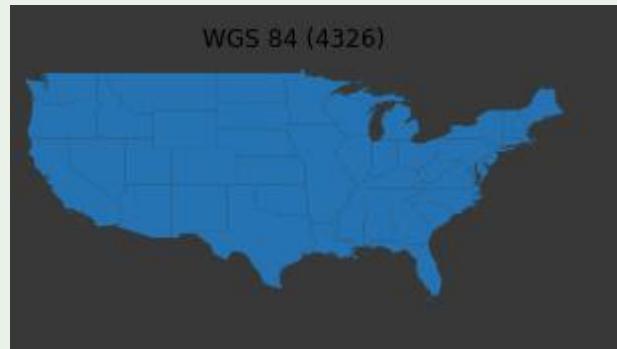
1. 多边形

➤ [(121,31),(122,31),(122,32),(121,32),(121,31)]

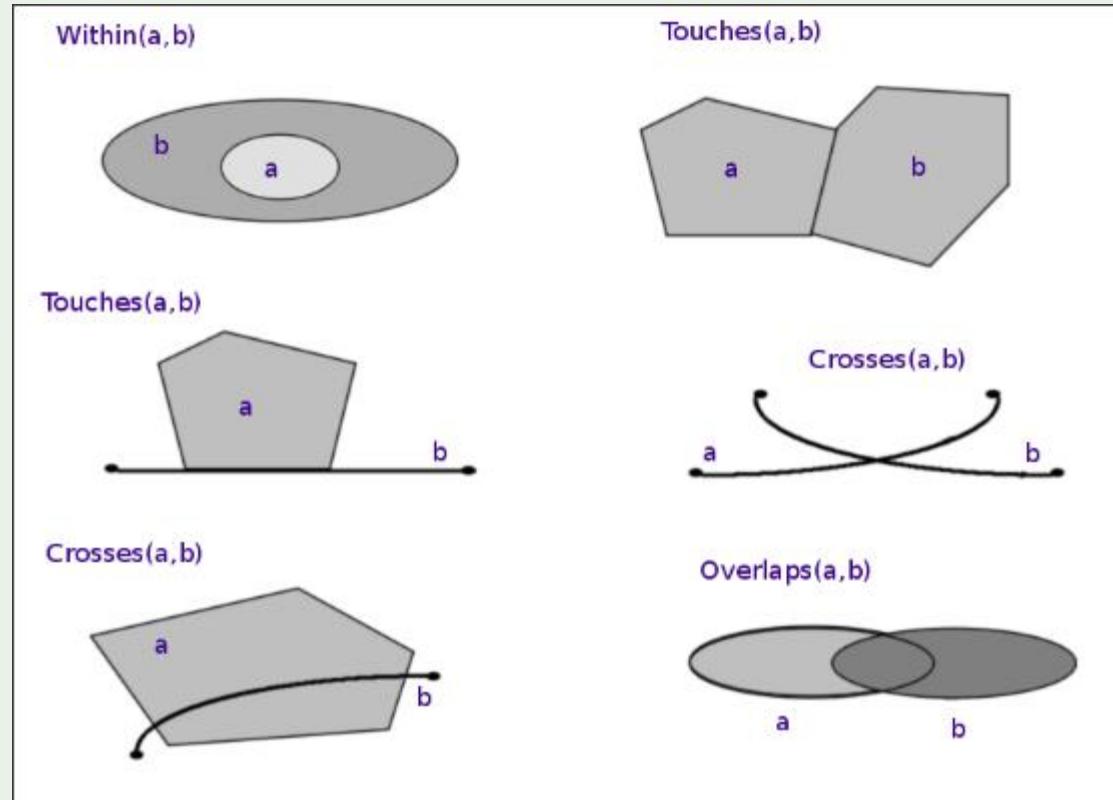
1. 多边形集(飞地)

➤ [[(121,31),(122,31),(122,32),(121,32),(121,31)],
[(101,31),(102,31),(102,32),(101,32),(101,31)]]

不同坐标参考系统产生的不同形状的地图



地理空间关系查询



地理空间测量

1. 面积
2. 周长
3. 距离

基于距离的地理空间关系查询

1. K-nearest
 1. 距离特定地理空间最近的K个其他地理空间
 2. join with specified distance

各CRS可支持的操作

1.WGS 84 (4326) 经纬度

- 地理空间关系查询

2.Web Mercator (3857)

- 地理空间关系查询

3.Albers EA (5070)

- 地理空间查询

- 面积，周长，距离

- 基于距离的地理空间关系查询

重要提醒

必须选用合适的坐标参考系统！！！



Part Three

参考资源

GeoSpark

Name	API	Spark compatibility	Introduction
Core	RDD	Spark 2.X/1.X	SpatialRDDs and Query Operators.
SQL	SQL/DataFrame	SparkSQL 2.1+	SQL interfaces for GeoSpark core.
Viz	RDD, SQL/DataFrame	RDD - Spark 2.X/1.X, SQL - Spark 2.1+	Visualization for Spatial RDD and DataFrame.
Zeppelin	Apache Zeppelin	Spark 2.1+, Zeppelin 0.8.1+	GeoSpark plugin for Apache Zeppelin

[GeoSpark Tutorial \(https://datasystemslab.github.io/GeoSpark/tutorial/rdd/\)](https://datasystemslab.github.io/GeoSpark/tutorial/rdd/)

GeoSpark示例代码

[GeoSpark Examples\(<https://github.com/yangshuxuan/GeoExamples>\)](https://github.com/yangshuxuan/GeoExamples)

Geospatial Fundamentals in Python with Geopandas

尝试地理空间数据各种操作

(<https://github.com/yangshuxuan/Geospatial-Fundamentals-in-Python>)

Processing Geospatial Data at Scale With Databricks

Databrick关于地理空间数据的讨论

(<https://databricks.com/blog/2019/12/05/processing-geospatial-data-at-scale-with-databricks.html>),

可以从

这篇文章开始获取更多的地理空间数据处理的inspire。

人流分析

人流分析

(<https://databricks.com/blog/2019/08/25/building-foot-traffic-insights-dataset.html>)

The End!