

Introducción a Sistemas Espaciales Distribuidos

Andrés Calderón, Ph.D.

December 2, 2024

Primero, una breve encuesta:

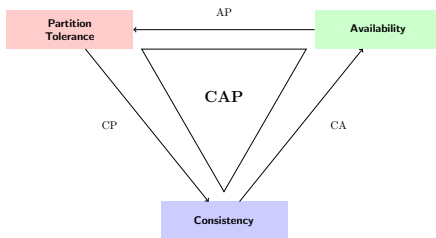


o ingrese a [menti.com](https://www.menti.com)
con el código **7492 8204**

- 1 Introducción
- 2 Motivación y Aplicaciones
- 3 Tecnologías y Herramientas
- 4 Particionamiento
 - Space-filling Curves
 - Árboles
- 5 Retos, Direcciones Futuras y Conclusión

• ¿Qué son los sistemas distribuidos?

- Sistemas con múltiples nodos interconectados.
- Propiedades clave:
 - Escalabilidad
 - Consistencia
 - Disponibilidad
 - Tolerancia a Fallos (Teorema CAP)



En sistemas distribuidos, solo dos propiedades pueden ser logradas al mismo tiempo.

- **¿Qué son los datos espaciales?**

- Datos relacionados con ubicaciones (ej.: mapas, imágenes satelitales, trazas GPS, ...).
- Representados como puntos, líneas, polígonos o rásters.

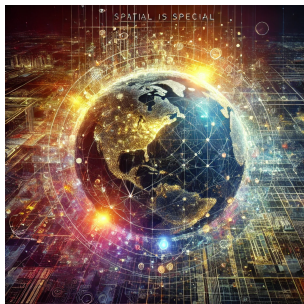
- **¿Por qué combinarlos?**

- Análisis eficiente de datos geoespaciales a gran escala.
- Aplicaciones en planificación urbana, ciudades inteligentes, monitoreo ambiental, etc.

Plan

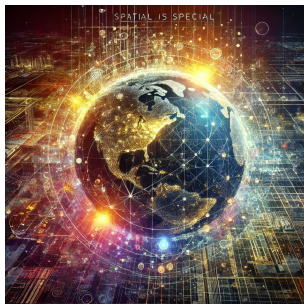
- 1 Introducción
- 2 Motivación y Aplicaciones**
- 3 Tecnologías y Herramientas
- 4 Particionamiento
 - Space-filling Curves
 - Árboles
- 5 Retos, Direcciones Futuras y Conclusión

Espacial es Especial



- Complejidad Multidimensional

Espacial es Especial



- Complejidad Multidimensional
- Topología y nuevos tipos de datos

Espacial es Especial



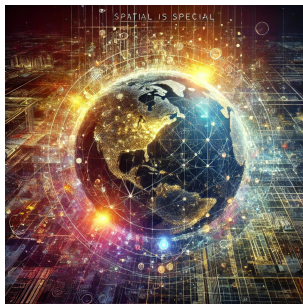
- Complejidad Multidimensional
- Topología y nuevos tipos de datos
- Interdependencia y Contexto

Espacial es Especial



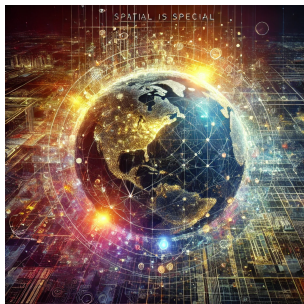
- Complejidad Multidimensional
- Topología y nuevos tipos de datos
- Interdependencia y Contexto
- Visualización e Interacción

Espacial es Especial



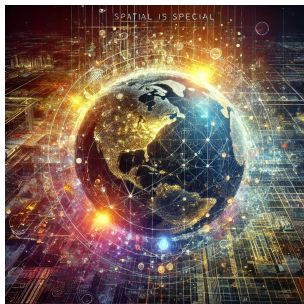
- Complejidad Multidimensional
- Topología y nuevos tipos de datos
- Interdependencia y Contexto
- Visualización e Interacción
- Integración con Big Data e IA

Espacial es Especial



- Complejidad Multidimensional
- Topología y nuevos tipos de datos
- Interdependencia y Contexto
- Visualización e Interacción
- Integración con Big Data e IA
- Desafíos de Escalabilidad

Espacial es Especial



- Complejidad Multidimensional
- Topología y nuevos tipos de datos
- Interdependencia y Contexto
- Visualización e Interacción
- Integración con Big Data e IA
- Desafíos de Escalabilidad

Motivación y Aplicaciones

- **Sistemas de Monitoreo de Tráfico Urbano:**

- California Performance Measurement System (PeMS): > 40K sensores.
- The Infra-Red Traffic Logger (TIRTL): Desplegado en varios países.
- INRIX Traffic App: 1.3 millones de vehículos.

- **Sistemas de Monitoreo de Tráfico Urbano:**

- California Performance Measurement System (PeMS): $> 40K$ sensores.
- The Infra-Red Traffic Logger (TIRTL): Desplegado en varios países.
- INRIX Traffic App: 1.3 millones de vehículos.

- **OpenStreetMap (OSM)**

- Planet.osm File: > 1.97 TB.
- Full History Dump: ≈ 3.97 TB.
- Tile Rendering: ≈ 54 TB. (Nov-2024).

● **Sistemas de Monitoreo de Tráfico Urbano:**

- California Performance Measurement System (PeMS): $> 40K$ sensores.
- The Infra-Red Traffic Logger (TIRTL): Desplegado en varios países.
- INRIX Traffic App: 1.3 millones de vehículos.

● **OpenStreetMap (OSM)**

- Planet.osm File: > 1.97 TB.
- Full History Dump: ≈ 3.97 TB.
- Tile Rendering: ≈ 54 TB. (Nov-2024).

● **GPS Data**

- Global Navigation Satellite Systems (GNSS): $1-20$ r / s / d.
- Vehicle GPS Systems: ≈ 200 GB / día para una flota mediana.
- Crowdsourced Apps (Google Maps, Strava, Garmin): ≈ 2 PB / mes.

Motivación y Aplicaciones

• Satélites de Observación de la Tierra

- Landsat 8 y 9 (USGS): ≈ 700 y 900 imágenes/día (desde 1972), 700 MB y 1 GB c/u.
- Sentinel-2 (ESA): ≈ 10 y 20 TB/día.
- GOES-R (NOAA) ≈ 3.5 TB/día.

• Satélites de Observación de la Tierra

- Landsat 8 y 9 (USGS): ≈ 700 y 900 imágenes/día (desde 1972), 700 MB y 1 GB c/u.
- Sentinel-2 (ESA): ≈ 10 y 20 TB/día.
- GOES-R (NOAA) ≈ 3.5 TB/día.

• Radar de Apertura Sintética (SAR)

- Sentinel-1: ≈ 1.6 TB / día.

- **Satélites de Observación de la Tierra**

- Landsat 8 y 9 (USGS): ≈ 700 y 900 imágenes/día (desde 1972), 700 MB y 1 GB c/u.
- Sentinel-2 (ESA): ≈ 10 y 20 TB/día.
- GOES-R (NOAA) ≈ 3.5 TB/día.

- **Radar de Apertura Sintética (SAR)**

- Sentinel-1: ≈ 1.6 TB / día.

- **Satélites Comerciales de Alta Resolución**

- Planet Labs o Maxar (constelaciones de satélites): ≈ 2 PB / año.

- **Satélites de Observación de la Tierra**

- Landsat 8 y 9 (USGS): ≈ 700 y 900 imágenes/día (desde 1972), 700 MB y 1 GB c/u.
- Sentinel-2 (ESA): ≈ 10 y 20 TB/día.
- GOES-R (NOAA) ≈ 3.5 TB/día.

- **Radar de Apertura Sintética (SAR)**

- Sentinel-1: ≈ 1.6 TB / día.

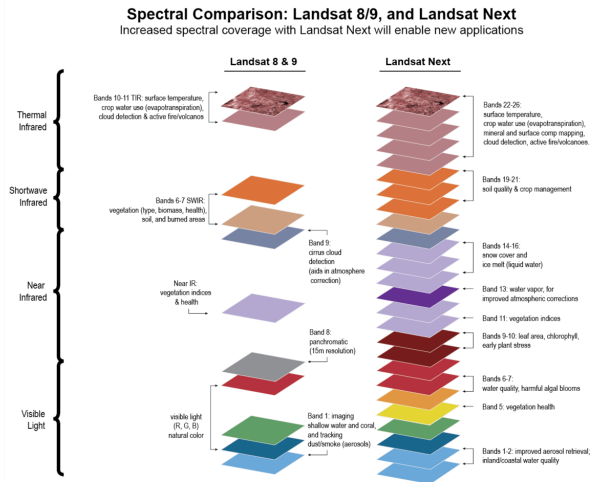
- **Satélites Comerciales de Alta Resolución**

- Planet Labs o Maxar (constelaciones de satélites): ≈ 2 PB / año.

- **MODIS (NASA)**

- Satélites Terra y Aqua: $\approx 5\text{--}10$ TB/día.

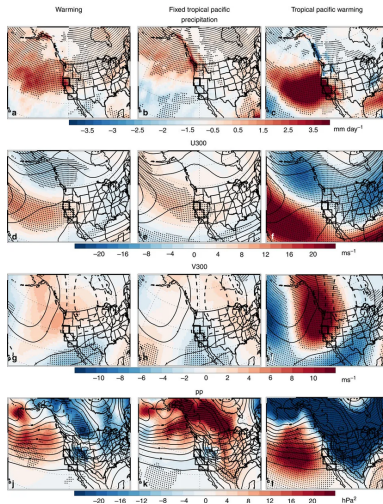
Motivación y Aplicaciones



usgs.gov

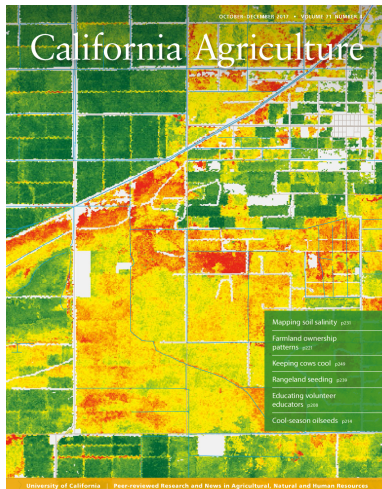
Motivación y Aplicaciones

Does El Niño increases California precipitation in response to warming?



(Allen and Luptowitz, 2017)

Can SAM recognize crops?



(Gurav et al., 2023)

We just raised \$21.5M in Series A funding led by Felicis → [Learn More](#)

 PRODUCTS ▾ SOLUTIONS ▾ RESOURCES ▾ COMPANY ▾ CONTACT SIGN IN TRY WHEROBOTS →

Home > [Announcing Our 21.5M Series A :: Unlocking Answers to Planetary-scale Questions.](#)

ANNOUNCING OUR \$21.5 MILLION SERIES A

🕒 5 MINS READ 20 NOV 2024

Led by  Felicis

Announcing Our 21.5M Series A :: Unlocking Answers to Planetary-scale Questions.

Each day, satellites, drones, applications, and GPS devices generate petabytes of spatial data that can be used to solve real-world problems. But the majority of such data is often stuck in siloed legacy systems or sits idle and disjointed. That's why we have dedicated our academic and professional careers to answer planetary-scale questions. We're on a mission to help companies make sense of their data so they can take on issues like how to manage their fleets of vessels and vehicles, where and how to build infrastructure, and determine the best methods to assess and mitigate risk of catastrophic natural disasters.

AUTHORS
Mo Sarwat / Jia Yu



wherobots.com

Plan

- 1 Introducción
- 2 Motivación y Aplicaciones
- 3 Tecnologías y Herramientas**
- 4 Particionamiento
 - Space-filling Curves
 - Árboles
- 5 Retos, Direcciones Futuras y Conclusión

- **Frameworks:**

- PostgreSQL + PostGIS (YugabyteDB)
- Apache Hadoop (SpatialHadoop)
- Apache Spark (Apache Sedona / Geotrellis / Geomesa)
- Google Earth Engine / Planetary Computer

- **Arquitecturas:**

- Bases de datos distribuidas con capacidades espaciales.
- Map Reduce para procesamiento de datos espaciales.

Tecnologías y Herramientas



SPATIAL



ArcGIS



GRASSGIS

Hadoop-GIS

Spatial Big Data Solutions



SECONDO

MD-Hbase



Google Earth Engine



Apache

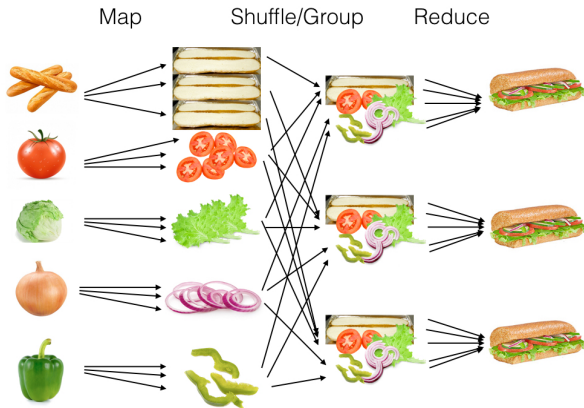
Sedona



BEAST



Map Reduce (a.k.a. Divide and Conquer)

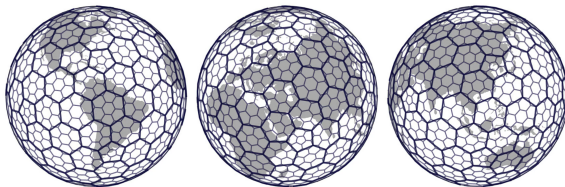


Plan

- 1 Introducción
- 2 Motivación y Aplicaciones
- 3 Tecnologías y Herramientas
- 4 Particionamiento**
 - Space-filling Curves
 - Árboles
- 5 Retos, Direcciones Futuras y Conclusión

● **Particionamiento en Sistemas Distribuidos Espaciales:**

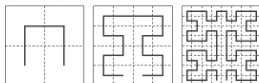
- Indexación espacial: Z-order, curvas de Hilbert, ...
- Particionamiento del espacio: Grillas, Kd-trees, Quadtrees, ...
- Otros o derivados: DCEL, H3, Geohash, ...
- Soportan operaciones distribuidas como: Uniones, intersecciones, consultas de proximidad, ...



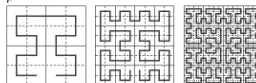
h3geo.org

Space-filling Curves

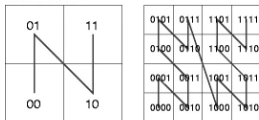
Hilbert-Moore curve:



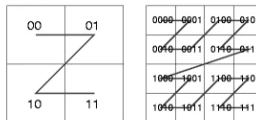
$\beta\Omega$ -curve:



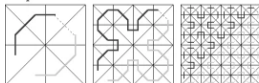
Morton order:



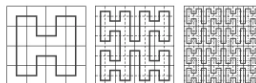
Z-curve:



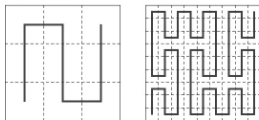
Sierpinski curve:



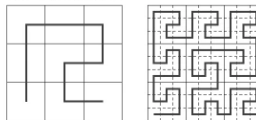
H-index:



Peano curve:

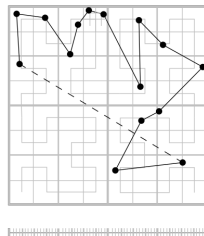
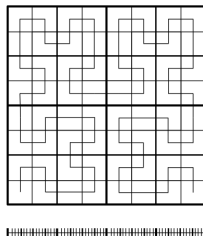
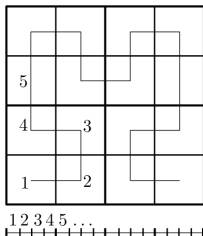
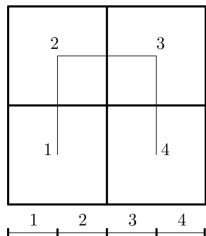


Peano-Meander curve:



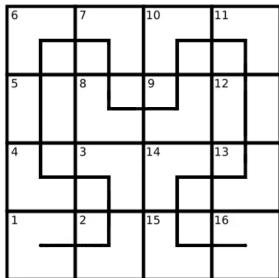
(Bader et al., 2011)

Space-filling Curves

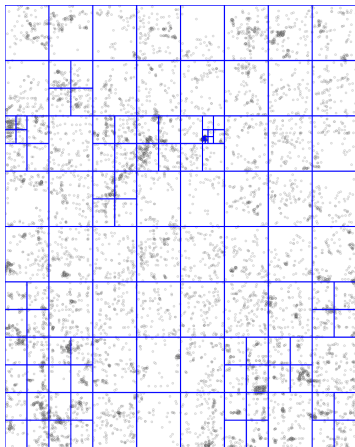


(Buchin, 2008)

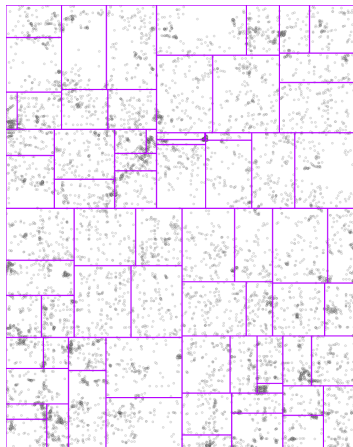
Space-filling Curves



Space-oriented vs Data-oriented

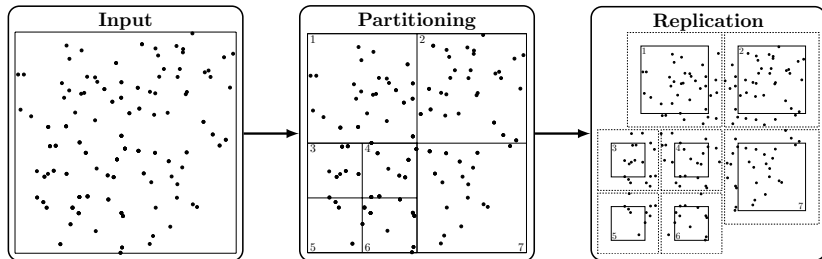


Quadtree



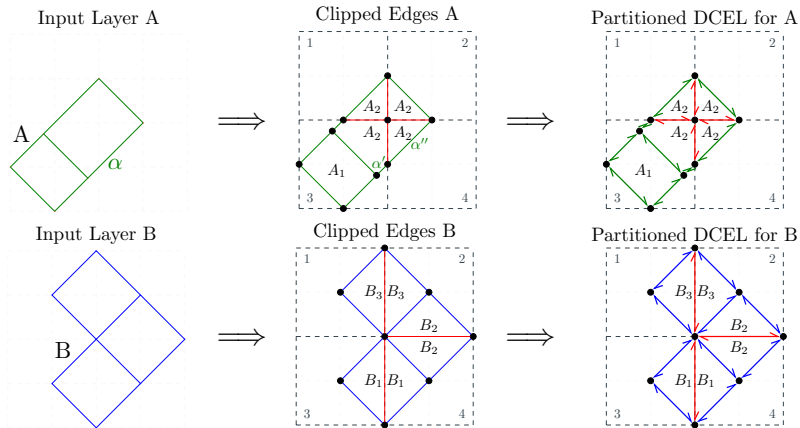
KD-tree

Arboles



(Calderon, 2025)

Arboles



(Calderon, 2023)

Comparación

Característica	SF Curves	Quadtree	Kd-tree
Localidad	Buena	Excelente	Buena
Dimensiones	Limitada	Principalmente 2D	Multi-dimensional
Eficiencia	Alta (range)	Alta (join)	Alta (k-NN)
Memoria	Bajo	Medio/Alto	Medio
Escalabilidad	Excelente	Moderada	Moderada
Densidad	Limitada	Excelente	Media
Complejidad	Baja/Media	Media	Alta

Plan

- 1 Introducción
- 2 Motivación y Aplicaciones
- 3 Tecnologías y Herramientas
- 4 Particionamiento
 - Space-filling Curves
 - Árboles
- 5 Retos, Direcciones Futuras y Conclusión

Retos y Direcciones Futuras

- **Retos Actuales:**

- Manejo eficiente de operaciones espaciales.
- Balanceo de carga durante el particionamiento.
- Flujos de datos espacio-temporales en tiempo real.

- **Retos Actuales:**

- Manejo eficiente de operaciones espaciales.
- Balanceo de carga durante el particionamiento.
- Flujos de datos espacio-temporales en tiempo real.

- **Tendencias Emergentes:**

- Integración con aprendizaje automático e IA.
- Análisis y minería de datos espacio-temporales.
- Procesamiento de imágenes satelitales de alta resolución.

- Los sistemas espaciales distribuidos combinan las fortalezas del cómputo distribuido y el análisis de datos geoespaciales.
- Las aplicaciones abarcan múltiples dominios, desde la planificación urbana hasta el monitoreo climático.
- Los avances continuos están impulsados por desafíos en escalabilidad, eficiencia y procesamiento en tiempo real.

¿Preguntas?

Haz el quiz
escaneando el código!

