

Spatial Data Management using Big Data Platforms

Yao-Yi Chiang

Computer Science and Engineering

University of Minnesota

yaoyi@umn.edu

CC-BY
Attribution



Spatial Data Can be Huge

OpenStreetMap Data Extracts

The OpenStreetMap data files provided on this server do **not** contain the user names, user IDs and changeset IDs of the OSM contributors. [Extracts with full metadata](#) are available to OpenStreetMap contributors only.

Welcome to Geofabrik's free download server. This server has data extracts from the [OpenStreetMap project](#) which are normally data download service is offered free of charge by Geofabrik GmbH.

Willkommen auf dem Geofabrik-Downloadserver. Hier gibt es Daten-Auszüge aus dem [OpenStreetMap-Projekt](#), die normalerweise vertraut zu machen, bevor Sie mit den Daten arbeiten.) Diese Downloads werden von der Geofabrik GmbH kostenlos angeboten.

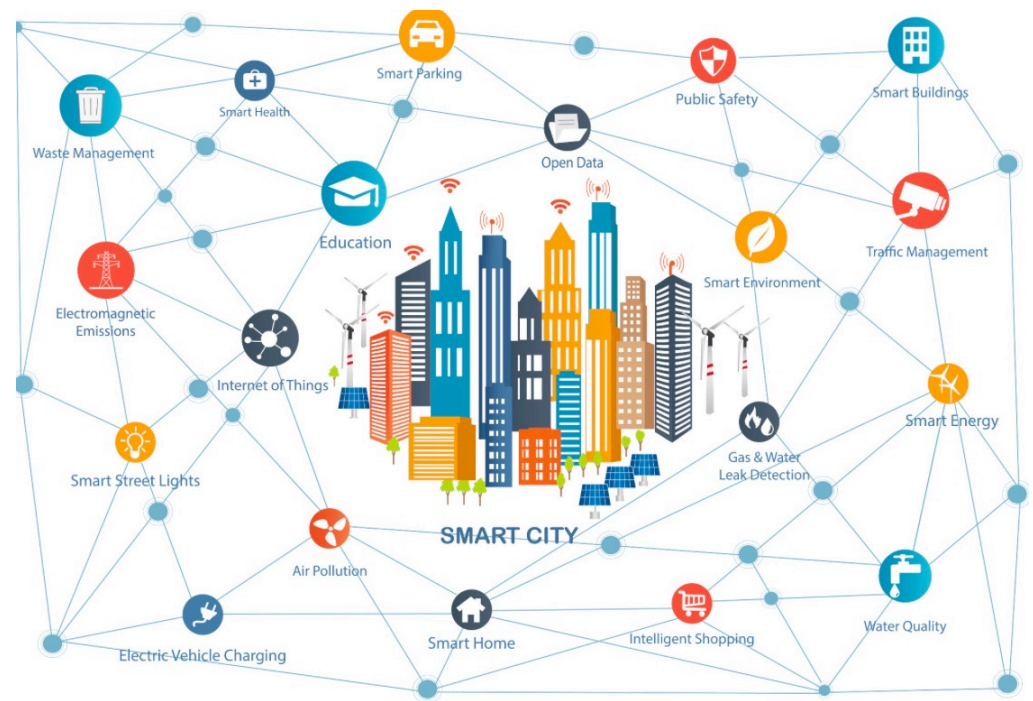
Click on the region name to see the overview page for that region, or select one of the file extension links for quick access.

Sub Region	Quick Links		
	.osm.pbf	.shp.zip	.osm.bz2
Africa	[.osm.pbf] (4.9 GB)	✘	[.osm.bz2]
Antarctica	[.osm.pbf] (31.0 MB)	[.shp.zip]	[.osm.bz2]
Asia	[.osm.pbf] (10.4 GB)	✘	[.osm.bz2]
Australia and Oceania	[.osm.pbf] (945 MB)	✘	[.osm.bz2]
Central America	[.osm.pbf] (501 MB)	✘	[.osm.bz2]
Europe	[.osm.pbf] (24.5 GB)	✘	[.osm.bz2]
North America	[.osm.pbf] (11.1 GB)	✘	[.osm.bz2]
South America	[.osm.pbf] (2.7 GB)	✘	[.osm.bz2]

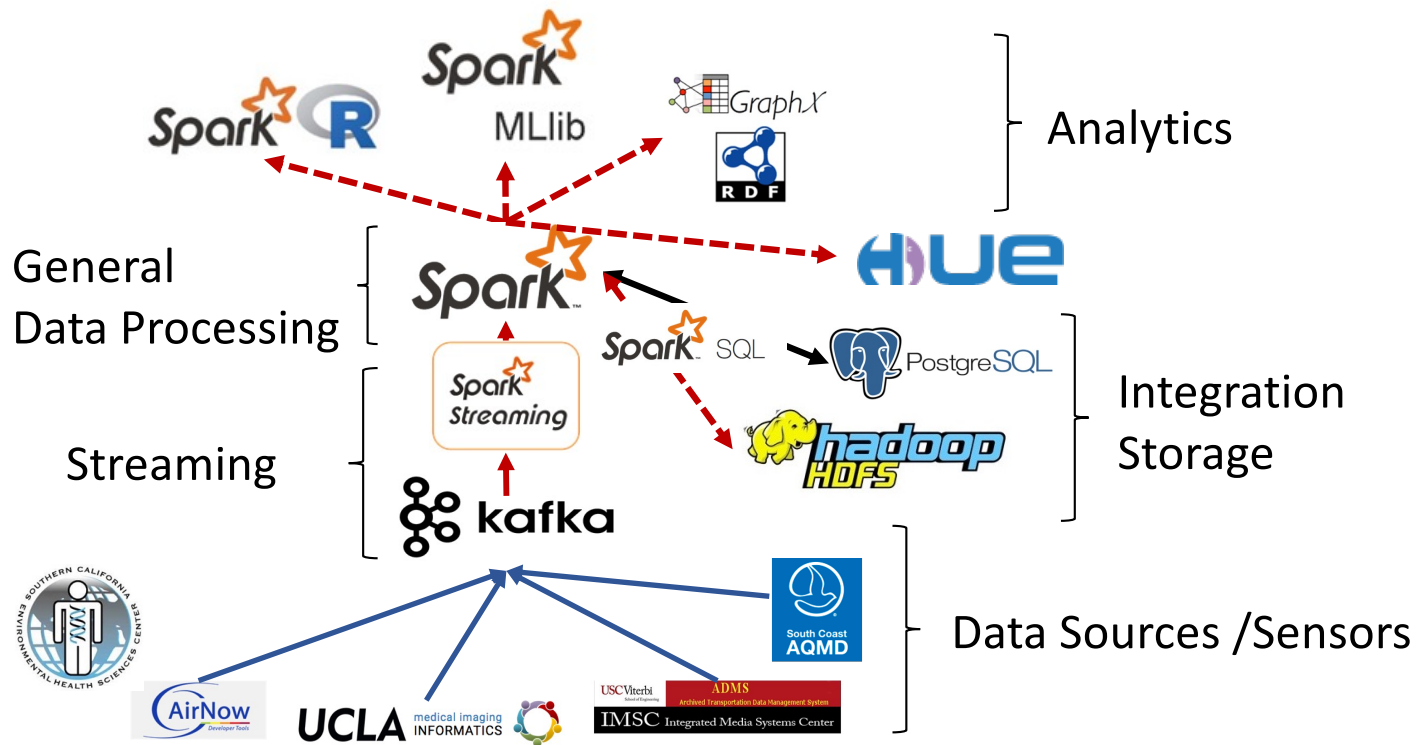
[Technical details](#) about this download service.

Spatial Data Can be Heterogeneous

- Fixed-site sensor data (e.g., traffic time series)
- Moving data (e.g., trajectories)
- Static data (e.g., road networks)



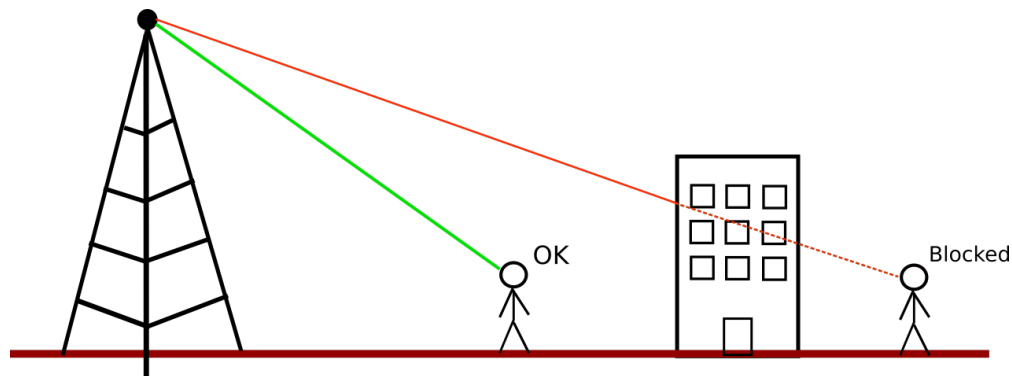
Scalable Data Integration & Analysis Architecture



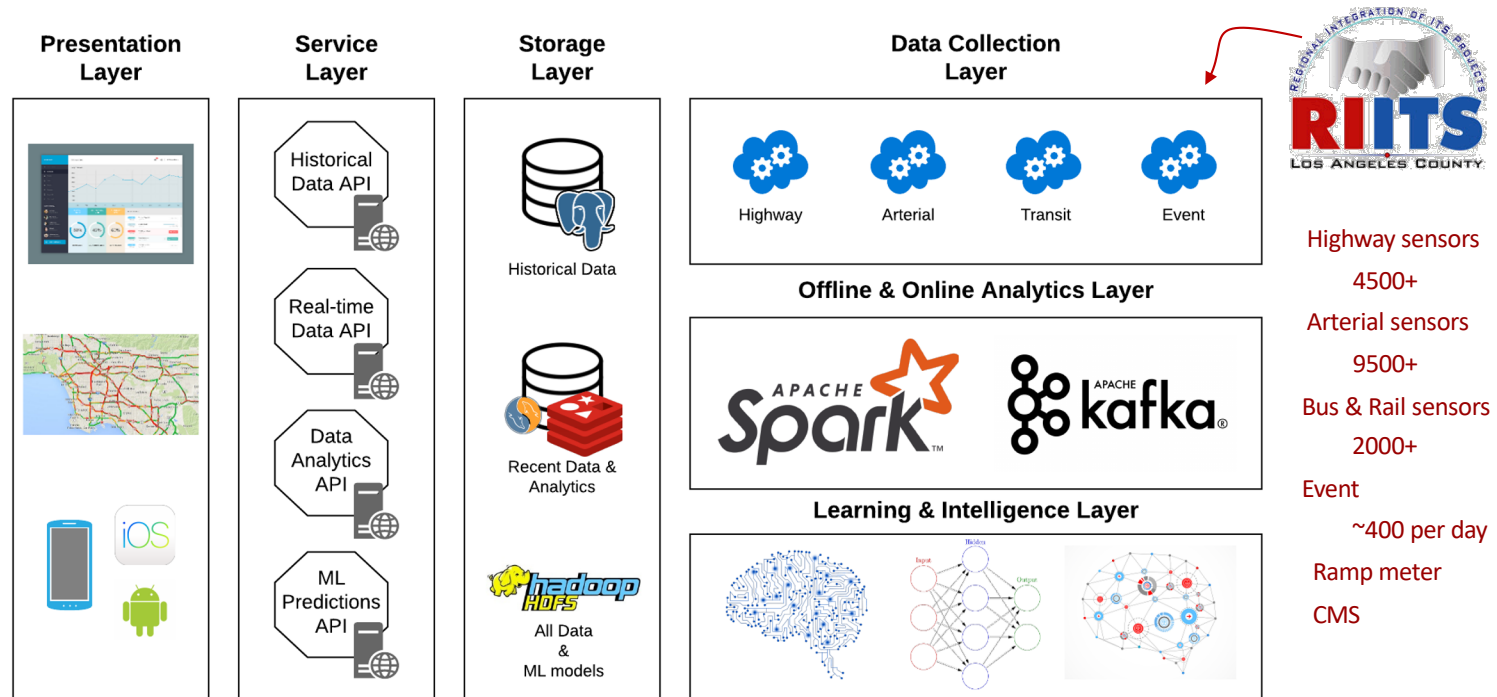
Stripelis, D., Ambite, J. L., Chiang, Y.-Y., Eckel, S. P., and Habre, R. (April 2017). A Scalable Data Integration and Analysis Architecture for Sensor Data of Pediatric Asthma, In *Proceedings of the 2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pp. 1407-1408, San Diego, CA, USA

Analysis Requirements Can Vary

- Real-time prediction of traffic using available traffic sensors and historical data from the last 10 years
- Real-time and offline trajectory mining from large amount of trajectories
 - e.g., next point-of-interest recommendation, moving behavior detection
- Offline line-of-sight analysis of cellphone towers for the entire US

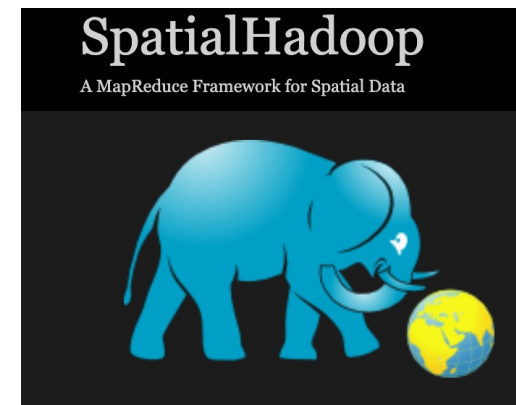
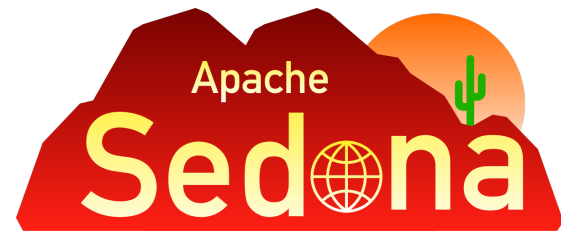


Archived Data Management System



Spatial Big Data Platform

- Like spatial databases, spatial Big Data platforms are software extensions on existing Big Data platforms
 - GeoSpark (Apache Sedona)
 - SpatialHadoop
 - GeoMesa
- Use A Common Programming Model
 - MapReduce



MapReduce

- A programming model for processing Big Data in a parallel distributed fashion
- Many implementations
 - Apache Hadoop
 - Apache Spark



Acknowledgements

- Gil, Yolanda (Ed.) Introduction to Computational Thinking and Data Science. Available from <http://www.datascience4all.org>



<https://creativecommons.org/licenses/by/2.0/>

These materials are released under a CC-BY License

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material
for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

*Artwork taken from
other sources is
acknowledged
where it appears.*

*Artwork that is not
acknowledged is by
the author.*

Please credit as: Chiang, Yao-Yi Introduction to Spatial Artificial Intelligence. Available from <https://yaoyichi.github.io/spatial-ai.html>

If you use an individual slide, please place the following at the bottom: “Credit: <https://yaoyichi.github.io/spatial-ai.html>”

We welcome your feedback and contributions.