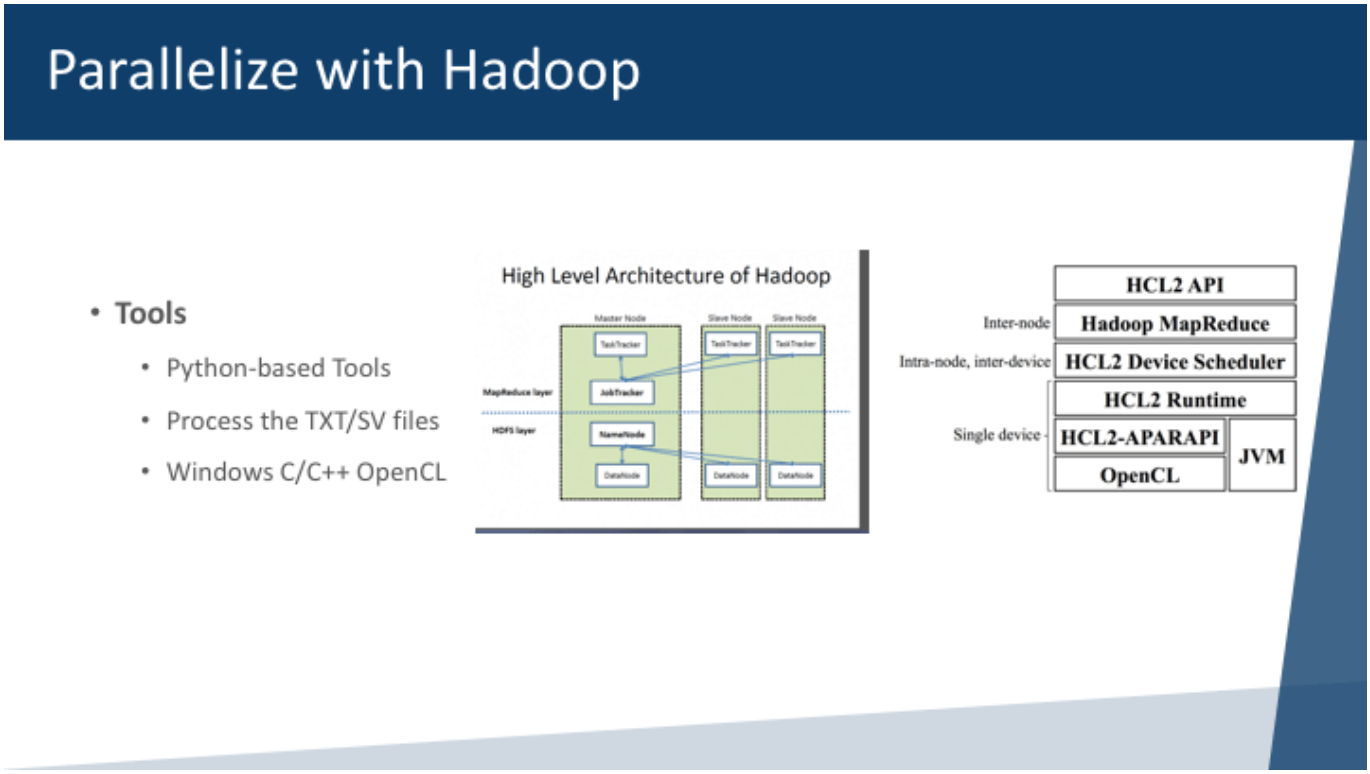
**Edwin Hernandez, PhD**
Innovator, Expert Witness, Engineer, Entrepreneur, Inventor
http://edwinhernandez.com

# Introduction Big Data in RF Analysis




# Big Data in RF Analysis

Big Data provides tools and a framework to analyze data, in fact, large amounts of data. Radio Frequency, RF, provides amounts of information that depending n how it is modeled or created, its analysis fits many statistical models and is in general predicted using passive filtering techniques.

The main tools for Big Data include statistical aggregation functions, learning algorithms, and the use of tools. There are many that can be purchased but many that are free but may require certain level of software engineering. I love Python and specially the main modules used in python are:

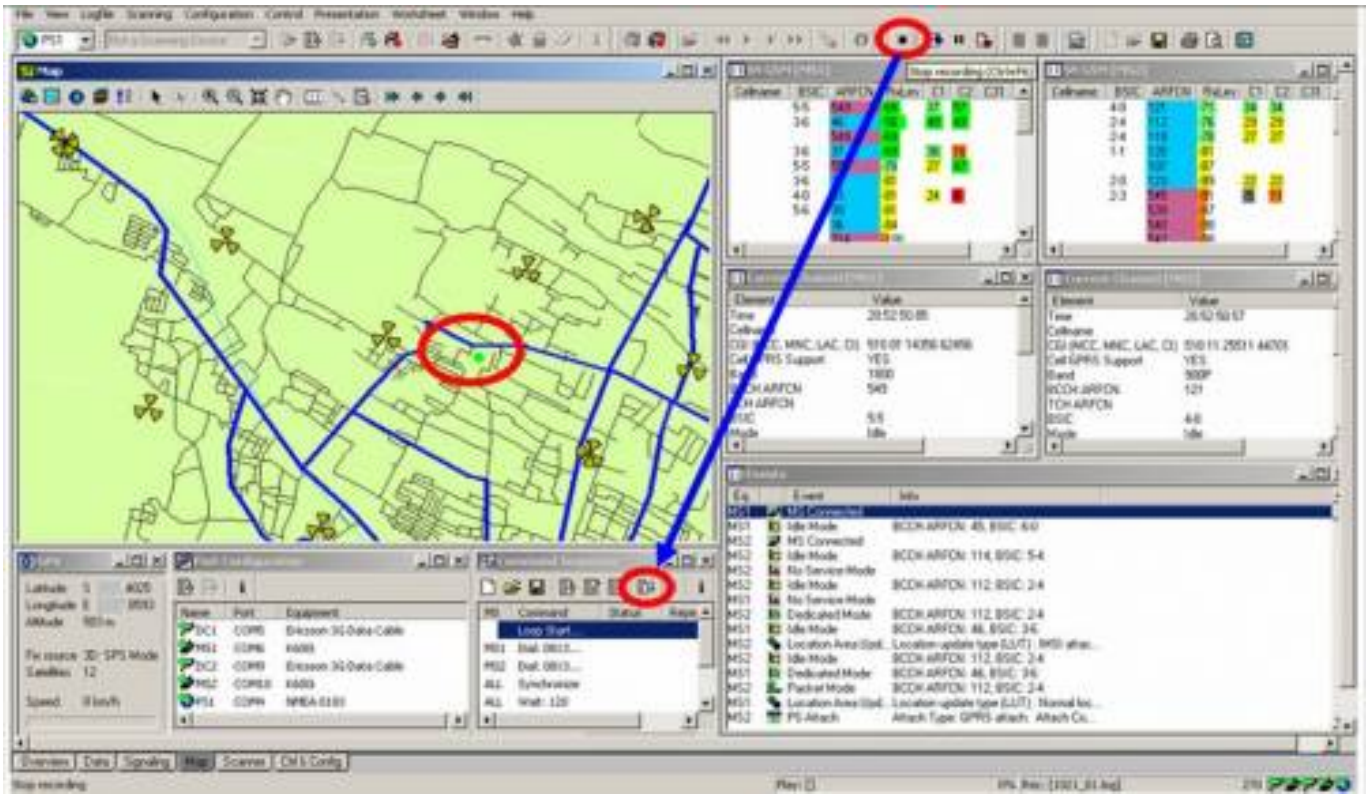
- Pandas
- SciPy
- NumPy
- SKLearn

and, there are many more used for the analysis and post-processing of RF captures.

## Drive Test and Data Simulation

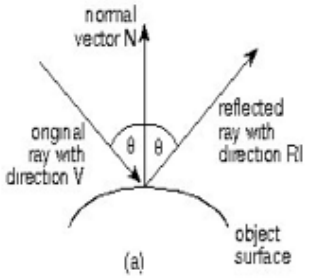
In general, many drive test tools are used to capture RF data form LTE/4G, and many other

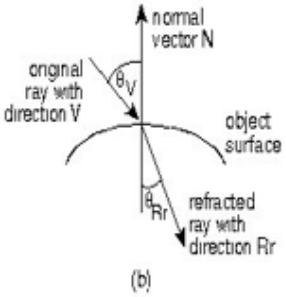
systems. As vendors, we can find Spirent, and many others, and we can capture RF from multiple base stations and map those to Lat/Long in a particular area covered by many base stations. It's obvious that drive test cannot cover the entire area, as expected extrapolation and statistical models are required to complete the drive test.

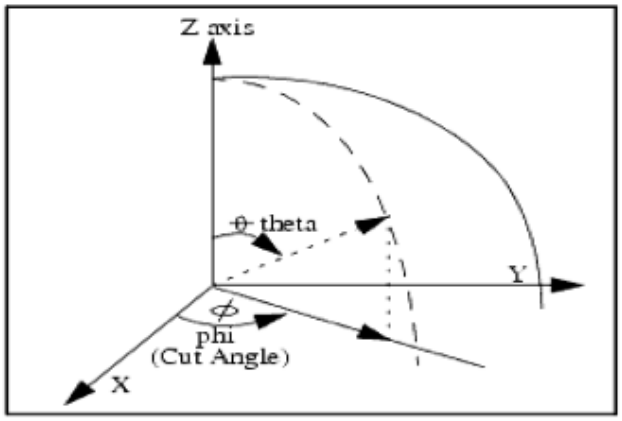


In a simulator, just as in MobileCDS and other simulators, specially those in "Ray Tracing," the simulator uses electromagnetic models to compute the RF received by an antenna.

**Edwin Hernandez, PhD**
Innovator, Expert Witness, Engineer, Entrepreneur, Inventor
http://edwinhernandez.com



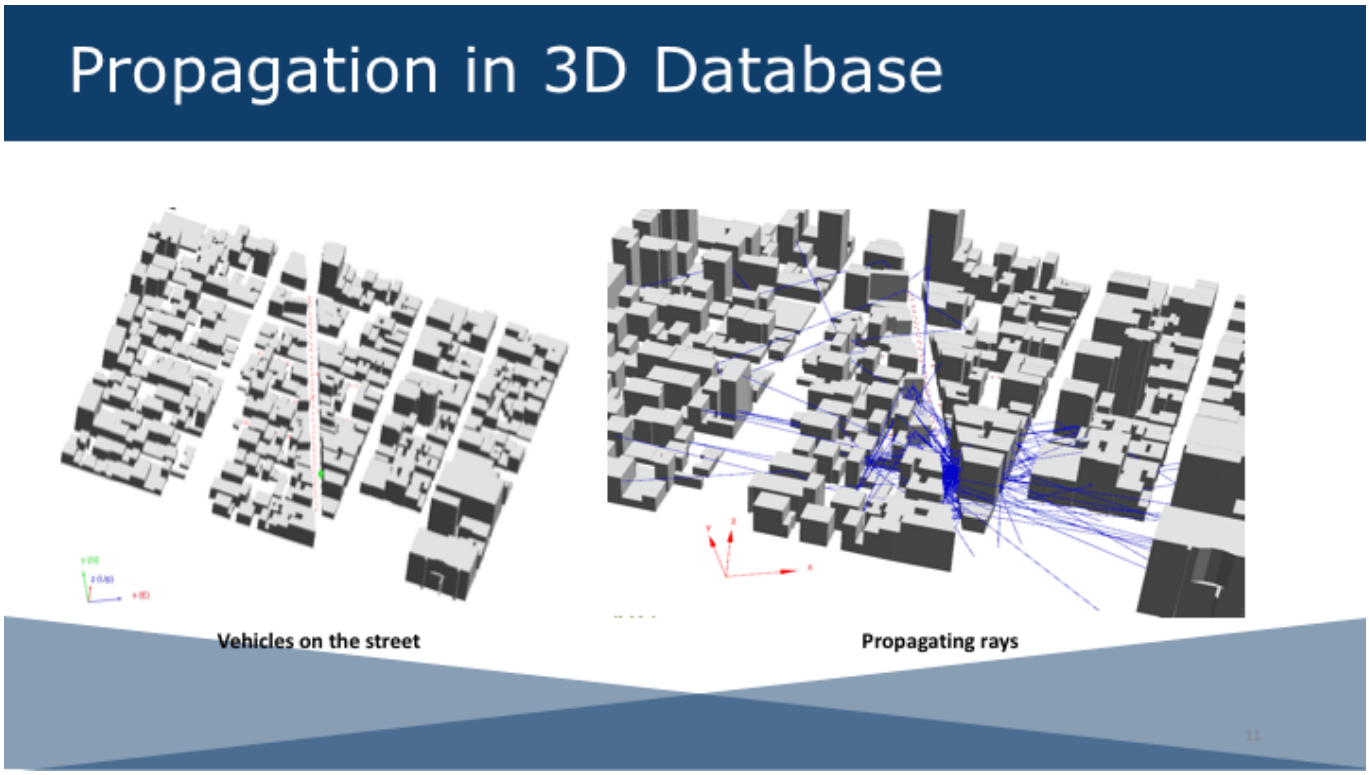




## Methods and Algorithms

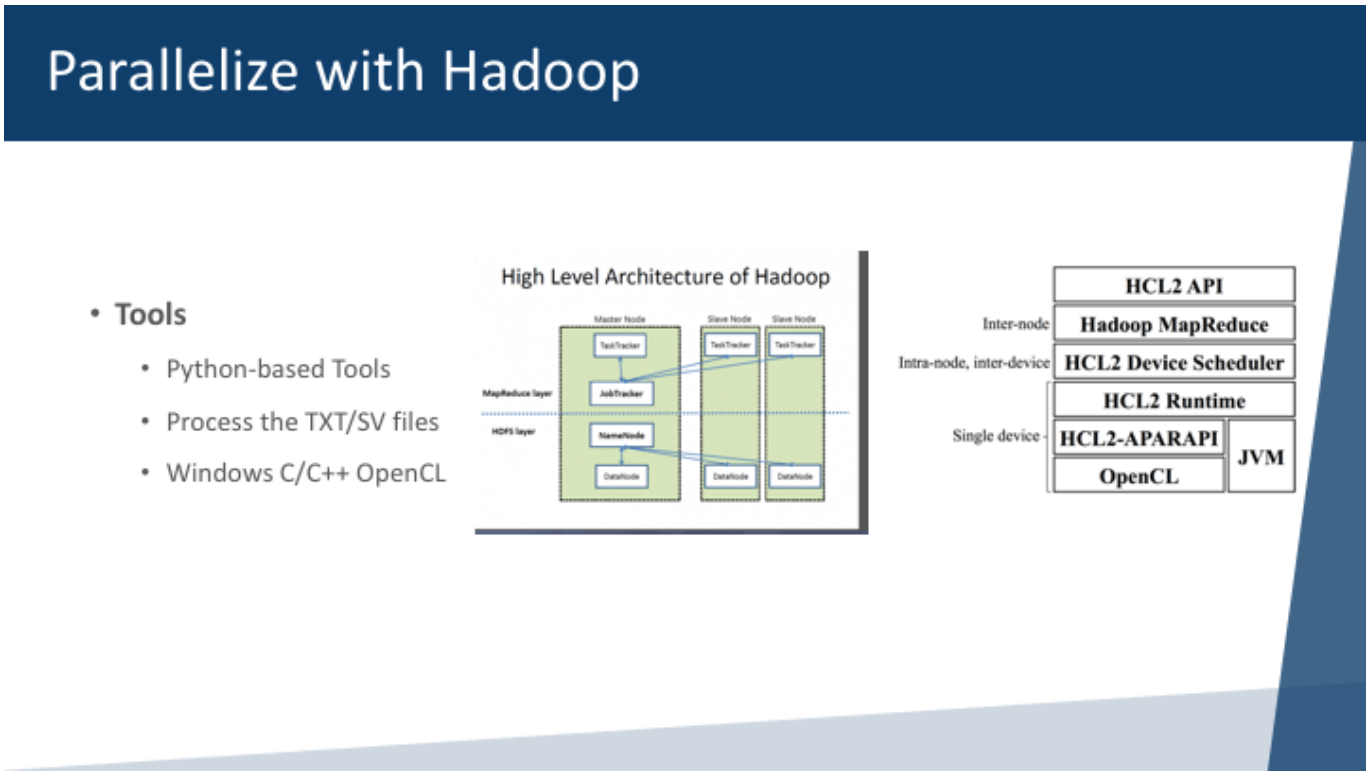
## Big Data Processing for a Massive Simulation

Unstructured data models are loaded with KML and other 3D simulation systems that include polygons and buildings that are situated on top of a google earth map or any other map vendor. The intersection of the model with the 3D database produces the propagation model that needs massive data processing, Map-Reduce and Hadoop to handle the simulation.

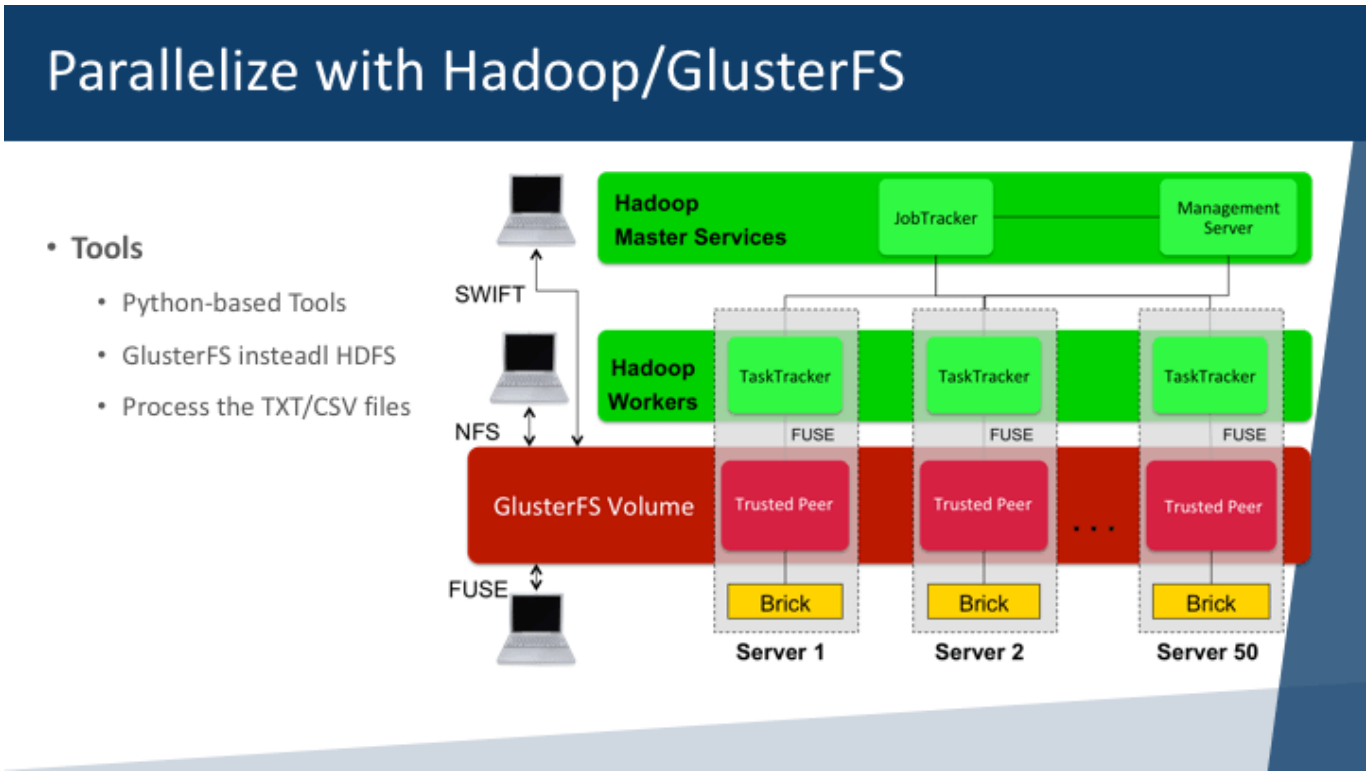



## HADOOP and MAP Reduce for RF Processing

The data is then stored in unstructured models with RF information, that include the Electromagnetic field, frequency, time, delay, error, and other parameters that are mapped to each Lat/Log or x,y, z coordinates in the plane being modeled. The tools are usually written in Python and parallelization can be done in multiple hadoop nodes and processing of CSV/TXT files with all the electromagnetic data and the 3D map being rendered.




As you can see the Hadoop/GlusterFS is our choice, as we don't see that much value for HDFS or the Hadoop Data File System are the ones that handle all the files and worker systems. As you can tell, we are fans of GlusterFS and processing of all Hadoop cluster nodes is managed in a massive processing network of high-performance networks and 10Gb Fiber network.

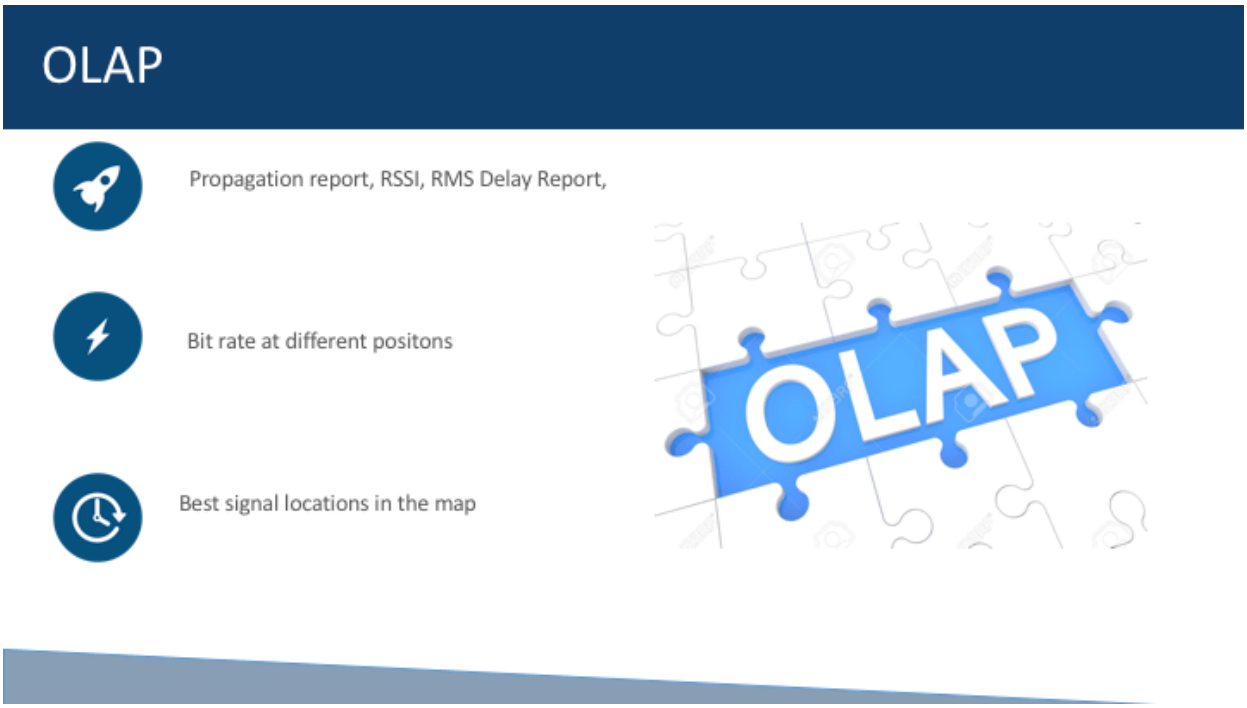



## Big Data models: OLTP and OLAP Processing

The OLTP and OLAP data models definitions can be found online:

> " - **OLTP (On-line Transaction Processing)** is characterized by a large number of short on-line transactions (INSERT, UPDATE, DELETE). The main emphasis for OLTP systems is put on very fast query processing, maintaining data integrity in multi-access environments and an effectiveness measured by number of transactions per second. In OLTP database there is detailed and current data, and schema used to store transactional databases is the entity model (usually 3NF).


- **OLAP (On-line Analytical Processing)** is characterized by relatively low volume of transactions. Queries are often very complex and involve aggregations. For OLAP systems a response time is an effectiveness measure. OLAP applications are widely used by Data Mining techniques. In OLAP database there is aggregated, historical data, stored in multi-dimensional schemas (usually star schema). "



## Conclusion

We have different research areas:

- Analysis of data for handover protocols,
- Data mining for better antenna positioning,
- Machine learning techniques for better PCRF polices and more