Spatial Distribution of Big Data

Report of the NIAS Wednesday Discussion held on 03 August 2016

Prof N Chandrasekaran is an adjunct faculty at the NIAS. Earlier, he served in the capacities of Director, CDAC, Deputy Director, DRDO, Scientist at NAL, OCAM, Cambridge and has also worked with IBM.

Big data is any collection of data sets that are large or complex. Traditional database management like RDBMS have huge limitations when it comes to processing big data. Data science is a new field that involves using methods to analyze massive amounts of data and extract knowledge it contains. Fig 1 captures the three generational changes that have taken place in the last few decades.

The traditional “Structured” data represented by ERP and CRM belongs to the first generation and had long been handled by RDBMS servers like Oracle, MS SQL Server, IBM DB/2, etc. Querying using SQL was fairly simple and straightforward. Web introduced its own data variety and complexity and the storage rose up to Terrabytes. Big data is a recent phenomenon and all that it encompasses are shown in top segment of the figure. Big data includes data transactions, the interaction of data as well as observations. One could see (Fig. 2) how the world of unstructured data (the big data) has for sure emerged in the recent past.

Unlike structured data, which can be handled by single type of RDBMS servers, it becomes much more difficult to store and analyze Big data, see Fig. 3 for variety of NoSQL (Not only SQL) DB servers. One has to be careful in choosing an appropriate DBS. The big giants like Amazon, LinkedIn, Google, Twitter, etc. have opted to employ NoSQL database servers like, Dynamo, Voldemort, BigTable, FlockdB, etc. and in that order. If
your Institution is looking to work on an appropriate Server, this figure should assist you to make that selection. Of course, one needs to match their type of Big Data with the type of data being handled by each one of these companies. This is fairly easy to achieve because most people are familiar with their offerings.

![Big Data types diagram](image)

**Fig. 4 exemplifies some typical SQL and NoSQL commands.**

**Several applications of Spacial distribution of big data were demonstrated during the lecture. The first pertained to a primitive application that used data available on a spread sheet with one of the columns, representing the Location details (like altitude and longitude or phrase which is understood by Google Search). This can be linked and used in conjunction with ‘Maps Engine of Google’ for superposition of the information that has been collected, see Fig. 5 and 6.**

**Video Synopsis features of the Israeli ‘BriefCam’ tool, which provides a synopsis that**
The most important aspect of the lecture related to CARTO Builder, which is a web-based drag-and-drop analysis tool that can use a variety of data sheets from public platforms.

Fortunately, the Builder can be effectively used with knowledge of just two of the features, viz., "Widgets and Predicting Functions" and without any coding knowledge. However, field experts can provide assistance to facilitate...
Carto for location analytics is like Dreamweaver for budding website builders with no coding knowledge. It will also be possible to perform in-depth querying through a SQL console and there is also a Carto-CSS console view for custom styling. The idea is to open up a broader skill set. Fig. 8 shows both the 'Map View'. On to its left is the 'Table' button, which will help you to drop tables for big data input. The beauty is that this product works on Google Maps. Some reference was made to the work on Solar, Wind and Refrigeration (also Global Warming) big data, as it applies to Mauritius. The same technique can be applied to interpret Heritage site data too.

Figure 8 identifies two key trends in the movement from mapping to map analysis. Traditional GIS treats geographic space in a manner similar to our paper map legacy. Points, lines and polygons are used to define discrete spatial objects, such as houses, streams and lakes. In turn, these objects are linked to attributes in a database that describe their characteristics and conditions. The result is a tremendously useful system enabling users to make complex geo-queries of the information and then map the results.