Big Data analytics in the Geo-Spatial Domain

Romulo Goncalves[1], Milena Ivanova[1], Martin Kersten[2], Henk Scholten[3], Sisi Zlatanova[4], Foteini Alvanaki[5], Pirouz Nourian[4], and Eduardo Dias[3]

[1]*NLeSC Amsterdam, The Netherlands* {{r.goncalves, m.ivanova}@esciencecenter.nl}

[2]*CWI Amsterdam, The Netherlands* {{martin.kersten@cwi.nl}

[3]*VU Amsterdam, The Netherlands* {{hscholten@feweb., e.dias@}vu.nl}

[4]*TU Delft, The Netherlands* {{s.zlatanova, p.nourian}@tudelft.nl}

[5]*MonetDB Solutions, The Netherlands* {foteini.alvanaki@monetdbsolutions.com}

Big data collections in many scientific domains have inherently rich spatial and geo-spatial features. Spatial location is among the core aspects of data in Earth observation sciences, astronomy, and seismology to name a few. The goal of our project is to design an efficient data management layer for a generic geo-spatial analysis system with focus on three dimensional (3D) city models.

Digital 3D city models play a crucial role in research of urban phenomena; they form the basis for flow simulations (e.g. wind streams, water runoff and heat island effects), urban planning, and analysis of underground formations. Urban scenes consist of large collections of semantically rich objects which have a large number of properties such as material and color. Modeling and storing these properties indicating the relationships between them is best handled in a relational database.

The provision of spatial and geo-spatial features in database systems needs to be extended and brought to maturity to fulfill the requirements of real-world scientific applications. A class of DBMSs, called column-stores, have proven efficiency for analytic applications on extremely large data sets. Column stores have become the de-facto standard for managing large data warehouses. Although column stores have a proven track record in business analytics, their pros- and cons- for GIS applications are not yet well understood.

Our goal is to have a spatial DBMS which iteratively loads data from different sources and converts it into a common format to enable 3D operations and analyses, such as 3D intersections, and semantic properties management.