Geospatial Big Data: Challenges and Opportunities ☆

Jae-Gil Lee *, Minseo Kang

Department of Knowledge Service Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

ARTICLE INFO

ABSTRACT

Geospatial big data refers to spatial data sets exceeding capacity of current computing systems. A significant portion of big data is actually geospatial data, and the size of such data is growing rapidly at least by 20% every year. In this paper, we explore the challenges and opportunities which geospatial big data brought us. Several case studies are introduced to show the importance and benefits of the analytics of geospatial big data, including fuel and time saving, revenue increase, urban planning, and health care. Then, we introduce new emerging platforms for sharing the collected geospatial big data and for tracking human mobility via mobile devices. The researchers in academia and industry have spent a lot of efforts to improve the value of geospatial big data as well as take advantage of its value. Along the same line, we present our current research activities toward the analytics of geospatial big data, especially on interactive analytics of real-time or dynamic data.

1. Introduction

Geospatial data has always been big data. In these days, big data analytics for geospatial data is receiving considerable attention to allow users to analyze huge amounts of geospatial data. *Geospatial big data* typically refers to spatial data sets exceeding capacity of current computing systems. McKinsey Global Institute says that the pool of personal location data was in the level of 1 PB in 2009 and is growing at a rate of 20% per year [1]. This estimation did not include the data from RFID sensors and those stored in private archives. According to the estimation by United Nations Initiative on Global Geospatial Information Management (UN-GGIM), 2.5 quintillion bytes of data is being generated every day, and a large portion of the data is location-aware. Also, in Google, about 25 PB of data is being generated per day, and a significant portion of the data falls into the realm of spatio-temporal data [2]. This trend will be even accelerated since the world becomes more and more mobile in these days. As in Fig. 1, in India, the internet traffic from mobile devices already exceeded that from desktop computers [3].

Along with this exponential increase of geospatial big data, the capability of high performance computing is being required greatly than ever, for modeling and simulation of geospatially enabled contents. However, because of limited processing power, it has been hard to fully exploit high-volume or high-velocity collection of geospatial data in many applications. Recently, distributed, parallel processing on a cluster of commodity computers or a cloud such as Amazon EC2[1] has been becoming widely available for use, breaking the existing limitations on processing power. In addition, big data platforms such as Hadoop [4], Hive [5], and MongoDB [6] have been developed such that users can implement big data analytics software very easily on a distributed, parallel computing platform. It is obvious that these recent improvements are providing us with a lot of opportunities for advanced analytics for geospatial big data [7–9]. According to Garner's hype cycle in Fig. 2, geospatial big data analytics belongs to the stage of *peak of inflated expectations* as of July 2012 [10].

Geospatial big data or simply spatial big data are societal opportunities [11,12]. The Millennium Project identified 15 global challenges that the human kind is facing as in Fig. 3 [13]. Many of them can benefit from geospatial big data. Shashi Shekhar [14], a renowned computer scientist, says that the seven challenges are related to geospatial big data, as indicated by boxes in the figure. For example, as for energy, eco-routing is one example that can save energy using geospatial big data. This technology minimizes fuel consumption rather than travel time or travel distance. For this purpose, eco-routing tries to find a route that avoids congestion, idling at red lights, turns and elevation changes, and so on. Compared to using the "Fastest Route" option, Ford researchers told that using the "Eco Route" option offered as much as 15% reduced

In India, Mobile Internet Traffic Surpassed Desktop Internet Usage in May, 2012 - Other Countries to Follow…

India Internet Traffic by Type, Desktop vs. Mobile, 12/08 – 11/12

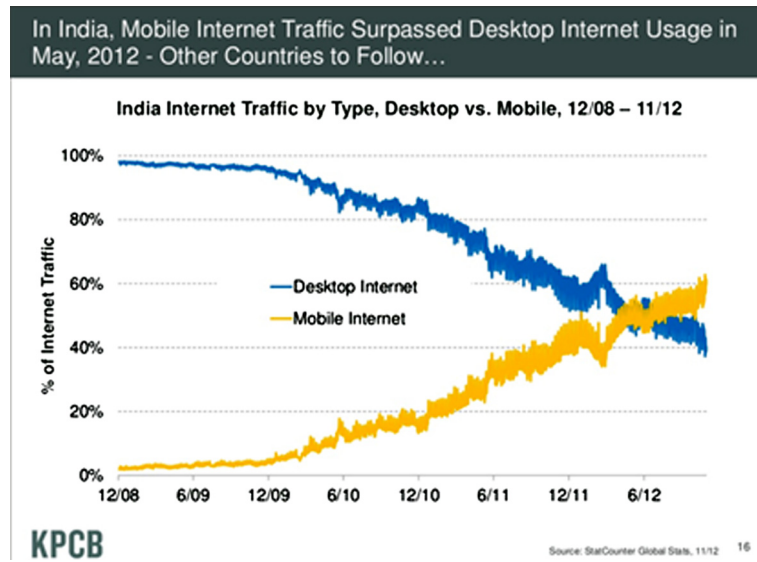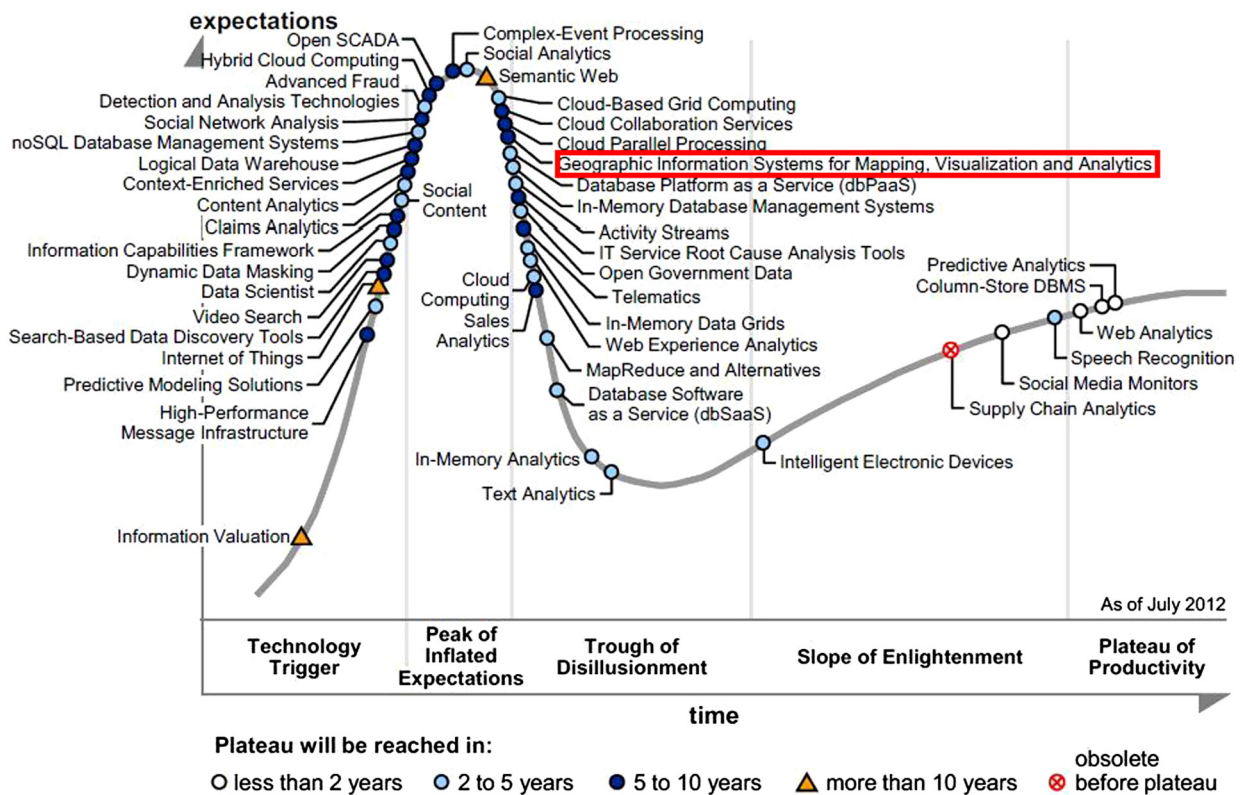**Fig. 1.** Mobile internet traffic in India [3].

Fig. 2. Gartner's hype cycle (source: Gartner) [10].

fuel consumption in some of their vehicles [15]. Now, in many Ford cars, we can find the eco-route option, as in Fig. 4.

McKinsey Global Institute conducted a study on how big data can innovate our world [16]. As for geospatial big data, the study says "the use of personal location data could save consumers worldwide more than $600 billion annually by 2020." One can find out users' current locations by tracking their mobile devices such as smart phones. The study mentioned geosocial networking services such as Foursquare used for locating friends and for finding nearby stores and restaurants, where many users check-in at various places and reveal their current location [17]. On the other

hand, according to the study, the biggest consumer benefit will be obtained from time and fuel saving thanks to location-based services that, by taking account of real-time traffic and weather data, help driver avoid traffic congestion and recommend alternative routes. Location tracking can be done by using a driver's smart phone or a global positioning system (GPS) equipped with a car.

2. Power of location

Sir Martin Sorrell [18], the CEO of WPP Group, says "Location targeting is holy grail for marketers." Big data analytics is an effec-