Virtual Observatory for twitter messages

Gergő Gombos[a], Zoltán Vincellér[a], Attila Kiss[a]

[a]Eötvös Loránd University
{ggombos,vzoli,kiss}@inf.elte.hu

Abstract

A new research direction has emerged as the investigation of On-line Social Networks. The Twitter is one of the most well-known social networks. Analysis of the Twitter is easier than other social networks because it provides the opportunity for collecting and downloading of a certain percentage of the messages without any restrictions. There are several researches on topics as detecting news and events, human behaviors, analyzing and mining of opinions. The on-line messages are available only through a continuous stream. To store the messages from the stream effectively and efficiently is serious challenge against software system design and architecture [3]. Every day about 10 GBs data are generated by this way and storing of this volume of data is not an easy task. In this paper we present a technique and architecture for collecting and storing the messages of the Twitter, and we present a prototype where data can be accessed for further analysis. Our system makes use specific techniques and methods of Oracle environment. Our software architecture approach is in contrast to previous solutions in which the systems use MS Sql [1] or MySql [2] DBMS . We demonstrate that indexing and Job scheduler of the Oracle provide advantages to retrieve and handle large amounts of data.

Keywords: Database, Architrecture, Social network

MSC: 68P20,68M01,68M11

References

[1] L. Dobos, J. Szüle, T. Bodnár, T. Hanyecz, T. Sebők, D. Kondor, Zs. Kallus, J. Stéger, I. Csabai, G. Vattay A multi-terabyte relational database for geo-tagged social network data, *CogInfoCom* (2013)

[2] M. Oussalah, F. Bhat, K. Challis, and T. Schnier. A software architecture for twitter collection, search and geolocation services., *Knowl.-Based Syst.*, (2013), 105–120.

[3] B.Molnár, Z. Vincellér Comparative study of Architecture for Twitter Analysis and a proposal for an improved approach, *CogInfoCom*, (2013)