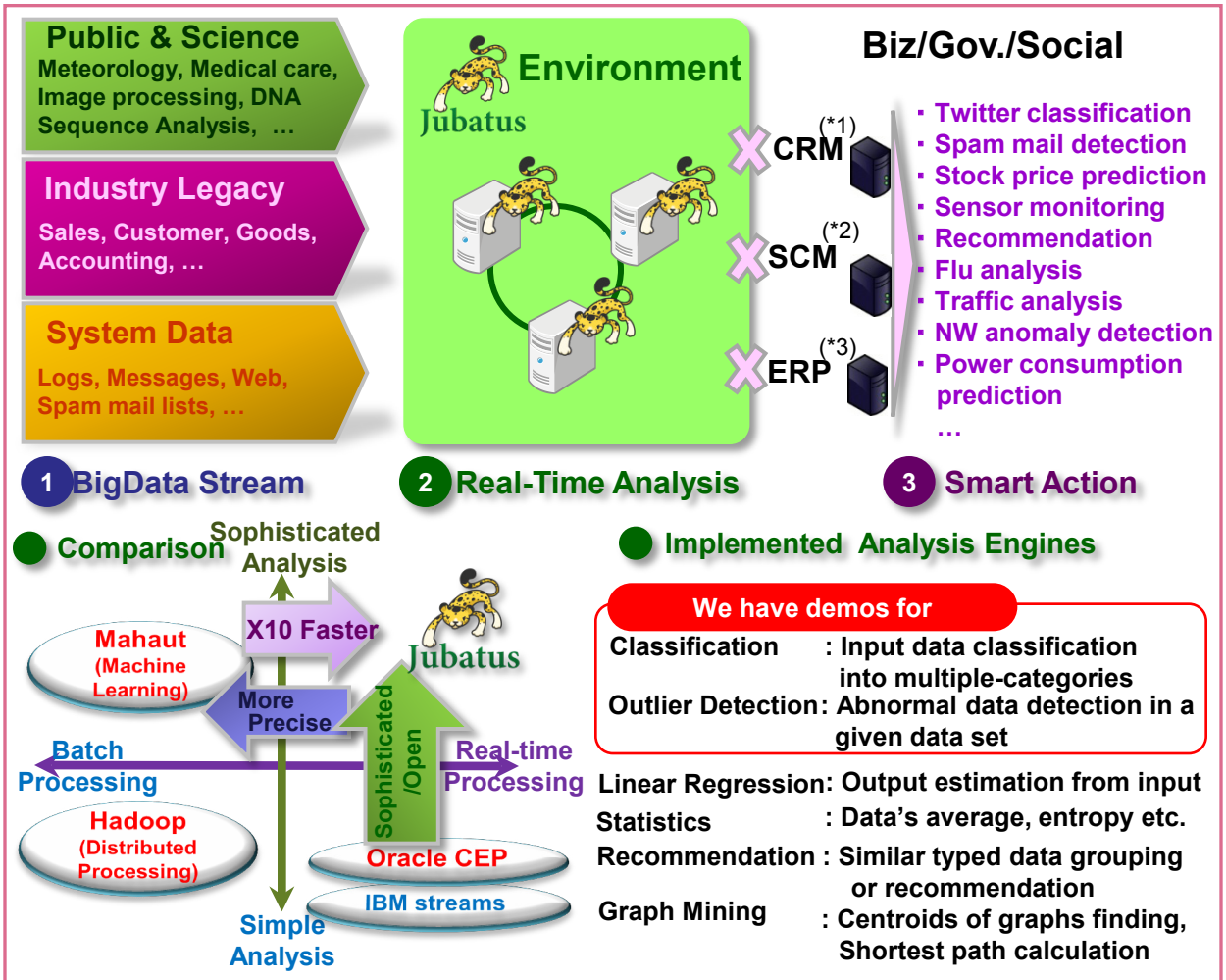
2 Smarter instant analysis of “Current” with big data

Streaming analysis with distributed online machine learning

***Abstract*— In the last decade, huge quantities of unstructured data such as micro-blog posts have been produced. Many studies and surveys have pointed out the potential benefits of the real-time analysis of these unstructured big data. Therefore, sophisticated data analysis, which can deal with unstructured data in a real-time manner, has become an emerging trend. However, real-time analysis and sophisticated analysis inherently have a trade off relationship. We studied a way of balancing these contradictory features at a high level. As a result of our study, we introduced Jubatus, which is a distributed real-time data analysis framework. With Jubatus, we can analyze and classify natural language data in a 16MB/s stream. By offering Jubatus as open source software, we are contributing to real-time marketing and smart social infrastructure management.**



* This framework is developed with Preferred Infrastructure , Inc. as an open source software. http://jubat.us
(*1)Customer Relationship Management, (*2)Supply Chain Management , (*3) Enterprise Resource Planning

Related works

[1] S. Oda, S. Nakayama, K. Uenishi, S. Kinoshita, “Jubatus: Distributed Processing Technique Enabling realtime Processing of Big Data”, *IEICE Tech. Rep.*, Vol. 111, No. 409, IN2011-126, pp. 35-40, 2012. (in Japanese)
[2] H. Makino, “Jubatus: Scalable Distributed Processing Framework for Realtime Analysis of Big Data”, in *Proc. XLDB2012,* 2012.
[3] K. Horikawa, Y. Kitayama, S. Oda, H. Kumazaki, J. Han, H. Makino, M. Ishii, K. Aoya, M. Luo, S. Uchikawa, “Jubatus in Action: Report on Realtime Big Data Analysis by Jubatus”, *NTT Technical Review*, Vol. 10, No.12, 2012.

Contact

Keitaro Horikawa Distributed Computing Technology Project, NTT Software Innovation Center
E-mail : horikawa.keitaro{at}lab.ntt.co.jp (Please replace {at} with @)