BiLoDEx

Big Log Data Exploitation

Industrial research project funded (2014 - 2017)

BiLoDEx (Big Log Data Exploitation) is a research project aiming at developing a software-based prototype for analyzing and visualizing data streams coming from different sources. In today's data centers a large amount of heterogeneous system, services and applications log-data is available. However, a well-known problem is that potentially valuable information is hidden among the "big data" and therefore not easy to be identified or correlated to be understood in the full context. In addition, large volumes of unstructured data cannot be processed automatically or manually. Starting point of this project is a big data warehouse filled with relevant log data.

Based on this we pursue the following research approach:

1. In order to exploit the data streams stemming from the data warehouse relevant analytics data will be semantically enriched, homogenized and filtered.
2. The semantically enriched log data will be the basis for composing several data streams, each for one of the source logs or combination of several logs.
3. The input streams will be harmonized, enhanced by several parameters and integrated into one main master data stream.
4. In case historical information is available steps 1 to 3 may also apply to historical log data.
5. The streamed data will be the source for visual analytics, monitoring, risks and performance influence prediction. BiLoDEx will experiment with different visualization techniques.
6. Basic alert management features will be integrated. In the case defined relevant parameters and/or predefined thresholds are reached a notification process will be initialized.

The BiLoDEx project outcome will be a scientific software-based prototype allowing users to carry out historical and real-time analysis of the preprocessed IT systems related data streams. Moreover, improved decision making and real time alerting in the domains of security-, trust-, risk- and performance management and business forensics will be offered. In this context, visual analytics techniques will combine the strengths of automated machine processing technologies and human cognitive capabilities. The prototype will be based on open-source technologies and open standards as far as possible.

The project started on March 1, 2014 and will run over a period of 3 years