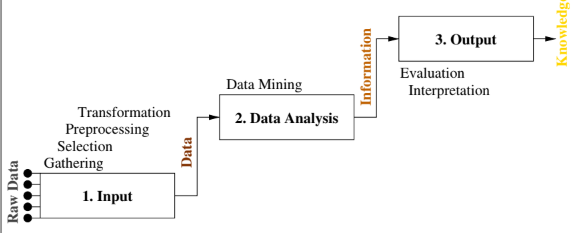
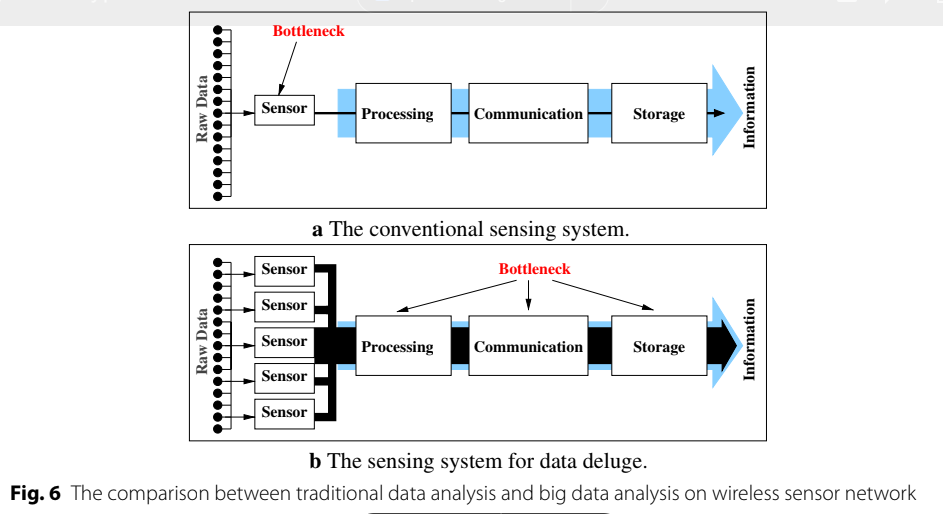
**[Big data analytics: a survey](https://drive.google.com/drive/folders/13iE9NU7ZXV5YtRgEab_uOtwFu0_z2xKm)**

**Abstract**

**The age of big data is now coming. But the traditional data analytics may not be able to handle such large quantities of data. The question that arises now is, how to develop a high performance platform to efficiently analyze big data and how to design an appropriate mining algorithm to find the useful things from big data. To deeply discuss this issue, this paper begins with a brief introduction to data analytics, followed by the discussions of big data analytics. Some important open issues and further research directions will also be presented for the next step of big data analytics. Keywords: Big data, data analytics, data mining.**

****

****

**Ham and Lee used the domain knowledge, B-tree, divide-and-conquer to filter the unrelated log information for the mobile web log analysis. A later study [75] considered that the computation cost of pre- processing will be quite high for massive logs, sensor, or marketing data analysis. Thus, Dawelbeit and McCrindle employed the bin packing partitioning method to divide the input data between the computing processors to handle this high computations of pre- processing on cloud system.**

**Jun et al. attempted to use the FPGA to accelerate the compression process.**

**Zou et al. [79] employed the tentative selection and predictive dynamic selection and switched the appropriate compression method from two different strategies to improve the performance of the compression process.**

**(1) Processing/Compute: Hadoop [83], Nvidia CUDA [84], or Twitter Storm [85], (2) Storage: Titan or HDFS, and (3) Analytics: MLPACK [86] or Mahout [87]. Although there exist commercial products for data analysis [83–86],**

**Huai et al. [88] presented a matrix model which consists of three matrices for data set (D), concur- rent data processing operations (O), and data transformations (T), called DOT.**

**generalized linear aggregates distributed engine (GLADE). The GLADE is a multi-level tree-based data**

**Since big data analysis is generally regarded as a high computation cost work, the high performance computing cluster system (HPCC) is also a possible solution in early stage of big data analytics. Sagiroglu and Sinanc [105] therefore compare the char- acteristics between HPCC and Hadoop.**

**Fan and Bifet pointed out that the terms “big data” and “big data mining” were first pre- sented in 1998,**

**++++++++++++++++++++++++++++++**

**Later to read .**

**algorithms—clustering, classification, associ-**

**ation rules, and sequential patterns—will**