

BIG DATA ANALYTICS

0 0 1 1 0 0 0 1 0 0 1 1 1 0 0 1 0 0 1 1 0 1 1 0 0 0 1 1 0 1 1 1

NC State Students are writing the book on “big data” - sifting useful information out of a sea of business, personal and other data available online and elsewhere. Or at least they’ve written about mining big data.



Big data analytics is the process of examining large amounts of data of a variety of types (big data) to uncover hidden patterns, unknown correlations and other useful information. Such information can provide competitive advantages over rival organizations and result in business benefits, such as more effective marketing and increased revenue.

The primary goal of big data analytics is to help companies make better business decisions by enabling data scientists and other users to analyze huge volumes of transaction data as well as other data sources that may be left untapped by conventional business intelligence (BI) programs. These other data sources may include Web server logs and Internet clickstream data, social media activity reports, mobile-phone call detail records and information captured by sensors. Some people exclusively associate big data and big data analytics with unstructured data of that sort, but consulting firms like Gartner Inc. and Forrester Research Inc. also consider transactions and other structured data to be valid forms of big data.

Big data analytics can be done with the software tools commonly used as part of advanced analytics disciplines such as predictive analytics and data mining. But the unstructured data sources used for big data analytics may not fit in traditional data warehouses. Furthermore, traditional data warehouses may not be able to handle the processing demands posed by big data. As a result, a new class of big data technology has emerged and is being used in many big data analytics environments. The technologies associated with big data analytics include NoSQL databases, Hadoop and MapReduce. These technologies form the core of an open source software framework that supports the processing of large data sets across clustered systems.

Potential pitfalls that can trip up organizations on big data analytics initiatives include a lack of internal analytics skills and the high cost of hiring experienced analytics professionals, plus challenges in integrating Hadoop systems and data warehouses, although vendors are starting to offer software connectors between those technologies.

Faculty

Dr. Nagiza Samatova

For More Information:

Dr. Nagiza Samatova
samatova@ncsu.edu
(919)513-7575

Room 2235
EB2, 890 Oval Dr.
Raleigh, NC 27695-8206

www.csc.ncsu.edu