BIG DATA TECHNOLOGY IN INVESTMENT BANKING

UNDERSTANDING THE FUNDAMENTAL CONCEPTS OF BIG DATA HADOOP

Distinguishing between the reality and the hype surrounding big data technology is challenging, and it is all the more confusing when technologists throw around jargon. Big data technology is not a mystery, and it can be understood once the basics are explained.

BIG DATA DEFINED

- Structured, semi-structured and unstructured data
- Includes 'messy data' – inconsistently formatted or poor quality data
- Supports analysis on the whole universe of data rather than a sample
- Frequently looks for correlations in data rather than attempting to determine causality

HADOOP

Hadoop is an open-source software framework for the distributed storage and processing of large data sets on clusters of industry standard hardware. Data is partitioned onto multiple servers for storage and processing, and the findings from each are drawn together as analysis of the complete dataset.

The Hadoop market is forecast to grow at a compound annual growth rate of

58%

> USD 1bn by 2020

DATA INGESTION

Data ingestion describes the process of transferring, loading and processing data for later use or storage in a database from a variety of sources as well as modifying and formatting individual files.

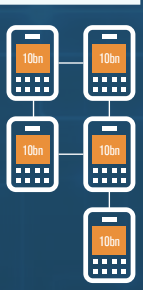
DATA MANAGEMENT

Data management describes the organisation, administration and governance of large volumes of both structured and unstructured data.

DATA PROCESSING

Data processing describes operations on data to retrieve, transform or classify information.

By 2020, there will be over



50bn

smart-connected devices in the world to collect, analyse and share data

The digital universe of data will increase from

4.4 ZETTABYTES

in 2015 to 10 times as much by 2020

BUSINESS BENEFITS

For banks, embracing Big Data technology can:

- Provide users with insights that traditional data analysis cannot do or cannot process as quickly;
- Produce results supported by a larger data population and that have more statistical certainty;
- Include a wider array of relevant information in analysis; and
- Eliminate the need for manually searching through unstructured data.

For more information please visit:
www.research.greyspark.com

capital markets intelligence
greysparkcmi