Co-funded by the Erasmus+ Programme of the European Union

What is Big Data

Big Data is ***very large***. In fact, its initially too cumbersome to obtain useful or valuable information. Some amount of pre processing is therefore required. Some cloud computing platforms come with tools that can do this, for example Google or Hadoop. Hadoop is a set of Open Source software tools which can help with the processing and storage of large amounts of data.

“Oracle defines big data as an aggregation of data from three sources: traditional (structured data), sensory (log data, metadata), and social (social media)”

The hooped diagram gives you an idea of the extent of the sources ***(variety)***, time scale ***(velocity)*** and ***volume*** (now regularly in the ***petabyte*** scale). One petabyte is 1,000,000 GB and one petabyte of average MP3 songs (for mobile, roughly one megabyte per minute), would require ***2000 years*** to play. A further difficulty is that much of this information will be uncertain, therefore we might question its ***veracity.***

MapReduce

The flow diagram indicates MapReduce - a programming model and implementation for processing and generating big data sets with a parallel, distributed algorithm on a cluster

A MapReduce program is composed of a Map() procedure that performs filtering and sorting (such as sorting students by first name into queues, one queue for each name) and a Reduce() method that performs a summary operation (such as counting the number of students in each queue, yielding name frequencies).

This project has been funded with support from the European Commission.
This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Machine learning is one of many subfields of artificial intelligence, concerning the ways that computers learn from experience to improve their ability to think, plan, decide, and act.

As Tekniker put it to us earlier in the year - “Analyse -> Predict -> Act”

Where Does the Data Come From - Horizontal

Data is generated at several places points along the production line (as shown) – including when the customer places the order where high levels of customisation are increasingly common – facilitated by the digitisation of the industrial process.

Where Does the Data Come From - Vertical

Here the data comes from the customers’ use of products and the support companies responsible for care and maintenance. With Industry 4.0 the manufacturers will assimilate the operational data with the objective of improving future designs – in real-time.

Smart Connected Products

Physical components comprise the product’s mechanical and electrical parts. In a car, for example, these include the engine block, tires, and batteries.

Smart components comprise the sensors, microprocessors, data storage, controls, software, and, typically, an embedded operating system and enhanced user

This project has been funded with support from the European Commission.
This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

interface. In a car, for example, smart components include the engine control unit, antilock braking system, rain-sensing windshields with automated wipers, and touch screen displays.

Connectivity components comprise the ports, antennae, and protocols enabling wired or wireless connections with the product. Connectivity takes three forms, which can be present together:

- ***One-to-one***: An individual product connects to the user, the manufacturer, or another product through a port or other interface—for example, when a car is hooked up to a ***diagnostic*** machine.
- ***One-to-many***: A central system is continuously or intermittently connected to many products simultaneously. For example, ***Tesla*** automobiles are connected to a system that monitors performance and accomplishes remote service and upgrades - the autonomous car will determine whether it is in need of repair, which may be a software update.
- ***Many-to-many***: Multiple products connect to many other types of products and often also to external data sources. An example shown during the LAIT week was agricultural: various types of farm equipment are connected to one another, and to GPS data, to coordinate and optimize the farm system. For example, automated tillers inject nitrogen fertilizer at precise depths and intervals, and seeders follow, placing seeds directly in the fertilized soil.

This project has been funded with support from the European Commission.
This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Rolls-Royce; Orona Lifts

Exabytes

(10^{18} bytes) Amount of data being generated is accelerating; exponential curve shown. As connected sensors are rolled out they are further contributing to the Internet of Things. The supporting technologies need to be capable of dealing with all of this:

- Data Acquisition: Wireless connected devices / sensors - low cost RFID; volume increase as videos are analysed - real-time
- Communication - Facebook and Microsoft working on 6,600km cable under the Atlantic
- Storage - cost per GB storage continues to decline; 90% reduction between 2002 - 2012; server cost also declining rapidly - 40% over the same timescale
- Processing - Moore's Law - processing power doubles every 2 years; physical limits of memory, contribution of quantum computing - illustration of uncertainty
- IoT Platforms and 'packaged' solutions

Have technologies advanced enough? – YES!

Cost reduction in acquisition, Communication, Storage & processing

…. But just a few concerns still not solved…

This project has been funded with support from the European Commission.
This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Autonomy

The objective is Autonomy. The capabilities of smart, connected products can be grouped into four areas: monitoring, control, optimization, and autonomy. Each builds on the preceding one; to have control capability, for example, a product must have monitoring capability.

Big Data Landscape 2017

Looks like a lot of work went into producing this? Jim Hao was probably helped by an app. - impressive all the same.

Shows the sheer scale and level of interest and developments surrounding Big Data - FirstMark is a New York venture capital firm.

Some of the big names are omnipresent: Oracle, Google, Microsoft, SAP (recently installed into Don and Low - now Greek owned - textiles factory in Forfar), Amazon. Others may be known to this audience - Jawbone, nVidia, the infamous Autonomy (purchased by HP then sold on to Micro Focus) and Matlab. There appear to be many niches, filled by specialists

Yes, it's intimidating but remember this is still a developing field; I imagine the numbers will come down as dominant players emerge. Intimidating perhaps but the variety and number of players indicates - for those who are interested - there will be

This project has been funded with support from the European Commission.
This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

sought after skills, particularly interdisciplinary - for example mechanical engineering + machine learning

For the moment there are Open Source operators and Tekniker guided us toward these as being more appropriate for our target audience - SMEs. We would be particularly appreciative of further assistance from our colleagues in the computing department should we be successful in the next stages of our endeavour.

This project has been funded with support from the European Commission.
This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.