e-COMMA

Data Specialist

TM 02: Technical Basics

Data Specialist|BM 02: Technical Basics

Training Contents
1. Data Specialist: introduction 1.1 Role description 1.2 Main tasks 1.3 AARRR
2. Data in e-Commerce 2.1 Overview 2.2 Big Data in e-Commerce
3. Databases 3.1 Overview 3.2 Tools

Data Specialist|BM 02: Technical Basics

Sources and further reading

Fan S. et al. (2015). Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix.
Fahl J. (2017). Data Analytics
Develleno M. (2017). Automate and Grow.
Schweidel D.A. (2017). Profiting from the Data Economy
Kotler P. (2017). Marketing 4.0: Moving from Traditional to Digital
http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data
http://www.business2community.com/big-data/big-data-and-analytics-value-chain-cross-section-0589031
https://www.linkedin.com/pulse/20140814193544-49814607-dynamic-pricing-the-future-of-ecommerce-in-india

1 Data Specialist: introduction

1. Data Specialist: introduction
1.1 Role description

The data scientist collects, prepares, compares and arranges sets of data that supports organisation development.

This professional is able to identify and research databases that are relevant for programmed actions. She/he manages and maintains the organisation's databases and selects, reduces, interprets and transforms data into relevant information.

1. Data Specialist: introduction
1.2 Main tasks

1. Identification, selection, organisation of open databases to be used in a marketing and sales development process.

Data Specialists typically begin any data conversion process by conducting a thorough data analysis of customer information. They research data sources and correct existing data as needed to ensure accuracy of the data recorded.

1. Data Specialist: introduction
1.2 Main tasks

2. Creation, implementation of marketing routines, using available data.

Data Specialists analyse existing systems and select a program suitable for his/her specific organisation. In some cases, Data Specialists may design a database or software program needed to convert data.

1. Data Specialist: introduction
1.2 Main tasks

3. Characterisation of the ROI of marketing and sales actions.

4. Definition and management of analytics.

5. Documentation and reporting.

Data Specialists must consistently provide reports regarding the progress of a conversion program to clients. They must present reports covering workflow and workflow disruptions, exceptions, costs and analysis results.

1. Data Specialist: introduction
1.2 Main tasks

6. Contribution to strategic analyses of the organisation (e-marketing strategy, community management strategy etc.).

7. Provision of Technical Support Assistance.

Data Specialists also maintain databases and answer any questions users might have regarding the system.

Keywords for Data Specialist are: AARRR – Acquisition, Activation, Retention, Reference, Revenues.

1. Data Specialist: introduction
1.3 AARRR

Acquisition: how many users yisit your website? where do they come from? who are they? how many views? how many clicks? bounce rate?

Activation: how is the users' first experience? what do they view? what do they do?

Retention: are users coming back?

Referral: do users share their experience?

Revenues: how much revenue is generated?

2 Data in e-Commerce

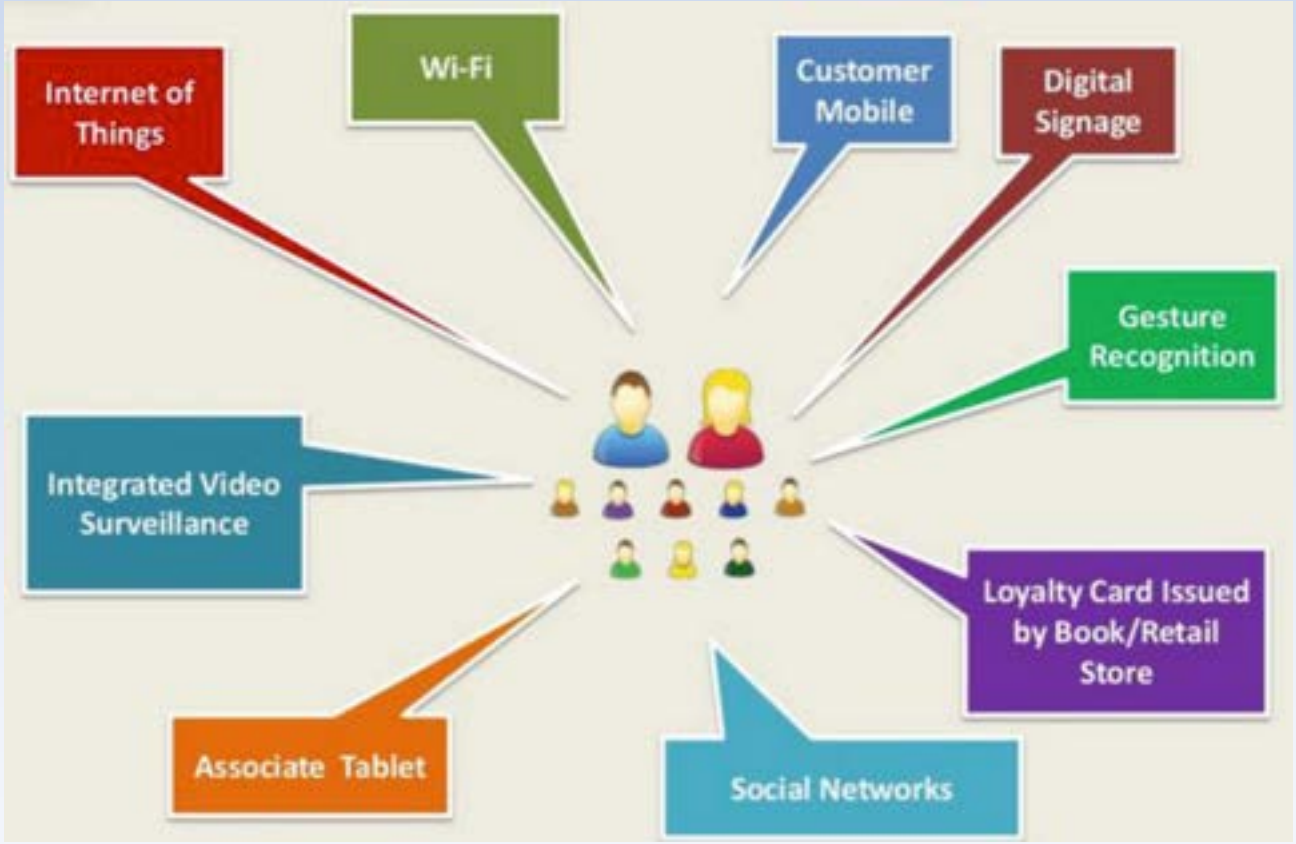
2. Data in e-Commerce
2.1 Overview: data sources

E-commerce:
Orders
Products
Baskets
Visits
Users
Marketing campaigns
Referring links
Keywords
Catalogues browsing

Social data:
Facebook
Twitter
Google

Traffic data:
Cookies
Google Analytics

2. Data in e-Commerce
2.1 Overview: Data Sources



2. Data in e-Commerce
2.1 Overview: Analytical Questions

- what are the best sellers in a category?
- is the most watched product at the same time the best selling one?
- which products sell best among the users who have already bought an item in the product category?
- how often does a given user group (e.g. new users) return to your shop?
- …

The problem is, however, that answering these questions does not lead directly to a bigger profit. Companies often get discouraged as the answers are difficult to apply in real life.

2. Data in e-Commerce
2.1 Overview: Actionable Data

- collaborative filtering;
- using the information on users' actions to automatically find the correlations between: elements on a website, a keyword and the link chosen;
- recommendations; products, offers;
- classification: users who continue shopping.
- regression: indicating trends or the lack of trends, predicting stocks, anticipating a product's future popularity, anticipating the future popularity of promotions, assessing the effect of marketing activities on sales or the number of users;
- categorisation and segmentation: customers, products.

2. Data in e-Commerce
2.1 Overview: Actionable Data Examples

If, thanks to Big Data, we can find the correlation between the social media and our system data, then taking into account that:

- 40% users purchased a product after liking or sharing it on social media;
- 71% users of social media buy mainly based on recommendations;
- we can prepare shopping recommendations for specific customers, based on their social media behaviour.

2. Data in e-Commerce
2.1 Overview: Actionable Data Examples

Example: T-Mobile

- billings, social media data;
- selecting clients for migration to premium models;
- detecting clients with high lifetime customer value.

2. Data in e-Commerce
2.1 Overview: Actionable Data Examples

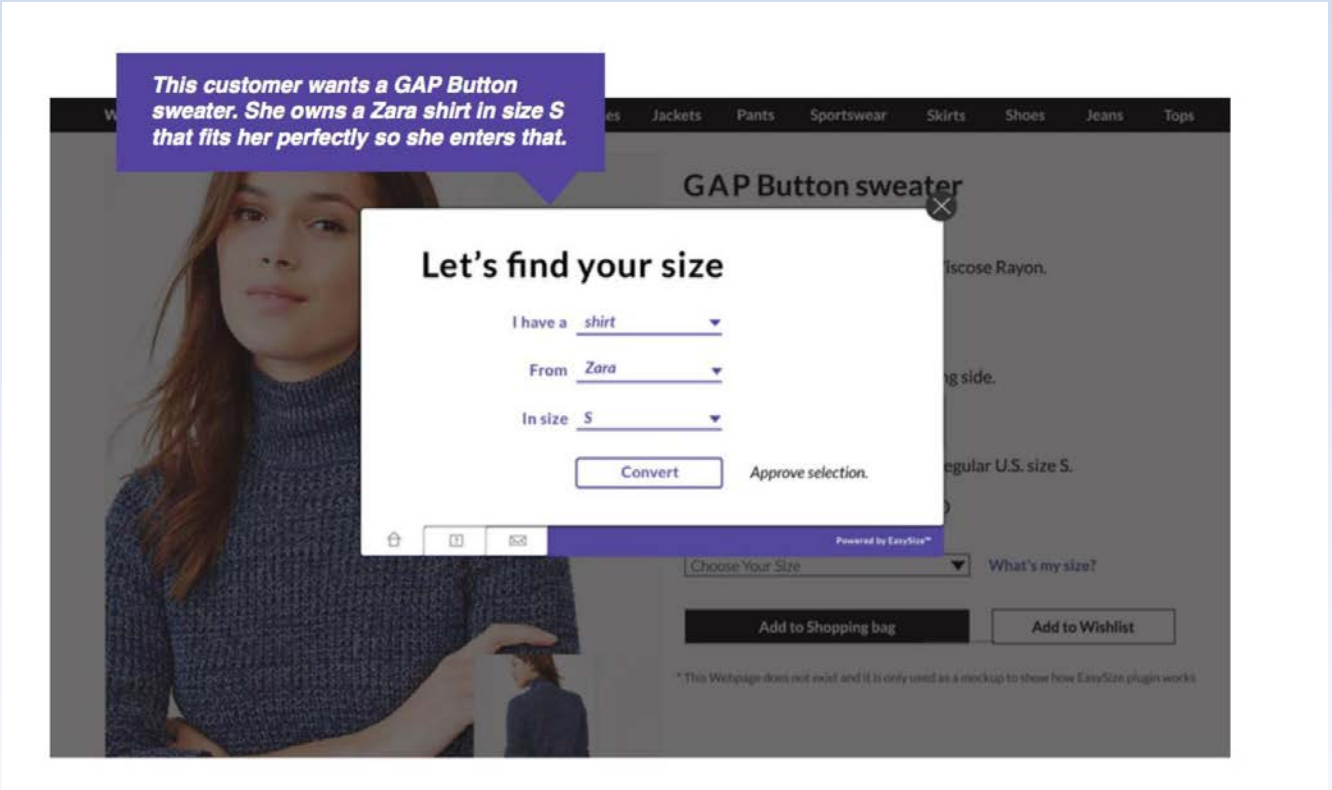
Example: STARBUCKS

- collecting the data about the customers' orders;
- personalising adverts;
- personalising vouchers;
- selecting the customers losing their interest in the offer;
- recovering lost customers.

2. Data in e-Commerce
2.1 Overview: Actionable Data Examples

Example: EasySize

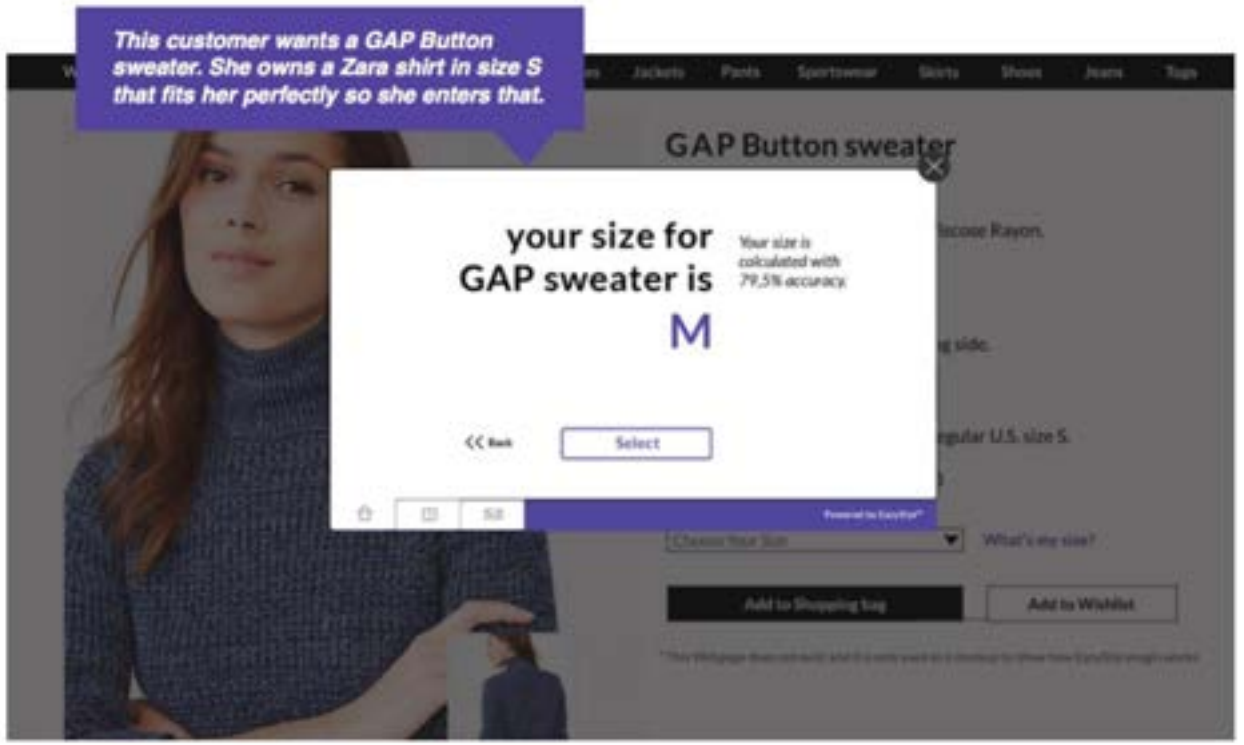
Analysing orders and returns – using the findings to decide which sizes in different brands would fit a given person.

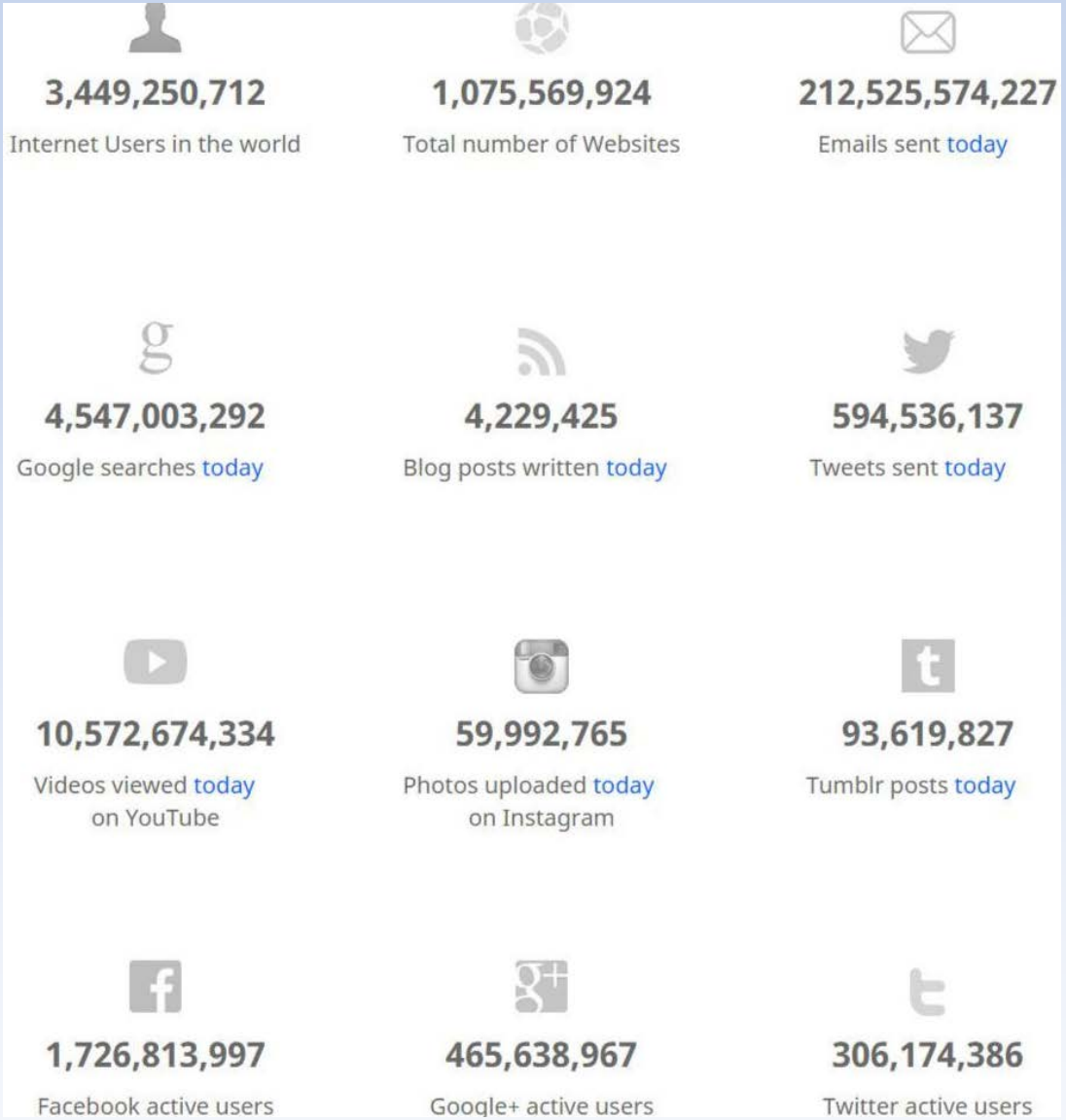


2. Data in e-Commerce
2.1 Overview: Actionable Data Examples

Example: EasySize

Results: decrease in returns by 35-40%.





2. Data in e-Commerce
2.2 Big Data in e-Commerce

Big Data

It describes how much data is part of our lives, precipitated by accelerated advances in technology

OPTIMISED

Big data benefits in e-commerce

- product portfolio;
- pricing;
- online/in store experience;
- advertising/marketing;
- budget customer service;
- inventory.

2. Data in e-Commerce
2.2 Big Data in e-Commerce

Big data benefits in eCommerce

By examining purchase history, online "travel", likes via social networks, geo-location, retailers can now create real-time, targeted promotions broadcast directly to customers' smart phones while they shop.

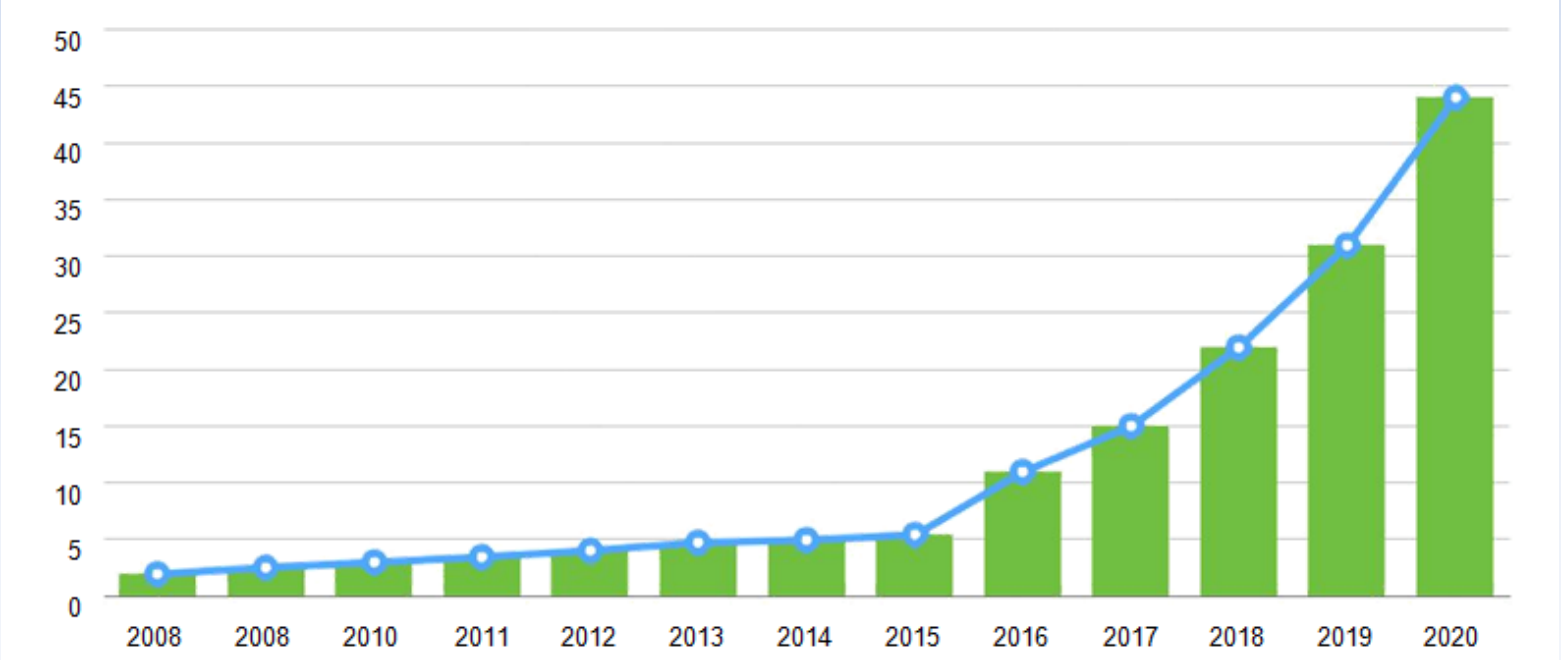
Supported by data coming from online sources, retailers can now pinpoint which merchandise should be stocked at specific locations and where items should be placed throughout the store (e.g. pregnant woman seeing baby products at the entrance in a shop).

By tailoring offers to each individual customer, retailers are seeing an increase in returning clients. Customers nowadays are looking for the easiest and most convenient way to shop and Big Data allows retailers to understand their customers' needs before they even enter a store.

2. Data in e-Commerce
2.2 Big Data in e-Commerce

BIG DATA IS GROWING AT 40% ANNUALLY

Data is growing at a 40 percent compound annual rate, reaching nearly 45ZB by 2020.



2. Data in e-Commerce
2.2 Big Data in e-Commerce

THEN…

Sales

NOW…

Data-driven pricing

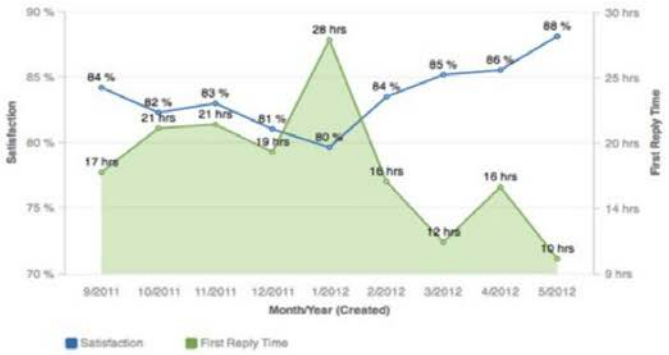
and recommendations

2. Data in e-Commerce
2.2 Big Data in e-Commerce

TRANSFORMATION OF CUSTOMER SERVICE

THEN…
Unhappy Customers

NOW…
Customer Insight



3 Databases

3. Databases
3.1 Overview

Databases are tools to manage all kinds of data generated by a business. In the field of e-commerce and online-marketing the data of online users and customers are most relevant.

The collected **data sets** are collected and stored. Data my refer to the origin of the user, personal data, what s/he viewed and clicked, the purchase details, time spent on the website etc.

The challenge is to manage the masses of data. Helpful tools are **database-management systems (DBMS).** These software applications are interfaces for data specialists. They assist in generating useful information out of that mass of data.

3. Databases
3.1 Overview: DBMS

Database-management systems (DBMS) help managing all kinds of data. They offer the opportunity to display specific data only and relate different sets of data.

DBMS use different programming standards as **SQL** or **ODBC**.

For further information on these programming languages see:

https://www.youtube.com/watch?v=7Vtl2WggqOg

https://www.youtube.com/watch?v=nWeW3sCmD2k

https://www.youtube.com/watch?v=VkMXJvaWeTE

Hadoop

The Apache distributed data processing software is so pervasive that sometimes the terms „Hadoop" and „Big Data" get used synonymously.

Hadoop is known for the ability to process extremely large data in both structured and unstructured formats reliably replicating chunks of data to nodes in the cluster and making it available locally on the processing machine.

Apache Foundation also sponsors a number of related projects that extend the capabilities of big data Hadoop.

MapReduce

If Hadoop is the big data mahout, MapReduce happens to be its lifeline.

A programming model and software framework for writing applications, MapReduce works to rapidly process vast amounts of data in parallel on large clusters of computer nodes.

Widely used by Hadoop, as well as many other data processing applications.

Grid Gain

GridGain is a Java based middleware for faster in-memory processing of Big Data in real time.

GridGain is compatible with the Hadoop Distributed File System.

Requires Windows, Linux or Mac OS X operating system.

HPCC Systems

Developed by LexisNexis Risk Solutions, HPCC is short for „high performance computing cluster”.

HPCC Systems delivers on a single platform, a single architecture and a single programming language for data processing.

Both free community versions and paid enterprise versions are available.

Storm

Storm differs from other tools with its distributed, real-time, fault-tolerant processing system, unlike batch processing systems of Hadoop.

Having real-time computation capabilities, it is fast and highly scalable, often being described as the "Hadoop of real-time".

It is fault-tolerant and works with nearly all programming languages, though typically Java is used.

Cassandra

Cassandra is a highly scalable NoSQL database for massive data across multiple data centres and the cloud.

Used by many organisations with large, active datasets, including Netflix, Twitter, Urban Airship, Constant Contact, Reddit, Cisco and Digg.

Its commercial support and services are available through third-party vendors.

HBase

HBase is the non-relational data store for Hadoop.

Being a column-oriented database management system, HBase is well suited for sparse data sets and is written in Java.

Supports writing applications such as Avro, REST and Thrift

Features include:

- linear and modular scalability;
- strictly consistent reads and writes;
- automatic failover support and much more.

MongoDB

MongoDB was originally developed by 10gen designed to support humongous databases.

It's a NoSQL database written in C++ with document-oriented storage, full index support, replication and high availability and scales horizontally without compromising functionality.

Commercial support is available through 10gen.

Neo4j

Neo4j boasts performance improvements of up to 1000x or more versus relational databases.

Stores data structured in graphs instead of tables and is a disk-based, fully transactional Java engine.

Organisations can purchase advanced and enterprise versions from Neo Technology.

CouchDB

CouchDB stores data in JSON documents that can be accessed via the web or query using JavaScript.

Offers distributed scaling with fault-tolerant storage.

Key features include:

- on-the-fly document transformation;
- real-time change notifications;
- easy-to-use web administration console.