Introduction to Big data

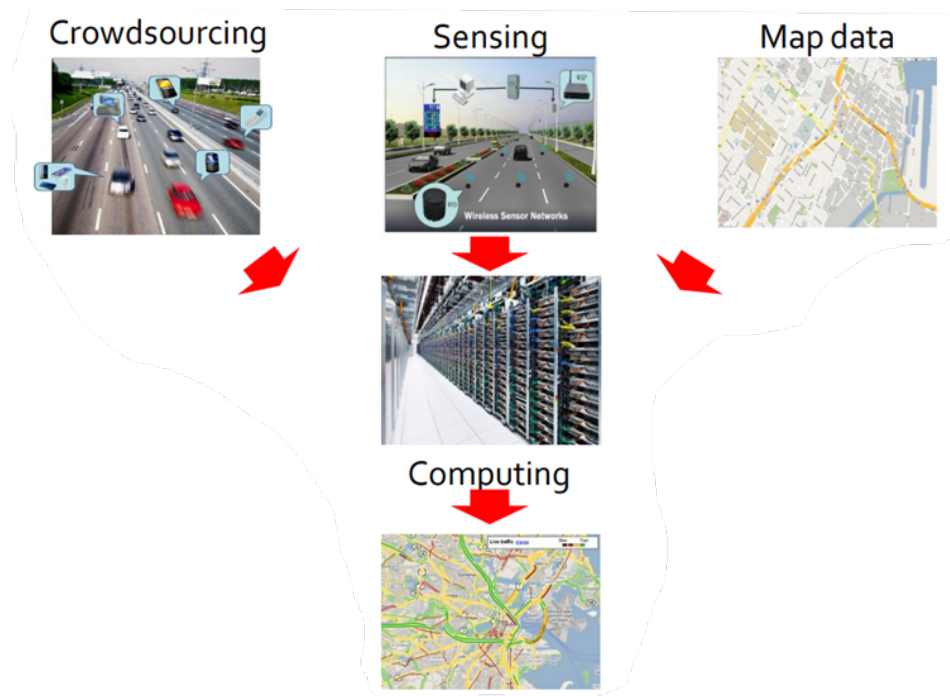## Table of contents

1 Who generates Big Data

- User generated content: social networks (web and mobile)
- Health and scientific computing
- Log files: web server log files, machine system log files
- Internet of Things (IoT): sensor networks, RFIDs, smart meters

1.1 Example of Big Data at work

2 The five Vs of Big Data

- Volume: scale of data
- Variety: different forms of data
- Velocity: analysis of streaming data
- Veracity: uncertainty of data
- Value: exploit information provided by data
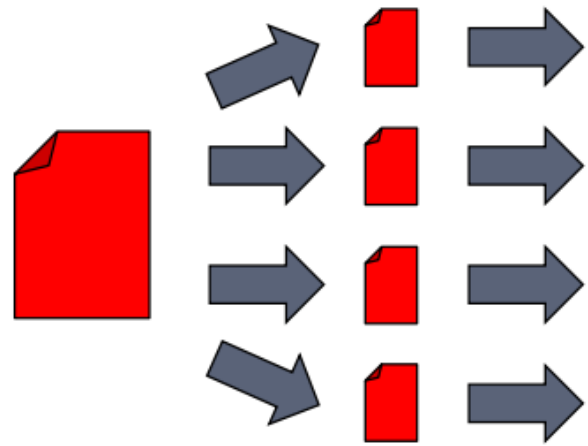
Figure 1: Bigdata example

3 The bottleneck and the solution

3.1 Bottleneck

- Processors process data
- Hard drives store data
- We need to transfer data from the disk to the processor

3.2 Solution

- Transfer the processing power to the data
- Multiple distributed disks: each one holding a portion of a large dataset

- Process in parallel different file portions from different disks