

## Table of contents

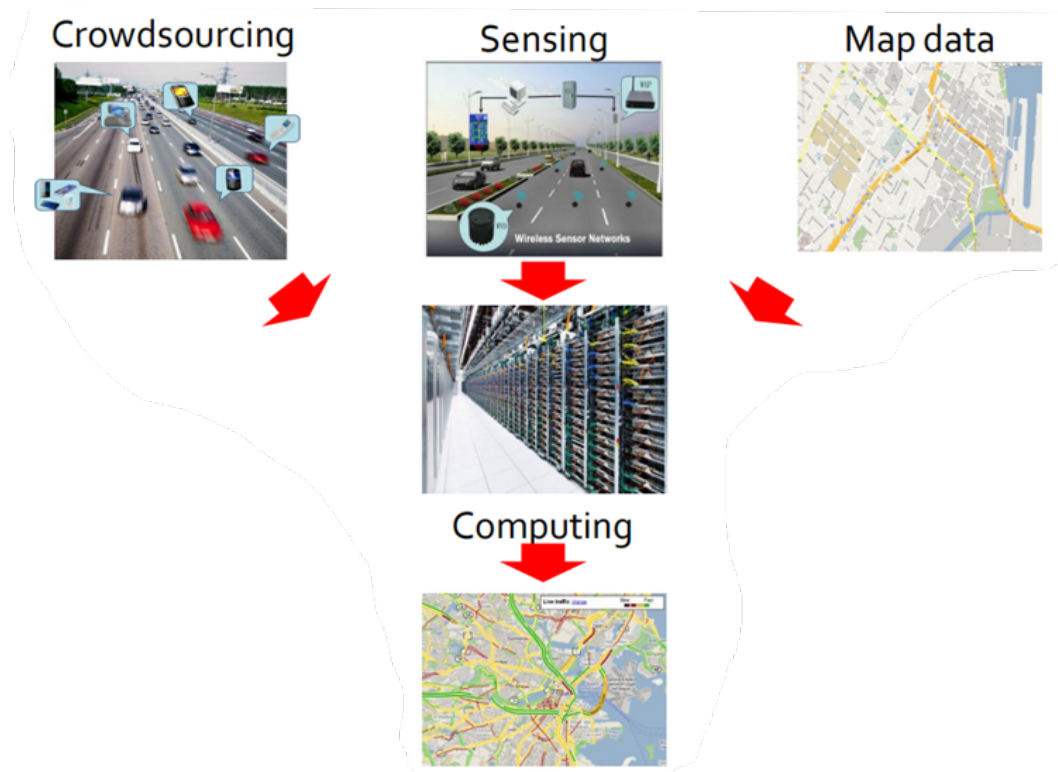
Who generates Big Data . . . . .	1
The five Vs of Big Data . . . . .	1
The bottleneck and the solution . . . . .	2

### 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

### Example of Big Data at work

Figure 1: Bigdata example



### 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

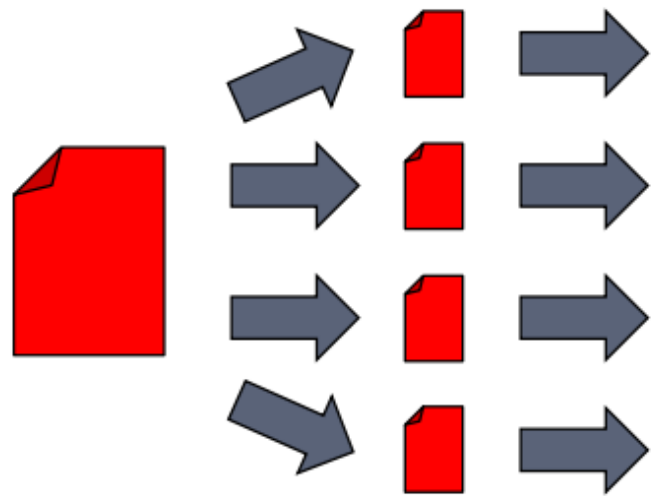
## The bottleneck and the solution

### Bottleneck

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

### 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