

# Functional Testing of MapReduce Programs

Jesús Morán

Software Engineering Research Group

<http://giis.uniovi.es>

University of Oviedo



# Agenda

- Big Data
- MapReduce paradigm
- Phd schedule
- Conclusions

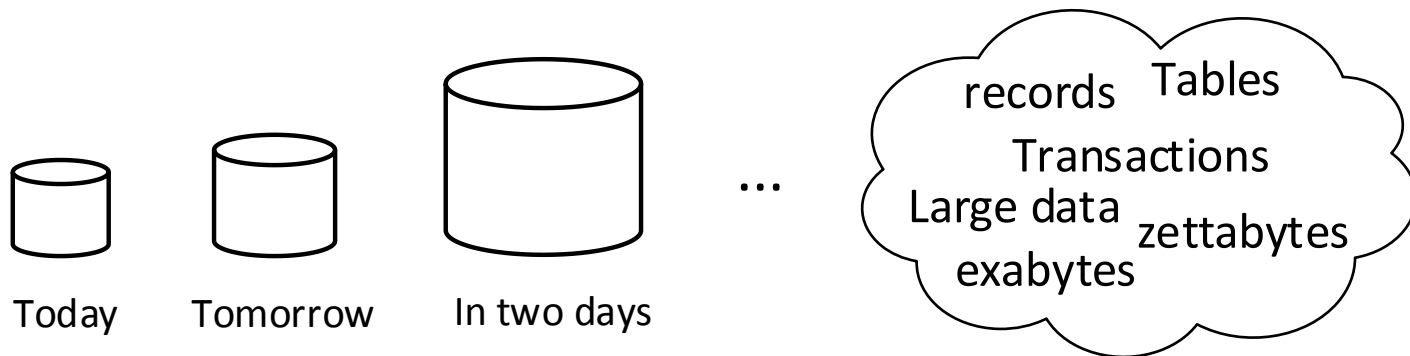
# Big Data

- Data processing programs that can not use traditional technology because

# Big Data

- Data processing programs that can not use traditional technology because

- Volume



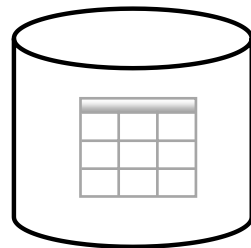
# Big Data

- Data processing programs that can not use traditional technology because
  - Volume
  - Velocity
    - is the program enough fast to process all possible information in a timely way and generates value?

# Big Data

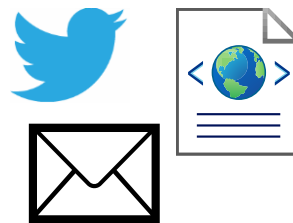
- Data processing programs that can not use traditional technology because

- Volume
- Velocity
- Variety



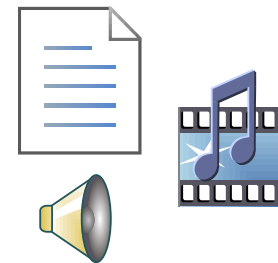
Structured  
data

VS



Semi-structured  
data

VS



Unstructured  
data

# MapReduce paradigm

- Distributed data processing
- “Divide and conquer” principle
- Very used in Big Data
- MapReduce runs in big clusters
- MapReduce runs over commodity hardware

# Phd schedule

- Hypothesis:
  - The MapReduce programs have a set of features (infrastructure failures, information without a data model, sometimes the code is re-run, and so on) that are not well covered by the traditional testing techniques. **New testing techniques or an adjustment of traditional testing techniques could guarantee the quality of MapReduce programs**
  - The testing design and execution in MapReduce are a complex process, including by the absence of a consistent data model and other factors that can modify the program execution (other programs running in the same time, net problems, and son on). **The automation of testing tasks that reduce the problems in MapReduce could improve the development and quality of the programs**



# Phd schedule

- Today:
  - Classification of MapReduce faults
    - Morán, J., de la Riva, C., Tuya, J. MRTree: Functional testing based on MapReduce's execution behaviour". International Symposium on Big Data Research and Innovation, 380-384, 2014.
  - Empirical study and state-of-art about testing in MapReduce (I am writing the paper)
  - Benchmark of programs in order to validate the research results
  - Testing technique for MapReduce programs
    - Moran, J.; De La Riva, C.; Tuya, J., "Testing data transformations in MapReduce programs" In the proceedings of International Workshop on Automated Software Testing (A-TEST), 2015

# Phd schedule

## ■ Future:

- New testing techniques for MapReduce programs
- Tools
- Testing in other Big Data technologies

# Conclusions

- The developers and CEOs do not trust in the quality of MapReduce programs
- The principal research effort is focused in performance testing

# Conclusions

- The developers and CEOs do not trust in the quality of MapReduce programs
- The principal research effort is focused in performance testing

This PhD tries to cover the gap

# Questions?

Jesús Morán

Software Engineering Research Group

<http://giis.uniovi.es>

University of Oviedo

