BIG DATA ANALYTICS AND REASONING

Master Program in Computer Science
University of Calabria

Prof. F. Ricca

Join the Bigdata team

- CODE: **tqxvhnz**

Course organization (provisional)

- ECTS Breakdown – 6 Credits >> 56 teach. + 94 homew.
  - *Introduction and basic notions*
  - *Foundations of modern database systems*
  - *Hadoop framework & Map reduce*
  - *Apache Spark & Map reduce*
  - *Advanced topics of research on Big Data*
  - *Development tools*
  - *Development of a project*

Foundations of modern databases

- Three Database Revolutions
- Big Data, and Hadoop
- The Birth of NoSQL
- Document Databases
- Graph Databases
- Column Databases
- In-Memory Databases

Hadoop framework & Map reduce

- Components of Hadoop
  - Architecture
  - HDFS
  - Map Reduce
  - Ecosystem
- Understanding Map Reduce
- Writing basic MapReduce programs
- Programming practices

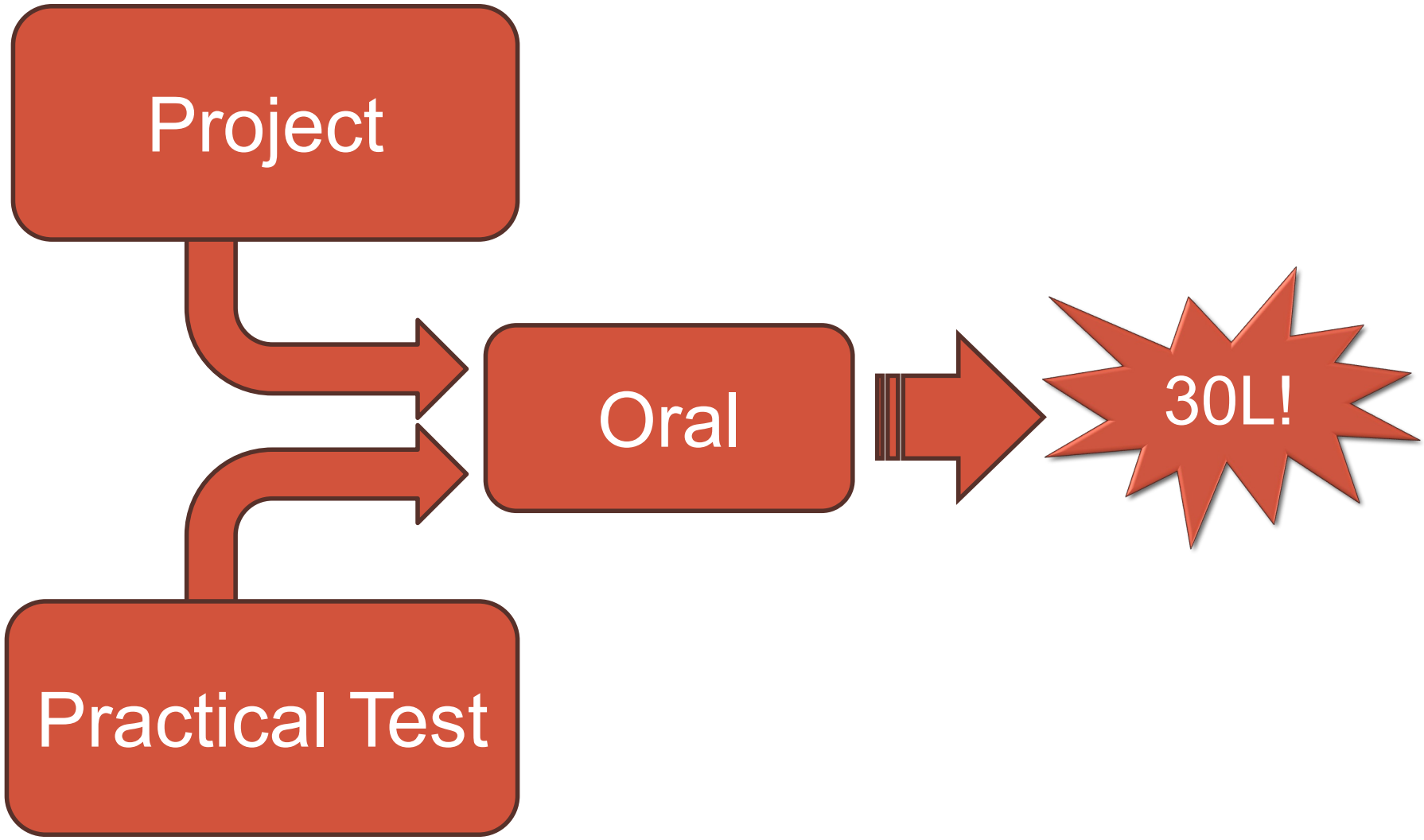
Development Tools (dr. Mazzotta)

- Hadoop
  - Cluster configuration
  - Hive
  - Hbase
  - Zookeper/Ambari
- Spark
  - Configuration
  - Spark SQL
  - Spark R
- Data Ingestion
  - Sqoop

Lectures & exercises

- Teaching
  - Basic and advanced notions
  - Configuring and installing tools
- Practice
  - Exercises during classes
  - Intermediate tests – *determining exemptions! (NEW!)*
  - Project assignment
- Attendance is MANDATORY!
  - You need 70% of presence
  - *…. Being present in the room is not sufficient!*

How to do the exam



How to do the exam

- Project + Oral
  - Preparation of a project (80% impact)
    - *Valid only for the first exam appointment*
  - Oral exam (20% impact)
- Practical Test + Oral
  - Live development of a small assignment (50% score)
  - Oral exam (50% score)

- *You can reach the maximum score with both modalities!*

Intermediate tests (NEW!)

- There will be 3-4 intermediate tests
  - At least two needed for counting active presence
- Generate exemptions on the project
  - Pass the test on Hive → No Hive in the project
- Generate exemptions on the Practical Test
  - Pass the test on Hive → No Hive in the test

- *Exemptions valid only for the first exam appointment!!*

  Gli esoneri sono validi solo per il I appello

Books and other material

- **Course Slides**
  - Available in our course in MS Teams
    - (Channel "Generale", tab "Files", open folder "Materiale del corso")
  - *They will be released/updated during the semester*

- **Recommended books & resources**
  - *"Next Generation Databases"* by Guy Harrison, Springer
  - *"Hadoop in Action"* by Chuck Lam, Manning
  - https://hadoop.apache.org/docs/
  - "Data Algorithms - recipes for scaling up with hadoop and spark" Mahmoud Parsian O'Really
  - "Spark in Action" Second Edition by Jean Jorges Perrin, Manning

- **Further readings**
  - *"Practical Hadoop Migration"* by Bhushan Lakhe, Springer
  - *"Big Data Principles and best practices of scalable realtime data systems"* by Nathan Marz and James Warren, Manning
  - BTW, *the interned is full of resources…*

QUESTIONS?