EXAM of SYSTEMS AND METHODS FOR BIG AND UNSTRUCTURED DATA

Marco Brambilla

marco.brambilla@polimi.it

@marcobrambi

Format

Close questions on all sessions

Exercises on

- Nosql & Data Models
- Data engineering
- Scalable computation

Questions
Foundations

Big Data

Data Science and Data Engineering

4 paradigmatic shifts

Questions
Nosql & Data Models

Features and differences between the presented models

ACID vs. BASE

The CAP theorem

Features of

- Graph DB
  - Property graphs (neo4j)
  - Direct labelled graphs (RDF)
- Key-value DB (Redis)
- columnar DB (Cassandra)
- Document NoSQL (Mongo)
- Search-based (ElasticSearch)

Questions

Taming Data Volume

map-reduce basics (be able to exemplify it)

Hadoop architecture and components

Horizontal vs Vertical Scalability

Questions
Data engineering

Data science

Streaming pipelines

Exercises
Nosql & Data Models

Given a specification of a data-centric project, the challenge is to specify

- which data is needed to address the project,
- the comprehensive conceptual model of the data,
- which data model families should be used for each part of the data,
- which design can be devised at the logical level for each part,
- and how it can be implemented .

Specification of the solution of a few queries (mainly for graph or document models) may be requested.

Exercises

Data engineering & Data science pipeline

Hadoop simple examples

Marco Brambilla.