CS-E4640 Big Data Plaforms

How to succeed in Big Data: Experiences from The Course
The Real Job & Assignment Questions

Tri Nguyen
PhD student at Department of Computer Science
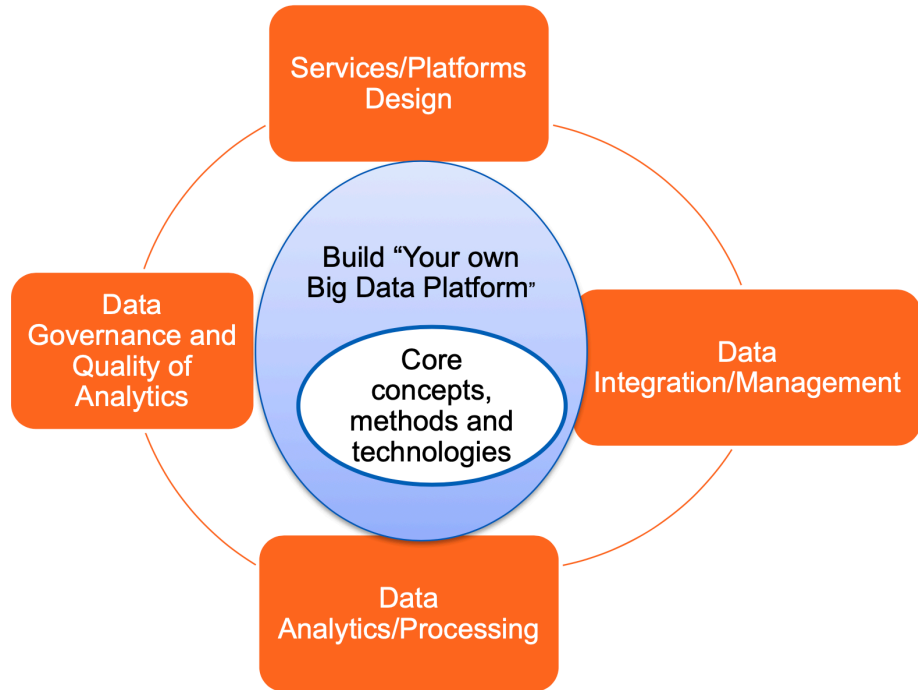*Researcher at AaltoSEA*

Assignment

- **Build your own Big Data Platform**
  - Theory:
    - *Design*
    - *Explanation*
  - Implementation:
    - *Coding*
    - *Testing*

Assignment

- **Don't have any idea?**
  - Inspired by your projects or real applications
    - *What are the big data problems?*
    - *Macro-thinking*
  - Technologies:
    - *Hands-on Tutorials*
    - *Tech Radar*
    - *Your favor*
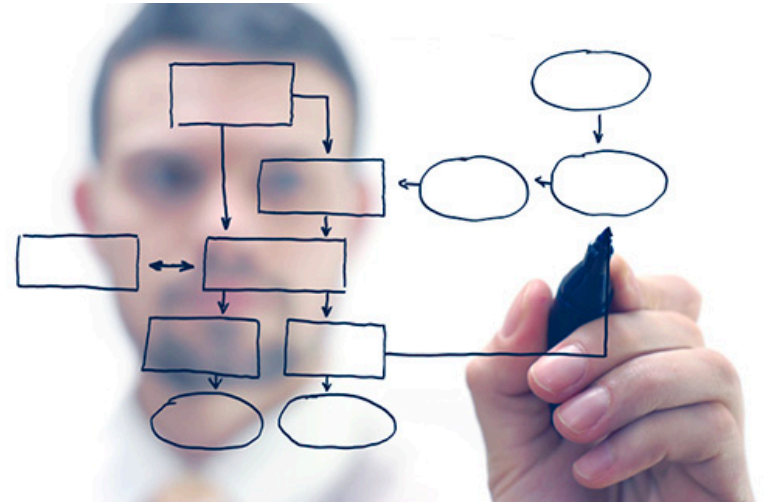   - ➢ *Why do you use these technologies but not others?*

Assignment

- **Design**
  - Clear Architecture
    - *Module/Service*
    - *Function*
    - ➢ *Realistic, reasonable, scalable, extendable, …*

  - ➢ Tips:
  - *- Stick to the assignment instructions*

Assignment

- **Implementation**
  - Coding:

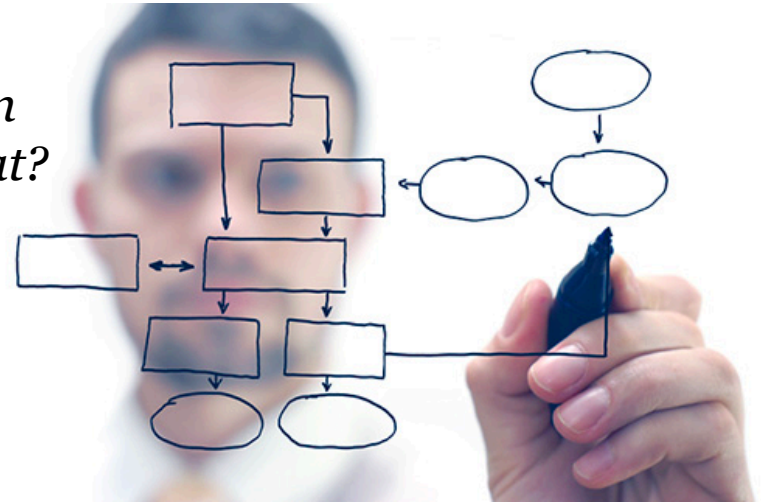   *- Each service/microservice should be implemented separately and can be run independently! Which service does what?*

  - Testing:

    - *Prove Big Data concepts, realistic scenarios*

    - *Test the system limits!*

  - ➢ Tips:

    - Try to use sample codes (e.g. official/public sources)

    - Try to make it reusable for later assignments

Assignment

- **Experiments**
  - Again! be realistic, macro-thinking
    - *Workload? Why they are big data problems?*
    - *Monitor/measure multiple metrics, different scenarios, failures,...*

  - ➢ Tips:
  - *- What are existing solutions, why/when they are applied?*
  - *- Simulate big workloads to see potential problems, interesting points,...*

Aalto University
School of Science

Don't panic

- **It's not as scary**
  - Tough but doable

- **You can succeed**
  - Time and effort

- **Be active**
  - Don't hesitate to ask for help

- **Focus on learning**
  - The course is rewarding, don't fixate on grade but actual learning

Don't procrastinate

- **Start working early**
  - A week – danger zone
  - 3 days – unlikely
  - Overnight - impossible
- **Buffer time**
  - Unexpected issue
  - Troubleshoot or workaround
  - Start over

Motivation

- **Applications**
  - ML/AI, Data mining, scalable apps
- **Job opportunity**
  - ML engineer, Data analysist, Cloud DevOps

➢ Great opportunity to learn new technologies! Don't limit yourself to something you already know!

**Aalto University
School of Science**

Assignment

- **How do we evaluate your works?**
  - Based on the question outlines
    - *Answer one by one in order following the question outlines to help us save grading time without missing your points!*
    - *Be clear, concrete and complete!*
  - Demo
    - *Demo & F2F meeting*
  - ➢ Tips:
  - *- Stick to the question outlines, don't mess among questions!*
  - *- Search for existing solutions and show your understanding*
  - *- Don't copy others' works without citation!*

Thank you!

Any Question?

Aalto University
School of Science