COEN 424/6313 – Programming on the Cloud

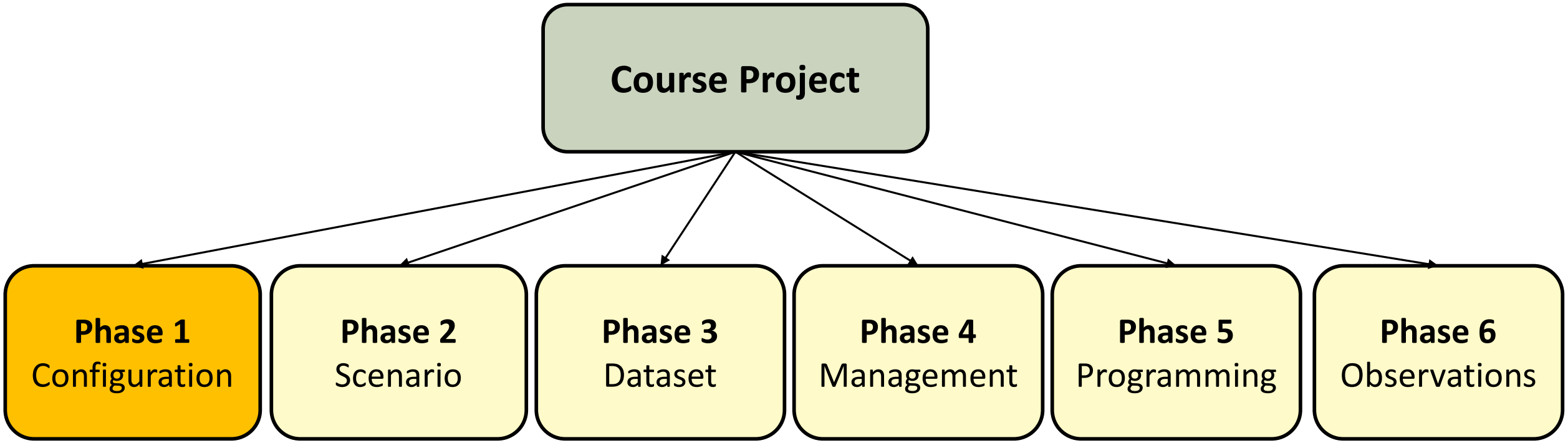
Course Project

COEN 424/6313 PROGRAMMING ON THE CLOUD

DESIGNED BY: OMAR ABDUL WAHAB

CONCORDIA UNIVERSITY, FALL 2018

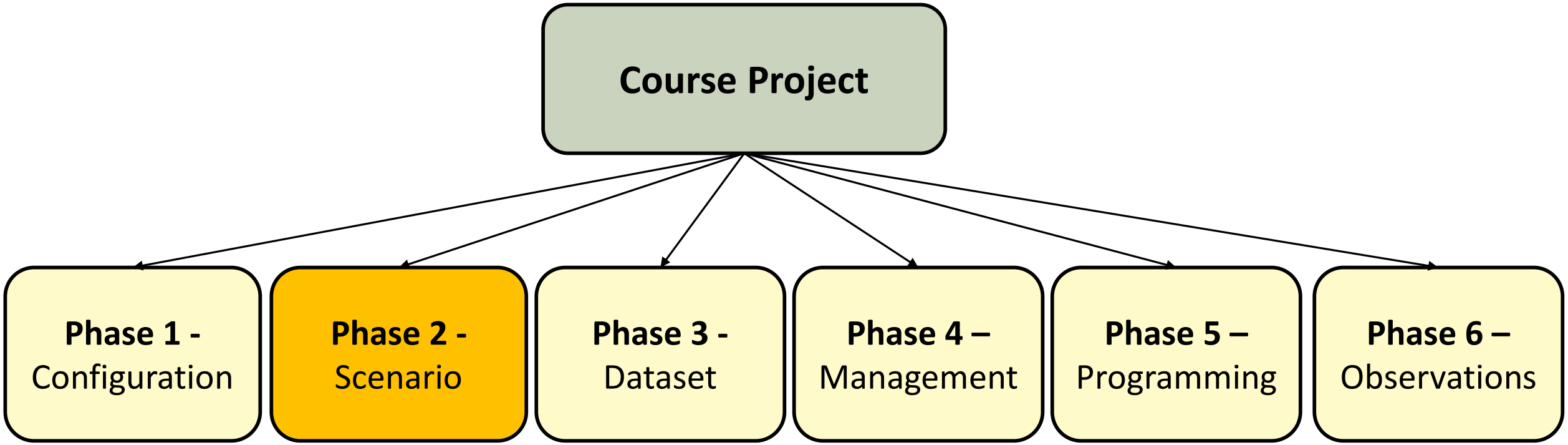
Project Phases



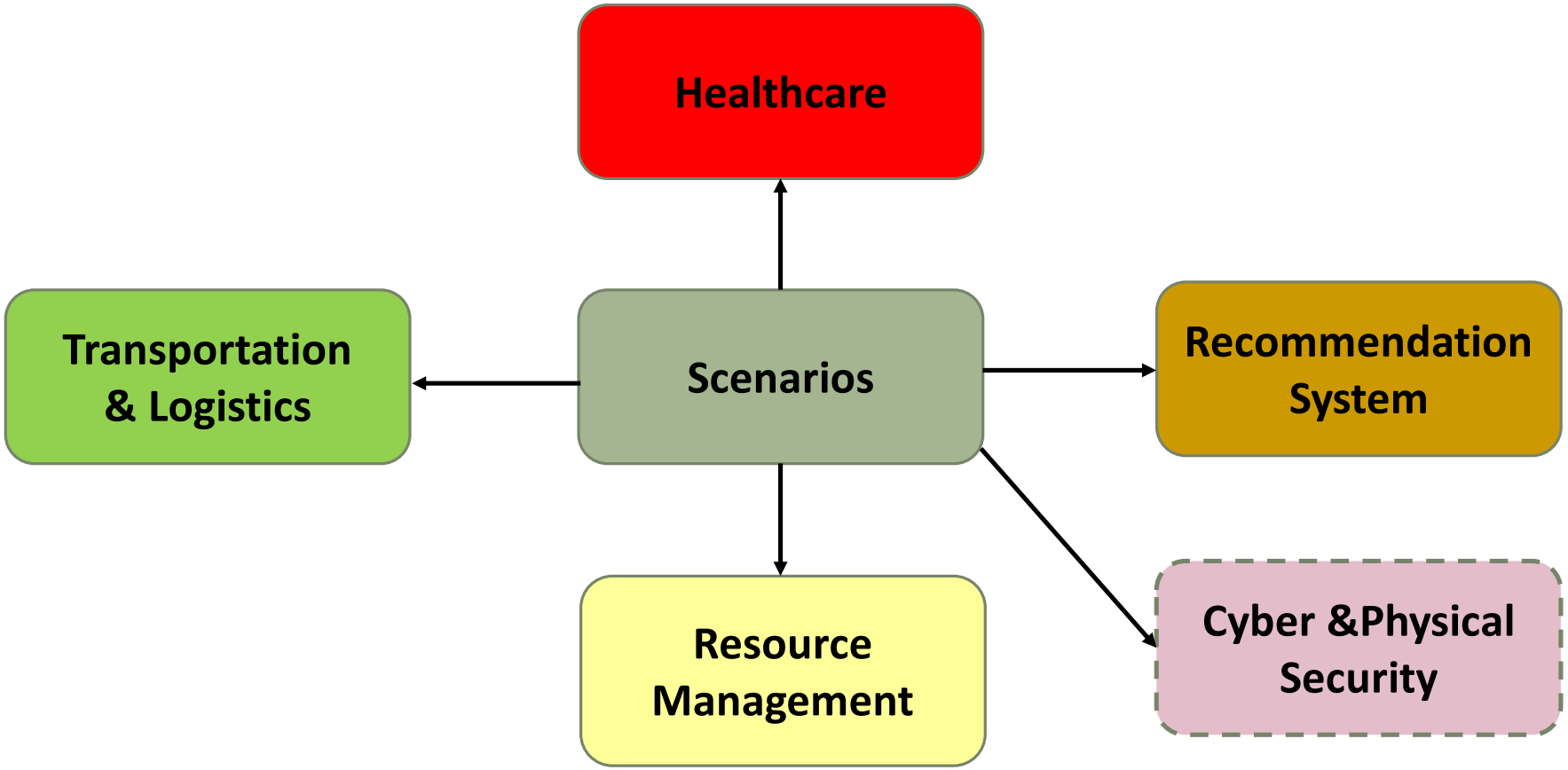
Phase 1 – Warm-up and Configuration

- Get familiar with a cloud computing platform (e.g., Amazon EC2, Google App Engine, Windows Azure, Rackspace, Eucalyptus) or set up a private cloud by yourself on your PC (e.g., VMWare, OpenStack).
- Provision a single instance manually and analyze its performance (e.g., CPU, RAM).
- Configure and utilize a Hadoop cluster on your (public or private) cloud
- Learn how to develop and deploy MapReduce or Spark jobs on your Hadoop cluster

Project Phases



Phase 2 – Pick a Scenario



Phase 2 – Pick a Scenario

Transportation and Logistics:

- Collision avoidance
- Route planning and delivery time optimization
- Parking control in smart cities
- Decision-making in logistics and procurement
- Inventory management
- Any other relevant topic

Phase 2 – Pick a Scenario

Resource Management:

- Energy and hardware minimization in cloud datacenters;
- Power quality management in smart grid;
- Buildings energy efficiency optimization;
- Any other relevant topic.

Phase 2 – Pick a Scenario

Healthcare:

- Personalized health care management:
  - Example: establish preventive medical interventions to avoid hospitalization and more serious health consequences;
- Inequalities that exist in healthcare & medicine;
- Electronic health records management;
- Any other relevant topic.

Phase 2 – Pick a Scenario

Recommender system:

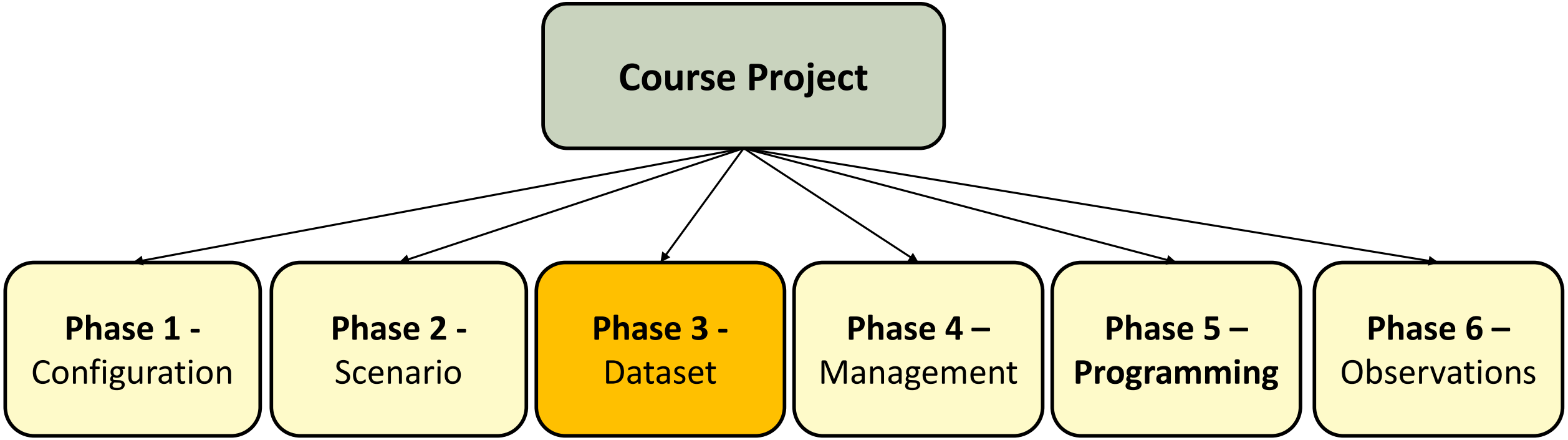
- Recommender System that can help recommend potential candidates for political/non-political positions based on factors such as: candidate's current political achievements, candidate's previous political achievements, candidate's educational background, candidate's financial status, candidate's family background, etc.
- **Career Counselling Recommender system:** recommending the combination of subjects that a student need to pass in order to study a particular course at a university.
- **Cold-start problem:** providing recommendations to newly registered users whose preferences are not known yet or partially known.
- Any other relevant topic.

Phase 2 – Pick a Scenario

Cyber & Physical Security:

- Intrusion detection/prevention systems in cloud computing, IoT, vehicular networks, etc.;
- Video monitoring and suspicious action recognition;
- Any other relevant topic.

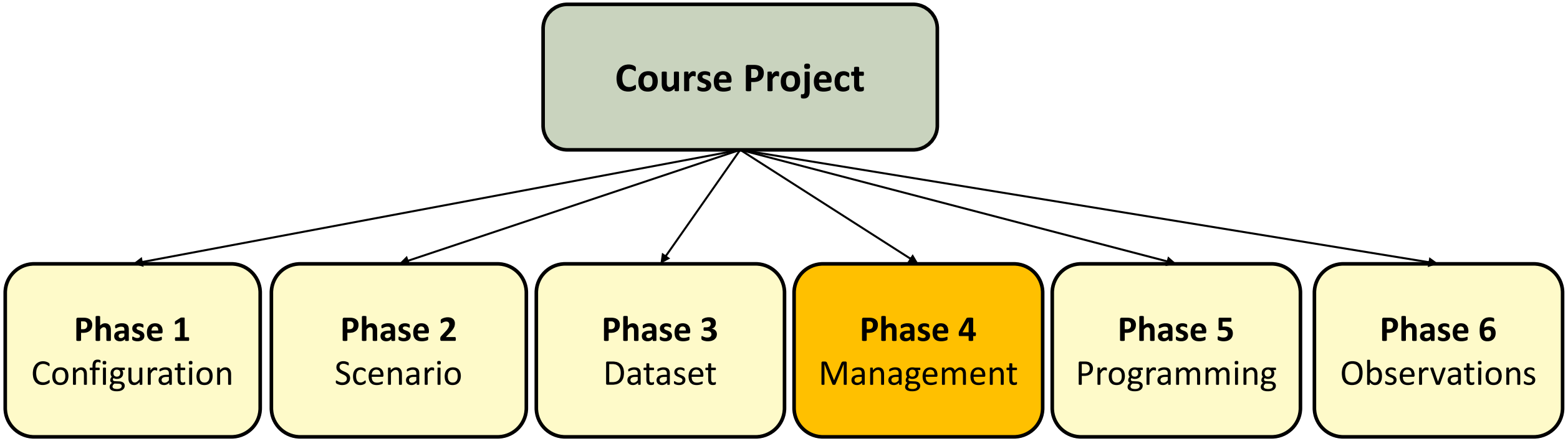
Project Phases



Phase 3 – Choose Your Dataset

- Choose the dataset you will be analyzing for the project.
- You can also create your own data model (JSON, XML, etc.).
- Some potential sources:
  - Kaggle
  - data.world
  - CRAWDAD
  - Etc.
- Do some preprocessing on your data (if needed):
  - **Data cleaning:** fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies.
  - **Data integration:** using multiple databases, data cubes, or files.
  - **Data transformation:** normalization and aggregation.
  - **Data reduction:** reducing the volume but producing the same or similar analytical results.
  - **Data discretization:** part of data reduction, replacing numerical attributes with nominal ones.
- **Useful link for data pre-processing:** http://www.cs.ccsu.edu/~markov/ccsu_courses/datamining-3.html

Project Phases



Phase 4 – Project Management

Goal:

- **Background:** provide a concise review of literature (existing works in the chosen field).
- **Scope:** What will you do? a concise statement to depict the work to be performed
- **Methodology:** How will you be performing the work?

Objectives:

- The end results that you aim to achieve by the end of the project.
- In the end of the project, you should include a brief description of how did you accomplish each of the specified objectives.

Phase 4 – Project Management

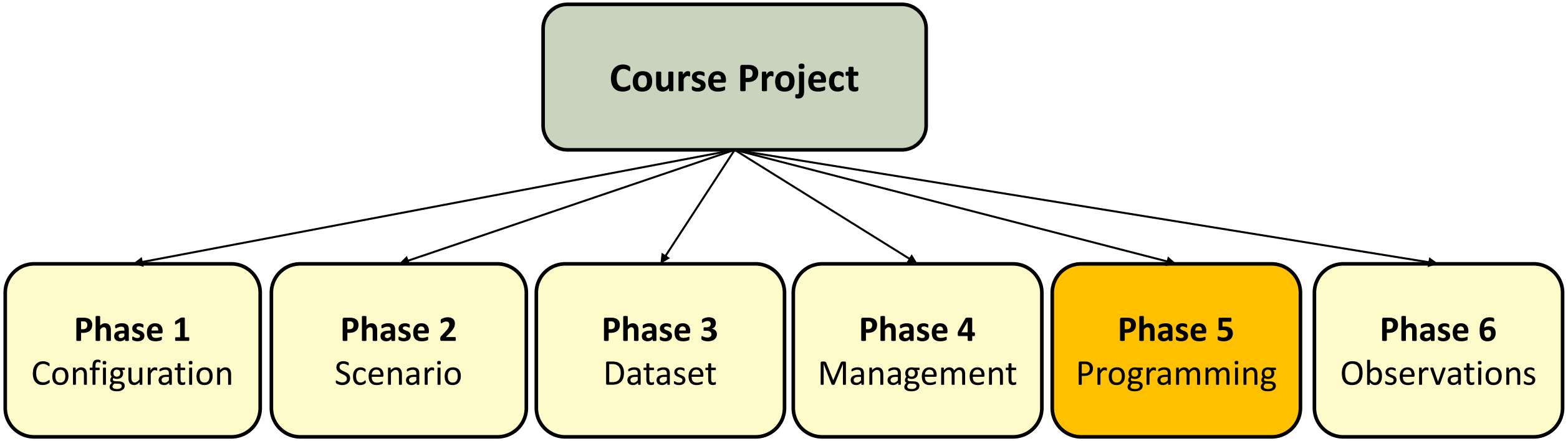
Assumptions:

- Describe the uncertain information you will take as facts as you conceive, plan and perform the project.

Schedules:

- Provide a timeline for each step in the project, i.e., how long each step in the project will take.
- Identify the order of the experiments.
- Describe the relationships among the different experiments (e.g., can they be performed in parallel or they must be executed sequentially).

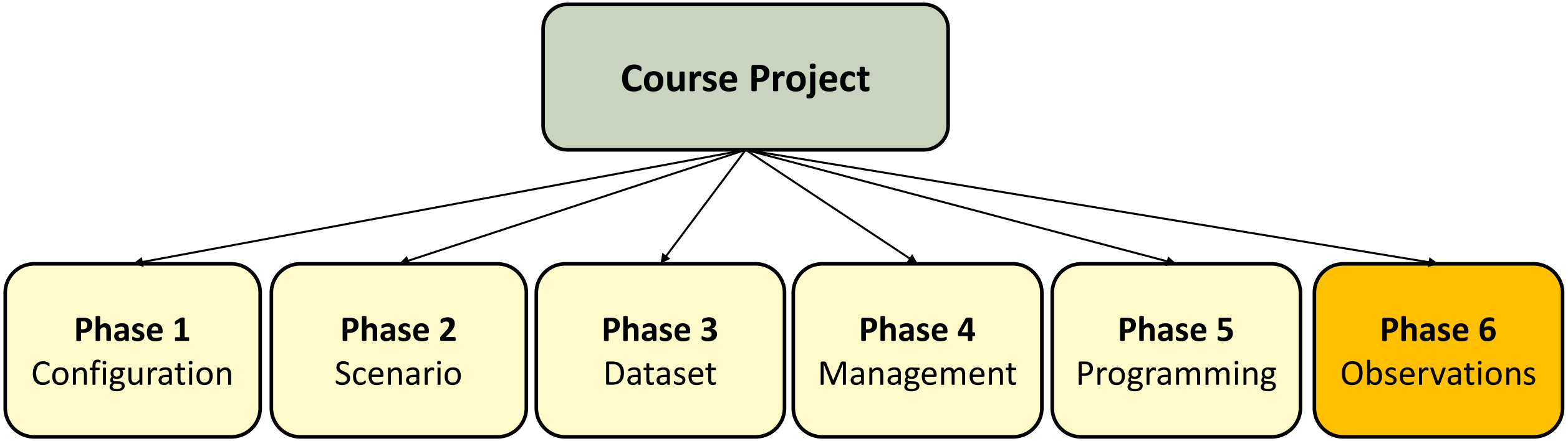
Project Phases



Phase 5 –Do the Programming!

- Use NoSQL to store and query your dataset.
- Develop a MapReduce or Spark application for the selected problem.
- Utilize or mash up external web services as part of your application (e.g., twitter).
  - A mash-up is a Web page or application that integrates complementary elements from two or more sources (Reference: https://whatis.techtarget.com/).

Project Phases



Phase 6 –Observations

Conduct some performance evaluation on your system:

- Execution time;
- Scalability (e.g., maximal input data size that the system can afford);
- CPU & memory utilization;
- Reliability & consistency (e.g., does the output remains the same if the size of the application grows up?);
- Security level;
- Etc.

Extra Features! Extra Points!

You must select at least **one** of the following features to implement:

- Create a mobile application that takes advantage of your web service (**5 bonus points**);
- Expose REST APIs for your web service (**5 bonus points**);
- Integrate deep learning methods to analyze your data or to augment your data size (**10 bonus points**)
- **Useful paper for data augmentation:** Data Augmentation Generative Adversarial Networks) ;

Deliverables

Week 2 Friday 14th by 23:55 : Group information submission

- Find your project partners and discuss your possible project ideas and which scenario you will be implementing.
- The size of a group is 3 or 4 (preferably consisting of a mixture of graduate and undergraduate students).
- Delegate a group representative (for communication purposes only!).
- The group representative should submit to the Moodle site a single **pdf** or **txt** file with the file name
  - D1-[SID of Member]- [SID of Member]- [SID of Member].pdf or
  - D1-[SID of Member]- [SID of Member]- [SID of Member].txt
- For each group, please indicate the following information on each member:
  - [SID] [First and Last Name] [Undergraduate or Graduate] [Project Representative: Yes/No]

Deliverables

Week 5 Thursday October 4th class time: Group Presentation [5 points]

- Present the details of your project till phase 4 (inclusive).
- Be prepared to answer questions like:
  - Who would be the end users of your web service?
  - How are the end results of your project useful for the society, users, businesses?
  - How would you meet the project requirements?
  - What would be the contribution of each group member in the project?
- This group presentation is **mandatory**.
- One representative of each group will give the talk (Other group members may be asked to answer questions).
- The presentation slides should be uploaded to Moodle by **Thursday October 4th at 23:55.**
  - **File name format:** D2-[GroupID].pdf or D2-[GroupID].ppt

Deliverables

Week 13 Thursday November 29th class time: Final presentation, demo, and report

- Each group will give a 15-minute presentation (including a demonstration) of the project.
- Each group member should participate in the presentation talk, live demo or Q&A.
- Students from other groups should prepare **at least one** question to ask for each other group.
- Be prepared to answer questions on all the criteria described in the project phases and how your project conforms to these criteria.
- **Demo:** in case of failure, be sure to prepare a backup video demo.
- By **November 29th Thursday 23:55**, each group should submit the report which must include discussions on all of the project phases.

Report Skeleton

- **Abstract**
- **Introduction** (including goal, objectives, problem statement, assumptions, methodology, and schedule)
  - Make sure to highlight each of these aspects in the introduction (e.g., use subsections).
- **Project Description:**
  - Function provided by the service
  - Architectural design
  - Technical implementation details
  - The URL to access your project source code & URL to access your web service
- **Discussion:**
  - Discussion on the experience, observations, and lessons learnt
  - Include Figures (if applicable): e.g., scalability, security level, quality of service, etc.
- **Contribution of each member:**
  - The role and technical contribution of each member
- **Reference to technical and academic articles**

Project Submission

- The report should follow the IEEE format described in: https://www.ieee.org/conferences/publishing/templates.html
- The file name should follow the naming convention: **D3-report-[GroupID].pdf**
- The presentation slides (ppt or pdf) should follow the naming convention:

  D3-presentation-[GroupID].pdf or D4-presentation-[GroupID].ppt

- All the source code of your project should be packed in a single package and should following the name convention: **D3-project-[GroupID].tar or D3-project-[GroupID].zip**