MAS DSE 260: Capstone Project

İlkay ALTINTAŞ, Ph.D.

Lecture 1: Getting Started

Today's Topics

1. What is a capstone project?

   - Understanding class objectives

   - Setting expectations

2. Roadmap of our 10-step project

3. STEP I: Understanding the Challenge

4. Report I Format : DUE 1/18/18Roadmap for the next 10 weeks

What is a Capstone Project?

- **Objective:** To complete an end to end analysis of a large dataset with big data characteristics.
  - Includes
    - data collection,
    - data preparation,
    - exploratory analysis,
    - model building,
    - visualization, and
    - reporting.

- **Products:**
  - Final report (preferred if publishable as a conference paper)
  - Output data products
  - Developed analytical tools/methods/workflows (if applicable)

# Milestones for the Capstone Project

- **First Year:**
  - Spring Quarter: Capstone project class is introduced.
  - Summer: Advisors provide short summaries of projects so that students can identify who they want to work with. Students start to form teams, define project and find advisor.

- **Second Year:**
  - Fall Quarter: Students finalize project teams and start collecting/exploring data.
  - Winter Quarters: Teams work on their projects and present monthly progress reports. Alternate meetings: once a month for 2 hours with advisor, twice a month for half an hour with capstone faculty (i.e. Altintas).
  - Spring quarter: Teams finalize their projects, including documentation and final report. Teams make open presentations to their peers, advisor and capstone faculty, and receive final grade.

Ultimate Goal

Insight = Data Product

"Big" Data + Exploratory Analysis and Modeling → Insight

Question
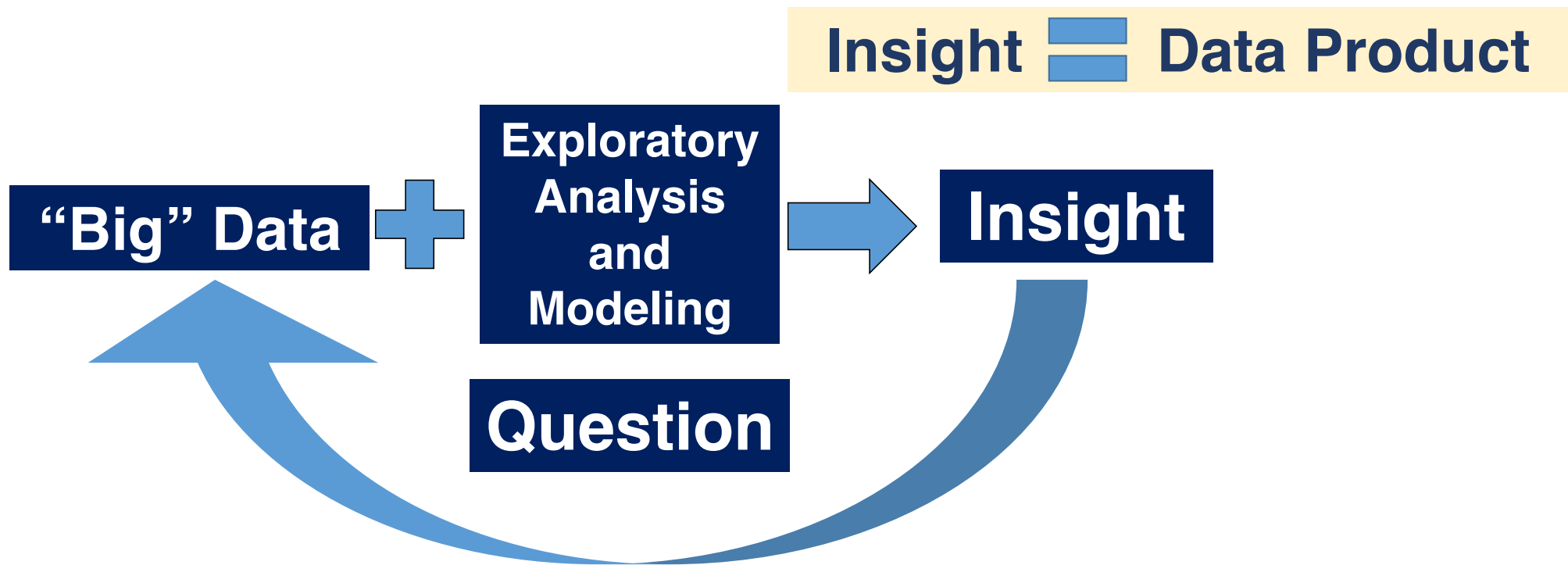
We will do it through
10 deliverables and 5 presentations!

Approach: Focus on Process and Team Work

"The" Team

Expertise and skills often overlap, but nobody has it all!

- Data engineer
- Data analyst
- Methods expert
- Scalability and operations expert
- Business manager
- Business analyst
- Visualization and dashboard developer
- Solution architect
- Story teller/coordinator
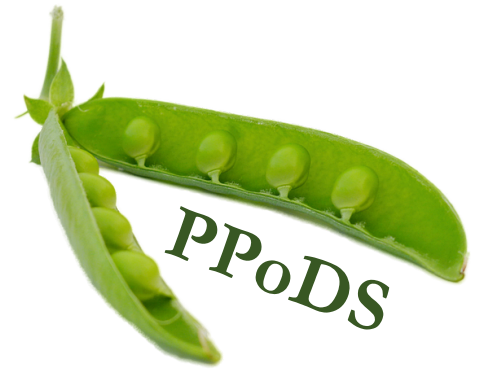- Project manager

PPoDS

People

?

Problem
or
Purpose

Process

Platforms
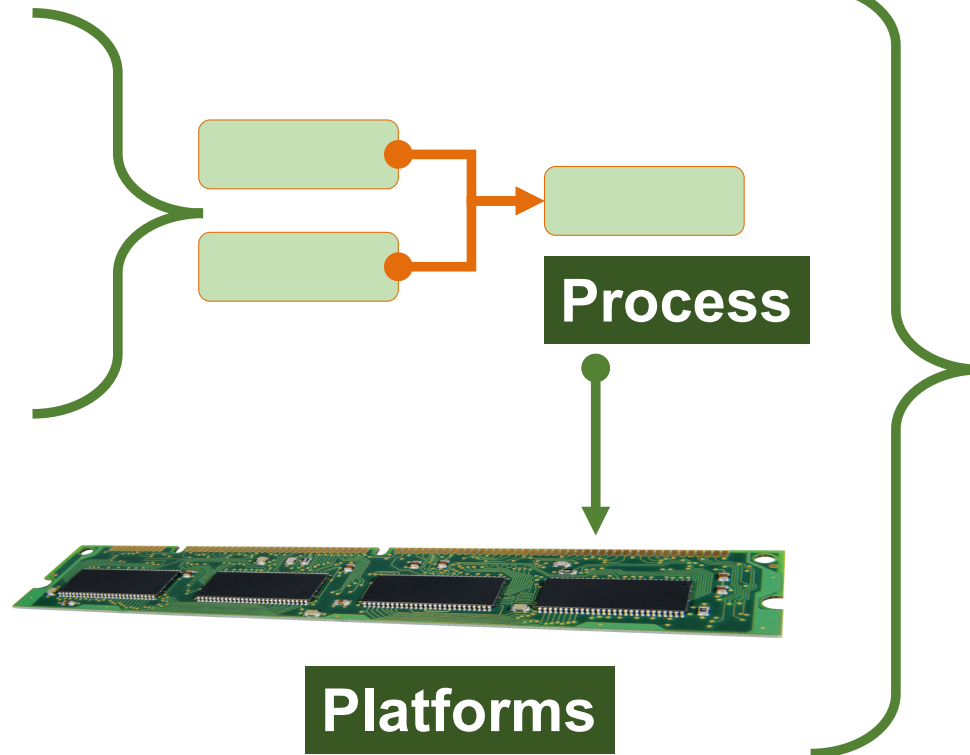
Programmability

PPoDS

Process Roadmap (260 A)

- Step 1: Understanding the Challenge
  - REPORT 1: due 1/18
- Step 2: Designing the Data Acquisition and Preparation Pipelines
  - REPORT 2: due 2/1
- Step 3: Exploring Data
  - PRESENTATION 1: 2/3
  - REPORT 3: due 2/15
- Step 4: Defining Your Hypothesis and Minimum Viable Modeling Product
  - REPORT 4: due 3/1
- Step 5: Creating a Solution Architecture for Modeling and Optimization
  - PRESENTATION 2: 3/3
  - FINAL WINTER REPORT: due 3/16

Process Roadmap (260 B)

- Step 6: Modeling and Visualization

- Step 7: Evaluating and Interpreting Modeling Results

- Step 8: Deploying a Robust and Scalable Solution

- Step 9: Developing a Communication Plan and Monitoring Dashboard

- Step 10: Optimization

Grading

- Reports: 5% each, total 50% over two quarters
- Presentations: 5% 5each, total 20% over two quarters
- Final presentation and demo: 15%
- Final report: 10%
- Final poster: 5%

Create an Ecosystem that Enables Needs and Best Practices

ACQUIRE → **PREPARE** → **ANALYZE** → **REPORT** → **ACT**
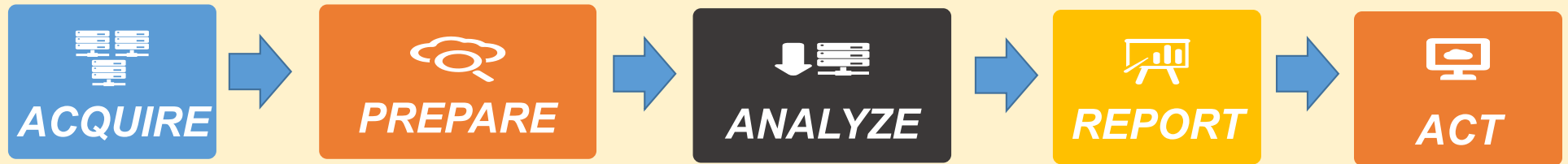
- data-driven
- dynamic
- process-driven
- collaborative

- accountable
- reproducible
- interactive
- heterogeneous

STEP 1: Understanding the Challenge

(a.k.a. the PLANNING Phase)

Objectives

- Specify the key challenge that makes the project important

- Identify relevant data sources

- Distill specific and concise questions related to the challenge that can be solved using the identified datasets

- Define the project team responsibilities

- Define a baseline approach and success metrics

START HERE

The project starts when a domain expert recognizes the opportunity and/or need.

Take Stock: Define vision and scope

- What is the exact need?

- What datasets are available?

- Who are the (current) stakeholders?

- What would you gain when the problem is solved?

- What are potential roadblocks? Think cultural, policy/privacy related, political, technical and data availability timelime.

- What is the timeline, resources and budget for the proposed project? i.e., function of 5 months + # of team members + $2000

Start simple, iterate often, until a joint vision is defined!

What defines a data driven problem?

It Depends
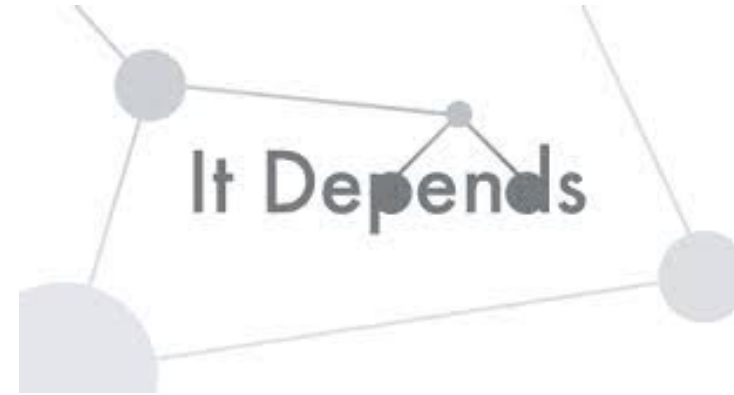
- When you know the question…
  - Look for factual answers
- When you have data or many potentially connected datasets…
  - Discover patterns
  - Look for known patterns
  - Explore connections and relationships
  - Derive questions

First focus on NEEDS to develop VISION!

YOU ARE HERE

Needs == why
Vision == how
Outcome == what

→ DESIGN APPROACH
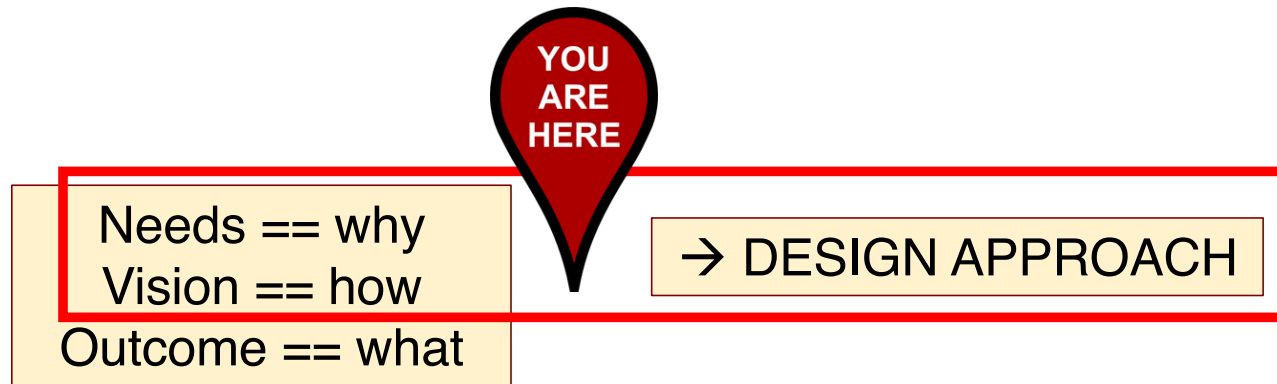
- Think of the first step as a design effort
- Every discussion needs to have a purpose driven by needs
- Ask small concise and purposeful questions about the data entities to start exploring with data
- Do not focus on what yet!

Success-Oriented Design

- Vision for how success happens
- Design baseline success metrics
- Develop a data strategy based on vision and metrics

Domain needs + questions + data → Vision + metrics → Data strategy

Division of Project Team Responsibilities

- Be flexible and ready to assume multiple roles
- Focus on your strengths, but also what you need to improve
- No bad tasks!
- Must assign:
  - a project coordinator/manager
  - a budget manager
  - a record keeper
- Each team member is expected to demonstrate both individual and collaborative work.

Report Guidelines

- Title, team members and advisor(s)
- Sections:
  - Challenge
  - Opportunities as a set of questions
  - Data sources
  - Approach
  - Team Roles and Responsibilities
  - Project Coordination and Communication Plan
  - Bullets for each team member's individual contributions in Step 1
- Keep it to 4-6 pages
- Due date: 1/18/2018 midnight

Questions?

Ikay Altintas, Ph.D.
Email: altintas@sdsc.edu