

Data Modeling Samples

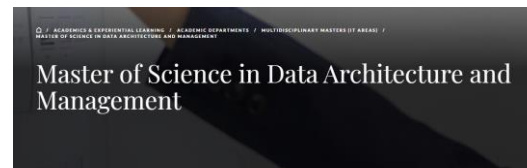
Graduate Student Data Model

- Examine Northeastern University graduate student course enrollment and performance in following MS programs:
 - [Information Systems](#)
 - [Data Architecture and Management](#)



The MSIS Degree Focuses on the Intersection of Business, IT Management, and Engineering

In Northeastern's Master of Science, MS, in Information Systems (MSIS) degree program, you will embark on an IT leadership pathway as a more evolved, dynamic software engineer. Through an innovative curriculum and a rigorous academic program that is both broad and deep, master's information systems students will become problem solvers who design groundbreaking software solutions that fit within the complexities of the business world and solve the humanistic needs of the people who use them.

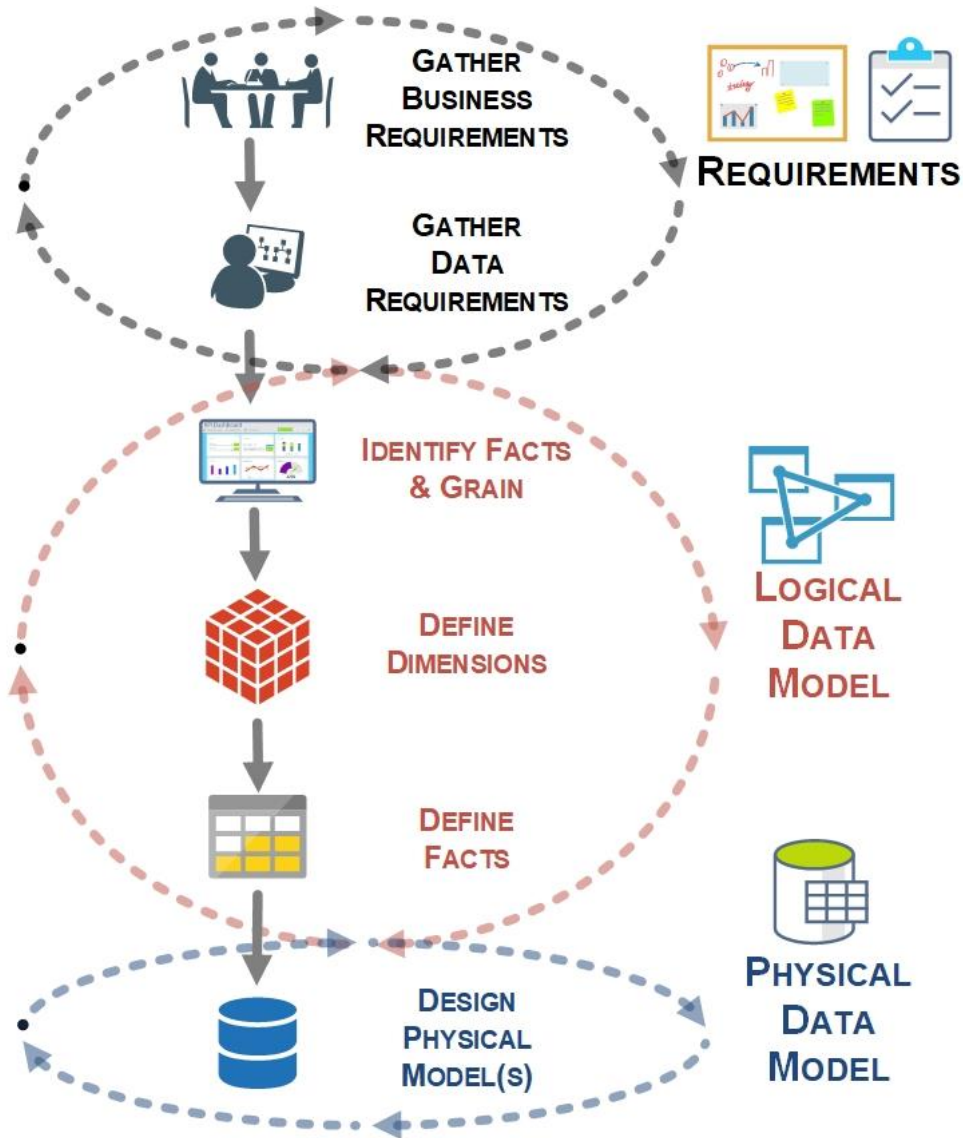


Overview

The Master of Science, MS, in Data Architecture and Management degree at Northeastern University is designed to prepare you for the evolving and growing field of data engineers. By providing you the holistic knowledge and technical skill to engineer big data systems, curate data, and integrate and process data so that data scientists and business analysts can access and interpret data in context to provide meaningful results to management. Businesses across industries have a rising need for data engineers—those who can integrate gigantic amounts of fragmented data, often from thousands of data sources, and transform it into useful information used to identify business opportunities and avoid risk.

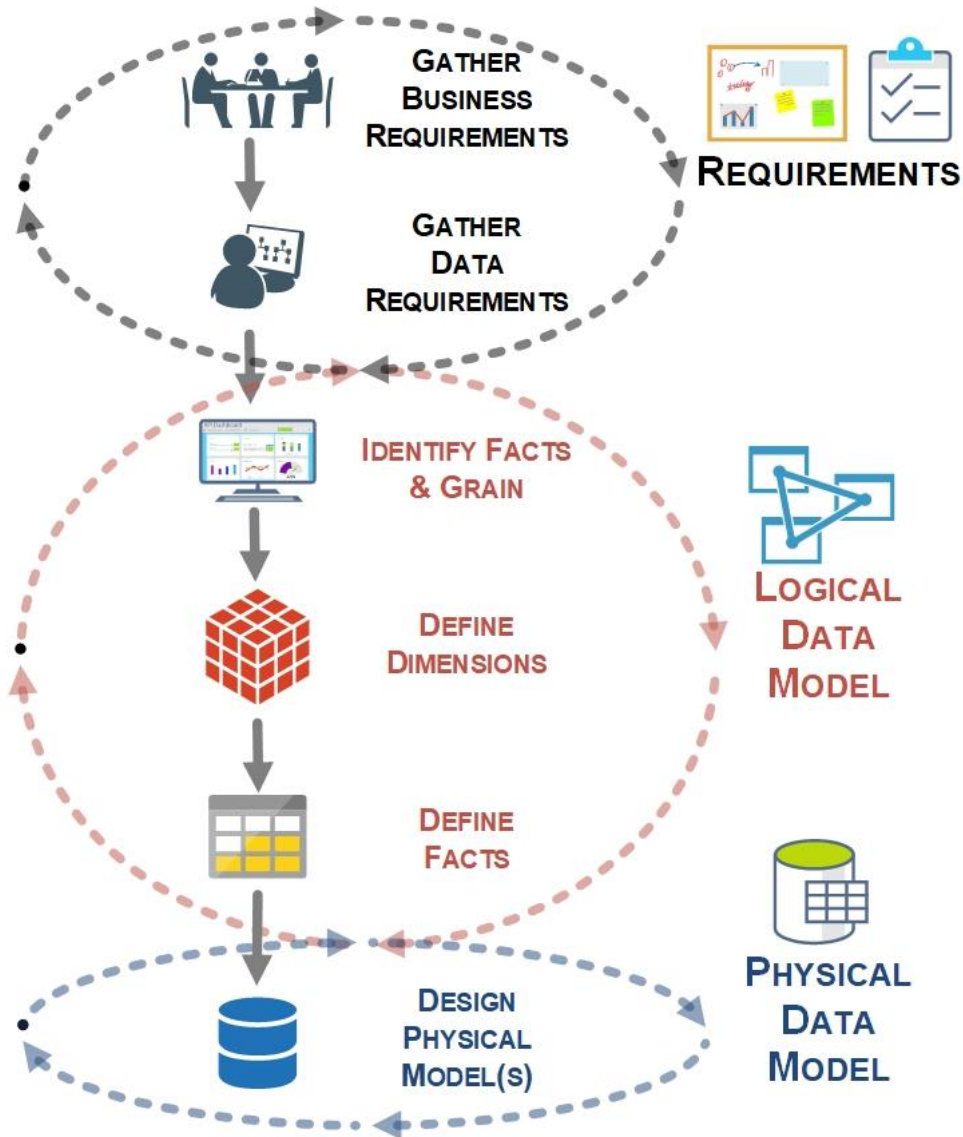


Data Modeling Lifecycle



- Three phases of data model design
 - Requirements
 - Business
 - Data
 - Logical Data Model
 - Identify Facts & Grain
 - Define Dimensions
 - Define Facts
 - Physical Data Model
 - Define SQL DDL to create tables & columns

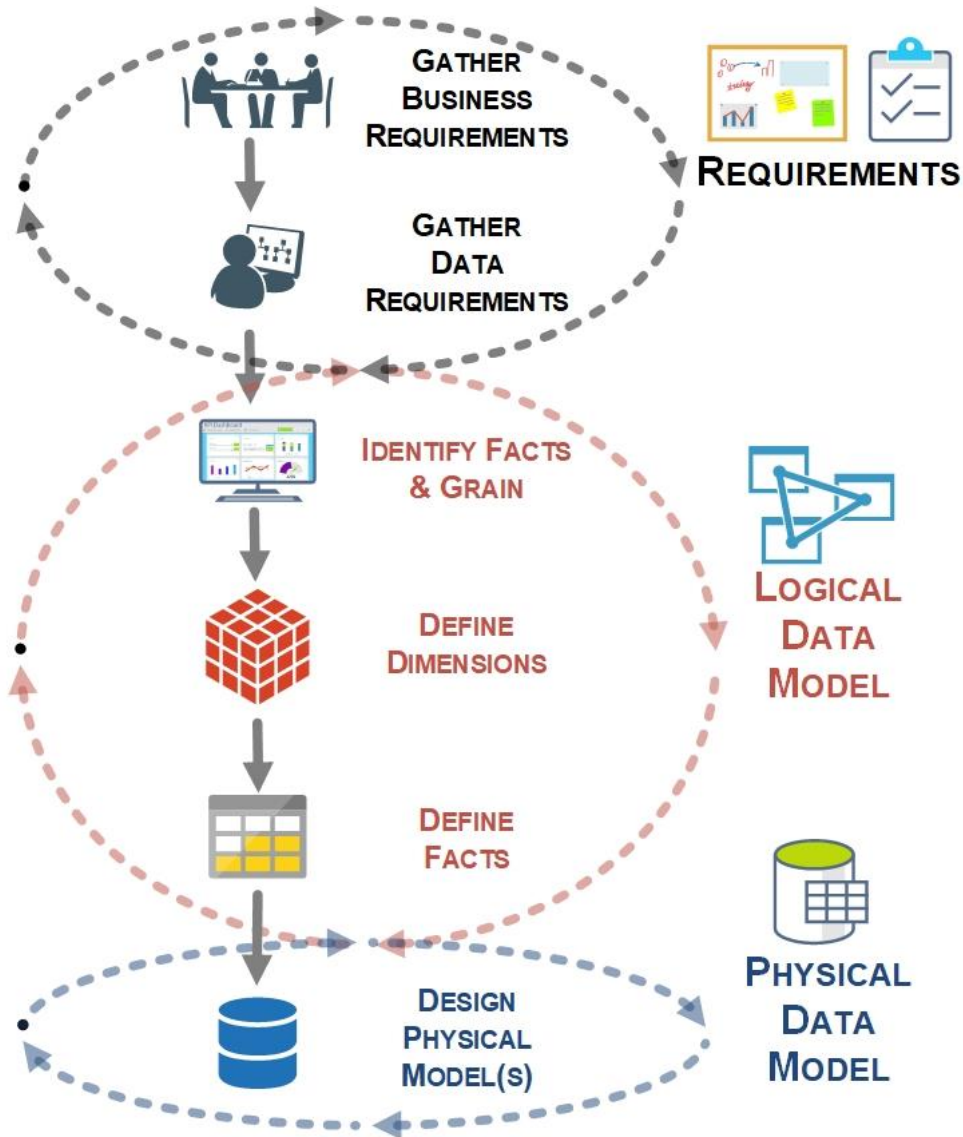
Data Modeling Lifecycle



Requirements

- Gather, analyze & prioritize business requirements
- Identify business processes
- Identify data sources
- Determine if data requirements is user-based or source-based
- If user-based (new business process)
 - Create conceptual model
- If source-based
 - Review existing data models
 - Perform Source systems analysis
 - Perform data profiling

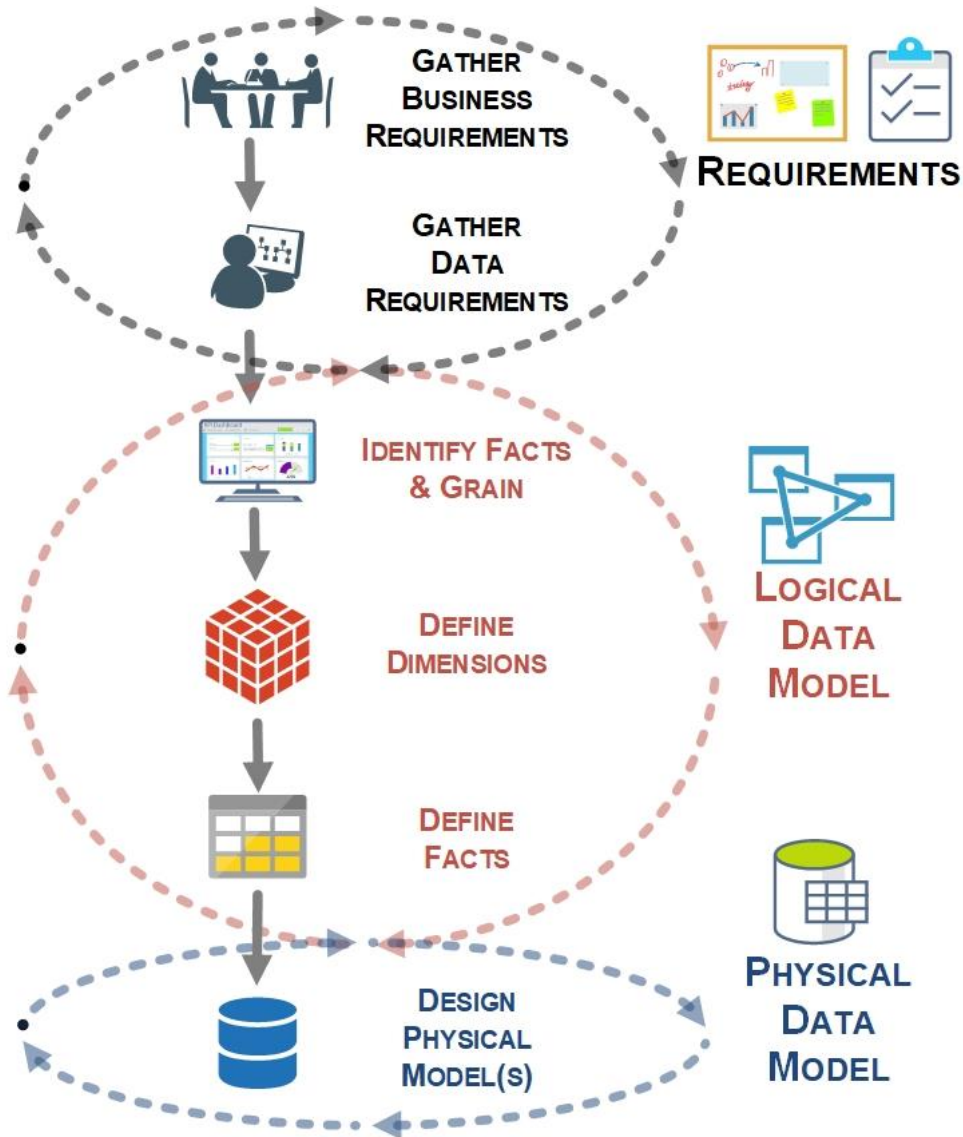
Data Modeling Lifecycle



Determine Grain

- Identify grain(s) in business processes
- Identify facts
 - Transaction, Periodic & Accumulating
- Identify fact table types
- Identify fact table granularity
- Identify preliminary dimensions

Data Modeling Lifecycle

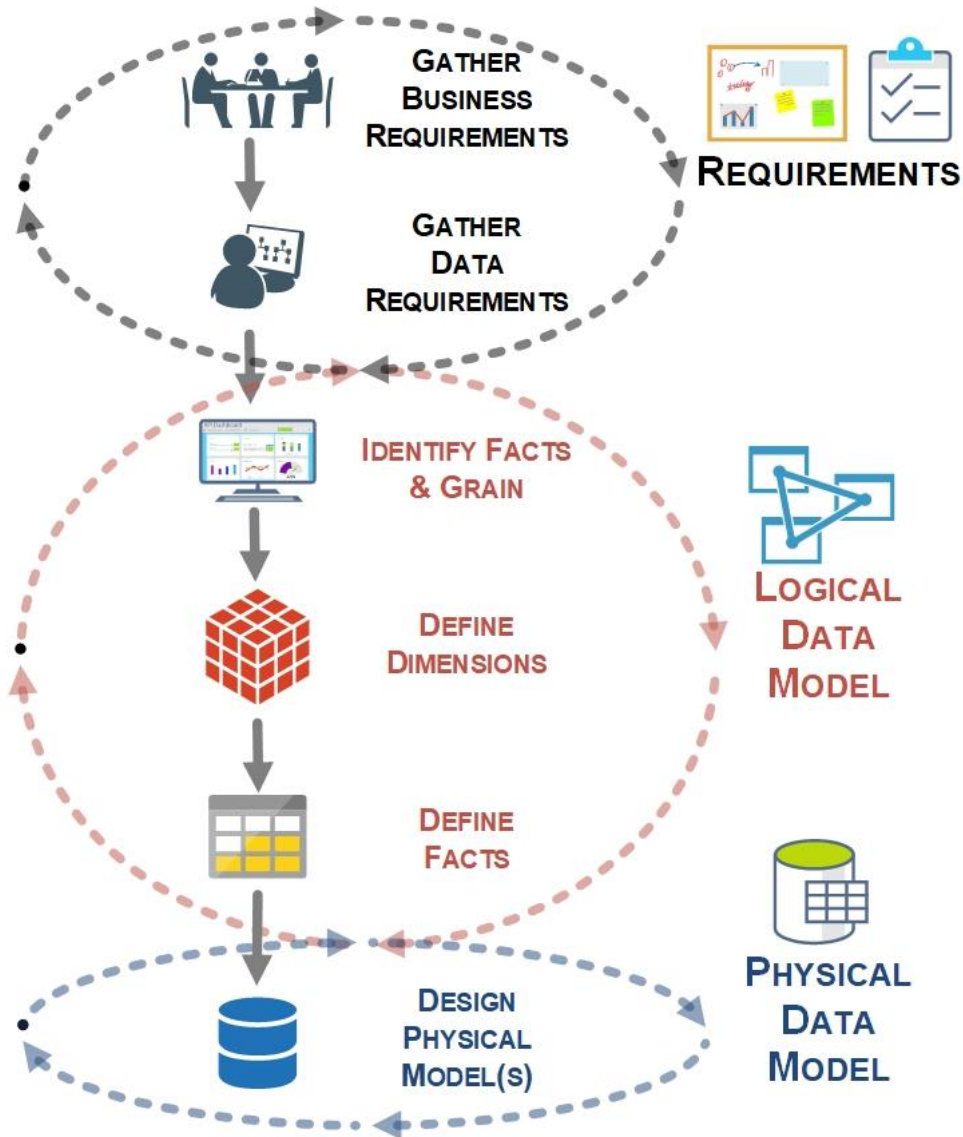


Define Dimensions

- Determine all dimensions
- Identify degenerate & conformed dimensions
- Identify dimensional attributes & validate granularity
- Identify hierarchies & attributes
- Identify date & time attributes
- Identify slowly changing dimensions (SCD) & types
- Identify multi-valued dimensions & define approach
- Identify role-playing dimensions
- Identify & classify specialized dimensions
 - Junk, Rapidly Changing, Hot Swappable, etc.
- Define keys & data relationships:
 - Surrogate keys (SKs), natural keys (NKs) and alternative keys (AKs)
 - Foreign keys (FKs)
- Define change data capture (CDC) attributes

Note: Not all listed may apply to a particular data model

Data Modeling Lifecycle

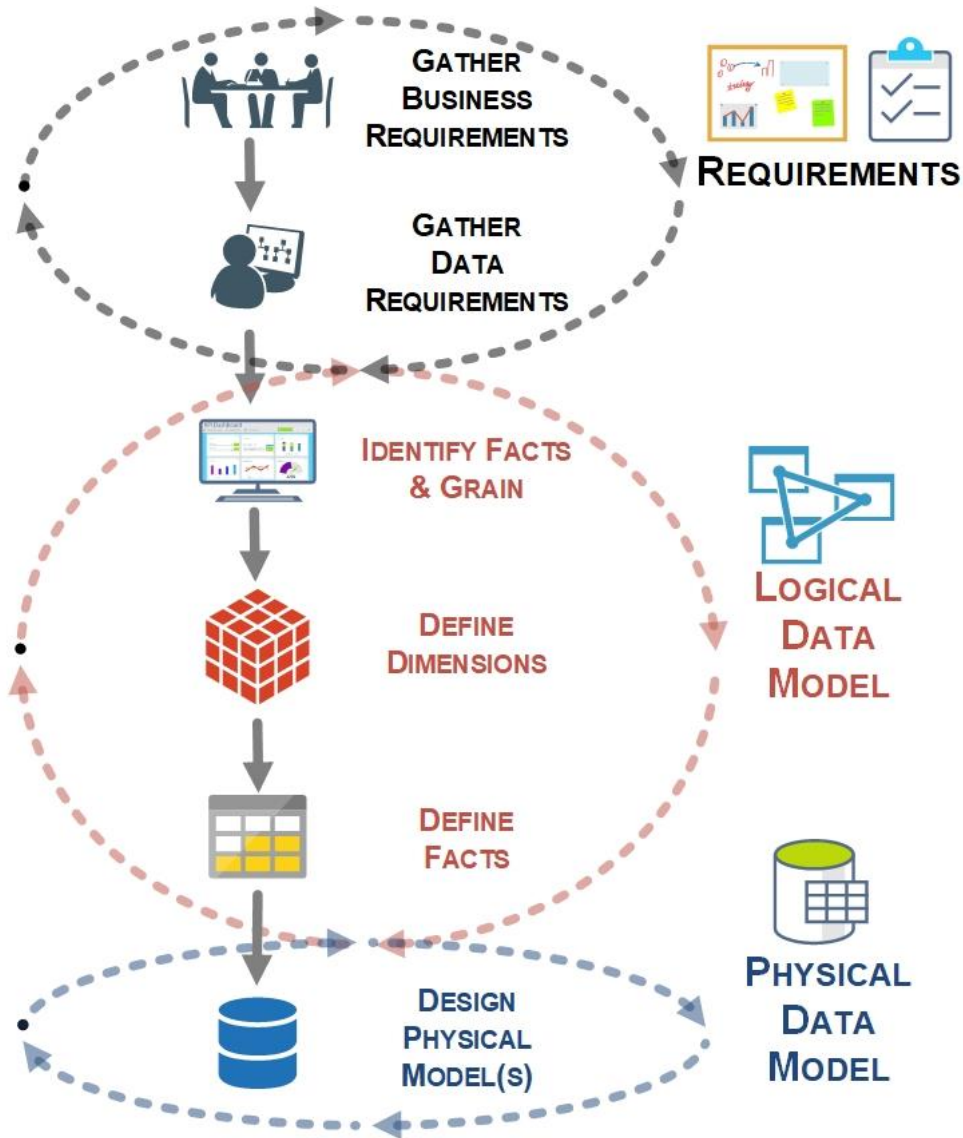


Define Facts

- Determine all facts
- Identify conformed facts
- Categorize fact table types & define approach
 - Transactional and event
 - Periodic
 - Accumulating
- Identify fact attribute types
 - Additive, semi-additive & non-additive
- Identify derived attributes & define approach
- Identify composite keys & design primary key (PK) approach
- Define dimensional hierarchies & define approach
 - Identify potential aggregates & define approach
- Identify composite keys & design PK approach

Note: Not all listed may apply to a particular data model

Data Modeling Lifecycle



Design Physical Data Model

- Estimate dimension & fact tables sizing & growth
- Determine target database(s)
 - DBMS type
 - Specific DBMS
- Define tables and attributes according to specific DBMS
- Define keys as appropriate - PKs, SKs, FKs
- Determine use cases for views and materialized views such as role-playing dimensions
- Define performance tuning approach
 - Different types of indexes, partitioning, etc.

Defining Dimensional Data Model

University Graduate Student

Data Modeling Lifecycle - Graduate Student Data Model

Data model design

- Requirements
 - **Business**
 - Data
 - Logical Data Model
 - Identify Facts & Grain
 - Define Dimensions
 - Define Facts
 - Physical Data Model
 - Define SQL DDL to create tables & columns
- Identify business analysis
 - Examine Northeastern University graduate student course enrollment and performance in following MS programs:
 - [Data Architecture and Management](#)
 - [Information Systems](#)
 - Analyze students' program performance
 - Courses taken including current enrolled, grades, GPA, etc.

Data Modeling Lifecycle - Graduate Student Data Model

Data model design

- Requirements
 - Business
 - **Data**
 - Logical Data Model
 - Identify Facts & Grain
 - Define Dimensions
 - Define Facts
 - Physical Data Model
 - Define SQL DDL to create tables & columns
- Gather, analyze & prioritize business requirement
 - What are the core requirements and elective courses for these programs
 - What classes (course & section) did students complete in a semester & what were their grades
 - What were the enrollment and drop dates if applicable for students in each class in a semester
 - Who were the teachers in each class above
 - What were the classes taught each semester
 - What teachers taught classes in a degree program in a semester
 - Who are the students enrolled in a degree program and attributes such as ID, email, date of birth (DOB), hometown, campus/off-campus address if applicable, etc.

Data Modeling Lifecycle - Graduate Student Data Model

Data model design

- Requirements
 - Business
 - Data
- Logical Data Model
 - **Identify Facts & Grain**
 - Define Dimensions
 - Define Facts
- Physical Data Model
 - Define SQL DDL to create tables & columns

Determine Grain

- Identify grain(s) in business processes
- Identify facts
- Identify fact table types
 - Transaction, Periodic & Accumulating
- Identify fact table granularity
- Identify preliminary dimensions

Data Modeling Lifecycle - Graduate Student Data Model

Data model design

- Requirements
 - Business
 - Data
 - Logical Data Model
 - Identify Facts & Grain
 - Define Dimensions
 - Define Facts
 - Physical Data Model
 - Define SQL DDL to create tables & columns
- Identify high level entities and measures (metrics)
 - College, Department
 - Degree Programs
 - Semesters
 - Courses
 - Classes (course, section, semester)
 - Students
 - Teachers
 - Course grades
 - ???

Data Modeling Lifecycle - Graduate Student Data Model

Data model design

- Requirements
 - Business
 - Data
- Logical Data Model
 - Identify Facts & Grain
 - **Define Dimensions**
 - Define Facts
- Physical Data Model
 - Define SQL DDL to create tables & columns

Define Dimensions

- Determine all dimensions
- Identify degenerate & conformed dimensions
- Identify dimensional attributes & validate granularity
- Identify hierarchies & attributes
- Identify date & time attributes
- Identify slowly changing dimensions (SCD) & types
- Identify multi-valued dimensions & define approach
- Identify role-playing dimensions
- Identify & classify specialized dimensions
 - Junk, Rapidly Changing, Hot Swappable, etc.
- Define keys & data relationships:
 - Surrogate keys (SKs), natural keys (NKs) and alternative keys (AKs)
 - Foreign keys (FKs)
- Define change data capture (CDC) attributes

Note: Not all listed may apply to a particular data model

Data Modeling Lifecycle - Graduate Student Data Model

Data model design

- Requirements
 - Business
 - Data
- Logical Data Model
 - Identify Facts & Grain
 - Define Dimensions
 - **Define Facts**
- Physical Data Model
 - Define SQL DDL to create tables & columns

Define Facts

- Determine all facts
- Identify conformed facts
- Categorize fact table types & define approach
 - Transactional and event
 - Periodic
 - Accumulating
- Identify fact attribute types
 - Additive, semi-additive & non-additive
- Identify derived attributes & define approach
- Identify composite keys & design primary key (PK) approach
- Define dimensional hierarchies & define approach
 - Identify potential aggregates & define approach
- Identify composite keys & design PK approach

Note: Not all listed may apply to a particular data model

Data Modeling Lifecycle - Graduate Student Data Model

Data model design

- Requirements
 - Business
 - Data
 - Logical Data Model
 - Identify Facts & Grain
 - Define Dimensions
 - Define Facts
 - Physical Data Model
 - Define SQL DDL to create tables & columns
- Create dimensional data model using data modeling tool
 - Create SQL DDL (data definition language) scripts
 - Create tables and columns with primary keys (PKs) and foreign keys (FKs) defined
 - Out of scope:
 - Indexes, triggers, constraints and other physical attributes other than what is necessary to define PKs and FKs

Graduate Student Data Model

Deliverables:

- Create a list of dimensions & facts
- Create tables with a sample list of attributes for the above
- Create a dimensional model using ER/Studio
 - Upload data model
- Create sample data & store in Excel, csv or tsv files
 - Note: can use actual course related data *BUT students & grade data should be fictitious*
- Create a SQL Server, MySQL, PostgreSQL or Oracle (on cloud or notebook) database with data model above - Upload SQL script
- Load the sample data using Alteryx data prep
- Create SQL queries answering some of the questions used for requirements
 - Upload SQL queries & results
- Bonus: Use PowerBI to display some of the data above
 - Upload PowerBI file