

Big Data Modeling and Management Systems

by University of California San Diego

About this Course

Once you've identified a big data issue to analyze, how do you collect, store and organize your data using Big Data solutions? In this course, you will experience various data genres and management tools appropriate for each. You will be able to describe the reasons behind the evolving plethora of new big data platforms from the perspective of big data management systems and analytical tools. Through guided hands-on tutorials, you will become familiar with techniques using real-time and semi-structured data examples. Systems and tools discussed include: AsterixDB, HP Vertica, Impala, Neo4j, Redis, SparkSQL. This course provides techniques to extract value from existing untapped data sources and discovering new data sources.

At the end of this course, you will be able to:

- * Recognize different data elements in your own work and in everyday life problems
- * Explain why your team needs to design a Big Data Infrastructure Plan and Information System Design
- * Identify the frequent data operations required for various types of data
- * Select a data model to suit the characteristics of your data
- * Apply techniques to handle streaming data
- * Differentiate between a traditional Database Management System and a Big Data Management System
- * Appreciate why there are so many data management systems
- * Design a big data information system for an online game company

This course is for those new to data science. Completion of Intro to Big Data is recommended. No prior programming experience is needed, although the ability to install applications and utilize a virtual machine is necessary to complete the hands-on assignments. Refer to the specialization technical requirements for complete hardware and software specifications.

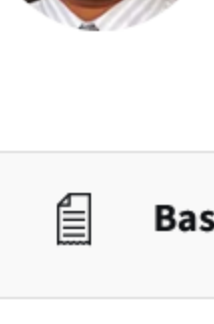
Hardware Requirements:

(A) Quad Core Processor (VT-x or AMD-V support recommended), 64-bit; (B) 8 GB RAM; (C) 20 GB disk free. How to find your hardware information: (Windows): Open System by clicking the Start button, right-clicking Computer, and then clicking Properties; (Mac): Open Overview by clicking on the Apple menu and clicking "About This Mac." Most computers with 8 GB RAM purchased in the last 3 years will meet the minimum requirements.You will need a high speed internet connection because you will be downloading files up to 4 Gb in size.

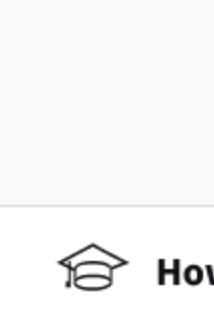
Software Requirements:

This course relies on several open-source software tools, including Apache Hadoop. All required software can be downloaded and installed free of charge (except for data charges from your internet provider). Software requirements include: Windows 7+, Mac OS X 10.10+, Ubuntu 14.04+ or CentOS 6+ VirtualBox 5+.




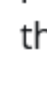

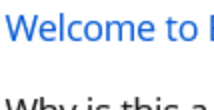
[Show less](#)



Taught by:
[Ilkay Altintas](#), Chief Data Science Officer
San Diego Supercomputer Center



Taught by:
[Amarnath Gupta](#), Director, Advanced Query Processing Lab
San Diego Supercomputer Center (SDSC)

 Basic Info	Course 2 of 6 in the Big Data Specialization
 Commitment	6 weeks of study, 2-3 hours/week
 Language	English, Subtitles: Arabic, French, Bengali, Ukrainian, Chinese (Simplified), Greek, Italian, Portuguese (Brazil), Vietnamese, Dutch, Korean, Oriya, German, Pashto, Urdu, Russian, Thai, Indonesian, Swedish, Turkish, Azerbaijani, Spanish, Dari, Hindi, Japanese, Kazakh, Hungarian, Polish
 How To Pass	Pass all graded assignments to complete the course.
 User Ratings	 Average User Rating 4.4

Syllabus

Module 1

Introduction to Big Data Modeling and Management

Welcome to this course on big data modeling and management. Modeling and managing data is a central focus of all big data projects. In these lessons we introduce you to the concepts behind big data modeling and management and set the stage for the remainder of the course.

 14 videos, 8 readings


1. **Video:** [Welcome to Big Data Modeling and Management](#)
2. **Video:** Why is this a New Course in the Big Data Specialization?
3. **Discussion Prompt:** Getting to know you: Tell us about yourself and why you are taking this course
4. **Video:** Summary of Introduction to Big Data (Part 1)
5. **Video:** Summary of Introduction to Big Data (Part 2)
6. **Video:** Summary of Introduction to Big Data (Part 3)
7. **Reading:** Slides: Summary of Introduction to Big Data
8. **Video:** Big Data Management "Must-Ask Questions"
9. **Video:** Data Ingestion
10. **Video:** Data Storage
11. **Video:** Data Quality
12. **Video:** Data Operations
13. **Video:** Data Scalability and Security
14. **Reading:** Slides: Big Data Management
15. **Discussion Prompt:** Let's discuss: What area of big data management interests you most?
16. **Reading:** Reading on Storage Systems
17. **Video:** Energy Data Management Challenges at ConEd
18. **Reading:** Slides: Energy Data Management Challenges at ConEd
19. **Video:** Gaming Industry Data Management: Q&A with Apmetrix CTO Mark Caldwell
20. **Video:** Flight Data Management at FlightStats: A Lecture by CTO Chad Berkley
21. **Reading:** Slides: Flight Data Management at FlightStats
22. **Discussion Prompt:** Let's discuss: What are the design criteria in the big data applications you have heard?
23. **Reading:** Downloading and Installing Docker Desktop Instructions
24. **Reading:** Instructions for Downloading Hands On Datasets
25. **Reading:** Basic terminal shell commands

[Show less](#)

Module 2

Big Data Modeling

Modeling big data depends on many factors including data structure, which operations may be performed on the data, and what constraints are placed on the models. In these lessons you will learn the details about big data modeling and you will gain the practical skills you will need for modeling your own big data projects.

 11 videos, 12 readings

1. **Video:** [Introduction to Data Models](#)
2. **Video:** Data Model Structures
3. **Video:** Data Model Operations
4. **Video:** Data Model Constraints
5. **Reading:** Slides: What Is A Data Model?
6. **Discussion Prompt:** Let's discuss: Modeling data in your daily life
7. **Reading:** Introduction to CSV Data (OpenOffice)
8. **Reading:** Introduction to CSV Data (Microsoft Excel)
9. **Video:** Introduction to CSV Data
10. **Video:** What is a Relational Data Model?
11. **Reading:** Slides: What Is A Relational Data Model?
12. **Video:** What is a Semistructured Data Model?
13. **Reading:** Slides: What is a Semistructured Data Model?
14. **Discussion Prompt:** Let's discuss: Utilization of XML or JSON on the Internet
15. **Reading:** Exploring the Relational Data Model of Comma Separated Values (OpenOffice)
16. **Reading:** Exploring the Relational Data Model of Comma Separated Values (Excel)
17. **Video:** Exploring the Relational Data Model of CSV Files
18. **Reading:** Installing Python
19. **Reading:** Creating a Python Virtual Environment
20. **Reading:** Exploring the Semistructured Data Model of JSON data
21. **Video:** Exploring the Semistructured Data Model of JSON data
22. **Reading:** Exploring the Array Data Model of an Image
23. **Video:** Exploring the Array Data Model of an Image
24. **Reading:** Exploring Sensor Data
25. **Video:** Exploring Sensor Data

[Show less](#)

 **Graded:** Practical Quiz for Week 2 Hands-On Lectures

Module 3


Big Data Modeling (Part 2)

These lessons continue to shed light on big data modeling with specific approaches including vector space models, graph data models, and more.

 5 videos, 5 readings

1. **Video:** [Vector Space Model](#)
2. **Reading:** Slides: Vector Space Model
3. **Video:** Graph Data Model
4. **Reading:** Slides: Graph Data Model
5. **Video:** Other Data Models
6. **Reading:** Slides: Other Data Models
7. **Reading:** Exploring Vector Data Models with Lucene
8. **Video:** Exploring the Lucene Search Engine's Vector Data Model
9. **Reading:** Exploring Graph Data Models with Gephi
10. **Video:** Exploring Graph Data Models with Gephi

[Show less](#)

 **Graded:** Data Models Quiz

Module 4

Working With Data Models

Data models deal with many different types of data formats. Streaming data is becoming ubiquitous, and working with streaming data requires a different approach from working with static data. In these lessons you will gain practical hands-on experience working with different forms of streaming data including weather data and twitter feeds.

 5 videos, 5 readings

1. **Video:** [Data Model vs. Data Format](#)
2. **Reading:** Slides: Data Model vs. Data Format
3. **Video:** What is a Data Stream?
4. **Reading:** Slides: What is a Data Stream?
5. **Video:** Why is Streaming Data different?
6. **Reading:** Slides: Why is Streaming Data Different?
7. **Video:** Understanding Data Lakes
8. **Reading:** Slides: Understanding Data Lakes
9. **Discussion Prompt:** Let's discuss: Streaming data applications
10. **Reading:** Exploring Streaming Sensor Data
11. **Video:** Exploring Streaming Sensor Data

[Show less](#)

 **Graded:** Data Formats and Streaming Data Quiz

Module 5


Big Data Management: The "M" in DBMS

Managing big data requires a different approach to database management systems because of the wide variation in data structure which does not lend itself to traditional DBMSs. There are many applications available to help with big data management. In these lessons we introduce you to some of these applications and provide insight into how and when they might be appropriate for your own big data management challenges.

 7 videos, 2 readings

1. **Video:** [DBMS-based and non-DBMS-based Approaches to Big Data](#)
2. **Reading:** Slides: DBMS-based and non-DBMS-based Approaches to Big Data
3. **Video:** From DBMS to BDMS
4. **Video:** Redis: An Enhanced Key-Value Store
5. **Video:** Aerospike: a New Generation KV Store
6. **Video:** Semistructured Data - AsterixDB
7. **Video:** Solr: Managing Text
8. **Video:** Relational Data - Vertica
9. **Reading:** Slides: From DBMS to BDMS


[Show less](#)

 **Graded:** BDMS Quiz

Module 6

Designing a Big Data Management System for an Online Game

In these lessons we give you the opportunity to learn about big data modeling and management using a fictitious online game called "Catch the Pink Flamingo".

 1 reading

1. **Reading:** A Game by Eglenec Inc. : Catch The Pink Flamingo
2. **Discussion Prompt:** Let's discuss: Analytical tasks to make Catch the Pink Flamingo better
3. **Discussion Prompt:** Let's discuss: Using the data model for Catch the Pink Flamingo

[Show less](#)

 **Graded:** Designing a Data Model for 'Catch the Pink Flamingo'

[View Less](#)

How It Works

General

What do start dates and end dates mean?

Once you enroll,

[▼ More](#)

Peer-graded assignments

Peer-graded assignments require you and your classmates to grade each other's work.

[▼ More](#)

Course 2 of Specialization

Unlock Value in Massive Datasets

Learn fundamental big data methods in six straightforward courses.



Big Data
University of California San Diego


[Learn More](#)

[View the course in catalog](#)

Related Courses



Big Data - Capstone Project
University of California San Diego




Introduction to Big Data
University of California San Diego



Graph Analytics for Big Data
University of California San Diego



Machine Learning With Big Data
University of California San Diego



Big Data Integration and Processing
University of California San Diego