Course Outline

General Information

COURSE ID (CB01A AND CB01B)

CIS 44F

COURSE TITLE (CB02)

Introduction to Big Data and Analytics

COURSE CREDIT STATUS

Credit - Degree Applicable

EFFECTIVE TERM

Fall 2024

COURSE DESCRIPTION

This course is an introduction to Big-Data deluge, management of unstructured and structured data and design of large scale database systems. Concepts covered include map-reduce parallel processing algorithms, real-time analytics, classification, and predictive analytics, attributes of Big-Data and related issues. The course also introduces large-scale file systems and operations and parallel processing algorithms.

FACULTY REQUIREMENTS

COURSE FAMILY

Not Applicable

Course Justification

This course is transferable to all California State University campuses and is required as part of the Database Design for Developers (Oracle) Certificate of Achievement. Big Data and Analytics stand today as the standard storage and analysis system. It introduces learners with language to access extremely large storage systems for creating and managing a database. It is beneficial for those with careers in IT, including Database Architects, Database Administrators, and Database Designers, to hold certification for a specific database software program.

Foothill Equivalency

DOES THE COURSE HAVE A FOOTHILL EQUIVALENT?

No

FOOTHILL COURSE ID

Formerly Statement

FORMERLY STATEMENT	(Formerly CIS 064F.)

Course Development Options

BASIC SKILL STATUS (CB08)

Course is not a basic skills course.

GRADE OPTIONS

- Letter Grade
- Pass/No Pass

REPEAT LIMIT

0

Transferability & Gen. Ed. Options

Information below is subject to change. For the official listing of courses, their approval dates, and transfer credit limitations, check the De Anza catalog (by academic year), ASSIST.ORG (https://assist.org/) and C-ID.NET (https://c-id.net/).

TRANSFERABILITY

Transferable to both UC and CSU

Units and Hours

Summary

MINIMUM CREDIT UNITS	4.0
MAXIMUM CREDIT UNITS	4.0

Weekly Student Hours

Type	In Class	Out of Class
Lecture Hours	4.0	8.0
Laboratory Hours	0.0	0.0

Course Student Hours

COURSE DURATION (WEEKS)

12.0

HOURS PER UNIT DIVISOR

36.0

Course In-Class (Contact) Hours

LECTURE

48.0

LABORATORY

0.0

TOTAL

48.0

Course Out-of-Class Hours

LECTURE

96.0

LABORATORY

0.0

NA

0.0

TOTAL

96.0

Prerequisite(s)

Corequisite(s)

Advisory(ies)

ESL 272 and ESL 273, or ESL 472 and ESL 473, or eligibility for EWRT 1A or EWRT 1AH or ESL 5

Limitation(s) on Enrollment

Entrance Skill(s)

General Course Statement(s)

Methods of Instruction

Lecture and visual aids
Discussion of assigned reading Discussion and problem solving performed in class Collaborative learning and small group exercises Collaborative projects Collaborative learning and small group exercises Homework and extended projects

Assignments

A. Readings from Text.
B. Documenting, coding, testing and debugging six to ten programs with guidance provided with clearly documented design, half completed in the computer lab, half completed as homework.

Methods of Evaluation

A. One or two midterm examinations requiring some programming, concepts clarification and exhibiting mastery of large scale database systems principles.
B. A final examination requiring concepts clarification and exhibiting mastery of large scale database system principles.
C. Evaluation of programming assignments, based on correctness, documentation, code quality, and test plan executions.

Essential Student Materials/Essential College Facilities

Essential Student Materials:

- None.

Essential College Facilities:

- None.

Examples of Primary Texts and References

Author	Title	Publisher	Date/Edition	ISBN
Danette McGilvray	Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (TM)	Academic Press	June 4, 2021 - 2nd edition	ISBN-13 L 978-0128180150

Examples of Supporting Texts and References

Author	Title	Publisher
None.		

Learning Outcomes and Objectives

Course Objectives

- Explore big-data technologies as means to solving key business analytical problems.
- Interpret and analyze techniques for setting up patterns for data analysis.
- Compare and contrast the data and relation algorithms.
- Examine data pre-processing and visualization techniques for enabling data analytic scenarios.
- Articulate the characteristics of regression, forecasting and classification techniques for predictive analytics.
- Interpret and analyze architecture of database clustering technologies.

CSLOs

- Design, implement and debug a large scale database system using technology like Hadoop or Cassandra.
- Perform data analysis using a large-scale database systems given a set of user requirements.

Outline

A. Explore big-data technologies as means to solving key business analytical problems.
    1. Data analytics, Data mining and knowledge discovery.
    2. Competitor, intelligence and big data.
    3. Business case studies: Electronic Health Records (EHR), US Dept of Transportation.
B. Interpret and analyze techniques for setting up patterns for data analysis.
    1. RDBMS Relational Modeling
    2. No-SQL DB Modeling
    3. Datawarehousing modeling, data mining and online analytical processing.
C. Compare and contrast the data and relation algorithms.
    1. Auto-Associator
    2. Component Analysis
    3. Diagrams
    4. Multidimensional Scaling
    5. Histograms
D. Examine data pre-processing and visualization techniques for enabling data analytic scenarios.
    1. Error Type and Error Handling
    2. Filtering
    3. Data Transformation
    4. Data Merging
    5. Linear Correlation, correlation and causality.
    6. Chi-square test for independence

E. Articulate the characteristics of regression, forecasting and classification techniques for predictive analytics.
1. Linear regression, linear regression with nonlinear substitution and robust regression.
2. Cross validation and feature selection.
3. Finite state machines, recurrent models and autoregressive models.
4. Classification criteria, naive bayes classifier and linear discriminant analysis.
5. Support vector machines, nearest neighbor classifier and learning vector quantization.
6. Decision Trees

F. Interpret and analyze architecture of database clustering technologies.
1. Hadoop
2. Oracle RAC
3. MySQL Clusters
4. Windows Clustering
5. Cassandra
6. Trackvia, nCluster from Teradata.