

DRAFT

Master of Business Administration (Data Analytics)

Syllabus for the MBA (Data Analytics) (2016 – 2018)



Centurion
UNIVERSITY

Centurion University of Technology and Management (CUTM)

Ramachandrapur, PO: Jatni, Dist: Khurda, 752050

www.cutm.ac.in

April 2016

Table 1: Semester-wise Academic Curriculum for the MBA (Data Analytics) (2016–2018)

Semester 1 (24 Credits of Compulsory Courses in the Class Room)
<ol style="list-style-type: none"> 1. Business Economics (4 Credit) 2. Financial Accounting (4 Credit) 3. Individual and Group Behaviour (2 Credit) 4. Managerial Analysis and Communication I (2 Credit) 5. Quantitative Analysis (4 Credit) 6. Basics of Business Intelligence & D/W(4 Credit) 7. Marketing Management & Analytics (4 Credit)
Semester 2 (26 Credits of Compulsory Courses in the Class Room)
<ol style="list-style-type: none"> 1. Managerial Analysis and Communication II (2 Credit) 2. Financial Management (4 Credit) 3. Machine Learning & Computational Statistics (4 Credit) 4. Web Analytics and Intelligence (4 Credit) 5. D/W Project Life Cycle Management (4 Credit) 6. ETL, Data Profiling and Data Modelling with Pentaho (4 Credit) 7. Data Visualisation with Advance XIs (4 Credit)
Special Project: 6 Credit in the Organisation
Semester 3 (24 Credits of Courses in Project Mode in the Organisation)
<ol style="list-style-type: none"> 1. Unstructured Data Analytics with Hadoop/Big Data/NoSql (4 Credit) 2. Advanced Visualisation – Tableau/ QlikView and D3js (4 Credit) 3. Advanced Data Modelling & Mining with R/Python (4 Credit) 4. Project - ETL (4 Credit) 5. Project - Visualisation (4 Credit) 6. Project – (Data Modelling & Mining) (4 Credit)
Semester 4 (26 Credits of Course in Project Mode in the Organisation)
<ol style="list-style-type: none"> 1. Internship (26 Credits)

Note:

- Total 106 Credit (24+26+6+24+26)
- 2nd year in Project Mode (in attachment with organisation)
- Students to select and work in the Domain of their choice in 2nd year
- Available Domains for Electives are **Financial Services, Utilities, Sales and Health & Pharma**



COURSE DESCRIPTIONS¹

Semester-I

1. Business Economics (4 Credit)

Course Objective: The primary objective of this course is to develop an understanding of different microeconomic and macroeconomic concepts and learn their applications in the process of development and business decision making.

Curriculum: The firm and its goal, Analysis of demand and supply, Analysis of elasticity, Estimation and forecasting of demand function, Estimation of production and cost function, Breakeven analysis, Market structure, conduct and performance including perfect competition and economic efficiency, monopoly, allocative efficiency, oligopoly; Pricing methods covering pricing of multiple products, employment inputs; Taxes and subsidies, and price controls.

National income accounting; growth and development; conceptualisation of inflation; unemployment and poverty as central endogenous macroeconomic variables; fiscal and monetary policies; understanding of government budgets; economic reform and structural adjustment programmes; implications of the policies of liberalisation; privatization and globalization; economic reforms

Text book: 1. Managerial Economics, 4th Edition, by H. Craig Petersen and W. Cris Lewis, Pearson Education.
2. Economics, 18th Edition, by Paul Samuelson, W. D. Nordhaus, Tata McGraw Hill, New Delhi.

2. Financial Accounting (4 Credit)

Course Objective: The objective of the course is to familiarise students with basic accounting principles, concepts and conventions for managerial decision making.

Curriculum: Business and accounting environment, objectives and uses of financial statement, assumptions, characteristics and definitions of financial statements, process of recognition, basis of measurement, requirements of companies' act, accounting standards, balance sheet, profit and loss account, cash flow statement analysis, accounting process, financial reports, inventory valuation, fixed asset accounting, accounting for depreciation, financial statement analysis.

Text book: 1. Financial Accounting for Management, by N. Ramachandran and Ram K. Kakani, Tata McGraw Hill, New Delhi.

3. Individual and Group Behaviour (2 Credit)

¹ For the text books, the latest year of reference and edition will be considered.

Course Objective: The objectives of the course are to understand the dynamic of individual, interpersonal and group behaviour in organizational setting, develop students' knowledge and competence to deal with human problems of management and developing students' awareness and insight for personal and professional growth.

Curriculum: Behaviour of individuals and small groups; learning; perceptual processes; beliefs; attitudes and values; personality; attribution; self-concept; motivation; leadership; active listening; interpersonal communication and interpersonal conflicts; and management of stress at workplace.

Text book: 1. Organisational Behaviour, by Steven L. McShane and Marry Ann Von, Tata McGraw Hill, New Delhi.

4. Managerial Analysis and Communication (2 Credit)

Course Objective: It will also help students to develop their written and oral communication ability.

Curriculum: Business communication including forms and media, types, barriers, ethical dilemma, three step process of written communication (planning, writing and completing), group dynamics, basics of team work, nonverbal and verbal communication, presentation skills, group discussion and interview techniques.

Text book: 1. Oxford Guide to Writing and Speaking, by John Seely, Oxford University Press

2. Business Communication Today by Bovee, Thill and Schatzman, Pearson Education

5. Quantitative Analysis (4 Credit)

Course Objective: The course is intended to help management students to appreciate the relevance of the use of quantified facts in the complex world of development and business decision making. The emphasis is on conceptual understanding of the basic mathematical and statistical tools and techniques and its meaningful applications.

Curriculum:

- Descriptive Statistics: Regression Analysis, Assumptions of OLS Regression, Regression Modelling. Correlation, ANOVA, Forecasting, Heteroscedasticity, Autocorrelation, Introduction to Multiple Regression.
- Statistical Inference & Regression Analysis: Need for Business Modelling, Regression – Concepts, Blue property-assumptions-Least Square Estimation, Variable Rationalization, and Model Building etc.
- Excel as a tool to analyze quantitative data
- Overview of R in statistical data analysis

Text book: 1. Statistics for Management, R. I. Levin and D. S. Rubin, 7th Edition,

Pearson Education.
2. Fundamentals of Statistics, S. C. Gupta, Himalaya.

6. Basics Of Business Intelligence & D/W (4 Credit)

- a) What is BI & DW
 - Data to Information Lifecycle
 - Brief History of Accessing, Reporting And Analyzing Data
 - Business Intelligence (BI) defined
 - Data Warehousing (DW) defined
 - DM VS EDW
- b) Where is BI & DW being used today
 - Business Drivers For BI
 - Business and IT Drivers For DW
 - Applications that use BI And DW
 - Data Shadow Systems
 - Industry terminology
- c) BI & DW - The Architectures
 - How do BI & DW fit together?
 - Components of D/W
 - OLAP Architectures (OLAP Vs OLTP)
 - Staging layer, core layer & semantic layer
 - Technology maps to various layers
- d) BI & DW - The Architectures
 - Data Integration & transformations (ETL Vs ELT)
 - Core layer (star schema /snowflake)
 - Data visualisations
 - Predictive modelling
- e) Data Integration & Transformations
 - Overview
 - Data modelling concepts
 - Data Integration Framework (DIF)
 - Transforming data to information
 - Process management
 - Data staging options
- f) D/W & Data Modelling
 - Data Stores
 - Data Warehouse, Data Marts, Operational Data Stores, Cubes
 - Star Schema
 - No SQL databases
 - Bigdata/Hadoop overview
- g) Data Visualisations
 - Prepare the data for Visualization,
 - Various visualisation techniques.
 - Standardized reporting & compliances, Decision Models
 - Pivot grid & chat
- h) Analytics
 - Descriptive vs Real-time vs predictive analytics
 - Excel and R

7. Marketing Management and Analysis (4 Credit)

Course Objective: The objective of the course is to make the students understand the basic concepts of marketing.

Curriculum: Introduction to Marketing Theory, managerial action frameworks- marketing concept; 4 Ps; and 4 Cs; components of marketing strategy- product; price; promotion and distribution; related illustrations; viewpoints and caselets; integrative concepts like positioning; segmentation; branding; advertising campaigns and marketing planning. Marketing processes such as supply chain, customer relationship, e-marketing or market knowledge management that are needed to achieve organisation's desired performance.

Text book: 1. Marketing Management – A South Asian Perspective, by Kotler, Keller, Koshy and Jha, Pearson Education, New Delhi.

Semester- II

1. Managerial Analysis and Communication (2 Credit)

Course Objective: The objective of this course is to help students to develop their written and oral communication ability.

Curriculum: Business communication including forms and media, types, barriers, ethical dilemma, three step process of written communication (planning, writing and completing), group dynamics, basics of team work, nonverbal and verbal communication, presentation skills, group discussion and interview techniques.

Text book: 1. Oxford Guide to Writing and Speaking, by John Seely, Oxford University Press

2. Business Communication Today by Bovee, Thill and Schatzman, Pearson Education

2. Financial Management (4 Credit)

Course Objective: The objective of the course is to familiarise students with basic concepts of financial management for managerial decision making.

Curriculum: Concepts and methods related to financial analysis; forecasting cash and working capital flows; capital investment analysis/capital budgeting techniques; cost of capital; capital structure; and operating and financial leverage; working capital management; time value of money and valuation of securities; and concepts of risk and return.

Text book: 1. Fundamentals of Financial Management, by Vanhorne and Wachowicz, Prentice Hall of India, New Delhi.

3. Machine Learning & Computational Statistics (4 Credit)

- a) Statistical Inference & Regression Analysis : Need for Business Modelling, Regression – Concepts, Blue property-assumptions-Least Square Estimation, Variable Rationalization, and Model Building etc (in detail)
- b) Reasoning: Logic and its Limits, Dealing with Uncertainty
- c) Mechanical Logic - The Semantic Web - Limits of Logic
- d) Description and Resolution - Collective Reasoning.
- e) Statistical Forecasting - Neural Networks
- f) Predictive Analytics - Sparse Memories - Sequence Memory - Network Science
- g) Data Analysis: Regression and Feature

4. Web Analytics and Intelligence (4 Credit)

- a) Inside the search engine - Examples of intelligent web applications - Basic elements of intelligent applications - Machine learning, data mining – Searching, Reading, indexing, and searching.
- b) Streams, Information and Language, - Statistics of Text - Analyzing Sentiment and Intent – Load - Databases and their Evolution, Big data Technology and Trends.
- c) An overview of clustering algorithms - Clustering issues in very large datasets - The need for classification - Automatic categorization of emails and spam filtering - Classification with very large datasets - Comparing multiple classifiers on the same data.
- d) Real time analysis
- e) Usage of NoSQL in Webanalytics
- f) Visualisation(big data)

5. D/W Project Life Cycle Management (4 Credit)

- a) PLANNING AND REQUIREMENTS
Planning Your Data Warehouse, The Data Warehouse Project, The Project Team, Project Management Considerations, Defining the Business Requirements, Dimensional Analysis, Information Packages—A New Concept, Requirements Gathering Methods, Requirements Definition: Scope and Content, Requirements as the Driving Force for Data Warehousing, Data Design, The Architectural Plan, Data Storage Specifications, Information Delivery Strategy
- b) ARCHITECTURE AND INFRASTRUCTURE
The Architectural Components, Understanding Data Warehouse Architecture, Distinguishing Characteristics, Architectural Framework, Technical Architecture, Infrastructure as the Foundation for Data Warehousing, Infrastructure Supporting Architecture, Infrastructure Supporting Architecture, Database Software, Collection of Tools, The Significant Role of Metadata, Why Metadata is Important, Metadata Types by Functional Areas, Business Metadata, Technical Metadata, How to Provide Metadata.
- c) DATA DESIGN AND DATA PREPARATION
 - Principles of Dimensional Modeling, From Requirements to Data Design, The STAR Schema, STAR Schema Keys, Advantages of the STAR Schema, Updates to the Dimension Tables, Miscellaneous Dimensions, The Snowflake Schema, Aggregate Fact Tables, Families of STARS, Data Extraction, Transformation, and Loading, Data Extraction, Data Transformation, Data

Loading, ETL Summary, Why is Data Quality Critical?, Data Quality Challenges, Data Quality Tools, Data Quality Initiative.

d) INFORMATION ACCESS AND DELIVERY

Matching Information to the Classes of Users. Information from the Data Warehouse, Who Will Use the Information?, Information Delivery, Information Delivery Tools, Demand for Online Analytical Processing, Major Features and Functions, OLAP Models, OLAP Implementation Considerations, Data Warehousing and the Web, Web-Enabled Data Warehouse, Web-Based Information Delivery, OLAP and the Web, Building a Web-Enabled Data Warehouse, What is Data Mining?, Major Data Mining Techniques, Data Mining Applications.

e) DW TESTING

Unit, IT, ST, Data volume, ETL and other test techniques

f) IMPLEMENTATION AND MAINTENANCE

The Physical Design Process, Physical Design Steps, Physical Design Considerations, Physical Storage, Indexing the Data Warehouse, Performance Enhancement Techniques, Data Warehouse Deployment, Major Deployment Activities, Considerations for a Pilot, Security, Backup and Recovery, Growth and Maintenance, Monitoring the Data Warehouse, User Training and Support, Managing the Data Warehouse

g) Scoping, Estimation, proposal writing and Risk planning

6. ETL, Data Profiling and Data Modelling with Pentaho (4 Credit)

a) ETL Overview

OLTP versus Data Warehousing , What Is ETL? , The Evolution of ETL Solutions , ETL Building Blocks , ETL ELT and EII , Virtual Data Integration , Data Integration Challenges , Methodology: Agile BI, ETL Design, Data Acquisition, Design for Failure, Change Data Capture, Data Quality, Data Profiling , Data Validation, ETL Tool Requirements, Connectivity, Platform Independence, Scalability, Design Flexibility, Reuse, Extensibility, Data Transformations , Testing and Debugging , Lineage and Impact Analysis, Logging and Auditing.

b) KETTLE/Pentaho ETL CONCEPTS

The Building Blocks of Kettle Design, Transformations, Jobs, Transformation or Job Metadata, Database Connections, Tools and Utilities, Repositories, Virtual File Systems , Parameters and Variables, Visual Programming, Installation and Configuration, Kettle Software Overview, Installation, Java Environment, Installing Kettle, Configuration, The Rental Star Schema, Prerequisites and Some Basic Spoon Skills, Setting Up the ETL Solution, Working with Spoon, The Sample ETL Solution, The Sample ETL Solution, Recurring Load.

c) Data Extraction, Cleansing and Conforming

Kettle Data Extraction Overview, File-Based Extraction, Database-Based Extraction, Web-Based Extraction, Stream-Based and Real-Time Extraction, Working with ERP and CRM Systems, Data Profiling Using eobjects.org Data Cleaner, CDC: Change Data Capture, Data Cleansing, Data-Cleansing Steps, Using Reference Tables, Data Validation, Error Handling, Auditing Data and Process Quality, Deduplicating Data, Scripting

- d) Handling Dimension Tables & Loading Fact Tables
 Managing Keys: Managing Business Keys, Generating Surrogate Keys, Loading Dimension Tables, Snowflaked Dimension Tables, Star Schema Dimension Tables, Slowly Changing Dimensions, Generated Dimensions, Junk Dimensions, Recursive Hierarchies, Loading in Bulk, STDIN and FIFO, Kettle Bulk Loaders, Dimension Lookups, Maintaining Referential Integrity, Late-Arriving Data, Fact Table Handling
- e) Scheduling, Monitoring, Versioning and Migration
 Scheduling, Operating System–Level Scheduling, Using Pentaho’s Built-in Scheduler, Monitoring, Version Control Systems, File-Based Version Control Systems, Kettle Metadata, Managing Repositories, Version Migration System, Lineage, Logging and Operational Metadata, Performance Tuning, Transformation Performance: Finding the Weakest Link, Improving Transformation Performance, Improving Performance in Reading Text Files, Improving Performance in Writing Text Files, Improving Database Performance, Sorting Data, Reducing CPU Usage, Improving Job Performance.
- f) ETL in bigdata/Hadoop scenario

7. Data Visualisation with Advance XIs (4 Credit)\

- a) Introduction to xls/powerpivot
 - Using a powerpivot on an Excel table, Understanding the data model, Using OLAP tools and converting to formulas, Understanding Excel 2013, Creating a Power View report, Loading data from external sources, Other tools like power pivot and Qlickview
- b) Detailing Power pivot
 - Understanding Powerpivot calculations, Understanding calculated columns and fields, Handling errors in powerpivot expressions, Formatting powerpivot code, Common powerpivot functions, Using basic powerpivot functions, Understanding over-denormalization, Understanding OLTP and data marts, Using advanced relationships
- c) Publishing & Loading Data using power pivot
- d) Understanding Calculate, hierarchies in power pivot
 - Why is CALCULATE needed?, Using CALCULATE inside a row context, Understanding circular dependencies, CALCULATE rules, Understanding ALLSELECTED, Understanding hierarchies, Building hierarchies, Using parent/child hierarchies, Power View basics, Understanding table, matrix, and cards, Using charts, Using maps, Understanding drill-down, Using tiles, Understanding multipliers.
- e) Shaping the reports & Performing date calculations in powerpivot
 - Key Performance Indicators (KPIs), Creating data models for Power View, Understanding Power View metadata, Defining sets, Creating dynamic sets with MDX, Using perspectives, Understanding drill-through, Building a calendar table, Aggregating and comparing over time, Closing balance over time, Computing moving averages, Banding, Ranking, Using many-to-many relationships, Implementing basket analysis

Semester- III

1. Unstructured Data Analytics with Hadoop/Big Data/NoSql (4 Credit)

- Design Data Architecture and manage the data for analysis, understand various sources of Data like Sensors/signal/GPS etc.
- Export all the data onto Cloud ex. AWS/Rackspace etc.
- Introduction to Big Data tools like Hadoop, Spark, Impala etc., Data ETL process, Identify gaps in the data and follow-up for decision making.
- Run descriptive to understand the nature of the available data, collate all the data sources to suffice business requirement, Run descriptive statistics for all the variables and observe the data ranges, Outlier detection and elimination.
- Hypothesis testing and determining the multiple analytical methodologies, Train Model on 2/3 sample data using various Statistical/Machine learning algorithms, Test model on 1/3 sample for prediction etc.
- Integrating Pentaho to Hadoop file structure
- Integration Relation DB with No SQL
- Integration of R

2. Advanced Visualisation – Tableau/ QlikView and D3js (4 Credit)

- Introduction to advance visualisation
- Matplotlib package – Plotting Graphs – Controlling Graph – Adding Text – More Graph Types – Getting and setting values – Patches.
- Introduction to tableau
- Introduction to Qlickview
- Custom Reporting and D3js
- Building and publishing reports in tableau & qlickview
- Realtime reporting/mobile analytics

3. Advanced Data Modelling & Mining with R/Python (4 Credit)

- Introduction to R & Predicting Algae Blooms
Starting with R, R Objects, Vectors, Factors, Generating Sequences, Sub-Setting, Matrices and Arrays, Data Frames, Creating New Functions, Objects, Classes, and Methods, Managing Your Sessions, Loading the Data into R, Data Visualization and Summarization, Unknown Values, Obtaining Prediction Models, Model Evaluation and Selection, Predictions for the Seven Algae, The Available Data, Defining the Prediction Tasks, The Prediction Models, From Predictions into Actions, Model Evaluation and Selection.
- Detecting Fraudulent Transactions & Classifying Microarray Samples
Problem Description and Objectives, The Available Data, The Available Data, Obtaining Outlier Rankings, Semi-Supervised Approaches, Brief Background on Microarray Experiments, Gene (Feature) Selection, Predicting Cytogenetic Abnormalities, The Modeling Techniques, Random Forests, k-Nearest Neighbors, Comparing the Models.
- PYTHON CONCEPTS, DATA STRUCTURES & DATA WRANGLING
Interpreter – Program Execution – Statements – Expressions – Flow Controls – Functions - Numeric Types – Sequences - Strings, Tuples, Lists and - Class Definition – Constructors – Inheritance – Overloading – Text & Binary Files - Reading

and Writing, Combining and Merging DataSets – Reshaping and Pivoting – Data Transformation – String Manipulation, Regular Expressions.

- d) DATA AGGREGATION, GROUP OPERATIONS ,TIMESERIES, WEB SCRAPING
Goup By Mechanics – Data Aggregation – Group wise Operations and Transformations – Pivot Tables and Cross Tabulations – Date and Time Date Type tools – Time Series Basics – Data Ranges, Frequencies and Shifting, Data Acquisition by Scraping web applications –Submitting a form - Fetching web pages – Downloading web pages through form submission – CSS Selectors.