upGrad | IIIT-B | LIVERPOOL JOHN MOORES UNIVERSITY

Master of Science in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

TRACK	COURSE	MODULE NAME	DESCRIPTION	SESSION
CURRICULUM	PRE-PROGRAM PREPARATORY CONTENT	Data Analysis in Excel	Taught by one of the most renowned data scientists in the country (S.Anand, CEO, Gramener), this module takes you from a beginner level Excel user to an almost professional user.	Introduction to Excel
				Data Analysis in Excel - I: Functions, Formulae, and Charts
				Data Analysis in Excel - II: Pivots and Lookups
		Analytics Problem Solving	This module covers concepts of the CRISP-DM framework for business problem-solving.	The CRISP-DM Framework - Business and Data Understanding
				CRISP-DM Framework - Data Preparation, Modelling, Evaluation and Deployment
	COURSE 1: DATA TOOLKIT	Data Analysis using SQL	Data in companies is definitely not stored in excel sheets! Learn the fundamentals of database and extract information from RDBMS using the structured query language.	Basics of SQL: Data Retrieval, Compound Functions, Relational Operators, and Sorting
				Advanced SQL: Aggregate Functions, Nested Queries, and Joins
		Introduction to Python	Build a foundation for the most in-demand programming language of the 21st century.	Understanding the upGrad Coding Console
				Data Structures in Python
				Control Structure and Functions
		Programming in Python	Learn how to approach and solve logical problems using programming.	Logic and Syntax Building
				Data Structures: Lists, Strings, Dictionaries, and Stacks
				Time Complexity
				Searching and Sorting
				Two Pointers
				Recursion
		Python for Data Science	Learn how to manipulate datasets in Python using Pandas which is the most powerful library for data preparation and analysis.	Introduction to NumPy
				Operations on NumPy Arrays
				Introduction to Pandas
				Getting and Cleaning Data
		Visualization in Python	Humans are visual learners and hence no task related to data is complete without visualisation. Learn to plot and interpret various graphs in Python and observe how they make data analysis and drawing insights easier.	Introduction to Data Visualization
				Basics of Visualization: Plots, Subplots and their Functionalities
				Plotting Data Distributions
				Plotting Categorical and Time-Series Data
		IMDb Movie Assignment	Reinforce the concepts learnt in data science through this rigorous assignment involving the past hundred years of movie data.	Problem Statement
				Evaluation Rubric
				Final Submission
				Solution
		Exploratory Data Analysis	Learn how to find and analyse the patterns in the data to draw actionable insights.	Data Sourcing
				Data Cleaning
				Univariate Analysis
				Segmented Univariate
				Bivariate Analysis
				Derived Metrics
		Inferential Statistics	Build a strong statistical foundation and learn how to 'infer' insights from a huge population using a small sample.	Basics of Probability
				Discrete Probability Distributions
				Continuous Probability Distributions
				Central Limit Theorem
		Hypothesis Testing	Understand how to formulate and validate hypotheses for a population to solve real-life business problems.	Concepts of Hypothesis Testing - I: Null and Alternate Hypothesis, Making a Decision, and Critical Value Method
				Concepts of Hypothesis Testing - II: p-Value Method and Types of Errors
				Industry Demonstration of Hypothesis Testing: Two-Sample Mean and Proportion Test, A/B Testing
		Credit EDA Case Study	Solve a real industry problem through the concepts learnt in exploratory data analysis.	Problem Statement
				Evaluation Rubric
				Final Submission
				Solution
	COURSE 2 - MACHINE LEARNING	Introduction to Machine Learning and Linear Regression	Venture into the machine learning community by learning how one variable can be predicted using several other variables through a housing dataset where you will predict the prices of houses based on various factors.	Simple Linear Regression
				Multiple Linear Regression
				Industry Relevance of Linear Regression
		Linear Regression Assignment	Build a model to understand the factors car prices vary on and help a Chinese company enter the US car market.	Problem Statement
				Evaluation Rubric
				Final Submission
				Solution
		Logistic Regression	Learn your first binary classification technique by determining which customers of a telecom operator are likely to churn versus who are not to help the business retain customers.	Univariate Logistic Regression
				Multivariate Logistic Regression: Model Building and Evaluation
				Logistic Regression: Industry Applications
		Unsupervised Learning: Clustering	Learn how to group elements into different clusters when you don't have any pre-defined labels to segregate them through K-means clustering, hierarchical clustering, and more.	Introduction to Clustering
				K-Means Clustering
				Hierarchical Clustering
				Other Forms of Clustering: K-Mode, K-Prototype, DB Scan
		Business Problem Solving	Learn how to approach open ended real world problems using data as a lever to draw actionable insights.	Introduction to Business Problem Solving
				Business Problem Solving: Practical Examples
		Assignment: Unsupervised + Supervised	Apply the machine learning concepts learnt to solve a real-life predictive analytics problem.	Problem Statement
				Evaluation Rubric
				Final Submission
				Solution
		Case Study: Lead Scoring	Help the Sales team of your company identify which leads are worth pursuing through this classification case study.	Problem Statement
				Evaluation Rubric
				Final Submission
				Solution

upGrad | IIIT-B | LIVERPOOL JOHN MOORES UNIVERSITY

Master of Science in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

TRACK	COURSE	MODULE NAME	DESCRIPTION	SESSION
SPECIALISATION 1: DEEP LEARNING	COURSE 3 - ADVANCED MACHINE LEARNING	Tree Models	*Learn how the human decision making process can be replicated using a decision tree and other powerful ensemble algorithms.*	*Introduction to Decision Trees* *Algorithms for Decision Tree Construction* *Truncation and Puning* *Random Forests*
		Model Selection & General ML Techniques	*Learn the pros and cons of simple and complex models and the different methods for quantifying model complexity, along with general machine learning techniques like feature engineering, model evaluation, and many more.*	*Principles of Model Selection* *Model Evaluation* *Model Selection: Best Practices*
		Bagging and Boosting	*Learn about ensemble modelling through bagging and boosting and understand how weak algorithms can be transformed into stronger ones.*	*Introduction to Boosting and AdaBoost* *Gradient Boosting*
		Advanced Regression	*In this module, take a more advanced look at regression models and learn the concepts related to regularization.*	*Generalized Linear Regression* *Regularized Regression*
		Advanced Regression Assignment	*Build a regularized regression model to understand the most important variables to predict the house prices in Australia.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
		Principal Component Analysis	*In this module, take a more advanced look at regression models and learn the concepts related to regularization.*	*Prinicipal Component Analysis and Singular Value Decomposition* *Principal Component Analysis in Python*
		Time Series Analysis	*In this module, you will learn how to analyse and forecast a series that varies with time.*	*Introduction to Time Series and its Components* *Working with Stationary Time Series* *End-to-End Analysis of Time Series*
		Telecom Churn Case Study	*Solve the most crucial business problem for a leading telecom operator in India and southeast Asia - predicting customer churn.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
	COURSE 4 - DEEP LEARNING AND NEURAL NETWORKS	Introduction to Neural Networks	*Learn the most sophisticated and cutting-edge technique in machine learning - Artificial Neural Networks or ANNs*	*Structure of Neural Networks* *Feed Forward in Neural Networks* *Backpropagation in Neural Networks* *Modifications to Neural Networks* *Hyperparameter Tuning in Neural Networks*
		Neural Networks Assignment	*Build a neural network from scratch in Numpy to identify handwritten digits.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
		Convolutional Neural Networks - Introduction and Industry Applications	*Build a neural network from scratch in Numpy to identify handwritten digits.*	*Introduction to Convolutional Neural Networks* *Building CNNs with Python and Keras* *Style Transfer and Object Detection* *Industry Demonstration: Using CNNs with Flowers Images* *Industry Demonstration: Using CNNs with X-Ray Images*
		Recurrent Neural Networks	*Ever wondered what goes behind machine translation, sentiment analysis, speech recognition? Learn how RNN helps in these areas having sequential data like text, speech, videos, and a lot more.*	*What Makes a Neural Network Recurrent* *Variants of RNNs: Bidirectional RNNs and LSTMs* *Building RNNs in Python*
		Gesture Recognition	*Make a Smart TV system which can control the TV with user's hand gestures as the remote control*	*Two Architectures: 3D Convs and CNN-RNN Stack* *Understanding Generators* *Starter Code Walkthrough* *Problem Statement and Final Submission*
	COURSE 5 - CAPSTONE PROJECT	Capstone Project	*Choose from a range of real-world industry woven projects on advanced topics like Recommendation Systems, Fraud Detection, Emotion Detection from faces, Social Media Listening, Speech Recognition among many others.*	*An Overview of the Domain and Associated Concepts* *Problem Statement* *Evaluation Rubric* *Mid Submission* *Final Submission* *Solution*

upGrad | IIIT-B | LIVERPOOL JOHN MOORES UNIVERSITY

Master of Science in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

TRACK	COURSE	MODULE NAME	DESCRIPTION	SESSION
SPECIALISATION 2: NLP	COURSE 3 - ADVANCED MACHINE LEARNING	**Tree Models**	*Learn how the human decision making process can be replicated using a decision tree and other powerful ensemble algorithms.*	*Introduction to Decision Trees* *Algorithms for Decision Tree Construction* *Truncation and Puning* *Random Forests*
		Model Selection & General ML Techniques	*Learn the pros and cons of simple and complex models and the different methods for quantifying model complexity, alongwith general machine learning techniques like feature engineering, model evaluation, and many more.*	*Principles of Model Selection* *Model Evaluation* *Model Selection: Best Practices*
		Bagging and Boosting	*Learn about ensemble modelling through bagging and boosting and understand how weak algorithms can be transformed into stronger ones.*	*Introduction to Boosting and AdaBoost* *Gradient Boosting*
		Advanced Regression	*In this module, take a more advanced look at regression models and learn the concepts related to regularization.*	*Generalized Linear Regression* *Regularized Regression*
		Advanced Regression Assignment	*Build a regularized regression model to understand the most important variables to predict the house prices in Australia.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
		Principal Component Analysis	*Understand important concepts related to dimensionality reduction, the basic idea and the learning algorithm of PCA, and its practical applications on supervised and unsupervised problems.*	*Prinicipal Component Analysis and Singular Value Decomposition* *Principal Component Analysis in Python*
		Time Series Analysis	*In this module, you will learn how to analyse and forecast a series that varies with time.*	*Introduction to Time Series and its Components* *Working with Stationary Time Series* *End-to-End Analysis of Time Series*
		Telecom Churn Case Study	*Solve the most crucial business problem for a leading telecom operator in India and southeast Asia - predicting customer churn.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
	COURSE 4 - NATURAL LANGUAGE PROCESSING	**Lexical Processing**	*Do you get annoyed by the constant spams in yor mail box? Wouldn't it be nice if we had a program to check your spellings? In this module learn how to build a spell checker & spam detector using techniques like phonetic hashing,bag-of-words, TF-IDF, etc.*	*Introduction to Natural Language Processing* *Basic Lexical Processing: Tokenization, Bag of Words, TF-IDF* *Advanced Lexical Processing: Canonicalization, Phonetic Hashing, Spell Corrector, Pointwise Mutual Information*
		Syntactic Processing	*Learn how to analyse the syntax or the grammatical structure of sentences with the help of algorithms & techniques like HMMs, Viterbi Algorithm, Named Entity Recognition (NER), etc.*	*Introduction to Syntactic Processing* *Parsing* *Information Extraction* *Conditional Random Fields (CRF)*
		Syntactic Processing Assignment	*Build a POS tagger for tagging unknown words using HMM's & modified Viterbi algorithm.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
		Semantic Processing	*Learn the most interesting area in the field of NLP and understand different techniques like word-embeddings, LSA, topic modelling to build an application that extracts opinions about socially relevant issues (such as demonetisation) on social media platforms.*	*Introduction to Semantic Processing* *Distributional Semantics* *Topic Modelling* *Social Media Opinion Mining: Semantic Processing Case Study*
		Chatbot Case Study	*Imagine if you could make restaurant booking without opening Zomato. Build your own restaurant-search chatbot with the help of RASA - an open source framework and deploy it on Slack.*	*Building Chatbots with Rasa* *NLP Course Project - Building a Chatbot: Problem Statment and Final Submission*
	COURSE 5 - CAPSTONE PROJECT	**Capstone Project**	*Choose from a range of real-world industry woven projects on advanced topics like Recommendation Systems, Fraud Detection, Emotion Detection from faces, Social Media Listening, Speech Recognition among many others.*	*An Overview of the Domain and Associated Concepts* *Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*

upGrad | IIITB | LIVERPOOL JOHN MOORES UNIVERSITY

Master of Science in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

TRACK	COURSE	MODULE NAME	DESCRIPTION	SESSION
SPECIALISATION 3: BUSINESS INTELLIGENCE/DATA ANALYTICS	COURSE 3 - SQL AND NOSQL DATABASES	**Introduction to Databases**	*Learn how data is stored and which database is optimal to use in a particular scenario.*	*OLAP vs OLTP* *Designing a database schema* *Creating ER Diagram*
		Advanced SQL and Best Practices	*Apply advanced SQL concepts like windowing and procedures to derive insights from data and answer pertinent business questions.*	*Stored functions and procedures in SQL* *Conditional constructs in SQL* *Query Optimisation*
		Schema Design and Retrieval Assignment	*Design a schema with a predefined objective and create a database and perform analysis.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
		NoSQL Databases and Best Practices	*Take your knowledge of query languages a step further by learning about MongoDB - a NoSQL database which is becoming more and more popular in the industry.*	*Introduction to NoSQL Database* *Configuring MongoDB* *CRUD and MongoDB Shell* *Indexing and Aggregation* *CouchDB/Redis/HBase/Neo4J*
		Introduction to Cloud and Hive	*Understand the basics of cloud and learn about the Hive Query Language.*	*Understanding Big Data* *Introduction to Apache Hive* *Hive Data Models - Partitions and Buckets* *File Formats in Apache Hive* *Data Analysis using Apache Hive*
		SQL Case Study: Analysing Big Data in Retail	*Understand how a project in the industry is taken up and solved through a comprehensive business case study.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
	COURSE 4 - NATURAL LANGUAGE PROCESSING	**Advanced Excel**	*Learn the advanced concepts in Excel and start to perform data analysis like a pro!*	*Advanced Functions in Excel* *Slicers with Pivots* *Map Visualization* *Data Analysis Toolpak: Descriptive Summaries, Correlation, Regression*
		Visualisation using Tableau	*Learn advanced visualisation techniques using the most in-demand visualization tool in the industry.*	*Data Exploration in Tableau* *Visualising and Analysing Data in Tableau* *Advanced Visualisations using Tableau*
		Interactive Marketing Campaign Analysis	*Apply the new found Excel and Tableau skills to solve an exciting business assignment.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
		Visualisation using PowerBI	*Take your visualization game a step forward by understanding how to operate PowerBI.*	*PowerBI: Introduction and Setup* *Visualising and Analysing Data in PowerBI* *Data Transformations using PowerBI* *Making Interactive Dashboards with PowerBI*
		Introduction to R and RShiny	*Get a brief introduction to another popular data science language and learn how to manipulate dataframes in R and learn to create attractive dashboards and web applications using Shiny®.*	*Introduction to R* *Data Structures and Programming Constructs in R* *Dataframe Manipulation in R* *Visualizations in R* *Creating Web Applications using RShiny*
		Effective Communication Strategies, Formats, and Templates	*Learn how to effectively strategise, communicate, and fine grain your data analysis projects.*	*Introduction to MS Powerpoint* *Identifying the Right Insights* *Documenting Insights* *Choosing the Right View for your Dashboard* *Storytelling with Visualization*
		Presentations to Technical and Non-Technical Stakeholders	*Understand how to optimally present your findings to technical and non-technical stakeholders and upgrade your storytelling skills.*	*Understanding the End Goal of your Presentation* *Understanding your Stakeholder and Stakeholder Empathy* *Levels of Details for Different Stakeholders - CXO/Leadership Vs Team Presentations* *Do's and Don'Ts of Making a Presentation* *Case Study: Making a Presentation for Different Stakeholders*
		Business Case Study	*Understand how a project in the industry is taken up and solved through a comprehensive business case study.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
	COURSE 5 - CAPSTONE PROJECT	**Capstone Project**	*Solve an end-to-end real-life industry problem from a wide variety of domains like Marketing, Retail, E-Commerce, Supply Chain, Healthcare, BFSI, and many more.*	*An Overview of the Domain and Associated Concepts* *Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*

upGrad | IIITB | Liverpool John Moores University

Master of Science in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

TRACK	COURSE	MODULE NAME	DESCRIPTION	SESSION
SPECIALISATION 4: BUSINESS ANALYTICS	COURSE 3 - ADVANCED MACHINE LEARNING	**Tree Models**	*Learn how the human decision making process can be replicated using a decision tree and other powerful ensemble algorithms.*	*Introduction to Decision Trees* *Algorithms for Decision Tree Construction* *Truncation and Puning* *Random Forests*
		Time Series Analysis	*In this module, you will learn how to analyse and forecast a series that varies with time.*	*Introduction to Time Series and its Com-* *Working with Stationary Time Series* *End-to-End Analysis of Time Series*
		Time Series Assignment	*Apply the concepts learnt in time series to solve a forecasting problem.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
		Model Selection & General ML Techniques	*Learn the pros and cons of simple and complex models and the different methods for quantifying model complexity, alongwith general machine learning techniques like feature engineering, model evaluation, and many more.*	*Principles of Model Selection* *Model Evaluation* *Model Selection: Best Practices*
		Telecom Churn Case Study	*Solve the most crucial business problem for a leading telecom operator in India and southeast Asia - predicting customer churn.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
	COURSE 4 - BUSINESS REQUIREMENTS	**Advanced SQL and Best Practices**	*Apply advanced SQL concepts like windowing and procedures to derive insights from data and answer pertinent business questions.*	*Stored functions and procedures in SQL* *Conditional constructs in SQL* *Query Optimisation*
		Advanced Excel	*Learn advanced visualisation techniques using the most in-demand visualization tool in the industry.*	*Advanced Functions in Excel* *Slicers with Pivots* *Map Visualization* *Data Analysis Toolpak: Descriptive Summaries, Correlation, Regression*
		Structured Problem Solving using Frameworks	*Apply the new found Excel and Tableau skills to solve an exciting business assignment.*	*Introduction to Structured Problem Solving* *Understanding the Business Problem - I: Interviewing and Basic Frameworks* *Understanding the Business Problem - II: SWI, SWHYS, SPIN Framework, etc.*
		Hypothesis Formulation	*The module will equip you with a stepwise process for understanding a business problem and building hypotheses around it.*	*Introduction to Hypothesis Formulation* *Issue Tree Framework* *Specialized Frameworks: 4Ps and 5C*
		Assignment	*Apply your learnings from the course to solve a real-life business problem.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
		Revenue and Operational Cost Modelling	*Learn about evaluation models covering revenue and operational cost modelling.*	*Introduction to Cost Modelling* *Modelling Frameworks* *Revenue and Operational Cost Modelling in Various Domains*
		Introduction to Economics and Financial Concepts	*Understand financial concepts such as revenue, cost of goods sold, profit, balance sheets, cash flow statements, and EBITDA. Also, get a brief introduction to economics concepts including supply and demand curves, cost curves, and a lot more!*	*Introduction to Financial Concepts* *Revenue, Balance Sheets, Cash Flow Statements* *EBITDA* *Introduction to Economics* *Supply-Demand Curves* *Basic Econometrics*
		Effective Communication Strategies, Formats, and Templates	*Learn how to effectively strategise, communicate, and fine grain your data analysis projects.*	*Introduction to MS Powerpoint* *Identifying the Right Insights* *Documenting Insights* *Choosing the Right View for your Dashboard* *Storytelling with Visualization*
		Presentations to Technical and Non-Technical Stakeholders	*Understand how to optimally present your findings to technical and non-technical stakeholders and upgrade your storytelling skills.*	*Understanding the End Goal of your Presentation* *Understanding your Stakeholder and Stakeholder Empathy* *Levels of Details for Different Stakeholders - CXO/Leadership Vs Team Presentations* *Do's and Don't's of Making a Presentation* *Case Study: Making a Presentation for Different Stakeholders*
		Business Case Study	*Understand how a project in the industry is taken up and solved through a comprehensive business case study.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
	COURSE 5 - CAPSTONE PROJECT	**Capstone Project**	*Solve an end-to-end real-life industry problem from a wide variety of domains like Marketing, Retail, E-Commerce, Supply Chain, Healthcare, BFSI, and many more.*	*An Overview of the Domain and Associated Concepts* *Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*

upGrad | iiit-b | LIVERPOOL JOHN MOORES UNIVERSITY

Master of Science in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

TRACK	COURSE	MODULE NAME	DESCRIPTION	SESSION
SPECIALISATION 5: DATA ENGINEERING	COURSE 3 - DATA ENGINEERING - I	**Introduction to Hadoop and MapReduce Programming**	*Understand the world of distributed data processing and storage with Hadoop. Learn to write MapReduce jobs in Python.*	*Concepts retailed to distributed computing* *Hadoop Distributed File System* *MapReduce Programming in Python*
		Data Management and Relational Database Modelling	*Understand the concepts of Data Management and learn to model data for a Relational Database.*	*Enterprise Data Management* *Relational Database Modelling* *Normal Forms and ER Diagrams*
		NoSQL Databases and Apache HBase	*Learn the concepts of NoSQL databases. Understand the workings of Apache HBase.*	*Concepts of NoSQL Databases* *Introduction to Apache HBase* *HBase Python API* *Comparision of NoSQL Databases*
		Data Warehousing (Optional)	*Understand the intricacies behind designing a data warehouse and a data lake for your use case.*	*Introduction to Data Warehouse and Data Lakes* *Designing Data Warehousing for an ETL Data Pipeline* *Designing Data Lake for an ETL Data Pipeline*
		Data Ingestion with Apache Sqoop and Apache Flume	*Get familiar with the challenges invoked in data ingestion. Use Sqoop and Flume to ingest structured and unstructured data into Hadoop.*	*Introduction to Data Ingestion* *Structured data ingestion with Sqoop* *Unstructured data ingestion with Flume*
		Assignment (Optional)	*Practice the concepts learnt so far with this comprehensive assignment.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
		Building and Querying Data Warehouse with Apache Hive	*Manage and query a data warehouse with Apache Hive. Learn to write optimized HQL for large scale data analysis.*	*Fundamentals of Apache Hive* *Writing HQL for Data Analysis* *Hive Query Optimization*
		Case Study: Ingestion & Warehousing	*Make use of Sqoop, Flume, Hive and HBase to design an ETL data pipeline.*	*Introduction and Problem Statement* *Grading Rubrics and Submission*
	COURSE 4 - DATA ENGINEERING - II	**Data Processing with PySpark**	*Get introduced to Apache Spark, a lighting fast big data processing engine. Use PySpark to create large scale data processing applications.*	*Introduction to Apache Spark* *Apache Spark Architecture* *PySpark APIs - RDDs, DataFrames, SQL* *Spark Job optimization*
		Real-Time Data Streaming with Apache Kafka	*Understand the producer-consumer architecture of Apache Kafka. Learn to set up a Kafka cluster for managing real-time data.*	*Fundamentals of Apache Kafka* *Setting up Kafka Producer and Consumer* *Kafka Connect API*
		"Real-Time Data Processing using Spark Streaming"	*Learn about the real-time data processing architecture of Apache Spark. Build Spark Streaming applications to process data in real-time.*	*Spark Streaming Architecture* *Spark Streaming APIs* *Building Stream Processing Application with Spark*
		Assignment (Optional)	*Use Kafka and Spark to develop a real-time data processing applicaiton.*	*Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*
		Building Automated Data Pipelines with Oozie/Airflow	*Automate your Data Pipeline with Apache Airflow.*	*Fundaments of Oozie/Airflow* *Workflow Management with Oozie/Airflow* *Automating an entire Data Pipeline with Oozie/Airflow*
		Analytics using PySpark	*Use PySpark to do EDA and Predictive Analysis of large datasets.*	*"Exploratory Data Analysis with PySpark"* *Predictive Analysis with Spark MLlib*
		Case Study: Kafka, Spark Streaming and PySpark	*Build an end-to-end real-time data processing application using Spark Streaming and Kafka.*	*Introduction and Problem Statement* *Grading Rubrics and Submission*
	COURSE 5 - CAPSTONE PROJECT	**Capstone Project**	*The capstone project will stich all the components of data engineering together.*	*An Overview of the Domain and Associated Concepts* *Problem Statement* *Evaluation Rubric* *Final Submission* *Solution*

Master of Science in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

RESEARCH METHODOLOGY

(DURATION: 2 MONTHS) — 10 CREDITS

WHAT IS RESEARCH?

Familiarise with different aspects of research

- Intro to research
- Importance of research
- Criticism in research and its importance
- Peer reviews in research and its importance

TYPES OF RESEARCH

Develop an understanding of various research design and techniques

- Descriptive vs Analytical
- Applied vs Fundamental
- Quantitative vs Qualitative
- Bayesian vs Frequentist Approach

RESEARCH PROCESS

Learn about the different steps in the research process and how to evaluate a literature

- Research question
- Hypothesis and aims
- Formulating a Problem
- Literature review

RESEARCH PROJECT MANAGEMENT

Learn how to plan the project timelines and arrange for data & software

- Understand the different steps involved in a project cycle
- Project Requirements on Data
- Identifying the milestones in a research project
- Learn how to track the progress using Gantt Chart

REPORT WRITING AND PRESENTATION

Master good scientific writing and proper presentation skills

- Art of writing papers
- Parts of a paper
- Tools to write papers
- Publishing papers: Journals + Seminars

SCIENTIFIC ETHICS

Develop an understanding of the ethical dimension in research

- Citation Methods and Rules
- Honor Code
- Research Claims
- IPh

MASTER'S DISSERTATION

(DURATION: 4 MONTHS) — 60 CREDITS

MASTER'S THESIS

Articulate, research and present your project.

- Monthly Checkpoints
- Submission

An example of project outlines is here:

- Investigate the risk factors for eye disease from complex longitudinal datasets
- Investigate a diagnosis of eye diseases using imaging ophthalmic data
- Multi-task learning for drug design and discovery
- Using stacking for brain tumour discrimination
- Investigate dietary patterns and metabolite fingerprints of takeaway (fast) food consumers using PCA and Clustering methods
- Longitudinal studies to investigate the complex link between corporate environment engagement, green disclosure, business model transformation and supply chain performance
- Preventing credit card fraud through pattern recognition
- Developing a recommender system for a Media giant
- Using social media feed to place tweets regarding natural disasters on a map