Data and Analytics - TechM

	MON	TUE	WED	THU	FRI	ENVIRONMENT
Week1-Python SQL **(Agile for Developers, Git - Fundamentals, Python-Fundamentals)**	Project0 Agile for Developers SDLC • Introduction To SDLC • Waterfall • Agile • Agile Vs Waterfall • Story Pointing • Scrum Ceremonies Git Fundamentals OS-Introduction • OS: Fundamentals • Unix/Linux: Demo Moving and Deleting Files (Using Git Bash) • Unix/Linux: Demo File Commands	Project0 Python-Fundamentals Python-Orientation • Interpreter vs Compiler • What is Python Why Python • Full Stack Overview Python-Basics • Python Syntax • Comments • Variables and Datatypes • Operators • User Input and Output Python-DataTypes • Strings	Project0 • Scope • If-Else • While • For • Function • Lambda • Arrays • Classes and Objects • OOP Concepts • Inheritance • Iterators	Project0 Python Coding Challenge Python-Modules • Math • Logging • JSON • Regex • numpy • pandas • pip and install pip • pylint • Connect SQL Python-Exception Handling • Error • Exception Handling	Project0 Python-File Handling • Read Files • Write Create Files • Delete Files • File Handling Review Topics	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
	(Using Git Bash) Git Introduction • Source Control Management(git,vcs, cvcs,dvcs) • Git Fundamentals • Initializing A Repository • Pushing To A Remote Repository • Git Commit, Branch, Merge, Push, Pull Git Exercises	• Casting • Boolean • Lists • Tuples • Range • Sets • Binary Type • Nontype • Dictionaries • Numbers • Namespaces		• try-except • Module		
Week2-Python SQL (SQL)	Project0 QC Audit MySQL SQL-Introduction • What Is A Database • What Is Sql • Consistency • Introduction To	Project0 SQL Coding Challenge Sub Languages • Overview Of Sublanguages • Ddl • Dml • Dql	Project0 SQL Joins coding test • Inner Join • Left And Right Joins • Outer Join • Cross Join • Equi And Theta Joins	Project0 Python Coding Challenge Advanced-SQL • Scalar Functions • Sequence • Trigger • Views • Window Functions	Project0 Review Topics	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
	Rdbms Structure • Schema • Table Structure • Sql Data Types • Normalization • Multiplicity • Data Modeling And Erd • Primary Key • Composite Key • Foreign Key • Unique Key • Secondary Alternate Key • Referential Integrity	• Dcl • Tcl DDL • Create Drop Truncate • Constraints • Auto Incrementing • Check • Default • Cascade DML • Insert • Update • Delete DQL • Queries • Aggregate Functions • Clauses • What Is A Subquery	• Aliases Transaction • What Is A Transaction • Acid Properties • Transaction Properties • Crud Operations • Transaction Commit Rollback Isolation Levels	(ROW_NUMBER, RANK, DENSE_RANK, LEAD, LAG, etc.) • CASE statement • COALESCE • What Is A Stored Procedure • What Is A User Defined Function • Indexes • Performance Tuning • Data Manipulation • Dynamic SQL • Advanced Data Types (JSON, XML, etc) • connecting to DB Using Python • Hierarchical Querying		

	MON	TUE	WED	THU	FRI	ENVIRONMENT
		• What Is A Join • Defining Schema				
Week-3-Hadoop (Hadoop, Hive, Spark)	Project0 QC Audit Hadoop Big Data Introduction • Big Data Fundamentals • Components Of Big Data • Architecture Benefits Challenges • Data lifecycle stages- Generation, collection, processing, storage, management, analysis, visualization, interpretation Hadoop Introduction • Big Data Fresher Hadoop Architecture • Hadoop	Project0 • Introduction To Mapreduce • Hadoop Vs Mapreduce Vs Spark Hive-Introduction • Introduction To Hive • Basic Hive Queries	Project1 Cloud Computing Cloud Introduction • Cloud Computing Model Types • Cloud Computing Service Types • Cloud Computing Definition GCP Introduction • Google Cloud Platform Overview • GCP Regions and Zones • IAM Basics • Pricing and Billing • Google Compute Engine • Google Cloud Storage	Project1 Spark Spark-Fundamentals • Introduction To Spark • Spark Ecosystem • Hadoop Vs Spark • Spark Setup • Local Vs Cluster Mode • Data Loading And Saving Through Rdds	Project1 Review Topics	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
	ecosystem Components of Hadoop • Introduction Hdfs • Evolution Of Hadoop • Hdfs Commands • Yarn Overview					
Week-4-PySpark (Cloud Computing, Spark Fundamentals)	Project1 QC Audit Spark-Operations-Pyspark • Introduction To Rdd • Basic Rdd Operations • Introduction To Pyspark • Entrypoint Sparksession • Shared Variables • Actions • Transformations	Project1 SQL Coding Challenge Spark-GCP • Cluster Modes Cluster Step Execution • Running Spark Job on Dataproc Spark- Advanced Spark-Advanced Concepts • Executors • Spark Caching Overview • Spark Jobs	Project1 Spark Coding Test Spark-SQL Spark-SQL Concepts • Introduction To Spark Sql • Introduction To Dataframes • Working On Dataframes • Narrow & Wide Transformations • Selecting, Renaming, Adding, Dropping columns	Project1 Python Coding Challenge • Sorting And Partitioning • Working With Json Datasets • Working With Parquet Files Spark-Streaming-Introduction • Introduction To Streaming • Spark Streaming • Spark Engine	Project1 Review Topics • General Interview Preparation	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
		Troubleshooting • Configure Memory Driver And Executors • Driver Class Configuration	• Filter, dropping rows • Using Dataframe Aggregate Functions • Expressions • Sorting • Null handling • Joins • UDF's • Spark caching / Persistence(All storage levels)	• Processing Data Stream Using Spark Streaming Tuning & Configuration • Spark Optimization		
Week5-Data Warehouse (Big Query)	Project1 QC Audit Data Warehousing • DataWarehousing-Introduction • Data Store Vendors • OLAP,OLTP Systems • DWH Vs. Data	Project1 Big Query Datasets • Creating Datasets • Public Datasets • Dataset Properties • Create and Query Clustered Tables • Create and Query External Tables Big Query Tables	Project1 Big Query Analyze • Introduction to BigQuery Analysis • Run a Query • Write Query Results • GoogleSQL ANSI standard • Querying with Arrays	Project1 Data Warehousing Test Big Query Routines • Manage Routines • User-Defined Functions • Table Functions • SQL Stored Procedures Big Query	Project1 Review Topics • Delta Lake Schema Evolution • Delta Lake Time Travel • Delta Lake Performance optimizations	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
	Lake,DWH Vs. Data Virtualization • DWH Architecture • Operational Data Store/Staging Area • Data Mart,Data Cleansing • Conceptual/Logical/ Physical • Dimensional Modeling • Star Schema & Snowflake Schema • Slowly Changing Dimensions • DWH Vendors, Cloud Vs. On-Premises Big Query Introduction • Introduction to BigQuery • Using The BigQuery sandbox	• Create and Use Tables • Table Schemas • Create, Manage, and Query Partitioned Tables	• Querying JSON data • Querying using Sketches • Multi Statement Queries • Recursive CTEs • Table Sampling • Multi Statement Transactions • Running Parameterized Queries • Creating and Running Saved Queries • Optimize Queries • Query External Tables • Logical Views • Materialized Views	Connections • Introduction to Connections • GCP GCS Connections • Manage Connections Load/Transform/Export Data • Creating a Search Index • Manage Search Indexes • Transfer GCS data • Schedule Transfers of Data with GCS • Load Avro, Parquet, CSV, JSON, and ORC batch data • Load externally partitioned data • Load data into partitioned tables • Transforming with DML and GoogleSQL		

	MON	TUE	WED	THU	FRI	ENVIRONMENT
	• BigQuery Dry Runs • gsutil and common bq commands			• Transforming data in Partitioned tables • Work with Change History • Export Data to a file • Export Data to GCS		
Week6-GCP Professional Data Engineer Review (Technologies)	Project2 QC Audit RDBMS • Google Cloud SQL • Spanner NoSQL • NoSQL Overview • Firestore • Datastore MemoryStore • Introduction to MemoryStore BigTable	Project2 Cloud PubSub • Introduction to PubSub • Cloud PubSub with Python • Cloud PubSup with Gcloud Apache Beam • Introduction to Apache Beam • Data Pipeline using Beam • Apache Beam Transformations	Project2 Dataflow • Introduction to Dataflow • Dataflow ML • Dataflow SQL • Creating Pipelines Data Fusion • Introduction to Data Fusion	Project2 Apache Airflow • Introduction to Airflow • Creating DAG Data Loss Prevention API • Introduction to DLP Data Catalog • Introduction to Data Catalog Data Analytics & ML • ML Basics • Data Preparation	Project2 Review	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
	• Introduction to BigTable • Creating an instance			with DataPrep • BigQuery ML • Datastudio		
Week7-GCP Professional Data Engineer Review (Topic Review)	Project2 GCP Data Engineering Review Designing Data Processing Systems • Identity and Access Management • Data security • Privacy • Regional considerations • Legal and regulatory compliance • Preparing and cleaning data (e.g., Dataprep, Dataflow, and Cloud Data Fusion) • Monitoring and orchestration of data	Project2 Ingesting and Processing the Data • Planning the data pipelines • Defining data sources and sinks • Defining data transformation logic • Networking fundamentals • Data encryption • Building the pipelines • Data cleansing • Identifying the services (e.g., Dataflow, Apache Beam, Dataproc, Cloud Data Fusion, BigQuery, Pub/Sub,	Project2 Storing the Data • Selecting storage systems • Choosing managed services (e.g., Bigtable, Spanner, Cloud SQL, Cloud Storage, Firestore, Memorystore) • Planning for storage costs and performance • Lifecycle management of data • Planning for using a data warehouse • Using a data lake • Designing for a data mesh	Project2 Maintaining and Automating Data Workloads • Optimizing resources • Designing automation and repeatability • Organizing workloads based on business requirements • Monitoring and troubleshooting processes • Maintaining awareness of failures and mitigating impact	Project2 Review	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
	pipelines • Disaster recovery and fault tolerance • Making decisions related to ACID compliance and availability • Data validation • Mapping current and future business requirements to the architecture • Designing for data and application portability • Data staging, cataloging, and discovery • Designing data migrations	Apache Spark, Hadoop ecosystem, and Apache Kafka) • Transformations • Data acquisition and import • Integrating with new data sources • Job automation and orchestration (e.g., Cloud Composer and Workflows) • CI/CD	Preparing and Using Data for Analysis • Preparing data for visualization • Sharing data • Exploring and analyzing data			
Week8	Project2 Recap QC Audit	Project2 Recap	Project2 Recap	Project2 Recap	Project2 Recap	

PROJECT	TECHNOLOGIES
Project1	
Recap	
Project0	
Project2	