



## Data and Analytics - TechM

	MON	TUE	WED	THU	FRI	ENVIRONMENT
<b>Week1-Python SQL</b> <b>(Agile for Developers, Git - Fundamentals, Python-Fundamentals)</b>	PO( Rest API Services )  Agile for Developers  <b>SDLC</b> <ul style="list-style-type: none"> <li>Introduction To SDLC</li> <li>Waterfall</li> <li>Agile</li> <li>Agile Vs Waterfall</li> <li>Story Pointing</li> <li>Scrum Ceremonies</li> </ul> Git Fundamentals  <b>OS-Introduction</b> <ul style="list-style-type: none"> <li>OS: Fundamentals</li> <li>Unix/Linux: Demo Moving and Deleting Files (Using Git Bash)</li> <li>Unix/Linux: Demo File Commands (Using Git Bash)</li> </ul> <b>Git Introduction</b> <ul style="list-style-type: none"> <li>Source Control Management(git,vcs, cvcs,dvcs)</li> <li>Git Fundamentals</li> <li>Initializing A Repository</li> <li>Pushing To A Remote Repository</li> <li>Git Commit, Branch, Merge, Push, Pull</li> </ul> Git Exercises	PO( Rest API Services )  Python-Fundamentals  <b>Python-Orientation</b> <ul style="list-style-type: none"> <li>Interpreter vs Compiler</li> <li>What is Python Why Python</li> <li>Full Stack Overview</li> </ul> <b>Python-Basics</b> <ul style="list-style-type: none"> <li>Python Syntax</li> <li>Comments</li> <li>Variables and Datatypes</li> <li>Operators</li> <li>User Input and Output</li> </ul> <b>Python-DataTypes</b> <ul style="list-style-type: none"> <li>Strings</li> <li>Casting</li> <li>Boolean</li> <li>Lists</li> <li>Tuples</li> <li>Range</li> <li>Sets</li> <li>Binary Type</li> <li>Nontype</li> <li>Dictionaries</li> <li>Numbers</li> <li>Namespaces</li> </ul>	PO( Rest API Services )  <ul style="list-style-type: none"> <li>Scope</li> <li>If-Else</li> <li>While</li> <li>For</li> <li>Function</li> <li>Lambda</li> <li>Arrays</li> <li>Classes and Objects</li> <li>OOP Concepts</li> <li>Inheritance</li> <li>Iterators</li> </ul>	PO( Rest API Services )  Python Coding Challenge  <b>Python-Modules</b> <ul style="list-style-type: none"> <li>Math</li> <li>Logging</li> <li>JSON</li> <li>Regex</li> <li>numpy</li> <li>pandas</li> <li>pip and install pip</li> <li>pylint</li> <li>Connect SQL</li> </ul> <b>Python-Exception Handling</b> <ul style="list-style-type: none"> <li>Error</li> <li>Exception Handling</li> <li>try-except</li> <li>Module</li> </ul>	PO( Rest API Services )  <b>Python-File Handling</b> <ul style="list-style-type: none"> <li>Read Files</li> <li>Write Create Files</li> <li>Delete Files</li> <li>File Handling</li> </ul> Review Topics	
<b>Week2-Python SQL (SQL)</b>	PO( Rest API Services )  QC Audit  MySQL	PO( Rest API Services )  SQL Coding Challenge  <b>Sub Languages</b>	PO( Rest API Services )  SQL Joins coding test <ul style="list-style-type: none"> <li>Inner Join</li> </ul>	PO( Rest API Services )  Python Coding Challenge  <b>Advanced-SQL</b>	PO( Rest API Services )  Review Topics	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
	<b>SQL-Introduction</b> <ul style="list-style-type: none"> <li>What Is A Database</li> <li>What Is SQL</li> <li>Consistency</li> <li>Introduction To RDBMS</li> </ul> <b>Structure</b> <ul style="list-style-type: none"> <li>Schema</li> <li>Table Structure</li> <li>SQL Data Types</li> <li>Normalization</li> <li>Multiplicity</li> <li>Data Modeling And ERD</li> <li>Primary Key</li> <li>Composite Key</li> <li>Foreign Key</li> <li>Unique Key</li> <li>Secondary Alternate Key</li> <li>Referential Integrity</li> </ul>	<ul style="list-style-type: none"> <li>Overview Of Sublanguages</li> <li>DDL</li> <li>DML</li> <li>DQL</li> <li>DCL</li> <li>TCL</li> </ul> <b>DDL</b> <ul style="list-style-type: none"> <li>CREATE DROP TRUNCATE</li> <li>Constraints</li> <li>Auto Incrementing</li> <li>CHECK</li> <li>DEFAULT</li> <li>CASCADE</li> </ul> <b>DML</b> <ul style="list-style-type: none"> <li>INSERT</li> <li>UPDATE</li> <li>DELETE</li> </ul> <b>DQL</b> <ul style="list-style-type: none"> <li>Queries</li> <li>Aggregate Functions</li> <li>Clauses</li> <li>What Is A Subquery</li> <li>What Is A Join</li> <li>Defining Schema</li> </ul>	<ul style="list-style-type: none"> <li>Left And Right Joins</li> <li>Outer Join</li> <li>Cross Join</li> <li>Equi And Theta Joins</li> <li>Aliases</li> </ul> <b>Transaction</b> <ul style="list-style-type: none"> <li>What Is A Transaction</li> <li>ACID Properties</li> <li>Transaction Properties</li> <li>CRUD Operations</li> <li>Transaction Commit Rollback Isolation Levels</li> </ul>	<ul style="list-style-type: none"> <li>Scalar Functions</li> <li>Sequence</li> <li>Trigger</li> <li>Views</li> <li>Window Functions (ROW_NUMBER, RANK, DENSE_RANK, LEAD, LAG, etc.)</li> <li>CASE statement</li> <li>COALESCE</li> <li>What Is A Stored Procedure</li> <li>What Is A User Defined Function</li> <li>Indexes</li> <li>Performance Tuning</li> <li>Data Manipulation</li> <li>Dynamic SQL</li> <li>Advanced Data Types (JSON, XML, etc)</li> <li>connecting to DB Using Python</li> <li>Hierarchical Querying</li> </ul>		
<b>Week-3-Hadoop (Hadoop, Hive, Spark)</b>	PO( Rest API Services )  QC Audit  Hadoop  <b>Big Data Introduction</b> <ul style="list-style-type: none"> <li>Big Data Fundamentals</li> <li>Components Of Big Data</li> <li>Architecture Benefits Challenges</li> <li>Data lifecycle</li> </ul>	PO( Rest API Services )  <ul style="list-style-type: none"> <li>Introduction To Mapreduce</li> <li>Hadoop Vs Mapreduce Vs Spark</li> </ul> <b>Hive-Introduction</b> <ul style="list-style-type: none"> <li>Introduction To Hive</li> <li>Basic Hive Queries</li> </ul>	P1(Data Science )  Cloud Computing  <b>Cloud Introduction</b> <ul style="list-style-type: none"> <li>Cloud Computing Model Types</li> <li>Cloud Computing Service Types</li> <li>Cloud Computing Definition</li> </ul> <b>GCP Introduction</b> <ul style="list-style-type: none"> <li>Google Cloud</li> </ul>	P1(Data Science )  Spark  <b>Spark-Fundamentals</b> <ul style="list-style-type: none"> <li>Introduction To Spark</li> <li>Spark Ecosystem</li> <li>Hadoop Vs Spark</li> <li>Spark Setup</li> <li>Local Vs Cluster Mode</li> <li>Data Loading And</li> </ul>	P1(Data Science )  Review Topics  Project Presentation	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
	stages- Generation, collection, processing, storage, management, analysis, visualization, interpretation  <b>Hadoop Introduction</b>  <ul style="list-style-type: none"> <li>Big Data Fresher</li> </ul> <b>Hadoop Architecture</b>  <ul style="list-style-type: none"> <li>Hadoop ecosystem</li> </ul> <b>Components of Hadoop</b>  <ul style="list-style-type: none"> <li>Introduction Hdfs</li> <li>Evolution Of Hadoop</li> <li>Hdfs Commands</li> <li>Yarn Overview</li> </ul>		Platform Overview  <ul style="list-style-type: none"> <li>GCP Regions and Zones</li> <li>IAM Basics</li> <li>Pricing and Billing</li> <li>Google Compute Engine</li> <li>Google Cloud Storage</li> </ul>	Saving Through Rdds		
<b>Week-4-PySpark (Cloud Computing, Spark Fundamentals)</b>	P1(Data Science )  QC Audit  <b>Spark-Operations-Pyspark</b>  <ul style="list-style-type: none"> <li>Introduction To Rdd</li> <li>Basic Rdd Operations</li> <li>Introduction To Pyspark</li> <li>Entrypoint Sparksession</li> <li>Shared Variables</li> <li>Actions</li> <li>Transformations</li> </ul>	P1(Data Science )  SQL Coding Challenge  <b>Spark-GCP</b>  <ul style="list-style-type: none"> <li>Cluster Modes</li> <li>Cluster Step Execution</li> <li>Running Spark Job on Dataproc</li> </ul> Spark- Advanced  <b>Spark-Advanced Concepts</b>  <ul style="list-style-type: none"> <li>Executors</li> <li>Spark Caching Overview</li> <li>Spark Jobs Troubleshooting</li> <li>Configure Memory Driver And Executors</li> <li>Driver Class Configuration</li> </ul>	P1(Data Science )  Spark Coding Test  Spark-SQL  <b>Spark-SQL Concepts</b>  <ul style="list-style-type: none"> <li>Introduction To Spark Sql</li> <li>Introduction To Dataframes</li> <li>Working On Dataframes</li> <li>Narrow &amp; Wide Transformations</li> <li>Selecting, Renaming, Adding, Dropping columns</li> <li>Filter, dropping rows</li> <li>Using Dataframe Aggregate Functions</li> <li>Expressions</li> <li>Sorting</li> <li>Null handling</li> <li>Joins</li> <li>UDF's</li> </ul>	P1(Data Science )  Python Coding Challenge  <ul style="list-style-type: none"> <li>Sorting And Partitioning</li> <li>Working With Json Datasets</li> <li>Working With Parquet Files</li> </ul> <b>Spark-Streaming-Introduction</b>  <ul style="list-style-type: none"> <li>Introduction To Streaming</li> <li>Spark Streaming</li> <li>Spark Engine</li> <li>Processing Data Stream Using Spark Streaming</li> </ul> <b>Tuning &amp; Configuration</b>  <ul style="list-style-type: none"> <li>Spark Optimization</li> </ul>	P1(Data Science )  Review Topics  <ul style="list-style-type: none"> <li>General Interview Preparation</li> </ul>	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
			<ul style="list-style-type: none"> <li>Spark caching / Persistence(All storage levels)</li> </ul>			
<b>Week5-Data Warehouse (Big Query)</b>	P1(Data Science )  QC Audit  <b>Data Warehousing</b> <ul style="list-style-type: none"> <li>DataWarehousing- Introduction</li> <li>Data Store Vendors</li> <li>OLAP,OLTP Systems</li> <li>DWH Vs. Data Lake,DWH Vs. Data Virtualization</li> <li>DWH Architecture</li> <li>Operational Data Store/Staging Area</li> <li>Data Mart,Data Cleansing</li> <li>Conceptual/Logical/Physical</li> <li>Dimensional Modeling</li> <li>Star Schema &amp; Snowflake Schema</li> <li>Slowly Changing Dimensions</li> <li>DWH Vendors, Cloud Vs. On-Premises</li> <li><b>Big Query Introduction</b></li> <li>Introduction to BigQuery</li> <li>Using The BigQuery sandbox</li> <li>BigQuery Dry Runs</li> <li>gsutil and common bq commands</li> </ul>	P1(Data Science )  <b>Big Query Datasets</b> <ul style="list-style-type: none"> <li>Creating Datasets</li> <li>Public Datasets</li> <li>Dataset Properties</li> <li>Create and Query Clustered Tables</li> <li>Create and Query External Tables</li> <li><b>Big Query Tables</b></li> <li>Create and Use Tables</li> <li>Table Schemas</li> <li>Create, Manage, and Query Partitioned Tables</li> </ul>	P1(Data Science )  <b>Big Query Analyze</b> <ul style="list-style-type: none"> <li>Introduction to BigQuery Analysis</li> <li>Run a Query</li> <li>Write Query Results</li> <li>GoogleSQL ANSI standard</li> <li>Querying with Arrays</li> <li>Querying JSON data</li> <li>Querying using Sketches</li> <li>Multi Statement Queries</li> <li>Recursive CTEs</li> <li>Table Sampling</li> <li>Multi Statement Transactions</li> <li>Running Parameterized Queries</li> <li>Creating and Running Saved Queries</li> <li>Optimize Queries</li> <li>Query External Tables</li> <li>Logical Views</li> <li>Materialized Views</li> </ul>	P1(Data Science )  Data Warehousing Test  <b>Big Query Routines</b> <ul style="list-style-type: none"> <li>Manage Routines</li> <li>User-Defined Functions</li> <li>Table Functions</li> <li>SQL Stored Procedures</li> <li><b>Big Query Connections</b></li> <li>Introduction to Connections</li> <li>GCP GCS Connections</li> <li>Manage Connections</li> <li><b>Load/Transform/Export Data</b></li> <li>Creating a Search Index</li> <li>Manage Search Indexes</li> <li>Transfer GCS data</li> <li>Schedule Transfers of Data with GCS</li> <li>Load Avro, Parquet, CSV, JSON, and ORC batch data</li> <li>Load externally partitioned data</li> <li>Load data into partitioned tables</li> <li>Transforming with DML and GoogleSQL</li> <li>Transforming data in Partitioned tables</li> <li>Work with Change History</li> <li>Export Data to a</li> </ul>	P1(Data Science )  Review Topics <ul style="list-style-type: none"> <li>Delta Lake Schema Evolution</li> <li>Delta Lake Time Travel</li> <li>Delta Lake Performance optimizations</li> </ul>	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
				file <ul style="list-style-type: none"> <li>Export Data to GCS</li> </ul>		
<b>Week6-GCP Professional Data Engineer Review (Technologies)</b>	P2(ETL Pipeline )  QC Audit  <b>RDBMS</b> <ul style="list-style-type: none"> <li>Google Cloud SQL</li> <li>Spanner</li> </ul> <b>NoSQL</b> <ul style="list-style-type: none"> <li>NoSQL Overview</li> <li>Firestore</li> <li>Datastore</li> </ul> <b>MemoryStore</b> <ul style="list-style-type: none"> <li>Introduction to MemoryStore</li> </ul> <b>BigTable</b> <ul style="list-style-type: none"> <li>Introduction to BigTable</li> <li>Creating an instance</li> </ul>	P2(ETL Pipeline )  <b>Cloud PubSub</b> <ul style="list-style-type: none"> <li>Introduction to PubSub</li> <li>Cloud PubSub with Python</li> <li>Cloud PubSup with Gcloud</li> </ul> <b>Apache Beam</b> <ul style="list-style-type: none"> <li>Introduction to Apache Beam</li> <li>Data Pipeline using Beam</li> <li>Apache Beam Transformations</li> </ul>	P2(ETL Pipeline )  <b>Dataflow</b> <ul style="list-style-type: none"> <li>Introduction to Dataflow</li> <li>Dataflow ML</li> <li>Dataflow SQL</li> <li>Creating Pipelines</li> </ul> <b>Data Fusion</b> <ul style="list-style-type: none"> <li>Introduction to Data Fusion</li> </ul>	P2(ETL Pipeline )  <b>Apache Airflow</b> <ul style="list-style-type: none"> <li>Introduction to Airflow</li> <li>Creating DAG</li> </ul> <b>Data Loss Prevention API</b> <ul style="list-style-type: none"> <li>Introduction to DLP</li> </ul> <b>Data Catalog</b> <ul style="list-style-type: none"> <li>Introduction to Data Catalog</li> </ul> <b>Data Analytics &amp; ML</b> <ul style="list-style-type: none"> <li>ML Basics</li> <li>Data Preparation with DataPrep</li> <li>BigQuery ML</li> <li>Datastudio</li> </ul>	P2(ETL Pipeline )  Review  Project Presentation	
<b>Week7-GCP Professional Data Engineer Review (Topic Review)</b>	P2(ETL Pipeline )  <b>GCP Data Engineering Review</b>  <b>Designing Data Processing Systems</b> <ul style="list-style-type: none"> <li>Identity and Access Management</li> <li>Data security</li> <li>Privacy</li> <li>Regional considerations</li> <li>Legal and regulatory compliance</li> <li>Preparing and cleaning data (e.g., Dataprep, Dataflow, and Cloud Data Fusion)</li> <li>Monitoring and orchestration of data</li> </ul>	P2(ETL Pipeline )  <b>Ingesting and Processing the Data</b> <ul style="list-style-type: none"> <li>Planning the data pipelines</li> <li>Defining data sources and sinks</li> <li>Defining data transformation logic</li> <li>Networking fundamentals</li> <li>Data encryption</li> <li>Building the pipelines</li> <li>Data cleansing</li> <li>Identifying the services (e.g., Dataflow, Apache Beam, Dataproc, Cloud Data Fusion, BigQuery, Pub/Sub, Apache Spark,</li> </ul>	P2(ETL Pipeline )  <b>Storing the Data</b> <ul style="list-style-type: none"> <li>Selecting storage systems</li> <li>Choosing managed services (e.g., Bigtable, Spanner, Cloud SQL, Cloud Storage, Firestore, Memorystore)</li> <li>Planning for storage costs and performance</li> <li>Lifecycle management of data</li> <li>Planning for using a data warehouse</li> <li>Using a data lake</li> <li>Designing for a data mesh</li> </ul> <b>Preparing and Using</b>	P2(ETL Pipeline )  <b>Maintaining and Automating Data Workloads</b> <ul style="list-style-type: none"> <li>Optimizing resources</li> <li>Designing automation and repeatability</li> <li>Organizing workloads based on business requirements</li> <li>Monitoring and troubleshooting processes</li> <li>Maintaining awareness of failures and mitigating impact</li> </ul> AI- Orientation <ul style="list-style-type: none"> <li>ML Introduction</li> <li>AI Introduction</li> </ul>	P2(ETL Pipeline )  Review  AI-Tooling  <b>AI-Tooling-Orientation</b> <ul style="list-style-type: none"> <li>AI Tooling Overview</li> <li>AI Pair Programming Overview</li> <li>Codeium Overview</li> <li>Using Copilot, Codeium, Code Whisperer (TBD which one)</li> <li>Integration with IDE</li> </ul> <b>AI-Tooling-Code-Generation</b> <ul style="list-style-type: none"> <li>Use Cases and</li> </ul>	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
	<p>pipelines</p> <ul style="list-style-type: none"> <li>Disaster recovery and fault tolerance</li> <li>Making decisions related to ACID compliance and availability</li> <li>Data validation</li> <li>Mapping current and future business requirements to the architecture</li> <li>Designing for data and application portability</li> <li>Data staging, cataloging, and discovery</li> <li>Designing data migrations</li> </ul>	<p>Hadoop ecosystem, and Apache Kafka)</p> <ul style="list-style-type: none"> <li>Transformations</li> <li>Data acquisition and import</li> <li>Integrating with new data sources</li> <li>Job automation and orchestration (e.g., Cloud Composer and Workflows)</li> <li>CI/CD</li> </ul>	<p><b>Data for Analysis</b></p> <ul style="list-style-type: none"> <li>Preparing data for visualization</li> <li>Sharing data</li> <li>Exploring and analyzing data</li> </ul>	<ul style="list-style-type: none"> <li>GenAI Overview</li> </ul> <p><b>LLM-Overview</b></p> <ul style="list-style-type: none"> <li>LLMs (GPT, BERT, Claude, Llama, Copilot, Codeium)</li> <li>Use cases for LLM</li> <li>LLM best practices</li> <li>Security considerations</li> <li>Hallucinations</li> <li>AI Review</li> </ul> <p><b>Prompt-Engineering</b></p> <ul style="list-style-type: none"> <li>Prompt Engineering Introduction</li> <li>Zero-shot Prompting</li> <li>Few-shot prompting</li> <li>Constraints</li> <li>Fine-tuning and Conditioning</li> <li>Interaction and Dialog State</li> <li>Instructions and Guidelines</li> <li>Hallucinations</li> <li>Responsible Usage</li> <li>Security</li> </ul> <p>Prompt Engineering Review</p>	<p>Best Practices for GenAI Code Generation</p> <ul style="list-style-type: none"> <li>Using GenAI for Code Generation</li> </ul> <p><b>AI-Tooling-UnitTest-Generation</b></p> <ul style="list-style-type: none"> <li>Use Cases and Best Practices for GenAI Unit Tests</li> <li>Using GenAI for Testing</li> </ul> <p><b>AI-Tooling-Documentation-Generation</b></p> <ul style="list-style-type: none"> <li>Use Cases and Best Practices for GenAI Documentation</li> <li>Using GenAI for Documentation</li> </ul> <p><b>AI-Tooling-Code-Analysis</b></p> <ul style="list-style-type: none"> <li>Use Cases and Best Practices for GenAI Code Analysis</li> <li>Using GenAI for Code Analysis</li> </ul> <p><b>AI-Tooling-Code-Optimization</b></p> <ul style="list-style-type: none"> <li>Use Cases and Best Practices for GenAI Code Optimization</li> <li>Using GenAI for Code Optimization</li> </ul> <p><b>AI-Tooling-Responsible-Use</b></p> <ul style="list-style-type: none"> <li>Responsible Uses Overview</li> <li>AI Tools for Code Review</li> <li>Searching Codebases with GenAI</li> <li>Assessing Generated Content</li> </ul>	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
					Quality  AI-Tooling-Security  • Overview of Security Benefits/Risks with GenAI  • GenAI Security Analysis  • Common Security Problems/Solutions with GenAI  • Gen AI Security Best Practices  AI-Tooling Review	
Week8	P2(ETL Pipeline )  Recap  QC Audit	P2(ETL Pipeline )  Recap	P2(ETL Pipeline )  Recap	P2(ETL Pipeline )  Recap	P2(ETL Pipeline )  Recap	

PROJECT	TECHNOLOGIES
P1(Data Science )	PySpark, BigQuery, Hadoop, Spark-SQL
Recap	
PO( Rest API Services )	Python, SQL, REST, Git
P2(ETL Pipeline )	GCP, Data Visualization, Apache Airflow, BigQuery ML



Copyright © 2024 Revature, LLC. All Rights Reserved.

By viewing this document, you agree that under copyright law all content displayed is the sole intellectual property of Revature, LLC, a technology advancement and consulting company based in Reston, VA. All content generated by a representative of Revature which is used for the company's advancement, development, or have otherwise been developed at the company's request, are the sole property of the company. No intellectual property may be reproduced, distributed, altered, or shared without the explicit permission from a representative of Revature.