BridgeLabz	**BridgeLabz Data Engineering - Python + PySpark Track**			Stages - Python Programming, OOP, Database, Big Data, Hadoop, Hive, PySpark, Spark SQL, Project Simulation
Fellowship Part	**Num of Days**	**Fellowship Stage**	**Concept Introduced**	**Functionality Developed**
Core Foundation	8	Core Programming & Object Oriented Programming with Python	Software Installations, GIT concepts(GIT Commits, Version History and Messages,), PyCharm, Proper Programming Hygiene, I/O operations, Basic Data Types, Conditional Statements, Loop Structures	Basic Core Programming Constructs using Employee Wage Example
			Expressions, Conditions, Iterations, REPL, Functions & lambda Python Best Coding Practices : Python	Solve Employee Wage Problem using Python
			JSON, Nested Data Handling, Tuples, List & Dictionary comprehensions	Programs for JSON, Datatypes using Python
			Exceptions, Operators for Data manipulation and Comparison	Functional Programs, Logical Programs
			OOP concepts, PyUnit, Exceptions, Importing Modules, File Handling Operations	Programs for Unit Testing, Object Oriented Programs
Core Foundation	4	Python Libraries	Core Libraries - NumPy, Pandas	Core Libraries - NumPy, Pandas : Solve problems using Numpy and Pandas.
			Visualization Libraries - Matplotlib, Seaborn	Visualization Libraries - Seaborn, Plotly : Plot various types of graph using visualization libraries like Matplotlib, Seaborn
Core Foundation	2	Database	MySQL - Import, Export using Python, Queries, Joins	1. Install MySQL 2. Practice Problems with different Queries, Joins 3. CRUD operation using Python and MySQL 4. Import data into MySQL 5. Export data from MySQL
Big Data and Hadoop Introduction	6	Big Data and Hadoop Introduction	Overview & Need of BigData	1. Install Hadoop cluster on local machine 2. Create a small file on your machine. Put that on HDFS. Check the number of blocks created. 3. Get a bigger file (1 GB plus). Put that on HDFS. Check the blocks created again. It should be in parts of 128 MB. 4. MR program using Python
			Hadoop Introduction, Data Pipeline, Its components and Architecture	
			Installing Hadoop and familiarity with HDFS Commands, Read/Write Data in HDFS, Architecture, Configuration Properties, Hadoop Commands, YARN Introduction	
			Big Data Ingestion Tools: Sqoop and its uses, Importing and exporting data to Hadoop from RDBMS to Hadoop using Sqoop	Install Apache Sqoop. Transfer Data both from and to RDBMS to Hadoop/HDFS using Sqoop Transfer real time twitter data to Hadoop/HDFS

BridgeLabz	**BridgeLabz Data Engineering - Python + PySpark Track**			Stages - Python Programming, OOP, Database, Big Data, Hadoop, Hive, PySpark, Spark SQL, Project Simulation
Fellowship Part	**Num of Days**	**Fellowship Stage**	**Concept Introduced**	**Functionality Developed**
Data Processing using Apache Hive	**5**	**Apache Hive & Visualization**	**Apache Hive, Architecture, Hlve Meta Store, HiveQL-DDL, DML, Dynamic Partioning in Hive, Bucketing, UDF and UDAF**	**1. Use the static data available with BridgeLabz - Some GBs of data 2. Store the data on HDFS using HDFS put command 3. Preprocess the data (Use Python + Hive) 2. Finding users with lowest number of average hours 3. Finding users with highest number of average hours 4. Finding users with highest numbers of times late comings 5. Finding users with highest numbers of idle hours 6. Store the clean data on HDFS**
Data Processing Using PySpark	**8**	**Programming Hadoop with PySpark ETL/ELT Processing using PySpark**	**PySpark, Resilient Distibuted Dataset (RDD), Spark SQL, Optimisation Techniques and Performance Tuning, Spark SQL**	**1. Finding users with lowest number of average hours 2. Finding users with highest number of average hours 3. Finding users with highest numbers of times late comings 4. Finding users with highest numbers of idle hours ETL Practice problem 1. Fetch data from hadoop using Spark 2. Process user logs to find average late hours, average time spent, total number of leaves 3. Store data in mysql database**
Project Simulation	**6**	**Work on a sample project**	**Learners will work on a sample project for Data Processing using Hadoop and Spark,**	