Advanced Data Analytics using Hadoop

**Unit 1**

HDFS practice using HdfsCLI, Pyhdfs

Mapreduce programming using Java and Python

Sources:

1. [PyHDFS · PyPI](https://pypi.org/project/PyHDFS/); [GitHub - jingw/pyhdfs: Python HDFS client](https://github.com/jingw/pyhdfs)
2. [Quickstart — HdfsCLI 2.5.8 documentation](https://hdfscli.readthedocs.io/en/latest/quickstart.html)
3. [[Apache-SVN] Index of /hadoop/common/trunk/hadoop-mapreduce-project/hadoop-mapreduce-examples/src/main/java/org/apache/hadoop/examples](http://svn.apache.org/viewvc/hadoop/common/trunk/hadoop-mapreduce-project/hadoop-mapreduce-examples/src/main/java/org/apache/hadoop/examples/)
4. [Contents — Pydoop 2.0.0 documentation (crs4.github.io)](https://crs4.github.io/pydoop/)
5. [mrjob · PyPI](https://pypi.org/project/mrjob/)

**Unit 2**

Pig practice using Python

Sources:

1. [User Defined Functions (apache.org)](https://pig.apache.org/docs/r0.17.0/udf.html#python-udfs)
2. [Extending Apache Pig with Python UDFs | Codementor](https://www.codementor.io/@sheena/extending-hadoop-apache-pig-with-python-udfs-du107mj6t)

**Unit 3**

Hive

Sources:

1. [Python UDF with Apache Hive and Apache Pig - Azure HDInsight | Microsoft Docs](https://docs.microsoft.com/en-us/azure/hdinsight/hadoop/python-udf-hdinsight)
2. [PyHive · PyPI](https://pypi.org/project/PyHive/)

**Unit 4**

Hbase:

Sources:

1. [HappyBase — HappyBase 1.2.0 documentation](https://happybase.readthedocs.io/en/latest/)

**Unit 5**

Spark using pyspark

Sources:

1. [PySpark Documentation — PySpark 3.2.1 documentation (apache.org)](https://spark.apache.org/docs/latest/api/python/)