

## **Tutorial Proposal**

Topics to be included in the tutorial:

- Spark overview
- How to build Spark
- Launch a Python interpreter that is configured to run PySpark applications
- A program to demonstrate how to use Spark
- A peek into spark libraries :

Reference : <https://spark.apache.org/docs/latest/api/python/index.html>

- Spark SQL
  - How to implement in python
  - Dataset – json file loaded from twitter (Spark Apache provides a single interface for efficiently working with structured data, including Apache Hive tables, parquet files and JSON files).
- MLlib
  - Try available algorithms in python
- GraphX
- Spark Streaming
- Discuss pyspark.SparkContext
- Discuss pyspark.RDD