

Ajana Sathian

Big Data Engineer

✉ ajanacs@gmail.com

🏠 ajanacs.com

🔄 github.com/ajanac

Relevant Experience

Big Data Engineer @ UBS

Sept 2019 – Present // Nashville, TN

- Working on building and maintaining a data processing platform on Azure Cloud Data lake with Java, Scala and Python APIs.
- Communicate and collaborate with multi-disciplinary teams of developers in UBS to migrate their existing data processing jobs into this new platform.
- Refactoring existing data processing jobs onto this well-formed Data lake platform that unlocks added value associated with a Big Data approach in a Cloud based environment improve the time-to-market, increase emphatic auditing artifacts finally enable dependencies to be loosely coupled to the data processing jobs.

Big Data Developer @ Bank of America

Feb 2018 – Aug 2019 // Chicago, IL

- Developed Scala application to compare the stock market data from different data providers in order to detect prohibited market activities.
- Developed a Python application to create an alert system if the internal data base within the bank was not able to collect the market data within a specified period of time from the external stock market data providers. This notification system helped to solve the data loss.

Big Data Intern @ Cars.com

May 2017 – Feb 2018 // Chicago, IL

- Implemented Linear Regression Spark Application in Scala to predict the number of days to sell the car based on features like price, mileage, and model. Spark Application ran in the Cloudera cluster.
- Implemented Sqoop script command that transfer Machine Learning output data from Hive table to Oracle table.
- Wrote several Spark Unit test cases using Shared Spark Context for the existing Spark Applications.
- Created interactive Tableau dashboard for Hadoop data catalog that lists the metadata information about the tables in the Hadoop cluster.

Web Developer @ University of Illinois at Chicago

Nov 2016 – Jan 2017

- Developed a broadcasting module using Socket.IO & Node.JS to notify the web portal users reported or ongoing service disruptions. This feature helped the users from reporting duplicated service requests through the portal.

Skills

Programming Languages

Scala, Python, HTML, CSS, Java Script

Libraries & Frameworks

Apache Spark, IntelliJ IDEA, PyCharm
Jupyter Notebook, Databricks Notebook

Tools & Platforms

Cloudera, Azure, Git, Jenkins, Bitbucket

Education

Elmhurst College

2017 – 2018 // Elmhurst, IL

Masters in Computer Information Systems

GPA : 3.78

University of Calicut

2010 – 2013 // Calicut, India

Masters in Computer Applications

GPA: 3.64

University of Calicut

2007 – 2010 // Calicut, India

Bachelors in Computer Science

GPA : 3.88

Projects

Beer Data, Data Engineering Project

Developed a Python application that interacts with a web API that stores information about many different beers. Extracted the information about the first 100 beers (by id), clean the data and organize into a CSV file.

Twitter Sentimental Analysis

Developed a Spark application for Naïve Bayesian Classification Algorithm to classify the tweets into positive and negative. Spark application is written in Scala and implemented in Hortonworks cluster.

Interests

Sewing, fashion designing.