Ram S
sjram1751@gmail.com
(978)-419 1751

Note:Looking for W2 roles.

Summary:

- Highly skilled in Data Cleaning, Machine learning, Deep learning, and Data Visualization with strong programming experience in Python and R acquired professionally and academically for 7 years.
- 7 years of professional IT experience which includes experience in Hadoop using Cloudera, Hortonworks, and Hadoop working environment includes MapReduce, HDFS, HBase, Zookeeper, Oozie, Hive, Sqoop, Pig, Spark and Kafka.
- Expertise in Installing, Configuring and using Hadoop EcoSystem Components like HDFS, Hadoop MapReduce, Yarn, Zookeeper, Sqoop, Flume, Hive, HBase, Spark, Oozie. in Hadoop (CDH3/CDH4 & Horton works) architecture, MapReduce programming using Hive, Python/Java.
- Experience in configuring various configuration files like core-site.xml, hdfs-site.xml, mapred-site.xml, yarn-site.xml based upon the job requirement.
- Excellent understanding of Hadoop architecture and different components of Hadoop clusters which include components of Hadoop (Job Tracker, Task Tracker, Name Node and Data Node).
- Exposure in analyzing large data sets using HiveQL, Pig Latin, HBase and custom MapReduce programs in Python.
- Hands on performing ad-hoc queries on structured data using Hive SQL and used Partition and Bucketing techniques and joins with HIVE for faster data access.
- Extensively worked on Hive, Pig and Sqoop for sourcing and transformations.
- Experience working on deploying a Hadoop cluster using Cloudera 5.X integrated with Cloudera Manager for monitoring and Alerting.
- Experience in implementing Real-Time event processing and analytic's using messaging systems like Spark Streaming.
- Hands on experience in data mining process, implementing complex business logic and optimizing the query using HiveQL and controlling the data distribution by partitioning and bucketing techniques to enhance performance.
- Experience building Spark applications using Python/R.
- Experience with parallel computing resources like Spark, Hadoop, Databricks, AWS EC2, and Nvidia GPU's.
- Maintaining ETL process using Amazon Glue.
- Experience with Spark frameworks Spark Sql, Spark Streaming, Mahout & Mllib(Machine Learning), Graphx(Computation), SparkR.
- Experience in the following areas is a big plus: algorithm performance optimization, distributed computing, parallel computing, machine/deep learning.
- Well versed in areas such as applied statistics, machine/deep learning, natural language processing, and data mining.
- Experience with Structured and UN-structured data for classification and regression analysis.
- Proficient with distributed systems like Hadoop and SPARK and have experience processing large datasets in such frameworks.
- Experience in Data Acquisation, Data Cleaning(Normal data, Image Data, Text Data), Model training & Building, model testing, and Model Deployment.

Technical Skills:
Big Data Frameworks: Hadoop: HDFS, MapReduce, Hive, Hbase, Pig, Avro, Oozie
                          Spark: Spark Sql, Spark Streaming, Mahout & Mllib(Machine Learning), Graphx(Computation), SparkR.
Machine learning/Computational Packages: numpy, scipy, pandas, scikit -learn and nltk.

Data Visualization Libraries: MatplotLib, Seaborn, ggplot, plotly, geaplotlib , Dashboards Visualization.
Deep Learning Frameworks: PyTorch, Tensorflow.
Data Analysis Languages: Python, R.
Cloud Technologies: AWS S3, Google Cloud, Nvidia GPU cloud.
ETL process: Amazon Glue.

Experience:

Data Scientist   **January 2018 - September 2018**
Client :CGI
Responsibilities:
- Ability to Identify patterns, risks, trends and opportunities in the data.
- Support and design entire data pipeline.
- Handling data set size ranges from 1TB – 1PB.
- Worked on all aspects of creating predictive intelligence including collecting requirements, establish analytics work plan, data exploration, cleansing, and preparation, identifying features, selecting algorithms, building and testing models, iteratively improving solutions .
- Designed and implemented various machine learning models and applications.
- Designed and populated the relational account marketing database.
- Developed record linkage algorithm for corporate clients by sourcing data from internal and external databases, API and web scraping.
- Developed web apps and design API's for internal usage.
- Performed Ad-hoc statistical data analyses.
- Create a datasource from Amazon S3, loading a CSV file of information about customers, and information about how they responded to marketing communications.
- Build a machine learning model from the datasource.
- Measure model accuracy and adjust score threshold accordingly.

Environment:Python, R, Jupyter, Nvidia Gpu's,Tableau, Git, SQL server, MLLiB, NLTK, Spark, Azure.

Data Scientist   **September 2016– December 2017**
Client: AT&T
Responsibilities:
- Architect of new machine learning model scoring service with ability to serve models developed in several different languages and frameworks.
- Implemented Python variants of various learning algorithms, such as Generalized Additive Models and Constrained Linear Models.
- Contributed to key algorithms to generate model insights and auditability for regulatory compliance.
- Supported Data Scientists with ad-hoc and production algorithms for feature analysis and selection.
- Provided dashboards and automated reports for business stakeholders.
- Developed and deployed several models for credit underwriting, including models for new products.
- Analyzed and integrated new data sources into production systems to increase data redundancy.

Environment:Python, R, Jupyter, Nvidia Gpu's,Tableau, Git, SQL server, MLLiB, NLTK, Spark, Azure.

Machine Learning Engineer – Research Intern   **June 2015 – August 2016**
AI and Machine Learning Lab
Responsibilities:

- Build machine learning pipelines for both supervised and unsupervised learning.
- Transformer/Estimator/Pipeline API.
- Use transformers to perform pre-processing on a dataset prior to training.
- Train analytical models with Spark ML's DataFrame-based estimators including Decision Trees, Random Forests, Gradient Boosted Trees, Linear Regression, K-Means, and Alternating Least Squares.
- Tune hyperparameters via cross-validation and grid search.
- Evaluate model performance.

Environment:Python, R, Jupyter, Nvidia Gpu's.

Teaching Assistant   September 2014 – April 2015
Data Visualization and Machine Learning Course Work
Responsibilities:
- Analyzed astrophysical data sets containing millions of data points in a high-dimensional space. Measured clustering and over-densities of source properties.
- Stacked data from tens of thousands of observations to improve statistical significance of results .
- Developed and carried out an efficient statistical experiment to quantify varying noise levels using random sampling.
- Automated model fitting to uncover deviations from expected results .
- Worked in a cross-functional team.
- Collaborated on tens of projects spanning a wide range of topics.
- Carried teaching assistant duties in two undergraduate classes
- Held class sections and office hours.

Tableau Data Visualization Developer   August 2012   - July 2014
Client:muthoot finance
- Present data in a way that can be understood by a variety of stakeholders, including effectively presenting information or recommendations both written and verbally.
- Presentation of material should include data trends, graphs, scatter plots along with a written explanation that tells the story of the data.
- Design and conduct statistical analysis. Analysing and synthesizing data from a variety of sources to identify impacts, potential outcomes and risks.
- Ability to analyse problems systematically, creating a path for achieving the desired outcome with minimal supervision and use of best practices in data analysis and research.

Environment: Python, SQL, Tableau, Git, SQL server, MLLiB, NLTK, Spark, Azure, R Studio, MangoDB, Java, Hive.

Data Analyst   August 2011 – July 2012
Client:ICIC I Bank
Responsibilities:
- Data manipulation, working with missing data and summarization.
- Building data transformation pipelines(Extract-Transform-Load (ETL)), using plain python primitives, and executing them in parallel.
- Story telling through Data Visualization using line charts, Multple Plots, Bar Plots.
- Data Cleaning, working with command line, piping and redirecting output, and Data munging with command line.
- Git and Version control.
- Building and organizing complex SQL queries, table's relations and normalization.

- *API and Web scraping.*
- *Finding the inference using probability and statistics such as frequency distribution, mean,*
- *mode, Z-scores, probability distribution, significance testing , chi-square test and A/B Testing.*
- *Create dashboards and visualizations of processed data, identify trends, anomalies.*
- *Enhancing data collection procedures to include information that is relevant for building analytic systems, Processing, cleansing, and verifying the integrity of data used for analysis.*

Software Programmer – Internship   **June-2009 -August 2009**
Technophilia System, Mumbai, India

Education:
Master's in Computer Science, University of Massachusetts (Sept 2015 – Dec 2017)
Department of Computer Science
Lowell, MA, USA.
Course Work: *Analysis of Algorithms, Computing II, Operating systems, Advanced Algorithms, Foundation of Computer Science, Internet Web systems, Computer networks & security, Machine Learning, Xinu Operating System II, Data Mining, Data Base-l, Heterogeneous Data Visual Analytic 's, Computer Graphics-I, Big Data Design(Deep Learning), and Human Computer Interaction.*

Bachelor of Engineering, Gitam University (2007 – 2011)
College of engineering
Year of Graduation: May 2011