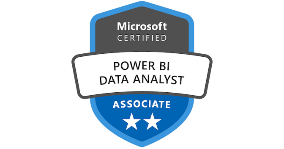
SRIKAR REDDY KONDAM (he/him) 

Senior Data Engineer

[**srikark1901@gmail.com**](mailto:srikark1901@gmail.com) **Mob: +1 (901) 610-2262**

**PROFESSIONAL SUMMARY**

* Demonstrates 10 years of robust experience in Data Analysis and Data Mining involving extensive datasets, both structured and unstructured. Proficient in Data Acquisition, Validation, Predictive and Statistical Modeling, as well as Data Visualization.
* Proficient in statistical programming languages like **R** and **Python**, **SAS**, and **Apache** **Spark**. Familiar with Big Data technologies including **Hadoop**, **Hive**, **Pig**, and **NoSQL** databases.
* Holds 3+ years of expertise in **Hadoop 2.0**, leading the development of enterprise-level solutions utilizing Hadoop tools such as **Spark**, **MapReduce**, **Sqoop**, **Pig**, **Hive**, **HBase**, and more.
* Extensive familiarity with Big Data algorithms and analytics through **Hadoop**, **MapReduce**, and **NoSQL**, showcasing distributed computing competence.
* Holds 3+ years of experience in **Hadoop 2.0**, successfully leading the development of enterprise-level solutions utilizing tools like **Spark**, **MapReduce**, **Sqoop**, **Pig**, **Hive**, and more.
* Proficient in crafting enterprise solutions with both batch processing (using Apache Pig) and streaming frameworks (**Spark Streaming**, **Apache Kafka**, **Apache Flink**).
* Expertly integrates Machine Learning, Predictive Analytics, and Big Data technologies, culminating in seamless integrated solutions.
* Proficient in Dimensional Data Modeling encompassing Relational Data Modeling, **ER/Studio**, **Erwin**, **Sybase** Power Designer, and Schema Modeling Power bi(Star Join Schema/Snowflake).
* Skilled in data cleansing and analysis using **HiveQL**, **Pig Latin**, and custom MapReduce programs in **Java**.
* Adept in creating custom User-Defined Functions (UDFs) and User-Defined Aggregation Functions (UDAFs) to extend Hive and Pig capabilities.
* Accomplished in importing/exporting data using **Sqoop**, effectively bridging HDFS and Relational Database Systems (RDBMS) such as Teradata.
* Strong expertise in Business Intelligence solutions via data modeling, Dimension Modeling, **ETL** Processes, Data Integration, and OLAP.
* Extensive involvement with Dimensional Modeling, Data Migration, Data Cleansing, Data Profiling, and ETL Processes in data warehousing contexts.
* Proficient in formatting and ETL processes for diverse data formats like **Avro**, **ORC**, Parquet, **CSV**, **JSON**, alongside Elasticsearch and MDM familiarity.
* Possesses deep knowledge of Machine Learning, Mathematical Modeling, Operations Research, and proficiency in **R**, **Python**, **SAS**, and Relational Databases.
* Utilizes **SQL**, **Numpy**, **Pandas**, **Scikit-learn**, **Spark**, and Hive extensively for Data Analysis and Model Development.
* Hands-on experience with processing frameworks such as **Spark** and **Spark SQL**.
* Experienced in data architecture, including data ingestion pipeline design, Hadoop information architecture, data modeling, mining, machine learning, and advanced data processing.
* Proficient in statistical analysis, Factor Analysis, validation using ROC plots, K-fold cross-validation, and adept in data visualization.
* Expertise in managing the complete data science project lifecycle, including data acquisition, cleaning, engineering, modeling, dimensionality reduction, and more.
* Strong background in Data Visualization using Tableau for various chart types, time series, and multiple axes.
* Skilled in Integration Services (**SSIS**), Reporting Services (**SSRS**), and Analysis Services (**SSAS**).
* Proficient in Normalization and De-normalization techniques for optimal performance in relational and dimensional databases.
* Well-versed in version control (Git, SourceTree, Bitbucket) and CI-CD tools, along with **AWS** products (S3, EC2, EMR, RDS).
* Versed in SDLC methodologies (**Agile**, **Waterfall**), with experience in Technical Design, Development, Testing, and Implementation of Enterprise-level Data marts and Data warehouses.

**EDUCATION**

* Bachelor of Technology in Computer Science from Jawaharlal Nehru Technological University, Hyderabad, India.

**TECHNICAL SKILLS**

**OLAP/Reporting Tools:**SSRS, SSAS, MDX, Tableau, Power BI

**Relational Databases:**SQL Server database 2014/2012/2008 R2/2005, Oracle 11g, SQL Server Azure, MS Access

**SQL Server Tools:**Microsoft Visual Studio 2010/2013/2015 , SQL server management studio

**Big Data Ecosystem:**HDFS, Nifi, Map Reduce, Oozie, Hive/Impala, Pig, Sqoop, Zookeeper and Hbase, Spark, Scala, Kafka, Apache Flink, AWS- EC2, S3, EMR.

**Other Tools:**MS Office 2003/2007/2010 and 2013, Power pivot, Power Builder, GIT, CI-CD, Jupyter Notebook.

**Programming languages:**C, SQL, PL/SQL, T-SQL, JAVA, Batch scripting, R, Python

**Data Warehousing & BI:**Star Schema, Snowflake schema, Facts and Dimensions tables, SAS, SSIS, and Splunk

**Operating Systems:**Windows XP/Vista/7/8 and 10; Windows 2003/2008R2/2012 Serve.

**PROFESSIONAL EXPERIENCE**

**Client:** US Bank, Minneapolis, MN **Feb 22 – Present**

**Role:** Senior Data Engineer

* Involved in gathering requirements, conducting analysis, designing solutions, overseeing development, managing changes, and facilitating deployments.
* Contributed to the development of real-time streaming applications on a distributed **Hadoop Cluster**, utilizing **PySpark**, **Apache Flink**, **Kafka**, and **Hive**.
* Employed Python and Apache Spark for the creation and execution of applications in Big Data Analytics and Machine Learning, including **Spark** **ML** and **Mllib** use cases.
* Extracted data from heterogeneous sources, applied intricate business logic to normalize raw network data, enabling BI teams to detect anomalies effectively.
* Designed and implemented Flink pipelines to process streaming data from **Kafka**, applying business logic for transformation and serialization.
* Developed a shared Flink module for AVRO data serialization and deserialization, incorporating schema definitions.
* Created a Spark streaming pipeline for batch processing real-time data, identifying anomalies through business logic, and storing them in an Hbase table.
* Implemented a layered Hadoop architecture for enhanced modularity, utilizing framework scripts to expedite development and standardized processes for error handling, logging, and metadata management.
* Indexed processed data, designed dashboards, and generated alerts in Splunk for support teams to utilize and take action upon.
* Responsible for managing operations and support across the Big Data Analytics platform, **Splunk**, and **Tableau** visualizations.
* Orchestrated the development and design of a customer and administrator dashboard control panel using Tableau, **PostgresSQL**, and RESTAPI calls.

**Client:** Bank Of America, Charlotte, NC **Dec 19 – Jan 22**

**Role:** Big Data Engineer

* Played a pivotal role in implementing novel statistical algorithms and operators on both Hadoop and SQL platforms. Leveraged optimization techniques, encompassing linear regression, K-means clustering, Naive Bayes, and other innovative approaches.
* Created a Spark batch job to automate the generation and metadata updates of external Hive tables built upon datasets residing in HDFS.
* Developed a shared Data Serialization module within Spark. This module effectively converts intricate objects into sequenced bits, utilizing formats like **AVRO, PARQUET, JSON,** and **CSV**.
* Engaged in ER Modeling, Dimensional Modeling (including Star Schema and Snowflake Schema), Data Warehousing, and proficient use of OLAP tools.
* Successfully populated both HDFS and **PostgreSQL** with substantial datasets using **Apache** **Kafka**.
* Architected and implemented a Rest API (Commerce API), facilitating seamless connectivity to PostgreSQL through Java services.
* Formulated a Batch Audit Process using batch and shell scripts. This process meticulously monitors each ETL job and provides detailed reporting, including table name, start and finish times, loaded row count, status, and more.
* Developed Spark jobs within the **PySpark** framework for efficient ETL operations, moving data from SQL Server to Hadoop.
* Held the responsibility for the continuous supervision and management of Elastic MapReduce (EMR) clusters through the AWS console.
* Designed and executed strategies for data acquisition, ingestion, and management of Hadoop infrastructure, along with the administration of other analytics tools like Splunk and Tableau.
* Possesses a solid grasp of build automation and the establishment of CI/CD pipelines.
* Crafted Python scripts to automate data ingestion pipelines across multiple data sources. Additionally, deployed Apache Nifi on AWS to streamline these processes.
* Designed and developed impactful Tableau visualizations, skillfully integrating calculations, parameters, calculated fields, groups, sets, and hierarchies to create insightful dashboards.

**Client:** Merck Corporation, Rahway, NJ **Oct 15 – Nov 19**

**Role:** Data Engineer/ Python Developer

* Executed Data Integration, Extraction, Transformation, and Loading (ETL) processes.
* During the preprocessing phase, utilized **Pandas** to handle missing data by either removal or replacement. Employed Over-sampling for the minority class and Under-sampling for the majority class, effectively balancing the dataset for a robust Data Cleaning process.
* Constructed models relying on a range of techniques, including **Linear Regression, Multiple Regression, Decision Trees, Random Forest, Logistic Regression,** and **Naive Bayes.**
* Evaluated and selected models through k-fold cross-validation, confusion matrices, and concentrated efforts on optimizing for a high recall rate. Implemented Ensemble models like Boosting and Bagging to enhance performance.
* Delivered data mining solutions to address challenges encompassing Association Rules, Classification, and Clustering problems.
* Utilized Python to identify trends and establish relationships within different data elements, subsequently deriving appropriate insights.
* Engineered a Rest API to serve predictions generated by the model, catering to other customers and teams.
* Produced intricate reports in varied formats, including list reports and summary reports, using advanced data manipulation techniques within **SAS** Enterprise Guide.
* Analyzed data and generated analytical reports using tools such as **SAS** and **MS SQL**.
* Developed compelling visualizations including charts, graphs, maps, area maps, dashboards, and storytelling features using **Tableau**.
* Established, maintained, and provided support for a Continuous Integration framework (CI/CD) based on Jenkins.
* Implemented, fine-tuned, and tested the model on **AWS** **EC2**, identifying the best-performing algorithm and optimal parameters for effective outcomes.

**Client:** Eka Software Solutions, Bangalore, India **May 13 – Aug 15 Role:** Python Data Developer

* Developed a data platform from scratch and took part in requirement gathering and analysis phase of the project in documenting the business requirements.
* Worked in designing tables in Hive, **MYSQL** using **SQOOP** and processing data like importing and exporting of databases to the HDFS, involved in processing large datasets of different forms including structured, semi-structured and unstructured data.
* Developed rest API's using python with flask and **Django** framework and done the integration of various data sources including **Java, JDBC, RDBMS, Shell Scripting, Spreadsheets,** and **Text files.**
* Worked with Hadoop architecture and the daemons of Hadoop including Name-Node, Data Node, Job Tracker, Task Tracker, and Resource Manager.
* Used AWS data pipeline for Data Extraction, Transformation and Loading from homogeneous or heterogeneous data sources and built various graphs for business decision-making using **Python matplot library**
* Developed scripts to load data to hive from HDFS and involved in ingesting data into Data Warehouse using various data loading techniques.
* Scheduled Jobs using crontab, run deck and control-m.
* Build Cassandra queries for performing various CRUD operations like create, update, read and delete, also used Bootstrap as a mechanism to manage and organize the html page layout
* Developed entire frontend and backend modules using Python on Django Web Framework and created User Interface (UI) using JavaScript, bootstrap, Cassandra with **MySQL** and **HTML5/CSS**
* Importing and exporting data jobs, to perform operations like copying data from HDFS and to HDFS using Sqoop and developed Spark code and Spark-SQL/Streaming for faster testing and processing of data.
* Analyzed SQL scripts and designed the solutions to implement using **PySpark**.
* Used JSON and XML SerDe's for serialization and de-serialization to load **JSON** and **XML** data into Hive tables.