MEGHA BASAVARAJU

DATA ENGINEER

[**adumpa@kiinfosystems.com**](mailto:adumpa@kiinfosystems.com) **|** +1 636-213-4566

# PROFESSIONAL SUMMARY

* With over **13** years of dedicated experience in the IT domain, I specialize in the Big Data ecosystem, encompassing Data

Acquisition, Ingestion, Modeling, Storage Analysis, Integration, Data Processing and Database Management. Proficient in crafting interactive dashboards, reports, and conducting ad-hoc analyses using Tableau and Matplotlib. My expertise extends to application development, implementation, deployment, and maintenance, leveraging **Hadoop** and **Spark**-based technologies like Amazon **redshift**, **Amazon EMR**, and Azure HDInsight. As an ardent Data Science enthusiast, I possess robust problem-solving, debugging, and analytical skills, actively aligning solutions with business requirements. Expertise in using **AWS S3** to stage data and to support data transfer and data archival. Experience in using **AWS** **Redshift** for large scale data migrations using **AWS DMS** and implementing **CDC** (change data capture). My extensive background includes hands-on experience with key **Big Data tools** such as Spark, Kafka, Hive, Impala, HBase and Sqoop. Proficient in reports generation using **Oracle**, **Unix shell scripting** and **Python.**

* Experienced in **leading** the team and **collaborating with a diverse** clientele to deliver tailored solutions. As a data engineer, I possess robust Problem-solving, debugging, and analytical skills, actively aligning solutions with business requirements.
* **Spark Application Development Expertise**: Experienced in developing Spark applications for interactive analysis, batch processing, and stream processing using PySpark.
* **Data Ingestion with SQOOP**: Extensive experience using Sqoop to ingest data from RDBMS systems like Oracle, MS SQL Server, Teradata, PostgreSQL, and MySQL.
* **Streaming Ingest Services and Processing**: Proficient in working with various streaming ingest services for both batch and realtime processing using tools like Spark Streaming, Kafka, Confluent, Storm, Flume, and Sqoop.
* **Teradata Expertise**: Extensively created and used various Teradata Set Tables, Multi - Set tables, global tables, volatile tables, and temp tables.
* **Advanced SQL Skills in Teradata**: Strong SQL experience in Teradata, including developing complex tuned queries with analytical functions and BTEQ scripts.
* **Data Migration Projects**: Involved in Data Migration projects from DB2 and Oracle to Teradata, including creating automated scripts using UNIX shell scripting, Oracle/TD SQL, TD Macros, and Procedures.
* **End-to-End ETL Pipelines**: Proficient in developing end-to-end ETL pipelines using Snowflake, Alteryx, and Apache NiFi for both relational and non-relational databases (SQL and NoSQL).

• **NoSQL Database Integration**: Strong working experience with NoSQL databases like. MongoDB, DynamoDB, and Cosmos DB, integrating them with Hadoop cluster.

* **Cloud-Based Solutions**: Experience with AWS and Azure cloud services for developing pipelines and Spark applications using EMR, LAMBDA, Redshift, HDInsight, and Databricks.
* **Windows and Web Application Development**: Extensive experience in Windows and Web application development utilizing a wide range of technologies including ASP.NET, C#, VB.NET, WPF, LINQ, Ajax, and various front-end technologies.
* **Business Intelligence Tools**: Developed advanced SAS Visual Analytics reporting and have experience in using Tableau and Power BI.
* **Cloud Services and Data Warehousing**: Experienced in setting up Data Lake/Data catalog on AWS Glue, implementing Data warehouse solutions in Confidential Redshift, and working with SQL Azure and Confidential Redshift/RDS.
* **Data Lake Architecture and Management:** Proficient in designing and managing Data Lake architectures, ensuring efficient storage, retrieval, and processing of large volumes of structured and unstructured data.
* **Real-time Data Processing with Apache Kafka**: Skilled in utilizing Apache Kafka for real-time data processing, enabling seamless ingestion and distribution of streaming data across various applications and systems.
* **Programming Language Proficiency:** Demonstrated expertise in programming languages such as Python and Scala, facilitating the development of custom applications, scripts, and solutions to address complex data processing and analysis requirements.

# TECHNICAL SKILLS

|  |  |
| --- | --- |
| Big Data Ecosystem | Spark |
| Cloud Environment | Amazon Web Services (AWS), Microsoft Azure |
| Databases | Oracle, MySQL, Postgre SQL, MS SQL Server, Teradata |
| NoSQL Database | Mongo DB, Dynamo DB |
| AWS | EC2, S3, Redshift, Lambda,, Data Pipeline |
| Azure | Databricks, Data Lake, Azure Data factory, SQL Database, SQL Data Warehouse,  Cosmos DB |
| Operating systems | Linux, Unix, Windows 10, Windows 8, Windows 7, Windows Server 2008/2003, Mac  OS |
| Software’s/Tools | Microsoft Excel, Shell Scripting, Linux, Putty, Jupyter Notebook, WinSCP, Servicenow,  Jira, Visual Studio |
| Reporting Tools/ETL Tools | Informatica, Talend, Data Stage, DMXpress, Precisely etc.. |
| Programming Languages | Python (Pandas, Scipy, NumPy, Scikit-Learn, Stats Models, Matplotlib, Plotly, Seaborn, Keras, TensorFlow, PyTorch), PySpark, SQL, PL/SQL, UNIX Shell Scripting |
| Version Control | GitLab |
| Job scheduling tools | Autosys, Control-M |

# PROFESSIONAL EXPERIENCE

**Symphony Health Solutions, Chesterfield, MO Jan 2022 - Present**

**Data Engineer HealthCare**

**Responsibilities:**

* Managed the team of five**.**
* Gathered business requirements from stakeholders, involved in designing and translated into optimal business solutions.
* Orchestrated project planning, including roadmaps, estimates, and execution, for multiple initiatives.
* Worked on **Apache Spark** data processing project to process data from **RDBMS** and several data streaming sources and developed **Spark** applications using **Python** on **AWS EMR.**
* Designed and deployed robust data pipelines leveraging Azure cloud services like **HDInsight, DataLake, Databricks, SQL, SQL** DB, DWH and Data storage explorer.
* Performed reporting analytics on data from AWS stack by connecting it to BI tools (**Tableau**, **Power Bi**).
* Designed and deployed multi-tier applications leveraging **AWS** services like (EC2, Route 53, S3, RDS, DynamoDB) focusing on high availability, fault tolerance, and **auto-scaling** in AWS **Cloud Formation**.
* Configured and launched **AWS EC2** instances to execute Spark jobs on **AWS Elastic Map Reduce (EMR).**
* Performed data transformations using Spark **Data Frames**, **Spark SQL**, **Spark File formats**, **Spark RDDs**.
* Transformed data from different files (Text, CSV, JSON) using **Python** scripts in **Spark.**
* Used **AWS** **Glue** to crawl the data lake in S3 to populate the Data Catalog.
* Worked with Data Analytics Services such as **Athena, Glue Data Catalog**. Used Quick Sight for dashboards.
* Developed **batch** and **streaming** processing apps using **Spark API**s for functional pipeline requirements.
* Automated data storage from streaming sources to AWS data lakes like **S3**, **Redshift** and **RDS** by configuring **AWS Kinesis (Data Firehose)**.
* Designed and Developed **ETL** jobs to extract data from Salesforce replica and load it in data mart in **Redshift.**
* Performed analytics using real time integration capabilities of **AWS Kinesis (Data Streams)** on streamed data.
* Stored data into various tiers of **AWS** **S3** based on business requirements and frequency of data access.
* Imported data from **AWS S3** into **Spark** **RDD** performed transformations and actions on RDD's.
* Worked with database administrating team on **SQL optimization** for databases like **Oracle**, **MySQL**, **MS SQL**.
* Assisted in configuring and implementing **MongoDB** cluster nodes on **AWS EC2** instances.
* Designed and implemented robust **CI/CD pipelines** using **Jenkins** for application to facilitate continuous integration and delivery.
* Managed Jenkins build agents and ensured their scalability to accommodate growing development teams and workloads.
* Performed statistical analysis using **SQL**, **Python**, R Programming and Excel.
* Build an **ETL** which utilizes spark jar inside which executes the business analytical model.
* Developed dashboards and ad-hoc reports using **MS Power BI** and **SSRS** for senior management team for analysis.
* Hands on expertise in **SQL reporting**, development, analytics, or BI.
* Used **JIRA** for Version Control needs, task check-ins.
* Experience in Working **Agile Methodolog**
* Knowledge of **SQL, PL/SQL**, and scripting languages.
* Writing complex **Sql** queries using various analytical and aggregate functions.
* Created **stored** **procedures**, **functions**, **packages**, **triggers**, v**iews**, **cursors, collections** and handling **exceptions.**
* Worked on performance tuningusing **Explain plan, SQL hints, parallelism etc..**
* Work Experience on **ETL** Tools.
* Developed **ETL** pipelines using **Spark** to extract, transform, and load data from various sources into a centralized data warehouse.

**Environment:** Apache spark, pyspark, SQL, PLSQL, AWS (EC2, S3, EMR, RDS, Lambda, Kinesis, Redshift, Cloud Formation) Azure (Databricks, Datalake), ETL, Python (Pandas, Numpy), GIT (version control), Snowflake, MongoDB, Autosys.

**Tata Consultancy Services, Bangalore, India Mar2010 - Jan 2022**

**Roles: Lead Data Analyst, PLSQL Lead, PLSQL developer**

**Clients: Hoffmann-La Roche, Credit Suisse Group AG, Switzerland, Christian Mutualite Domains: Health care, Banking and Financial services, Insurance etc..**

**Responsibilities:**

* **Database Development and Management:** Designed and developed database management systems for seamless data access, storage, and retrieval. Played a pivotal role in the end-to-end development of applications PL/SQL, Oracle, and SQL for robust functionality and seamless database operations.
* **Creation of PLSQL objects**: Worked on creation and generation of reports using PLSQL packages, stored procedures**,** Functions, Cursors, Triggers, Views, Materialized views, Temporary tables.
* **Error Handling**: Worked on Exception handling debugging PL/SQL Code, DDL, DML and display the error messages in the application.
* Created **materialized** **views** and **indexes** for better performance of tables.
* Worked on **dynamic SQL** and **autonomous transactions.**
* Created **Cursors** to pull data from the Database.
* **Performance Tuning and Optimization**: Demonstrated expertise in fine-tuning complex scripts and strategically redesigned tables to eliminate bottlenecks in the system, resulting in enhanced overall system performance. Worked on query optimization/ tuning using **explain plan,** SQL **HINTS**, parallelism etc.
* Worked on **external tables, Partitions, SQL loader**.
* **Efficient Data Handling with Linux and SQL:** Demonstrated proficiency in Linux and SQL scripting, enabling seamless data extraction from diverse sources such as Excel and Flat files, contributing to efficient data management.
* Worked on code deployment.
* Travelled to **Switzerland** for requirement gathering.
* Developed and modified **SAS Macros** for generation of reports.
* **Design Architecture and Documentation**: Developed comprehensive application designs, including class diagrams and sequence diagrams, ensuring clarity and coherence in project execution.
* **Agile Methodology and SCRUM Integration:** Engaged in **Agile** practices, actively participating in daily SCRUM meetings, sprint planning, ensuring seamless alignment with team goals. Contributed to daily Scrum calls and timely deliverables. Leveraged **JIRA** as a tracking tool to streamline sprint progress.
* **Data Mapping and System Testing**: Spearheaded data mapping specifications and executed comprehensive system test plans to ensure accurate data extraction, transformation, and delivery to external entities.
* **ETL Solution Development**: Engineered custom-built ETL solutions, batch processing, and real-time data ingestion pipelines using Shell Scripting for seamless data movement within database.
* **Preproduction and Postproduction Support**: Provided critical support in both preproduction and postproduction phases, ensuring robust application performance and user satisfaction.
* Applications: Worked on various applications like **Arisg**, **agxchange**, **SMD**, **ROMA**, **CMS**, **Asset** **control** and **Sienna**.
* Worked for multiple clients across health care, Banking and Financial services, Media, Insurance with above said responsibilities.

**Environment:** Oracle, SQL, PL/SQL, Unix, Python, SAS, Sybase, SQL Developer, PL/SQL developer, TOAD, Putty, ServiceNow,

Jira, Git, Control-M