

PRATIK SHINDE - DATA ENGINEER

Edison, NJ US | (747) 356-1617 | pratikshindecareer@gmail.com

Professional Summary

4+Years of experience as a Data Engineer in design, development, deploying, and large-scale supporting large-scale distributed systems. Proficient in the Software Development Life Cycle (SDLC) including requirements analysis, design, Implementation, & testing of major systems. Good experience in developing Web Services like R, and Python programming languages. Well-Versed in using major components of Hadoop ecosystem components like Hadoop, MapReduce, Hive, Apache Spark, Pig, Sqoop. Proficiency in developing various Packages to extract, and load (ETL) data into data warehouses and data marts from heterogeneous sources.

Skills

- **Methodology:** SDLC, Agile, Waterfall
- **Languages:** Python, R, Scala, SQL
- **Frameworks:** Kafka, Airflow, Snowflake, Django, Docker
- **Packages:** NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow
- **Reporting Tools:** Tableau, Power BI, SSRS
- **ETL Tools:** SSIS, SSAS, Informatica PowerCenter, Azure Data Factory
- **Bigdata Eco System:** Hadoop, MapReduce, Hive, Apache Spark, Pig, Hive, HDFS
- **Application Server:** Apache Tomcat, Nginx
- **Databases:** MS SQL Server, PostgreSQL, MongoDB, MySQL
- **Build /Other tools:** Anaconda, Sitecore, CMS, Jira, Confluence, Jenkins, Git, MS Office, Jenkins
- **IDEs:** PyCharm, Jupyter Notebook, Visual Studio Code
- **Operating Systems:** Windows, Linux

Work Experience

Data Engineer

05/2023 to Current

Humana – CA, USA

- Using the Agile methodology to build the different Software development life cycle phases.
- Working with Amazon Web Services (AWS) using EC2 for computing and S3 as a storage mechanism.
- Demonstrated expertise in building scalable data pipelines with AWS services (Glue, Athena, Lambda, Redshift, EMR) for efficient data processing and informed decision-making.
- Leveraged Python-based tools like Apache Kafka or AWS Kinesis to handle real-time data streaming. Built data pipelines that process and analyze data as it arrives, enabling real-time insights and decision-making.
- Integrated Python-based logging libraries with AWS CloudWatch and other monitoring tools to keep track of the system's health and performance in real-time.
- Implemented data masking techniques in Snowflake to enhance data privacy and security, ensuring sensitive information remains protected and accessible only to authorized users.
- Implemented data integration solutions using SSAS and Informatica to support OLAP and OLTP databases, enhancing reporting and analytics capabilities and followed Agile Methodology for application Implementation.
- Implemented cloud-based ETL solutions, utilizing GCP Dataflow and AWS Lambda functions for real-time data processing, resulting in reduced data latency by 40%

Data Engineer

08/2018 to 07/2021

Hexaware Technologies – India

- Developed various Python scripts to find vulnerabilities with SQL Queries by doing SQL.
- Utilize Apache Spark to process and analyze large datasets, optimizing performance and resource utilization.
- Worked with Amazon Web Services (AWS) cloud infrastructure services and was involved in ETL, Data Integration, and Migration.
- Hands on experience in processing different file types such as Parquet, Avro, JSON, XML etc.
- Experience with building and deploying applications in AWS (S3, Glue, Dynamo DB, Redshift, CloudWatch, Lambda)
- Developed custom Dashboards and Visualizations in Excel and Power BI.
- Ability to work in a dynamic environment with some level of ambiguity to implement customer and project requirements with minimal supervision.
- Build continuous integration/deployment (CI/CD) pipelines to accelerate development and improve team agility.
- Implemented MapReduce programs to perform joins using secondary sorting and distributed cache.
- Worked with MongoDB and utilized NoSQL for non-relation data storage and retrieval.
- Utilized advanced analytics techniques, such as regression and clustering, to uncover insights from sales data, resulting in a 10% increase in marketing campaign effectiveness.

Education

Master of Science: Computer Science

05/2023

California State University, Northridge - Northridge

Bachelor of Engineering: Computer Engineering

05/2016

Savitribai Phule Pune University - Pune