Paras Wadekar

paraswadekards@gmail.com | +1 (703) 665-9743 | linkedin.com/in/paras-wadekar | www.paraswadekar.com

Professional Summary

8+ years of experience in **Data Engineering** and **Big Data Analytics**, with a strong focus on designing and implementing scalable data pipelines, enterprise data governance strategies, and cloud-based analytics solutions.

Expertise in **Hadoop ecosystem technologies** such as **Spark**, **Kafka**, **Hive**, **HDFS**, and **MapReduce**, delivering efficient batch processing and real-time streaming data solutions.

Proficient in **Azure cloud services**, including **Azure Data Factory**, **Databricks**, **Data Lake**, and **Synapse Analytics**, for orchestrating complex ETL workflows and managing large-scale data pipelines.

Hands-on experience with **AWS services** like **Glue**, **Redshift**, **Lambda**, **S3**, **Athena**, and **CloudWatch**, enabling seamless data integration and real-time analytics.

Strong skills in **SQL** and **SparkSQL** for querying, transforming, and optimizing datasets, ensuring high performance and scalability across distributed environments.

Skilled in **Python**, **Scala**, and **Shell scripting** for automating ETL pipelines, transforming data, and building APIs for seamless data workflows.

Proficient in **Infrastructure as Code (IaC)** using **Terraform** and containerization with **Docker**, ensuring reliable and consistent deployment of data solutions.

Extensive experience in building interactive **Power BI** and **Tableau** dashboards to provide actionable insights and drive business decision-making.

In-depth understanding of **data governance frameworks** and tools like **Collibra** and **Informatica**, ensuring compliance with organizational policies and regulatory standards such as GDPR.

Proven ability to manage CI/CD pipelines using **Git**, **Azure DevOps**, and **GitLab CI/CD**, enhancing development and deployment efficiency for data engineering workflows.

Experienced in **data modeling**, including Star and Snowflake schemas, and optimizing data warehouses for analytics and reporting.

Adept at collaborating with cross-functional teams and stakeholders to deliver robust, business-focused data engineering solutions.

Key Skills

Data Governance: Enterprise data governance, metadata management, compliance frameworks (GDPR, HIPAA).

Data Tools: Collibra, Informatica, Alation, Apache Atlas.

Big Data Technologies: Hadoop (HDFS, MapReduce), Apache Spark, Hive.

Cloud Platforms: AWS (Lambda, SageMaker, Redshift, Glue), Azure (Data Factory, Synapse, AI Studio).

Documentation & Communication: Technical and business documentation, training delivery, presentations.

Programming & Scripting: Python, SQL, Shell scripting.

Infrastructure & IaC: Terraform, Docker, Kubernetes.

Education

Drexel University — Philadelphia, PA
PhD in Computer Science (In progress)

Drexel University — Philadelphia, PA
Master of Science in Computer Science (GPA: 3.9/4.0)

COEP Technological University — Pune, India
Bachelor of Technology in Electronics & Telecommunication Engineering (GPA: 8.6/10.0)

Professional Experience

Amazon Web Services (AWS) Herndon, VA
Data Engineer II *Aug 2022 – Present*

- Designed and implemented data governance policies and standards, ensuring compliance with GDPR and internal regulations.
- Created metadata management solutions to improve accessibility for technical and business stakeholders.
- Developed AWS Glue-based ETL pipelines, reducing data processing times by 40%.
- Built scalable Redshift data warehouses for analytics, optimizing query performance and storage.
- Conducted training sessions on data governance frameworks, improving team alignment on best practices.
- Automated infrastructure deployment using Terraform, reducing manual errors and enhancing consistency.
- Optimized Lambda functions for real-time processing, ensuring data integrity across pipelines.
- Utilized CloudWatch for real-time monitoring and troubleshooting of AWS services and workflows.

Amazon Web Services (AWS) Herndon, VA
Data Engineer I *Jun 2020 – Aug 2022*

- Created and managed S3-based data lakes to centralize and organize diverse datasets for analytics.
- Leveraged Amazon Kinesis for real-time data streaming, ensuring timely insights for critical business operations.
- Used AWS EMR with PySpark to process large-scale data, enhancing performance for complex analytical workloads.
- Developed Python scripts to integrate data from multiple sources into AWS EMR, ensuring seamless data flow.
- Built scalable, cost-efficient data lakes on S3 and processed them with AWS Glue and PySpark for advanced analytics and machine learning.
- Collaborated with cross-functional teams to resolve AWS infrastructure issues, improving platform reliability.
- Designed and documented data pipelines to enable self-service analytics for non-technical teams.

Mobicule Technologies Pune, MH, India
Data Engineer *Jun 2013 – Jul 2015*

- Developed data governance frameworks and policies to standardize data management practices across the organization.
- Built interactive Tableau dashboards to provide stakeholders with actionable insights and KPI tracking.
- Automated ETL pipelines using SQL and Shell scripting, reducing manual intervention by 50%.
- Designed and implemented data quality assurance processes, enhancing data accuracy and consistency.
- Managed data ingestion workflows from diverse sources into HDFS for analysis and reporting.
- Supported the integration of new systems into existing infrastructure with minimal downtime.
- Conducted performance tuning for SQL queries, improving report generation speed by 30%.
- Provided technical documentation on data governance processes to ensure alignment across teams.
- Worked closely with the business analytics team to deliver solutions tailored to business needs.
- Enhanced Tableau reports with dynamic filtering and real-time updates for better decision-making.

Coriolis Technologies Pune, MH, India
Data Analyst *Jun 2011 – Jun 2013*

- Prepared and analyzed large datasets using SQL to derive insights and generate comprehensive reports.
- Created technical documentation to guide data users on best practices for data governance and analytics.
- Designed and maintained interactive dashboards using Tableau, improving data visualization.
- Conducted data quality checks and implemented workflows to standardize incoming datasets.
- Implemented process automation scripts in Python to streamline routine reporting tasks.
- Collaborated with stakeholders to define reporting requirements and ensure data-driven decision-making.
- Integrated diverse datasets from multiple sources into a centralized reporting framework.
- Applied statistical methods to identify trends and anomalies, supporting operational improvements.
- Reduced data processing times by 20% through workflow optimization and query enhancements.
- Provided actionable insights from data analysis that contributed to cost savings and efficiency improvements.

Projects

Azure DE Project | ***Azure Data Factory, Azure Databricks, Microsoft Power BI*** *Jan 2025*

- Built an end-to-end data engineering solution using Azure Data Factory and Databricks for seamless ETL workflows.
- Designed a metadata-driven architecture to automate data ingestion from multiple sources such as HTTP, Blob Storage, and Data Lake.
- Developed complex ADF pipelines for batch processing, error handling, and triggering workflows based on dependencies.
- Implemented Delta Lake architecture in Databricks to enable ACID transactions and improve data consistency.
- Optimized data transformation workflows using PySpark to handle large-scale datasets efficiently.
- Created Power BI dashboards with dynamic visualizations, enabling real-time analytics and key performance indicators.
- Ensured data security and compliance by integrating Azure Active Directory and role-based access controls.
- Conducted thorough performance tuning for Databricks notebooks, reducing processing time by 30%.

AI Psychologist | ***Python, Generative AI, API Integration*** *Dec 2024*

- Developed an AI-based journaling assistant powered by GPT models to encourage user engagement and self-reflection.
- Implemented React.js components with Tailwind CSS for an interactive, responsive, and visually appealing user interface.
- Built a real-time text analysis pipeline to detect sentiment and provide personalized journaling prompts.
- Created dynamic features like markdown support, theme toggling, and chat animations for enhanced user experience.
- Integrated cloud-based API services for real-time generative AI responses and sentiment analysis.
- Ensured scalability and reliability by deploying the application on AWS using Lambda and API Gateway.
- Conducted A/B testing on UI/UX features to optimize user engagement and retention.

Tableau to AWS QuickSight Migration | ***AWS QuickSight, Tableau, AWS Glue, Redshift*** *Oct 2023*

- Migrated the existing Tableau reporting and analytics environment to AWS QuickSight for improved scalability, cost-efficiency, and performance.
- Configured data pipelines to ensure real-time or scheduled data updates from source systems to QuickSight.
- Recreated Tableau dashboards in AWS QuickSight by leveraging its built-in features such as ML Insights, interactive visuals, and custom SQL support.
- Set up user groups and roles for data access and dashboard sharing based on organizational requirements using AWS IAM to manage user access and permissions securely.
- Implemented security best practices including encryption of data at rest (Amazon S3) and in transit (QuickSight), and ensure compliance with data regulations.
- Conducted thorough testing to ensure data accuracy, report functionality, and performance benchmarks.

AWS Data Pipeline Project | ***AWS Glue, Redshift, Lambda, Terraform*** *Jun 2022*

- Architected ETL pipelines with AWS Glue, enabling seamless data ingestion and transformation for Redshift.
- Automated real-time data validation workflows using Lambda, ensuring high-quality and accurate datasets.
- Developed partitioning strategies in Redshift for optimized query performance over large datasets.
- Integrated Terraform for Infrastructure as Code (IaC), reducing manual setup and enhancing deployment consistency.
- Designed scalable data lake architectures on S3 to manage structured and unstructured data effectively.
- Used CloudWatch metrics and logs for monitoring Glue jobs and troubleshooting ETL failures.
- Implemented AWS Key Management Service (KMS) to encrypt sensitive data and ensure compliance with security standards.
- Conducted regular unit testing for Glue scripts and Lambda functions to ensure accuracy in data workflows.

Technical Skills

Programming Languages: Python, Scala, SQL, Bash

Big Data Technologies: Apache Spark, Kafka, Hadoop (HDFS, YARN, MapReduce), Hive, Pig, Sqoop

Databases: SQL Server, MySQL, PostgreSQL, DynamoDB, Cosmos DB, HBase

ETL Tools: Azure Data Factory, AWS Glue, Talend, Informatica, Apache NiFi, Data Stage

Cloud Platforms: Azure, AWS

Data Visualization: Power BI, Tableau, Matplotlib, Seaborn

Machine Learning: TensorFlow, Scikit-learn, Spark MLlib

Version Control & CI/CD: Git, GitHub, GitLab CI/CD, Azure DevOps

Scripting & Automation: Python scripts for automation of AWS tasks and ETL processes

Data Modeling: Dimensional modeling, schema design (Star, Snowflake)

Publications

Geometric modeling of complex knitting stitches using a bicontinuous surface and its offsets in the journal Computer Aided Geometric Design, Elsevier, 2021.

An optimized yarn-level geometric model for Finite Element Analysis of weft-knitted fabrics in the journal Computer Aided Geometric Design, Elsevier, 2020.

Geometric modeling of knitted fabrics using helicoid scaffolds in the Journal of Engineered Fibers and Fabrics, SAGE Publishing, 2020.

Block Division based CAMShift Algorithm for Real-time Object Tracking using Distributed Smart Cameras in the International Symposium on Multimedia, 2013.

Certifications

AWS Certified Solutions Architect - Professional

AWS Certified Data Engineer Associate

Tableau Data Analyst

Microsoft Power BI Data Analyst Associate

AWS Certified Machine Learning Specialty