Overview: A highly skilled Data Engineer with 4 years of experience specializing in cloud data architecture and integration. Expertise in leveraging AWS cloud services to build scalable data pipelines, ensuring seamless integration with cloud-based ERP systems like Microsoft Dynamics, Salesforce, and Oracle Fusion. Adept at handling large-scale data processing, transformation, and storage solutions. Strong communicator and problem solver with a focus on delivering efficient, reliable, and high-performance data workflows that support business decision-making processes across departments.

Key Responsibilities:

- **Data Engineering & Cloud Architecture:** Designed and optimized scalable data pipelines in AWS, integrating ERP systems like Dynamics 365, Salesforce, and Oracle Fusion.
- **Python & PySpark Development:** Developed custom Python and PySpark scripts for automating data transformations and processing large datasets efficiently.
- **ETL Process Development:** Managed ETL processes to integrate data from ERP systems into AWS data lakes/warehouses, ensuring timely and accurate data flow.
- **AWS Cloud Services:** Utilized AWS services (S3, Redshift, Lambda, Glue) to store, process, and analyze data, optimizing for scalability and cost-efficiency.
- **Data Validation & Cleansing:** Built validation routines using Python/PySpark to ensure data accuracy, consistency, and integrity, addressing discrepancies proactively.
- **Data Modelling & Transformation:** Transformed raw data into structured formats using Python, PySpark, and AWS, optimizing models for fast querying and insights.
- **SQL & NoSQL Optimization:** Wrote and optimized complex SQL and NoSQL queries for efficient data extraction and transformation at scale.
- **Automation & Workflow Management:** Automated data workflows using Python and AWS Lambda, reducing manual intervention and minimizing errors.
- **Performance Monitoring & Troubleshooting:** Monitored and optimized data pipelines and cloud infrastructure to identify and resolve bottlenecks and improve performance.
- **Documentation & Reporting:** Maintained detailed documentation for data systems, ETL processes, and pipeline performance, providing updates to stakeholders.

Qualifications:

- Bachelor's degree in computer science, Information Technology, or a related field.
- 4+ years of experience in data engineering, focusing on cloud-based data architectures, programming in Python and PySpark, and ERP system integrations.
- AWS Cloud Expertise: Advanced knowledge of AWS services, including S3, Redshift, Lambda, Glue, RDS, and Athena for building and optimizing data pipelines.
- Python & PySpark: Extensive experience with Python and PySpark for building efficient data processing solutions, automating workflows, and processing large datasets.
- ERP System Integration: Demonstrated expertise in integrating cloud-based ERP systems such as Microsoft Dynamics 365, Salesforce, and Oracle Fusion with AWS data pipelines.
- ETL Process Development & Management: Proven ability to develop, maintain, and optimize robust ETL processes that handle large volumes of data efficiently.
- SQL & NoSQL Optimization: Proficient in writing optimized SQL queries for relational databases and working with NoSQL systems (such as MongoDB or Cassandra) for data processing and transformation.
- Data Quality Assurance & Validation: Strong experience in implementing data quality checks, validation, and cleansing routines to ensure high integrity in the data pipeline.
- Problem Solving & Optimization: Strong troubleshooting skills with the ability to identify and resolve data-related performance issues and optimize workflows for better scalability and speed.
- Project Management & Leadership: Ability to manage multiple data engineering projects, ensuring that all tasks are completed on time, within scope, and according to requirements.

Skills:

- AWS Cloud Services (S3, Redshift, Lambda, Glue, RDS, Athena)
- Python & PySpark (Data Transformation, Big Data Processing, Automation)
- ETL Process Development & Optimization
- ERP System Integration (Microsoft Dynamics 365, Salesforce, Oracle Fusion)
- Data Pipeline Development & Workflow Automation
- SQL & NoSQL Query Optimization
- Data Modelling & Data Transformation
- Data Validation, Cleansing & Quality Assurance
- Performance Monitoring & Troubleshooting

- Project Management & Documentation