

Vamsi Kilaru

Location: Sunnyvale, CA

LinkedIn profile Link: <https://www.linkedin.com/in/vamsikilaru/>

Summary:

- Over 10+ years of professional IT experience, specializing in **Python**, Java, SnapLogic, Pentaho Suite, Salesforce, Big Data, AWS, and Machine Learning.
- Proven expertise in developing end-to-end business application data pipelines using cloud-based technologies and management tools like Salesforce, AWS, and Azure.
- Skilled in designing data architecture for high-tech infrastructure sales and subscription models.
- Proficient in creating insightful reports and dashboards utilizing visualization tools such as Tableau, Power BI, and Pentaho.
- Published in the Journal of Electronic Imaging (J. Electronic Imaging 24(1):013028, 2015).
- Holds a Master's degree with a strong academic record and exceptional Communication Skills

Data Warehouse Modernization | Data Infrastructure Modernization | Cloud-Based Data Platform | Data Analysis
Data Warehouse Migration | Data Modeling | Data Pipeline Architecture, Design & Optimization | Data Strategy
Cross-functional Collaboration | Data Quality | **Data Governance** | Data Security | Business Insight Generation
Process Improvement | Change Data Capture | Data Engineering Processes | Project Management | **ETL/ELT Processes**
Performance Tuning | Data Transformation | Cloud Migration | Data Quality Management | **Adobe Analytics** | **MarTech**

Technical skills:

Programming: Python, Java, SQL, MySQL, PostgreSQL
Data Engineering: Redshift, Snowflake, HDFS, MapReduce, Hive, Pig, HBase, Sqoop, Flume, Kafka, Airflow, Oozie
Cloud: AWS, Azure, Google Cloud Platform (GCP)
Data Visualization: Tableau, Power BI, Pentaho
Salesforce Development: Workflows, data migration, data cleansing
Big Data Technologies: Spark, Google Big Query (GBQ)
Other: MS-SQL Server, Oracle, DB2, Reltio

Professional Experience:

Workday Inc, Pleasanton, CA

Nov 2021 - Present

Data Engineer

- Improved throughput by 250% for outbound events by building efficient data management processes between master data management system Reltio and Salesforce platform, significantly enhancing data flow.
- Built federated data pipelines with unstructured data for MDM system, enabling robust data governance capabilities for over 8 million contacts, facilitating better data management and quality control.
- Enabled reporting and analytics for business users through Tableau, building self-service capabilities with cloud data sources such as Databricks, **Google BigQuery**, and Amazon Redshift, improving data accessibility and usability.
- Creating pipelines, data flows and complex data transformations and manipulations using ADF and PySpark with Databricks.
- Implemented a pub/sub model using Kafka for change event processing within pipelines, optimizing real-time data processing.
- Conducted ad analytics to evaluate the effectiveness of advertising strategies, leveraging **data from Adobe Analytics** and other marketing platforms.
- Applied advanced **customer segmentation** techniques using marketing data pipelines to enhance personalization and improve campaign outcomes.
- Created, provisioned different Databricks clusters, notebooks, jobs and autoscaling.
- Designed, constructed, and maintained scalable data pipelines for data ingestion, cleaning, and processing using Python and SQL, with the Databricks platform.
- Created Scheduled data analysis notebooks in databricks which gets fed into table Tableau for reporting and analytics

- Monitored and tracked campaign performance metrics, utilizing Adobe Analytics to deliver detailed and accurate reports to stakeholders
- Collaborated with data scientists, ML engineers, business, and IT stakeholders to define and enforce data governance policies, standards, and business analytics, leading to more cohesive data practices.
- Involved in designing lead scoring algorithm implemented in Databricks platform using Reltio and Salesforce data and **Adobe analytics**

Hitachi Vantara, Santa Clara, CA

April 2019 - Oct 2021

Data Management Lead

- Automated and delivered the ETL pipeline for global services data creation that resulted in improving quoting team efficiency by 83%
- Developed Python scripts for extracting data from web services API's and loading into databases.
- Prepared scripts to automate the ingestion process using Python and Scala as needed through various sources such as API, AWS S3, Teradata and snowflake
- Designed and Developed Spark workflows using Scala for data pull from AWS S3 bucket and Snowflake applying transformations on it.
- Created several Databricks Spark jobs with PySpark to perform several tables to table operations.
- Created, provisioned multiple Databricks clusters needed for batch and continuous streaming data processing and installed the required libraries for the clusters.
- Developed and managed all the data transformation processes related to asset and contract information across the organization used for business transactions and forecasting using Databricks platform
- Designed reports and dashboards for the business using Power BI resulted in 30% improvement in data quality for the data transformation done in databricks platform
- Worked on Salesforce development activities such as workflows rules

GSK Solutions, CA/TX

May 2015 - Dec 2018

Associate

- Worked on projects for five clients developing data pipelines with a wide range of technologies such as Pentaho Suite, Big Data Ecosystem, Java and AWS in domains related to retail, tech, aviation
- Migrated data with 4B opportunity value from Siebel to Salesforce platform
- Developed various data cleansing applications using pentaho data integration which improved the data quality by over 40%
- Built 20+ reports and dashboards using pentaho for the business to monitor ETL processes for error handling automation processes

Key Business Solutions, Sunnyvale, CA

Oct 2014 – March 2015

Data Analyst

- Part of Testing team to test HDS EscalFS which was built based on Hadoop Hortonworks platform
- Tested hundreds of jobs written in python for various functionalities of Hadoop eco system components such as HDFS, MapReduce, Hive, HBase, Zookeeper, Sqoop, Flume
- Analyzed the test results and worked with Hortonworks support team
- Documented all the test results for Management and Reporting

Education: Master of Science, Electrical Engineering Villanova University, Villanova, PA 12/2014