PRABHU BALAN JOSEPH

LEAD DATA ENGINEER

Lead Data Engineer with 14 plus years of experience with a proven track record in Architecting and implementing robust data solutions. Specialized in Apache Spark, Scala, Python, Tableau and a diverse range of Cloud technologies (Google & AWS). Adept at leading cross-functional teams and driving innovative data strategies to meet business objectives. Having Expertise knowledge in Insurance & Retail domain.

Informatica / Talend Certified Developer
AWS & Google Certified

📱 860-709-5688

✉ prabhubalanjoseph@outlook.com

in Bentonville, AR

📍 Linkedin.com/in/prabhubalan joseph

EDUCATION

B.Tech (IT)
Anna University, Chennai
2005 – 2009

EXPERTISE

TECHNICAL

- Spark, Scala & PL/SQL
- Talend, Informatica and Nifi
- BigData, Delta Lake, HUDI, Presto, Snowflake, BigQuery & Oracle
- Google & AWS Cloud
- Cloud break, CloudSQL, GCS, S3, EC2.
- Autosys, Airflow & Automic
- JIRA, Confluence, Rally.
- CICD (Looper & Concord)
- Tableau & Looker

PROFESSIONAL

- Collaboration
- Adaptability
- Problem Solving
- Strategic Thinking
- People Skills

WORK EXPERIENCE

Lead/Sr Data Engineer
Walmart, Texas | Jan 2020 – Present

- Design & Built various Spark/Scala pipelines for Data commercialization product of Walmart Data ventures data asset that helps Merchant and Suppliers community to take better decisions on item inventory and provide better Customer behavior insights and Product performance analysis into actions.
- It helps the business to utilize the Walmart data asset to generate new profitable revenue of 200 million plus every year.
- Implemented the Data Quality rules framework using Pyspark to calculate the key metrics and publish in Data Governance tool Collibra for Data owners and Business team.
- Lead the implementation of CCPA, CPRA and MHMD(My Health) privacy laws into Data engineering solutions for the application to stay be complaint.
- Lead the Data engineering team of 8+ people and working closely with Product team and Business team on Roadmap, LOE, data requirements and come-up with Architecture solution for new initiatives.
- Helping the Business team on re-writing any existing and New Third party data assets like Dun & Bradstreet, Nielson across the Global data teams and negotiate contracts by creating centralized 3P Data asset for Walmart teams for cost reductions etc.
- Building the functionality that helps the Data science team on building the models on top of Third-party data assets and Data commercialization projects.
- Expertise experience in Agile methodology and facilitates the scrum ceremonies like Sprint planning, backlogs etc.
- Maintaining the CICD process and Approval of code migration/review process end to end.
- Experienced in handling/coordinating between teams across multi-region efficiently.

- Experienced in handling multi-tera bytes of Data processing and transferring to external third parties via Google Cloud.
- Ownership of taking care end-to-end pipeline executions in Production and coordinating with Upstream teams on any data quality issues etc..

Module Delivery Lead (American Insurance Group)
Mphasis Corporation, Connecticut | Oct 2019 – Dec 2019

- Leading the team on building Life information factory platform for Data management and Governance process to handling the policies information and used by Actuarial system and finance team.
- Built Talend Bigdata jobs for Data standardization service by transforming data according to the Business rules and loading to Clean Lake (Amazon EMR).
- Developing Talend jobs for Data Acquisition services by pulling the flat files, fixed width separated files and XML's from S3 bucket to Murky Lake environment (Amazon EMR)
- Implemented the Data quality framework as part of data pipeline automation which saved manual data validations.

Sr Talend/Informatica Lead (The Hartford Insurance)
Cognizant, Connecticut | Mar 2010 – Oct 2019

- Built end to end Data pipelines using Informatica and Talend tool for Group Benefits and Commercial domain and store the data in On-Prem and Amazon S3 for downstream consumption.
- Hands on experience on Informatica 10.2 BigData Management tool to ingest Data into Hadoop platform.
- Build and maintaining the Master metadata using Informatica MDM tool for data lineage and reporting purpose.
- Building the Talend Ingestion framework to extract the Clickstream Data from BigQuery and store in Hadoop environment for Data science team modeling.
- Involved in building Nifi process to convert JSON files into AVRO format in Hadoop environment for Data Science team consumption.
- Leading the team of 10+ members (Onshore/Offshore) and coordinating with Application owners on new initiatives etc.
- Extensively used Talend components tMap, tDie, tConvertType, tFlowMeter, tLogCatcher, tRowGenerator, tOracleInput, tOracleOutput, tFileList , tReplicate, tSortrow, tDenormalize, tNormalize etc
- Creating resuable Audit framework for the Data commercial domain using Talend tool using context and global variables.
- Played active role in carrying out source data analysis, volumetric analysis, performance testing of the Datamart.
- Actively participating in the Tableau dashboard development along with the business team for data gathering and profiling.
- Handling Maintenance and Production support teams in Onshore and responsible for Application availability etc.
- Expertise in Ticketing tools like Rally, JIRA and creating knowledge sharing documents for L1 support team.
- Lead the Conversion project of Informatica to Talend tool and converting 1000+ mappings to Talend jobs in Bigdata which reduced the Licensing cost to Business and decommission of separate compute for Informatica tool.