

# SIBY ABIN THOMAS

## Senior Data Engineer

Kerala, India | [www.sibyabin.tech](http://www.sibyabin.tech) | +91 984-764-5476 | sibyabin@gmail.com

---

Customer-focused data engineer with experience in delivering cloud-based data lake projects with a focus on data quality, maintainability, and operational costs.

- 13 years of IT expertise, with a strong background in creating cloud-based data products.
- Crafted high-performance data pipelines with Apache Spark.
- Implemented idempotent and manageable workflows using Apache Airflow.
- Delivered reusable data processing solutions built on an open source technology stack.
- Proficient in AWS data services, with greater expertise in Spark, AWS, Apache Airflow, Python, and SQL.
- Worked in data teams of Fortune 500 companies from onsite & remote.

## SKILLS

**Big Data:** Apache Spark, HDFS, Pig

**AWS Services:** S3, Athena, Glue, EMR, RDS, EC2,

Data Pipeline, Step Functions, Redshift,

DMS, Lambda, Cloudwatch

**Databases:** MySQL, Redshift, Snowflake

**Other Technologies:** Pandas, boto3, Terraform,

Databricks, Deltalake, Dremio, DBT

**Orchestration:** Apache Airflow, Control-M

**Programming Languages:** Python

**Version Control:** Github, Bitbucket, Code commit

**Cloud:** AWS

**Scripting:** Bash

**ETL:** Pentaho

**Agile Tools:** Confluence, Jira

## WORK EXPERIENCE | PROJECTS

**SENIOR DATA ENGINEER - TATA CONSULTANCY SERVICES, KOCHI, AUG 2021 - PRESENT**

**Project : Data platform product development and support for a leading financial services company in UK**

- Authored data lake pipelines using Apache Spark in AWS Cloud.
- Collaborated with data science teams to improve model accuracy by implementing additional data quality checks using Python, pandas, Jupyter and Apache Airflow.
- Increased the efficiency of support services by building a dashboard on top of audit information.

**SENIOR DATA ENGINEER - INFOSYS, TRIVANDRUM, MAR 2013 - JULY 2021**

**Project 1: Deltalake development for a leading fashion brand in USA (Feb 2021 - July 2021)**

- Prepared the Bronze, Silver, and Gold layer delta tables for customer data from Segment processed using Pyspark.
- Streamlined the development workflow in the data science team by helping them orchestrate the notebooks using Apache Airflow, which resulted in eliminating the need for manual interventions, enhancing data quality, and improving reuse of code.
- Automated the spark code deployments to Databricks using python packaging and wheel files.

**Project 2: Re-architecture & cloud migration of Data platform for major US client (Jan 2019 - Jan 2021)**

- Crafted over 100 spark jobs using Pyspark to perform complex data transformations between ingestion and the refined layer (S3).
- Reduced cloud operational costs by 30% by moving workloads to transient clusters and implementing best practices for spark memory utilization.
- Created AWS CLI-based utilities to scale the EMR clusters up or down, create glue tables, and run crawlers.
- Improved operational efficiency through the use of dynamic airflow dags to orchestrate spark jobs.
- Managed 3 engineers' deliveries and mentored them in data engineering, Airflow, and Apache Spark.

**Project 3: Data Product development for a leading Loyalty Analytics company in UK ( Jun 2016 - Dec 2018)**

- Built a reusable ETL framework based on metadata ingestion that allowed the client to make data processing without having an in-depth knowledge of Pentaho, saving the customer 40% on development costs.
- Enhanced data product sales by building integrations with campaign management products from IBM.

- Migrated historical loyalty & campaign data to Redshift for a renowned Australian bank using AWS DMS.
- Supervised data platform releases and wrote the framework documentation.
- Received STAR award from client for individual contribution to Data Platform design and architecture.

#### **Project 4: Datawarehouse modernization for a leading US retail client ( Apr 2013 - May 2016)**

- Gathered requirements, designed, and designed highly customizable and maintainable ETL solution using an open technology stack (Pentaho/MySQL/Python).
- Rearchitected analytical solution with open source technologies, assisting the organization in lowering operational/licensing costs.
- Delivered dashboards and reports with actionable key performance indicators(KPI) to stakeholders, assisting them in making decisions to drive profitability for the business.
- Trained and grew both vendor and client offshore teams on Pentaho

#### **SOFTWARE ENGINEER - SATMETRIX , BANGALORE, SEP 2010 - MAR 2013**

- Accelerated customer implementations by 3x using tools/plugins created in python and automating the workflows using them.
- Contributed to organizations talent pool by recruiting and mentoring interns.
- Delivered NPS program for key customers, ING & Experian.

#### **SOFTWARE ENGINEER - RYANCOM INFOTECH, BANGALORE, APRIL 2009 - SEP 2010**

- Developed several modules for the CRM product using VB6 and managed 100+ client implementations, support and customer training .
- Developed a POS (Point of Sale) product in 40 days using C#.net, DevExpress tools and MSSQL.

#### **OTHER PROJECTS**

- Built a personal technical blog (<https://blogs.sibyabin.tech>) using GitHub pages and Jekyll.
- Delta lake website (<https://github.com/delta-io/website>) contributor.

#### **CERTIFICATIONS**

- **AWS Certified Solution Architect - Associate:** Amazon Web Services, Dec/2022
- **Databricks Certified Associate Developer for Apache Spark 3.0:** Databricks, August/2022
- **Databricks Lakehouse Fundamentals:** Databricks, July/2022
- **Apache Airflow DAG Authoring:** Astronomer, June/2022
- **Amazon Web Services Cloud Practitioner:** Amazon Web Services, May/2021
- **Apache Airflow Fundamentals:** Astronomer, March/2021

#### **AWARDS**

- **Most Valuable Player:** Awarded 3 times at Infosys for excellence in project delivery.
- **STAR Award - 2018:** Awarded by client AIMIA, considering the individual contribution to its global data platform design & architecture.
- **Customer Focus Award:** Awarded by Satmetrix in 2011 for successful implementations of their key customers ING & Standard Life in a short span and making them valuable customers.

#### **EDUCATION**

- **Master of Computer Application** - Bharathiar Univesity, India with 83.5%
  - **Bachelor of Science** - M.G University, India with 85.4 %
-