JOHN FRANCIS

Data Engineer | Databricks Certified | ETL | AWS | AZURE|
Datalakes| Lakehouse| Spark | SQL | Python | Databricks

📞 +91-9488835026   @ johnfrancis95815@gmail.com
🔗 https://www.linkedin.com/in/john-francis-526999148/   📍 India

SUMMARY

I'm a Data Engineer who would like to pursue a highly rewarding career, seeking for a job in challenging and healthy work environment where I can utilize my skills and knowledge efficiently for organizational growth.
Programming Languages: Python, JavaScript, SQL, HTML, CSS
Tools & Frameworks: Pyspark, Spark, Databricks, Delta Live Tables, Unity Catalog, AWS Glue, AWS S3, Azure Data Factory, Azure Storage Gen2, Fivetran, NumPy, Pandas, AWS ,Docker, GIT
Experienced Data Engineer: I have been in the core development team of more than 7 projects as a Data Engineer
Experienced Full stack developer: I have been in the core development team of 4 projects as a full stack developer

EXPERIENCE

Data Engineer
03/2024 - Present

BatchService
Tempe, AZ (Remote)

- Optimized ETL processes for US real estate data, handling over 100 million records per batch
- Focused on cost-efficiency and delivery speed
- Implemented complex business logic to ensure the most reliable and up-to-date data for customers and data consumers

Senior Data Engineer
10/2023 - 03/2024

Loyalytics AI
Bengaluru, India

GMG Retail (UAE) – Led the design and implementation of a robust data lake architecture for GMG, a global FMCG leader. By integrating data from diverse sources such as SAP, SFTP, and REST APIs, we created a centralized, real-time data platform leveraging AWS Glue and Databricks. This initiative streamlined data access and reporting, enabling better decision-making and operational efficiency across the organization.

Work-Management Tool – Loyalytics Internal Project - Spearheaded the development of an internal ETL pipeline to consolidate data from various sources like GitHub and ClickUp, utilizing Azure Databricks and Fivetran. This solution enabled the team to monitor sprint metrics in real-time, and automated Slack notifications to keep teams aligned and responsive to critical tasks. The tool significantly improved workflow visibility and accelerated project timelines.

- Led the design and implementation of a robust data lake architecture for GMG, a global FMCG leader.
- Integrated data from diverse sources such as SAP, SFTP, and REST APIs to create a centralized, real-time data platform leveraging AWS Glue and Databricks.
- Streamlined data access and reporting, enabling better decision-making and operational efficiency across the organization.
- Spearheaded the development of an internal ETL pipeline to consolidate data from various sources like GitHub and ClickUp, utilizing Azure Databricks and Fivetran.
- Enabled the team to monitor sprint metrics in real-time, and automated Slack notifications to keep teams aligned and responsive to critical tasks.
- Significantly improved workflow visibility and accelerated project timelines.

KEY ACHIEVEMENTS

💎 **ETL Optimization Success**
Reduced ETL processing time by 50% through optimization techniques.

💎 **Data Processing Efficiency**
Achieved a 30% cost reduction in data processing using cloud solutions.

💎 **Project Acceleration**
Implemented a data framework boosting project timelines by 20%.

💎 **Robust Data Management**
Managed a data lake architecture supporting over 20 TB of data.

EDUCATION, SUPPLEMETARY COURSES & CERTIFICATIONS

Bachelors Degree in Engineering

Databricks Data Engineering Associate Certification

Data Science Boot-Camp, Guvi, Chennai

Full Stack Development Boot-Camp, Guvi, Chennai

EXPERIENCE

Product Data Engineer
04/2023 - 10/2023

Loyalytics AI
Bengaluru, India

Swan AI – Lead Data Engineer, Cenomi – Data Engineer

- Led the ETL setup for unstructured retail business data across all Loyalytics clients, building a robust data lake architecture using Delta Live Tables and Apache Kafka streaming services.
- Developed and self-managed a scalable, reusable data lifecycle framework that ingests a variety of file types (XML, JSON, CSV, etc.) and integrates seamlessly with Unity Catalog and Delta Live Tables.
- The framework supports data streaming, automated quality checks, and governance, simplifying ETL setup for developers with minimal coding.
- This tool was used in high-profile projects like GMG and Cenomi, improving operational efficiency and reducing time-to-market for data processing.
- Environment isolation and parallel processing ensure scalability across different environments, enabling smoother transitions from development to production.
- Worked on setting up ETL pipelines for diverse data sources, including text files, zip files, JSON, XML, CSV, web scraping, Google Sheets, and Oracle Fusion REST API.
- By leveraging Delta Live Tables on Azure Databricks and Azure Data Factory, contributed to building a robust data lake architecture that enabled Cenomi to gain a unified view of their retail and shopping center operations.

Data Engineer
08/2022 - 04/2023

Loyalytics AI
Bengaluru, India

- Led the ETL setup for diverse data sources like MySQL, Oracle DB, SQL Server, and Oracle Fusion for Panda Retail.
- Migrated from Oracle Data Integrator to Azure Databricks and Azure Data Factory, optimizing data processing workflows and improving real-time analytics capabilities.
- Oversaw the setup of ETL pipelines for SQL Server and Microsoft SharePoint data for Urban Foods LLC.
- Established a robust data lake architecture using Azure Data Factory and Azure Databricks, enabling Urban Foods to consolidate critical business data and streamline reporting processes.
- Contributed to the ETL setup for unstructured retail business data for LuLu Group International.
- Transformed large XML-format POS data from SFTP into structured Delta Tables and pushed it to Kafka streams, helping LuLu Group build a scalable data pipeline.

Software Engineer
09/2020 - 09/2021

LiftOff LLC
Bengaluru, India

Various full stack and backend development projects focused on creating robust, user-friendly web applications.

- Contributed to both frontend and backend development for OkayHumans, a mental healthcare platform.
- Built on the Node.js framework (Hapi.js), implemented SQL transactions to ensure integrity in the booking flow and developed a robust membership system with failure refunds integrated via Stripe.
- Leveraged AWS services, CI/CD pipelines with GitHub Actions, and AWS Elastic Beanstalk to deploy the platform efficiently.
- Developed new APIs and features for KicksNation, a personalized sneaker styling service.
- Implemented a fraud detection system to identify and blacklist fraudulent users, helping safeguard business transactions.
- Integrated new sneaker brands into the existing database to expand product offerings and improve customer personalization.
- Managed the development of APIs and features for ThinkingNation, a US-based Ed-Tech business.
- Created a new submission flow, optimizing user experience and updating automated systems for invalid character mapping.

SKILLS

Medallion Architecture ·

Data Manipulation ·

Knowledge Sharing ·

Troubleshooting ·

Creative Problem Solving · AWS Glue ·

Databricks · Delta tables · DLT ·

Delta live tables ·

Apache Spark Streaming ·

Role-Based Access Control (RBAC) ·

Big Data · Azure Databricks ·

Microsoft SQL Server · Data Lakes ·

Extract, Transform, Load (ETL) ·

HTML ·

Object-Relational Mapping (ORM) ·

JSON · Quality Control ·

Quality Assurance ·

Amazon Web Services (AWS) ·

Amazon EC2 · Amazon S3 ·

Amazon Relational Database Service (RDS) ·

AWS Lambda · REST APIs · CSS ·

PySpark · Microsoft Azure ·

Azure Data Factory · JavaScript ·

Node.js · Python · MongoDB ·

Full-Stack Development · HTML5 ·

SQL · Pandas · Jupyter · NumPy ·

PostgreSQL