Data Engineer

Arpit Humane

Contact Info

E-mail : arpit.humane@gmail.com

Location: Tampa, FL

Professional Profile

Core Skill

- **Data & Cloud Platforms**: Databricks (Delta Lake, Unity Catalog, Workflows, Mosaic AI), AWS(S3, EC2), Apache Iceberg, Trino, RedHat OpenShift, Docker
- **Languages & Frameworks**: Python (PySpark), SQL, PL/SQL, Java, Apache Camel, Spring Boot, UNIX/Linux Shell Scripting
- **Generative & Agentic AI**: Multi-Agent Orchestration, Tool Calling, RAG (Retrieval-Augmented Generation), LLM Integration
- **Enterprise ETL & Databases**: Ab Initio (GDE, Express>it, Conduct>it, EME), Oracle, Relational Data Warehousing

Executive Summary

- Data Engineer with 10 years of experience designing enterprise data platforms, ETL pipelines and lakehouses. Hands-on experience building Databricks solutions (Delta Lake, Unity Catalog, Workflows) and open lakehouse architectures (Apache Iceberg, Trino), plus generative AI applications with multi-agent orchestration, tool calling and RAG. Strong ETL and SQL background, combined with Python-based data engineering and cloud orchestration.

Certifications

Ab Initio Graph Developer Certification

Professional Highlights

Dec 2022 till Present

Role: *Specialist - Data Engineer*
Company: LTIMindtree
Project: Global Trading Profit & Loss
Location: Tampa, FL

Project Description:
GTPL is a product processor that evaluates Profit & Loss for global trades. The project involved receiving real time data globally from Tibco queues and performs calculations based on business logic. These calculated transactions are further published in reports and posted to ledger database and downstream systems.
Extended with a Databricks lakehouse platform and an AI assistant for business-user trade queries and reconciliation.

Responsibilities:

- **Requirement Gathering -**
  > Worked with BAs to convert business requests into technical solutions and enhance BRDs/FRDs.
- **Developer -**
  > Built a Databricks lakehouse (Delta Lake, medallion bronze/silver/gold layers) for trade, trade-lot and ledger data using PySpark and SQL, with Iceberg tables queried through Trino for interoperability.
  > Developed ingestion and transformation pipelines for P&L reporting, with data quality checks and trade/lot balance reconciliation against Prod.
  > Orchestrated pipelines with Databricks Workflows and secured data with Unity Catalog access controls.
  > Tuned Spark jobs and Delta tables for performance and cost (partitioning, file compaction, cluster sizing).
  > Built a multi-agent AI assistant on Databricks, with tool calling and RAG over P&L documentation, so business users can query balances and reconcile in natural language.
  > Maintained Ab Initio batch and continuous flows and migrated legacy Unix scripts to Ab Initio.
  > Migrated Tibco BW to Apache Camel and built the CI/CD pipeline for containerized (Docker) deployment on OpenShift.
- **Leadership -**
  > Led lakehouse design and supported the team on technical roadblocks.

Environment: Databricks, Delta Lake, Apache Iceberg, Trino, AWS, PySpark, Unity Catalog, Workflows, SQL, Oracle, PL/SQL, LLM/RAG, Ab Initio, Unix, Java, Apache Camel, Redhat Openshift Dockers, Spring Boot.

May 2019 – Dec 2022

Role: *Consultant*
Company: Exusia Inc
Project: Rating Content Distribution

Project Description:

- RCD is a migration project to transition mainframe legacy system to Ab Initio Software.
- The project involved unloading the data from various databases, create mapping and transformation rules on the data and stag the data in ICFF files, further the project heavily relies on continuous flows utility to identify the transitioned components and route the user request accordingly.

Responsibilities:

- **Requirement gathering**- worked closely with Client to gather complex requirements and translate them into high quality code.
- **Developer**-
  > Designed & developed continuous services to deliver data to the clients in real-time.
  > Created generic metaprogramming code to convert mapped data to staging area for real-time data delivery.
  > Created various automated modules for integration & regression testing, component extraction, ad-hoc reporting processes for Clients and published them with comprehensive User Guides.
  > Worked independently on modules with none to minimum oversight. Took end-to-end ownership of assigned module right from requirement gathering, design, development, testing to prod support.
- **Leadership**-
  > Managed new joiners; mentored & guided to become valuable and responsible employees of the organization. Conducted performance improvement program for an Analyst to help them align with organization goals.
  > As part of the Client project, I have helped put together a KT & training plan for Clients. And guided team members as well as client developers on various development & utility efforts.

Environment: Ab Initio, Continuous Flows, Unix, SQL, Oracle, Conduct>it, SQL.

Nov 2018 – May 2019

Role: Senior Analyst
Company: Exusia Inc
Project: ProvX Framework, Multiple Projects.

Project Description:

- The framework is created to enable the business users to create ETL processes using user friendly Express>It UI.
- The framework is divided in various step to mimic the standard ETL process.
- The framework provides functionality to extract/load using multiple types of sources/targets- flat file, DB, Queue, CSV, Excel, AWS S3 etc.
- Various transform operations can be performed, CDC, lookups, joins, filters, removing duplicated etc.

Responsibilities:

- **Developer -**
  > Helped enhance the existing project to accommodate SCD type II. Created generic graphs and Express>it templates to unload & load data from/to new sources such as new DBs & Amazon S3.
  > Created python scripts to automate the configuration creation process. Used advanced components in Express>it to make the existing template generic for code maintenance.
- **Trainer -**
  > Continued to train new joiners on Ab Initio hands-on.
  > Provided training to use ProvX framework to company Employees, conducted theory as well lab sessions.

Environment: Ab-Initio, Unix, Express>it, Conduct>it, Oracle SQL Developer, Python.

Jan 2018 – Oct 2018

Role: Senior Analyst
Company: Exusia Inc
Project: Touch-Point Architecture Project.

Project Description:

- The primary object is to differentiate OptumRx through our unique ability to engage members at every touch point in the member journey.
- The process that will ingest opportunities from multiple sources(engines), append (If needed) the necessary individual/member identifiers and transform into common format for ingestion by C360 for generating the right offer.
- In addition, all opportunities will be persisted in the opportunities master database to enable reporting and data mining on the opportunities.

Responsibilities:

- **Developer** -
  > Worked right from the very initial stages of the project. Helped create project structure.
  > Created the initial generic graphs & plans using metaprogramming. Helped onboard new project members.
  > Created shell scripts for files archival process and other unix specific activities.
- **Trainer** -
  > I was part of the new-joiner Training program. Provided a training sessions on Unix.
  > Conducted hands on sessions for Ab Initio & SQL to entry level employees

Environment: Ab-Initio, Unix, Express>it, Conduct>it, SQL Developer.

Sept 2016 – Dec 2017

Role: Analyst
Company: Exusia Inc
Project: Low Cost Opportunity Project.

Project Description:

- Low Cost Opportunity (LCO) is a project to provide low cost alternative to patients. It is divided into LCA (Low Cost Alternative) and RMO (Retail to Mail Order).
- LCA is an initiative to inform the alternative drugs to the prescribed drugs based on lower cost whereas RMO provides an option to patients to opt for mail delivery of drugs in bulk in turn saving costs on drugs.
- The project was developed from scratch.

Responsibilities:

- **Trainee** - Underwent the trainings on Ab Initio, SQL & Unix. Quickly learned the basic concepts by actively participating the training sessions.
- **Developer** -
  > Created graphs, plans & Unix scripts from scratch.
  > Unit tested the code along with comprehensive documentation. Performed Code migration.
  > Provided developer support for Production issues.

Environment: Ab Initio, Unix, Putty, Express>it, Conduct>it, SQL Developer.

Education Summary

Diploma

Diploma in Advance Software Development Methodologies (2016),
Centre for Development of Advanced Computing.

Degree

Bachelor of Technology(B.Tech) in Chemical Engineering (2015),
National Institute of Technology - Trichy, India.