

Linkedin:ahmedamron  
Github: ahmedamron  
UVa: AhmedAmrOn

# Ahmed Amr Abdulfattah

+201155128882  
ahmedamron@gmail.com  
Open to Relocation

## Summary

a Staff Data Engineer with 12+ years of coding experience, and 8+ years of professional experience, significantly contributed to over 10 major initiatives focused on developing robust data infrastructures for global enterprise products in the US, UK, and Europe.

My experience allowed me to excel in creating scalable and efficient data architectures, directly impacting the operational capabilities of firms like EnergyAspects, Incorta, and InsideTrack. Among my achievements, centralizing org-wide analytical models, architecting a comprehensive real-time data infrastructure for a distributed social network, and building scalable data providing APIs stand out as testaments to my ability to innovate and implementing software solutions.

Seeking a position in a global enterprise tech organization where I can leverage my proven track record in software/data engineering to foster significant growth, further develop my leadership and technical skills, and contribute to groundbreaking projects that resonate on a global scale

## Skills

- **Programming Languages** Python, Java, JavaScript, Haskell, and C
- **Databases** PostgreSQL, Shooju, Redis, MongoDB, Memgraph, AWS Document DB, Pinecone
- **Data Warehouses** Snowflake, BigQuery, AWS Redshift
- **Google Cloud (GCP)** Artifact Registry, BigQuery, Build, Cloud Functions, Cloud Run, Composer, Compute, Data Studio, Firestore, Kubernetes Engine, Pub/Sub, Secret Manager, SQL, Storage
- **Amazon Web Services (AWS)** Athena, CloudWatch, Database Migration, Document DB, EC2, ElastiCache, Forecast, Glue, Glue DataBrew, IoT Core, Lambda, Lightsail, OpenSearch, RDS, S3, SageMaker
- **Deployment/CI/CD** CircleCI, Cloud Build, Docker, Github Actions, Kubernetes, Liquibase, terraform, Doppler, Unix/Linux
- **Data** Dremio, Hadoop, Spark, Tableau, data build tool(dbt), dimensional modeling, Stitch, Fivetran, Apache Kafka, Apache Airflow, FastAPI

## Professional Experience

Staff Data Engineer	Rubikal (Clutch B2B Leader)	Jan 2022 – Now
Directed the entire data architecture, development, and stakeholder engagement for 6 key initiatives in 2 key UK/US partnerships, enhancing their team capabilities through strategic team augmentation		
1. <b>EnergyAspects</b>	<b>(\$61M/y Energy Research Firm)</b>	<b>(London)(Remote)</b>
<ul style="list-style-type: none"><li>• Centralized BigQuery analytics with dbt, improving data accessibility and consistency across multiple sources.</li><li>• Streamlined internal APIs into a scalable, centralized API using Python and Cloud Services, enhancing system integration and scalability.</li><li>• Processed 25TB of Spire weather data into Shooju using Docker and AWS, enabling efficient data analysis.</li><li>• Developed scalable Models API for on-demand model processing, facilitating dynamic utilization.</li><li>• Maintained Composer Airflow Environment, resolving issues and performing upgrades for optimal workflow efficiency.</li><li>• Redesigned legacy infrastructure and enhanced data architecture, building a centralized data warehouse for daily insights.</li><li>• Implemented core internal and public APIs for data modeling and delivery, using FastAPI and Cloud Services.</li></ul>		

- Automated multiple ETL processes to relational and time-series databases, streamlining data processing.
- Improved data engineering practices with mandatory code reviews and quality checks, fostering a culture of excellence.
- Created and deployed org-wide dash app templates with Auth0 integration, enhancing app accessibility and security.
- Managed environment and DAGs migration during Cloud Composer/Airflow upgrades, ensuring smooth transitions.
- Built data-related CI/CD pipelines with Cloud Build and GitHub Actions, streamlining deployment and quality control

## 2. Akawin (Distributed Social Network) (Oregon)(Remote)

- Solely architected and built real-time data infrastructure that collects data from a distributed social network following nostr protocol then integrate into GraphDB, AWS DocumentDB, Elasticsearch, and AWS ElastiCache.
- Engaged in comprehensive project planning, establishing and meeting critical deadlines within a highly constrained timeline.
- Coordinated meetings with stakeholders to ensure project alignment with business objectives.

---

## Lead Data Engineer Rubikal (Linkedin Top Startup) Jan 2020 – Jan 2022 1. InsideTrack (\$54M/y US EdTech Leader) (Oregon)(Remote)

- Maintained our internal outsourced ETL tool (Pemi), enhancing efficiency through Python, Pandas, and Pytest.
- Developed ETL pipelines to process data from over 40 US universities at varying frequencies, integrating with UCoach data, and distributed to 5 microservices via Python, Pandas, Pemi, SQL, and SFTP, significantly improving data consolidation and accessibility.
- Oversaw real-time syncing tools utilizing Fivetran, Stitch, and Alooma, ensuring seamless data integration.
- Created dimensional models in the dbt Data Warehouse, employing AWS Redshift, Snowflake, and dbt to enhance data analysis capabilities.
- Managed InsideTracks Internal Workflow Manager/Scheduler, optimizing internal operations.
- Upgraded data deployment infrastructure using CircleCI, Kubernetes, Docker, and Ansible, bolstering deployment efficiency.
- Established Quarterly Team OKRs, aligning team efforts with strategic goals.

## 2. Concentrical (EdTech) (Oregon)(Remote)

- Played a pivotal role in the division of the company's data infrastructure into two distinct entities, InsideTrack and Concentrical, while simultaneously providing comprehensive support across product development and management tasks.
- Architected, implemented data deletion service RESTful APIs of five microservices and other AWS resources/backups, ensuring compliance with data privacy regulations.

---

## Senior Software/Data Engineer Rubikal (Consultancy) Aug 2018 – Jan 2020 1. Endless Studios (EdTech/Game Development) (Arizona)(Remote)

- Architected, implemented all ETL data pipelines targeting sources like Google Analytics, Firestore, Google Tag Manager, and PostgreSQL to Google BigQuery using Python, pandas, docker, Compute Engine, and Cloud Storage.
- Developed 20% of dimensional models using dbt in BigQuery.

- Constructed the entirety of CI/CD pipelines and cloud scripts, automating code releases and data warehouse model updates.

## **2. IMAGO Images** (Global Content Provider) (Berlin)(Remote)

- Developed an optimized data processing product that uses FFMPEG to process videos and generate a watermarked set of videos with different resolutions.

## **3. College Confidential** (World's Largest College Forum) (Oregon)(Remote)

- Led development of forums tagging project by tagging 1.5M posts with college/university/educational institutions name tags based on content.
- Engineered ETL processes to construct a global index of universities and college name synonyms, leveraging web crawling techniques and consolidating data within PostgreSQL databases.

## **4. Benchmark** (Business Intelligence and AI) (Washington)(Remote)

- Orchestrated ETL processes for data sources (GDELT v1/v2, XSub, QoG, SENDAI), building a data lake with tools including Python, AWS Glue, AWS Athena, PySpark, and AWS EMR.
- Developed time-series forecasting models with Amazon Forecast, enhancing predictive capabilities.

---

## **Data Engineer** BADR (Software Consultancy) Jun 2017 – Aug 2018

- Originated, project-managed, and developed a product analyzing Wuzzuf(LinkedIn like) job posts, serving both recruiters and job seekers by:
  - Unveiling essential skills per job title to streamline hiring processes and assist in crafting job descriptions.
  - Identifying current market-demanded skills for active job positions, enhancing job search strategies.
- Collaborated closely with leadership to align product features with strategic goals, ensuring successful implementation and market relevance.

---

## **Full-Stack Software Engineer** BADR (Software Consultancy) Mar 2017 – Jun 2017 1. Incorta (\$54.8M/y Data Platform) (Silicon Valley)(Remote)

- Spearheaded the development of a novel visualization library using d3.js as the sole engineer, initiating a project that led to the replacement of Tableau and integrating with a cross-functional team to maintain leadership across key development phases.
- Drove 60% of the development and pioneered 90% of the advanced features of the visualization library, significantly contributing to an estimated \$100K/year reduction in operational costs.
- Adapted to challenging requirements changes mid-project, seamlessly delivering new product features under evolving conditions.

## **2. WhoKnows** (AI-Powered Lead Generation) (Silicon Valley)(Remote)

- Revamped six complex report pages by creating all necessary data providing RESTful endpoints under a tight deadline.
- Inherited leadership and project management roles mid-project due to the long medical unavailability of the team lead and project manager, ensuring project continuity.
- Conducted client meetings, led team code reviews, and made critical delivery decisions to maintain project momentum.
- Delivered the entire project with positive feedback from the client, demonstrating adaptability and leadership in challenging circumstances.

3. **Elite HRV** (\$14.3M Funded Health Startup) (North Carolina)(Remote)
- Implemented tagging feature, and breathing options to a cross-platform mobile app using ionic.

- 
- University Teaching Assistant**      **AAST (CS Department)**      **Feb 2017 – Aug 2019**
- Taught on-campus courses of Computing Algorithms, OOP, and Data Visualization

## Education

- 
- B.S in Computer Engineering**      **Alexandria University, Egypt**      **2010 – 2015**
- A top 4% university globally in QS ranking.
  - Undergraduate Coursework: Operating Systems; Databases; Algorithms; Programming Languages; Comp. Architecture; Systems Programming; Computer Vision.
  - **College Projects:**
    - **Wikipedia Daily Hot Topics.** App shows top N topics visited. (Python, Hadoop, JS)
    - **Distributed Multi-Agent Classifier.** Classify items based on collected data by agents from multiple distributed locations. (Java, JADE)
    - **Twitter Sentiment Analysis.** An app that infers the sentiment of tweets. (Python, Machine Learning, JS)
    - **XO Game with AI Using Assembly.** XO game with MinMax algorithm. (X86 Assembly)

## Awards

- 
- **Google Hall of Fame:** for a successful report of a security bug in Google BigQuery that allowed unauthorized users to drop datasets.