Sarthak Haldar

sarthakhaldarofficial@gmail.com • +1 (520) 474-6788
linkedin.com/in/sarthak-haldar/ • Tucson, Arizona 85705

Entrepreneur and Big Data Engineer

Data Engineer and Application Developer with 4 years of experience in designing and implementing robust data solutions. Skilled in ETL processes, data architecture, and data integration, with a focus on data warehousing and pipelines. Proven expertise in software development, product design, business analysis, and managerial accounting, complementing technical proficiency. Adept at resolving complex data challenges, delivering business-critical insights, and leveraging cutting-edge technologies to drive innovation and optimize data-driven decision-making.

Technical Proficiencies

Cloud :	Amazon AWS, Azure
Languages :	Python, R, Apache Spark, HTML, Bootstrap CSS Framework, JavaScript, Node.js runtime
AI/ML :	Scikit-learn, Spacy, NLTK, TensorFlow, Keras, PyTorch
Tools :	Pandas, NumPy, SciPy, Matplotlib, Microsoft Excel, Multimedia tools (Photoshop, Premiere Pro, Procreate), GIS Leaflet JS, ArcGIS Online web mapping, API Gateway, building RESTful API's to enable data consumption, Beautiful Soup for Web Scraping
DBMS:	MySQL, Oracle PL/SQL, T-SQL SQL Server, MongoDB, Cypher query language for Neo4j graph DB, SparQL
Data Warehousing :	Apache Spark, Amazon Redshift, Databricks
Data Integration ETL :	AWS Lambda, AWS Glue, AWS Athena, Azure Data Factory, Azure Databricks, Azure Synapse Analytics
BI Tools :	Tableau, Power Query M functions, DAX functions for Microsoft Power BI, Microsoft Excel
Planning Tools :	Jira Work Management, Confluence, Miro, Lucid Chart
MLOps CI/CD :	Git, Github, Github Actions, Jenkins, Docker container, Airflow for job scheduling orchestration, Azure Data Factory
Microservices :	Kafka, Postman, REST API, Kubernetes
UI/UX:	Figma complying Jakob's Law
App Framework:	Flask, Shiny, Expo React Native and JavaScript for Mobile app
Library	Vite React JSX, Expo React Native

Core Competencies

- UI / UX Product Design
- Data Mining with Predictive Modelling
- Data Modelling and Statistical Analysis
- Data Integration ETL
- Business Intelligence
- Staging Big Data
- Hypothesis Testing A/B tests
- SQL, Python, R, JS
- Object Oriented Programming

Professional Experience

AcharyaG, https://acharyag.in/
Founder & CEO   **02/2025 – Present**

Spearheading AcharyaG, a platform that helps students connect with qualified in-person tutors for personalized learning. Overseeing business strategy, platform development, and market expansion to improve access to quality education.

WorkOrderWizard, Suwanee, GA, US
Principal Data Engineer   **05/2024 – 05/2025**

Developed and maintained data ELT pipelines with Azure and created analytics dashboards using react native front-end. Demonstrated proficiency in SQL, Python, and gained hands-on experience with Spark and Airflow. I effectively managed workflows using Git and Jira, and consistently brought a proactive, solution-oriented approach to solve complex challenges, particularly in driving customer management solutions in the hotel industry. Led product design with Figma and prototyped the front-end of a hotel management application using HTML, CSS, JavaScript, and React, enhancing user experience and achieving a 30% improvement in design iteration speed through streamlined workflows. Applied Object Oriented Programming Principles and SOLID concepts to implement critical features..

University of Arizona, Tucson, Arizona, US, https://viz.datascience.arizona.edu/barriomap/
Graduate Research Assistant Data Engineer   **01/2023 – 05/2024**

Assisted in conducting research related to urban planning, data science, and geospatial analysis. Collaborated on open-source research projects, implemented advanced features like vector export, and developed a geo mapping GIS application using R programming, Shiny web app package, and Leaflet JS library in collaboration with Dr. Cristian Roman Palacios, Assistant Professor of Practice from Department: School of Information. Collaborated on a Systems Engineering project funded by DoD involving Graph Database, Web Ontology Language (OWL) and RDF integration with spacecraft data, leveraging SPARQL for querying, and

successfully created a comprehensive dashboard to visualize the combined datasets. Used regression, Bayesian methods, tree-based learners, SVM, RF, XGBOOST, time series modeling, dimensionality reduction, SEM, GLM, GLMM, clustering, Deep learning. machine learning, probability theory and statistics according to business requirements.

Nvidia Corporation, Santa Clara, CA
Data Scientist Intern   **08/2022 – 01/2023**

Organize structured data – Perform Transformations like Aggregation/ normalization on Sales and Inventory records from Oracle ERP OLTP systems and Load data into warehouse; Performed Power Query M function for Get-data transformations; add modeling using DAX and created Sales visualization Reports with Power BI. Created calculated bridge-table to handle models with relationship cardinality many-to-many. Use Power BI Performance analyzer to tune the new measures in relationship models. Design data model that uses star schema instead of snowflake by creating a central fact table linking other tables, containing only the attributes needed. Accomplished implementing End-to-End Ops Data platforms and deploying scalable algorithms for supply planning through statistical analytics, ML, and anomaly detection. Proficient in leading discussions, identifying trends, and presenting actionable insights to drive business solutions.

Vodafone Intelligent Solutions, Pune, India
Big Data Engineer Senior Executive   **01/2020 – 12/2021**

Delivered successful implementations by working across a scrum team of 15 using agile approaches.

- Transformed semi-structured JSON data to structured data using Databricks platform and Spark for distributed data processing. Stored data in AWS S3 and Databricks DBFS, improving data accessibility and processing efficiency
- Performed ETL with PySpark + SQL to organize ingested batch file (JSON) and stream data to bronze, silver and gold table and maintain multi-hop Delta Live Table DLT pipeline or maintain a medallion architecture. Used Spark structured streaming processing engine to query infinite data source for incremental data processing.
- Process Change Data Capture CDC logs made to data using DLT. CDC involves processing of identifying and capturing changes made to data to the data source and then delivering those changes to the target.

Education

Masters Data Science
University of Arizona, College of Information, Tucson, Arizona, US   **08/2022 – 05/2024**

Courses: Data Visualization, Data Mining, Machine Learning, AI, NLP, Data Warehousing, Management Information Systems MIS Big Data

Bachelors Computer and Information Science
SRM University, Chennai, Tamil Nadu, India   **05/2016 – 05/2020**

Courses: Data Structures and Algorithms, Object Oriented Programming, Python, Database Management SQL, Computer Architecture, Theory of Computation, Computer Networking

Certificate & Certification

AWS Certified Cloud Practitioner (CLF-C01) Credly
Databricks Accredited Lakehouse Fundamentals Credential

Conference Publications

Design and Implementation of an Image Classifier using CNN, Conference name: IEEE, 2019
Critical Analysis on Cryptocurrency, International Journal of Computer Sciences and Engineering, 2018

Projects

Snake Game AI, 2023 Github   AWS Glue Pipeline, 2023 Link
Wine Quality Prediction, 2022 Github   Azure ETL Pipeline, 2024 Github

Portfolio

Website: https://www.sarthakhaldar.com/
Github: https://github.com/rog-SARTHAK
Tableau: https://public.tableau.com/app/profile/sarthak.haldar
Linkedin: https://www.linkedin.com/in/sarthak-haldar/