Zeeshan **Hooda**

SENIOR DATA ENGINEER · CANADIAN CITIZEN

☎ +1 (587) 568-6239 | ✉ zee@sparselabs.org | ⌂ github.com/zhooda

Experience

Data Solutions Architect
Calgary, Canada

SPARSE LABS
Oct 2022 - Present

- Designed and implemented a **greenfield clinical data model** for a large **healthcare client** from loosely scoped requirements
- Deployed **AWS DMS to capture change data** from on-prem and cloud **SQL Server and Oracle** databases into S3
- Implemented cloud data lake using **open-source Delta Lake, Spark, Glue ETL, Step Functions, and Redshift**
- Optimized Delta Lake tables and Spark Jobs to **reduce Redshift Spectrum resource consumption by 60%**

Lead Data Scientist
San Francisco, USA

SMART ACCESS
Apr 2020 - Oct 2022

- Engaged in **customer-facing pre-sales** activities to **close a $3.6M contract** with a Fortune 500 company
- Orchestrated a production **migration of in-house analytics infrastructure** to ensure integrity of downstream analytics products
- Architected and implemented a self-serve data platform in GCP using **BigQuery, Dataproc, Debezium, Airflow, and Superset**
- Established **SOPs for data and analytics products** and internal **SLAs for data availability and freshness**
- **Built and supervised a 5-person team** of data scientists, engineers, and analysts

Data Scientist
Calgary, Canada

SENSORUP
Jun 2018 - Apr 2020

- Developed multi-tenant data pipelines using **Airflow, Pandas, and Redshift** to ingest and transform data from fluid hauling sources
- **Trained and deployed machine learning models** to predict live facility conditions and latencies for intelligent vehicle dispatching
- Created an **edge computer vision service** to identify hazards on remote industrial sites, **reducing hazard reaction time by 75%**

Software Engineer
Edmonton, Canada

REDMAN TECHNOLOGIES
Dec 2016 - Jun 2018

- Designed **dimensional models** for neighborhood statistics and real estate data to be used in a **data warehouse**
- Developed **ETL pipelines using JavaScript** to ingest and transform data from **MongoDB, CouchDB, PostgreSQL, and MySQL**
- Created a **front-end visualization platform** for neighborhood statistics **used by Realtors across Canada**

Compiler Engineering Intern
Portland, USA

MOZILLA FOUNDATION
Apr 2016 - Dec 2016

- Implemented a **type-partitioned memory allocator** for the Rust compiler to increase application memory security
- **Contributed to the open-source LLVM project** to improve the performance of ARM64e translation to intermediate representation

Education

University of California, Berkeley
Berkeley, USA

DOCTOR OF PHILOSOPHY (PHD) IN COMPUTER SCIENCE
Apr 2025 (Expected)

- **IntentsRT**: Real-Time Intent Modeling on Sparse Clickstreams with Markovian Layer Routers

MASTER OF SCIENCE (MSC) IN COMPUTER SCIENCE
Apr 2023

- **MC^{2}Net**: A Scalable Neural Architecture for Monte Carlo Inference with Markov Chains

BACHELOR OF SCIENCE (BSC) IN COMPUTER SCIENCE
Apr 2022

- **Specialization**: Statistics and Machine Learning

Skills

ETL + Orchestration	Spark, Airflow, Beam, Dataproc, Kafka, Pub/Sub, Kinesis, Debezium, Trino, Athena
Data Lakes + Warehouses	BigQuery, Redshift, Snowflake, Pinot, Delta Lake, Hudi
Analytics	Superset, Tableau Online + Server, Metabase
Cloud + Operations	AWS, GCP, Docker, CI/CD, Terraform, CDK, Cloud Functions, Lambda
Databases	PostgreSQL, MySQL, Oracle, SQL Server, MongoDB, Cassandra
Languages + Frameworks	Python, Scala, Java, Go, Rust, C, Assembly (x86, ARM), LaTeX, Flask, FastAPI, GraphQL, Flutter, Svelte