

ANANT SHARMA

- www.datasorcerer.com
- anant.sharma2040@gmail.com

- [LinkedIn Profile](#)
- Phone: +1-778-881-4911



Summary

- **11+ years** of experience in **Data Engineering & Warehousing** across different nations for G-SIBs banks (BFSI domain) & **Telecom** industry
- [Watch my tech talk from Snowflake Data Summit in San Francisco where I discuss scaling data engineering for 2.5 PB migrations!](#)
- **AWS Solution Architect & Data Analytics** certified
- **SnowPro Core** certified along with **Azure Data Fundamentals** & other **Data Science & Python** certifications.
- Enabling my firm to become **Advance Tier partner** with AWS & earning a **Service Delivery Badge** in EMR
- Proficient with AWS services such as **EC2, EMRs, RDS, Redshift, S3, Athena, Lambda, Data Pipeline, Glue, Cloudfront, API Gateway** etc.
- Worked as a **Data Architect** for implementing **Data Lake** & Data warehouse in **Snowflake & Teradata**
- Deployed **Machine learning** models in production using AWS stack
- **Business knowledge** in **RDAR BCBS 239, CCPA (Privacy Act)**, Investment banking, Core Banking, Health Insurance, Mutual Funds, Retail Banking & Telecom sector
- Technical expertise in building **Data pipelines, Azure Data Factory, Apache Airflow, FSLDM, SAS Suite, Hadoop, Hive, Impala, Python, PySpark, ETL, SQL Scripting and Data Modeling**
- Wrote **research papers** on Data modeling & data migration using different methodologies
- Implemented end-to-end entire **website** for Stock Market [Stock 360 Watcher Website](#) hosted on **AWS**
- Created **android app** published on Google Play Store [Stock 360 Watcher App](#)

SKILLS

- | | |
|--------------------------------|--|
| • Amazon Web Services | - EC2, EMR, Data pipeline, RDS, Redshift, S3, Glue, CloudWatch, Athena, QuickSight, Lambda, Kinesis, DynamoDB, SageMaker, Redshift Spectrum, SQS, SNS, SES |
| • Azure | - Virtual Machines, Azure Data Factory , Databricks, Azure DevOps, Blob |
| • Other Cloud Services | - Google Firebase , phpMyAdmin, IBM Cloud |
| • Data Modeling | - ERwin Data Modeler, Hackolade |
| • ETL tools | - Glue, Apache Airflow, Azure Data Factory , SAS DI studio, SAS Enterprise |
| • Databases | - Redshift , RDS, DynamoDB, Snowflake , SQL Server, Oracle, Teradata , Hive |
| • Business Intelligence | - Tableau, Microsoft Power BI, SAS Web Report studio, SAP BO |
| • Programming Languages | - Python , PHP, Hive-QL , SQL, Base SAS, Java, HTML |
| • Platforms | - Hadoop, MS Windows NT/2000/XP/Vista and Unix |
| • Repository Services | - Github, Gitlab |
| • Productivity Tools | - Android Studio, PyCharm, MS Office Suite |

WORK EXPERIENCE

[Senior Data Architect](#)

ProCogia – Vancouver, British Columbia – January 2021 to present

Projects – [Telecom - Device Intelligence & Analytics](#), [AWS Partnership Building](#), [Bringing new business](#)

- For Clients -

- Saved **~\$1M annually** for the client by migrating from AWS to **Snowflake** architecture
- Built datalake architecture from scratch on **AWS + Snowflake** ingesting data from Kafka with landing zone and processed datalake.
- Implemented anonymization process to structure and standardize datalake using aws **EMR + Snowpark** scheduled as batch job
- ETL orchestration using ADF & Airflow.
- Integrated **Athena** with the processed anonymized datalake for the analyst to exploration of data using simple sql queries
- Refracted and deployed **Machine learning model** on AWS EMR after increasing performance 4000 times better which saved thousands of dollars on cloud cost by rewriting Scikit learn based python code with **Sedona** & Spark ML based **Pyspark** code.
- Created **Glue** job to create near real time data quality and analytics dashboard on AWS datalake for quickly detecting issues
- Created external tables in **Impala** from S3 & HDFS
- Worked with **Apache NiFi** to schedule AWS S3 files to HDFS using AWS SQS
- Installed, configured and created workflows on **Apache Airflow** hosted on ec2 instance to orchestrate pipelines and centralize ETL processes.
- Created **ETL monitoring process** and dashboard to check the overall health of the data being pushed from different ETL flows which monitors Lambda, Airflow jobs, EMRs, Snowflake etc using audit database in **RDS postgres**.
- Implemented telecom data warehouse from scratch on **Snowflake** including preliminary work for hosting Snowflake on Azure.
- Brought **thousands of dollars cost reduction** by simplifying the process on both cloud and on-prem.
- Created dimensional model on Snowflake for **data warehousing** using Kimballs architecture principles
- Installed, configured and integration **PowerBi** gateway on cloud to schedule reports incremental refresh on cloud
- Created web application for easy EMR use through few clicks from custom web interface, with an idea of replicating **Databricks features with zero cost** which saved **millions of \$\$** for the customer
- Created **cloud architecture for Analytics team** to orchestrate their flow with minimalistic effort which they can create on their own.
- Repeatedly recognised by the client every quarter as **cultural hero** for the contributions.
- Hosted client's public facing **android app** on cloud with the help of Aws cloudfront.
- Created various **Rest APIs** for scheduling jobs, getting data, etc using Aws Api gateway and Aws lambda
- Created serverless architecture for getting TCP based **5g latency metrics**
- Implemented **file compactor tool** which compresses the datalake incrementally using Spark on EMR
- Created **cost sharing pipeline and dashboard to monitor cost** across different users and internal customers
- Established **center of excellence** body for the client with best practices for cloud services and usage
- **Automating data dictionary population** to correspond schema changes in data lake
- **Containerization** of different projects like web applications, airflow deployment, flask apis etc using **docker**
- Involved in getting requirements & solving pain points for Analysts, data scientists, other teams and stakeholders on regular basis
- For ProCogia -
 - Submitted **20 opportunities & 2 EMR use cases** which is enabling ProCogia to become **Advance Tier partner with AWS** & earn a **service delivery badge in EMR**

- Involved in numerous sales meeting to convert prospective customers into highly valued clients
- **Designed and implemented website** which is the bible of Data Engineering practices in the firm

Lead Consultant

HCL Technologies – Kuala Lumpur, Malaysia – January 2019 to December 2020

Projects – [Regional Enterprise Data Warehouse \(REDW\)](#), [Avalog Data Migration](#), [Risk Evaluation](#)

- Created Data pipelines using **PySpark** to ingest data from various sources to AWS S3
- Using **AWS EMR Hive** for Data wrangling & analysing the data residing in AWS S3
- Data Modeling for Fraud Data & Group Risk Model for one of the leading banks of Malaysia
- Data Modeling for Group Risk Model to automate the regulatory reporting for Bank Negara
- Creating **Machine Learning** models using AWS **SageMaker** for customer churn analytics
- Automating the data validation **process reducing time from days to hours** using Python
- **Automating DDL creation script & Data migration script** using Python

Senior Consultant

PricewaterhouseCoopers – Mumbai, India – April 2016 to January 2019

Projects – [Insurance Data engineering](#), [CTR – Automation](#), [RDAR BCBS 239](#), [Asset Management Company](#)

- Implementation **RDAR BCBS 239** using **Teradata FSLDM**
- Implemented Data lake in AWS using **S3** as the storage, **Glue** as the ETL & **Athena** for Ad-hoc reporting.
- Involved in building Data lake flow (Ingestion → Cleansing → Transforming → Loading) using **Data pipeline & Python** scripts
- Data lake is being utilized for dashboard creation using **QuickSight** and to feed data to data science team
- Created flow for real-time data ingestion from **Kinesis Firehose & Google Firebase** to S3
- Implementation of NoSQL data model for **MongoDB** using **Hackolade**
- Allocated on-site and assisted PwC – South Africa in **CTR – Automation & Assurance** project
- Implementation of a **PwC data model** on **ERwin Data Modeler** for BFSI domain

Project Engineer

Wipro Technologies – Mumbai, India – April 2014 to April 2016

Projects – [Enterprise Data Warehouse with Teradata FSLDM & SAS](#)

- ETL lead for a big Data warehousing project which involved 39 source system having 20+ years of data for of the leading & oldest government bank in India
- Extensive use of Erwin Data Modeler, Base SAS Programming, SAS Enterprise Guide, SAS DI Studio, SAS Web Report Studio, Teradata, Teradata FSLDM
- **Provide reusable components** to load data from various databases such as Oracle, DB2, SQL Server, etc for which honored by **SPOT Award**
- Designed a script which was **selected as BVM process across all EDW projects** in the organization

CERTIFICATION ([Click here to view badges](#))

- [AWS Certified Solution Architect - Associate](#)
- [AWS Certified Data Analytics - Speciality](#)
- [Microsoft Certified - Azure Data Fundamentals](#)
- [Snowflake - Hands on Essentials - Data Warehouse](#)
- [IBM Data Science Professional Certificate](#)

EDUCATION

B.Tech in Information Technology

- Dr. A. P. J. Abdul Kalam Technical University – 70%

PUBLICATIONS

- **Tutorial & Blogs website** www.datasorcere.com
- **Stock Market Analysis Website** [Stock360Watcher - Click to view the website](#)
- **Stock Market Analysis App** [Stock 360 Watcher App - Click to download app](#)
- **%UniCreate A SAS Macro for Enterprise DWH Dynamic Implementation** [See PDF paper online here](#)
- **% BigDataLoader: A SAS Macro to Migrate Big Data 99% Faster** [See PDF paper online here](#)