Amar Petla

Staff Data &AI Engineering

Peoria, United States 85382

(480) 619-1143

amarnadh87@gmail.com

https://www.linkedin.com/in/amar-petla/

Senior Data & AI Architect | Snowflake Cortex | GCP Big Data Stack | Cloud-Native & API-Driven Platforms

Results-driven Staff Data Engineer and Architect with **13+ years of experience** designing and implementing enterprise-scale data, AI, and analytics platforms across **GCP, Azure, and AWS**. Proven expertise in building robust, scalable big data solutions using **BigQuery**, **DataProc**, **Airflow**, **Pub/Sub**, **Spark**, **Hive**, and **Kafka**, with strong proficiency in **Python**, **Java**, and **Shell scripting**.

Hands-on experience developing **microservices and RESTful APIs** using **Java and Spring Boot**, enabling flexible, service-oriented data platforms. Deep technical leadership in delivering AI/ML-driven systems leveraging **Snowflake Cortex**, **Snowpark**, **LLMs**, **RAG pipelines**, **semantic search**, and **vector embeddings** to enable intelligent, real-time decisioning.

Architected and optimized large-scale data pipelines, CI/CD workflows (**Jenkins**, **GitHub Actions**, **XLR**), and observability frameworks supporting 99.9% system uptime. Strong background in **data modeling**, **distributed systems**, and **event-driven architecture**, with a continuous focus on performance, governance, and operational excellence.

Trusted advisor and agile leader, known for mentoring teams, aligning business and technical strategies, and driving innovation across 8+ concurrent high-impact projects. Passionate about building end-to-end cloud-native platforms that turn business vision into data-driven, AI-enabled solutions.

EXPERIENCE HIGHLIGHTS

- Over the past 3+ years, I have led cross-functional teams and driven organizational goals by leveraging IT shared services. My collaboration with businesses and other domains ensured that data and architectural solutions met customer needs through reporting, analytics, digital marketing, and other resources. I developed a comprehensive data and AI strategy aligned with the organization's objectives, identified areas where data analytics and AI could drive innovation and improve decision-making, and created a roadmap for these initiatives. I led the development of data engineering and machine learning models, collaborated with software development teams to integrate AI solutions, and led the CoE for Technology innovation and product development to enhance enterprise data assets.

Skills

- **Cloud Platforms** - Azure, GCP, RedHat OpenStack and OpenShift, AWS
- **Cloud Services** - Snowflake,Google Big Query, Google cloud storage, google Composer, Google Vertex AI,Google Cloud Build,S3, Glue, EC2,lambda functions
- **Knowledge ETL Tools** - Ab Initio, Teradata
- **Monitoring** – Splunk,Grafana,Datadog,Airflow
- **Scripting Languages** – Shell Script
- **Programming Languages and Frameworks** - Java, J2EE, Spring boot, REST, Python
- **Big Data** - Hadoop, Map Reduce, Sqoop, Hive, Apache Spark
- **NoSQL & RDBMS** - Apache HBASE, Mapr-DB, Mem-SQL, Couchbase, PostgreSQL, MySQL,Mariadb
- **Streaming Technologies** - Solace, Kafka, Spark Streaming
- **Scheduling Tools** - AutoSys, Oozie,Event Engine, RMJ, DBT.
- **Data quality and data governance** - Collibra
- **Data Management:** Data Modeling,Data Warehousing,Data Pipeline Design,Data Quality Assurance,Data Governance
- **ML / AI :** Machine Learning, AI
- **Data Science Tools:** Python, R, Jupyter, TensorFlow, PyTorch, Scikit-Learn, Pandas, NumPy, SQL
- **Leadership**: Technology leadership work streams

Work History

2023-09 - Current

Staff Engineer

Freeport Mcmoran, PHOENIX

- Architected scalable data platforms across Azure, Snowflake, and GCP, supporting enterprise initiatives including PROD ACCT, FINWIP, HTOS, and HELENA.
- Designed hybrid Snowflake schemas with clustering, materialized views, and Query Acceleration Service (QAS) to optimize performance and cost.
- Built ELT pipelines using Azure Data Factory, Snowflake Tasks/Streams, and Python orchestration, ingesting 100M+ daily records from Event Hub, APIs,

and on-prem systems.
- Developed Spark-based ingestion flows using the Snowflake Spark Connector for schema-aware, secure data movement between Databricks and Snowflake.
- Leveraged Snowpark Python and Snowpark ML for scalable ML pipelines and feature engineering within Snowflake.
- Created Java microservices for real-time ingestion from Azure Event Hub to Snowflake using Kafka Connect + Snowflake Sink Connector.
- Implemented Snowflake Streams and Dynamic Tables for near real-time analytics and automated materialization.
- Integrated Snowflake Cortex functions (SUMMARIZE, CLASSIFY, EMBED_TEXT) into GenAI prototypes for intelligent automation and LLM-based insights.
- Built Retrieval-Augmented Generation (RAG) pipelines to enhance LLM-based retrieval and summarization.
- Established observability frameworks using Azure Monitor, Snowflake Account Usage, and custom logging pipelines.
- Tuned Snowflake workloads with auto-suspend/resume, multi-cluster warehouses, and cost diagnostics.
- Developed Spring Boot REST APIs and Java-based metadata services, secured via OAuth 2.0 / Azure AD.
- Led CI/CD pipeline design using GitHub Actions, Jenkins, and Terraform with automated testing and SonarQube scanning.
- Mentored a team of 6+ engineers, aligning delivery to quarterly OKRs and roadmap priorities.
- Contributed to Freeport's Analytics 4 Everyone (A4E) program and Snowflake Technical Syncs, shaping architecture and training.

2022-06 - 2023-09

Staff Data Engineer

Achieve (Freedom Financial Network), Tempe, AZ

- **Built and optimized large-scale data pipelines** on **GCP**, handling **10B+ daily events** across batch and streaming systems using **BigQuery**, **DataProc**, **Airflow**, and **Pub/Sub**.
- **Designed and implemented Spark-based ETL frameworks** and **Python workflows**, reducing end-to-end pipeline latency by **35%** and significantly improving SLA compliance.
- **Led reengineering of legacy workflows** to adopt **event-driven architecture** using **Kafka and GCP Pub/Sub**, enhancing system throughput and resilience.
- **Developed and managed Airflow DAGs** using custom Python operators to orchestrate **GBQ ETL operations** with precise scheduling, monitoring, and error handling.
- Engineered **BigQuery models** to support cost analytics, client engagement metrics, and retail performance across structured and semi-structured datasets (CSV, JSON, Cloud Storage, relational sources).
- Delivered **interactive Tableau dashboards and data sources**, enabling self-serve analytics and supporting executive-level reporting on **marketing costs (Google Ads)** and campaign performance.
- Created robust **data ingestion pipelines** to import/export data between GBQ and external systems, ensuring schema alignment, data cleansing,

and metadata propagation.

- **Analyzed large datasets** and wrote complex SQL queries to extract key insights from GBQ, enabling real-time decision support for business and product teams.
- Built and deployed **end-to-end machine learning pipelines** using **Vertex AI**, **Kubeflow**, and **BigQuery ML datasets**, including model training, hyperparameter tuning, and model evaluation workflows.
- Developed **data preparation pipelines** for ML training, performing advanced **data cleaning**, transformation, feature selection, and labeling in Python.
- Applied advanced **SQL partitioning, clustering, and materialized views** to optimize GBQ model performance and reduce query costs.
- Partnered with product owners and business stakeholders to translate roadmap features into scalable technical solutions with measurable business impact.
- Supported enterprise-wide data governance efforts by ensuring **lineage tracking**, **data quality monitoring**, and **observability metrics** across critical pipelines.
- Championed Agile development methodologies—led **sprint planning**, **peer code reviews**, and enforced **test-driven development (TDD)** within a **Jenkins-based CI/CD** deployment framework.
- **Mentored junior data engineers**, conducting knowledge-sharing sessions on GCP architecture, Python performance optimization, Spark tuning, and modern data modeling techniques.

2021-12 - 2022-06

Sr.Bigdata Engineer

JPMorgan Chase, Plano, TX

- Design and Developed Rest API using spring boot where upload custom Data dictionaries to Collibra-Digital Transformation tool for NoSQL databases Using Collibra Spring boot Integration.
- Design and developed recon framework which monitor Inbound/outbound dataflow using pyspark.
- Working on AWS migration from existing platorm (hadoop) to AWS.
- Created CICD pipelines.

2015-10 - 2021-11

Sr.Big Data Engineer I Solution Architect

American Express, Phoenix, AZ

- **Project: Trinity**
- This project is joint effort to support seamless electronic solution from Buyer Purchase Order to payment settlement processes for e-Catalog transaction on Marketplaces such as Amazon Business. The requirements within this document are specific to processes among Buyer, Amazon Business and American Express.
- **Responsibilities:**
- Developed Spark code using python and Spark-SQL for processing of data
- Load Order Confirmation, Purchase order and Invoice cXML in PostgreSQL Database using Spring.
- Experienced in cloud Technologies like OpenShift for deploying and maintaining jobs.

- Experienced with PostgreSQL and working with and maintaining in a production environment.
- Experienced with cloud Technologies (OpenShift) and development/deployment in cloud infrastructure.
- Developed Kafka producers and consumer using spring boot.
- Experienced in handling large datasets using Partitions, Spark in Memory capabilities, Effective & efficient Joins, Transformations during ingestion process from spark dataframes.
- Implemented Partitioning, Dynamic Partitions, Buckets in HIVE.
- Exposure in Spark program performance improvement and tuning of Spark SQL, Hive queries.

Project: Jungle Lid

- The project JungleLid objective is Enable reporting of b2b(amazon) transactions with line-Item Detail data to card members via MYCA (UI) in support of E2E Jungle program. Reporting accomplished via CSV file at cycle cut for billed transactions to be downloaded from MYCA by client/card member for import to quick books or similar client side software. Also solving for service to MYCA so LID can be displayed on user screen when viewing transactions for both billed and unbilled and requesting to see LID detail.
- **Responsibilities:**
- Developed Spark code using Scala and Spark-SQL for processing of data.
- Load complex Json data in to Hbase table from Spark Dataframe using SparkHbaseConnector (SHC), SQL insert into Hive table with HBase storage and RDD.saveAsNewAPIHaoopDataset.
- Worked on converting Hive queries into Spark transformations using Spark RDDs.
- Experience in Job management for scheduling and developed job processing scripts using EVENT ENGINE (Internal).
- Designed and developed test framework for Data comparison in hive tables and which generates report using Jenkins tools (Internal tool using Jenkins)
- Load data into Spark Dataframes and performed in-memory data computation to generate output response.
- Implemented Scoverage for Scala code coverage and Scalastyle for coding standard with automated control Using Maven tool.
- Developed Watch Dog scripts to monitor applications and send notifications if required.
- Performed Data Orchestration for transactional, incremental tables by dividing these into different categories and joining them to form large buckets.
- Implemented Hive Generic UDF's to implemented business logic around custom data types.
- Implemented Partitioning, Dynamic Partitions, Buckets in HIVE.

Project: MYDATA

- The project MyData objective is to build functional application with user-friendly UI where users can manage content, format and delivery of data files that American Express provides to Global Commercial Payments

customers. MyData has been designed in way that back-end data file generation engine can be reused for all Global files (expense reconciliation files) with little code change. MyData exposes subset of data found in Cornerstone (data lake), American Express' Big Data Platform, for both internal and external consumption. MyData is designed to make customized data sets available to American Express customers via self-service portal replacing existing legacy system.

Responsibilities:

- Good knowledge and worked on Spark SQL, Spark Core topics such as Resilient Distributed Dataset (RDD) and Data Frames.
- Exposure in Spark program performance improvement and tuning of Spark SQL, Hive queries.
- Worked on improving Hive queries performance by rewriting in Spark.
- Developed Spark code using java and Spark-SQL for faster testing and processing of data.
- Import data from different sources like HDFS/Hbase into Spark RDD.
- Migrated all data and tables from Spark 1.2 to Spark 1.4. Created parallel process for data all ingestions which uses spark 1.4 and then comparing and validating data to 1.2 there by killing processes in 1.2.
- Migrated all data and tables from Spark 1.4 to Spark 1.6. Created parallel process for data all ingestions which uses spark 1.6 and then comparing and validating data to 1.4 there by killing processes in 1.4
- Designed HBase tables for time series data. Designed row key to avoid region hotspotting and accommodate desired read access/query patterns, used SingleColumnfilter for fast key search across Hbase regions.
- Integrated real-time streaming technologies for accurate monitoring of critical business metrics.

2015-05 - 2015-09

Big Data Engineer

Comcast, West Chester, PA

- **Responsibilities:**
- Experienced in handling large datasets using Partitions, Spark in Memory capabilities, Broadcasts in Spark, Effective & efficient Joins, Transformations and others during ingestion process itself.
- Experienced in performance tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.
- Developed Spark scripts by using Scala shell commands as per the requirement.
- Developed Watch Dog scripts to monitor the applications and send notifications if required.
- Proactively involved in tuning the complex Hive queries
- Developed framework to import and export data from various sources like Teradata, Oracle, SQL server and Flat-files into HDFS.
- Extensively used Sqoop to connect with various databases to import data into Hive.
- Proactively monitored systems and services, architecture design and implementation of Hadoop deployment, configuration management, backup, and disaster recovery systems and procedures.

- Designed and developed scalable custom Hadoop solutions as per dynamic data needs.
- Assisted in designing, development and architecture of Hadoop and HBase systems.
- Supported technical team members in management and review of Hadoop log files and data backups.
- Participated in development and execution of system and disaster recovery processes.
- Worked on batch processing data using Apache Hadoop, Map Reduce and Apache Pig.

2014-05 - 2015-04

Software Engineer

Think Finance, Texas City, TX

- **Responsibilities:**
- Extensively used Sqoop to connect with various databases to import data into Hive.
- Proactively monitored systems and services, architecture design and implementation of Hadoop deployment, configuration management, backup, and disaster recovery systems and procedures.
- Designed and developed scalable custom Hadoop solutions as per dynamic data needs.
- Assisted in designing, development, and architecture Hadoop and HBase systems.
- Supported technical team members in the management and review of Hadoop log files and data backups.
- Participated in the development and execution of system and disaster recovery processes.
- Worked on batch processing data using Apache Hadoop, Map Reduce, and Apache Pig.
- Experience in using DML statements to perform different operations on Hive Tables.

Education

2013-12 - 2015-03

Master of Science: Computer Engineering

International Technological University - San Jose, CA

2009-07 - 2012-04

Bachelor of Science: Electronics And Communication Engineering

VIGNAN'S INSTITUTE oF INFORMATION TECHNOLOGY - Vizag

Accomplishments

Patent :

A system for automated supplier invoice reconciliation is disclosed. The system may receive an order confirmation associated with a purchase order (PO) from a supplier system. The system may receive the PO associated with the order confirmation from a buyer system. The system may receive a first invoice associated with the PO and the order confirmation from the supplier system.

The system may reconcile between the PO, the first invoice, and the order confirmation to generate a second invoice. The system may pass the second invoice to the buyer system.
https://patents.justia.com/inventor/amar-petla

Certifications

- AI FUNDAMENTALS
- Microsoft Certified: Azure Solutions Architect Expert
- Snowflake Squad
- Building Business Acumen Graduate
- Snowflake Snowpark DataFrame Programming

2025-07

- SnowPro Core Certification