
VENGATESWARAN ARUNACHALAM

Santa Clara, CA 95051 ♦ (408)784-9915 ♦ tagnev.vengat@gmail.com ♦

WWW: <https://www.linkedin.com/in/vengateswaran-arunachalam/> ♦ WWW: <https://github.com/tagnev> ♦

WWW: <https://medium.com/@tagnev.vengat>

PROFESSIONAL SUMMARY

Professional Summary:

Accomplished Full Stack Data Engineer & Data Architect with 14+ years of expertise in BI solutions. Proficient in Spark, Hadoop, Python, ReactJS, and the complete SDLC. Solid experience in cloud services with AWS & GCP, database solutions both in the cloud and on-premise, and mastery of BI tools including Tableau and MicroStrategy. Demonstrable skills in customer-focused open-source projects, content creation, and community engagement. Excel at advanced database solutions, ETL processes, and enhancing data infrastructure. Well-versed in partnering with diverse teams to foster data-driven strategies.

SKILLS

- | | |
|----------------------------------|-----------------|
| • Docker | • Tableau |
| • DBT | • MicroStrategy |
| • Kubernetes | • Looker |
| • Git | • Cassandra |
| • AWS & GCP architecture | • NoSQL |
| • Python(Django, Flask, FastAPI) | • Pyspark |
| • Shell scripting | • Airflow |
| • MySQL/Relational DB | • Streamlit |
| • Snowflake | • openAI |
| • Postgresql | • Incorta |
| • Hadoop | |

WORK HISTORY

Data Architect/Senior Data Engineer, 10/2021 - Current

Keysight Technologies – Santa Clara, CA

Project 1: Data Migration and Legacy System Decommissioning

Stack : Incorta, Spark, Kafka, ClickHouse, DBT, Snowflake, Tableau, PowerBI

- Managed the data migration process during a merger, which included transitioning data from legacy systems of both companies into a single data warehouse.
- Developed and executed migration scripts and ETL processes to ensure data integrity and consistency.
- Collaborated with business stakeholders to prioritize data migration efforts, identifying critical data elements and dependencies.
- Orchestrated the decommissioning of legacy systems post-migration, reducing operational costs and streamlining data management.

Project 2: Serverless Data Pipelines on AWS

- Architected and developed serverless data pipelines on AWS using services such as AWS Lambda, AWS Step Functions, and Amazon S3.
- Created event-driven data processing workflows to automate data ingestion, transformation, and loading.

Key Achievements:

- Reduced data pipeline execution time by 40% through optimization and parallelization of data processing tasks.
- Achieved a 99.5% data pipeline success rate, minimizing data processing failures and improving data reliability.
- Enabled cost savings of \$50,000 annually by leveraging serverless architecture, optimizing resource utilization, and reducing operational overhead.
- Implemented end-to-end monitoring and automated alerts, ensuring prompt response to any pipeline issues.

Project 3: Real-time Stream Processing with Apache Kafka and Spark

- Designed and implemented a real-time stream processing solution using Apache Kafka for data ingestion and Apache Spark Streaming for data processing.
- Utilized Scala and Python to develop custom stream processing logic for complex event processing and anomaly detection.

Key Achievements:

- Reduced data processing latency by 50% through optimization and fine-tuning of the Apache Spark Streaming jobs.
- Improved data quality by implementing data validation checks, resulting in a 20% reduction in data errors.
- Achieved a 30% increase in real-time analytics capabilities, providing faster insights for operational decision-making.
- Monitored and maintained pipeline health, ensuring a 99.9% data processing uptime, and implemented automated alerting for any anomalies.

Full Stack Data Engineer/Startup Founder, 01/2020 - 12/2022

Teckiy – Santa Clara, CA

- Collaborated with a cross-functional team to design, develop, and deploy a comprehensive end-to-end data

product leveraging Django as the backend framework and Google Cloud Platform (GCP) for cloud infrastructure.

- Designed and implemented a robust data pipeline on GCP, incorporating services like Google Cloud Storage, BigQuery, and Cloud Dataflow for data ingestion, storage, transformation, and analysis.
- Developed a user-friendly web application using Django for data visualization and interaction, providing stakeholders with real-time insights into critical business metrics.
- Implemented authentication and authorization mechanisms within the Django application to ensure data security and access control, utilizing Django's built-in authentication system.
- Leveraged Django's ORM (Object-Relational Mapping) to model and manage the application's database, optimizing data storage and retrieval.
- Integrated third-party APIs and external data sources into the Django application, enabling seamless data integration and enrichment.
- Collaborated with front-end developers to ensure a seamless user experience and responsive design, incorporating modern web development technologies.
- Implemented CI/CD pipelines for the Django application, automating the deployment process and ensuring continuous delivery of new features and updates.
- Managed the cloud infrastructure on GCP, optimizing resource usage and cost-efficiency through monitoring and scaling strategies.
- Conducted regular code reviews, testing, and documentation to maintain code quality and facilitate knowledge sharing within the team.

Data Engineer Lead, 04/2018 - 10/2021

Keysight Technologies – Santa Clara, CA

Project 1: Data Lake Architecture and Optimization

- Designed and implemented a data lake architecture on AWS S3, providing a scalable and cost-effective solution for storing and processing massive volumes of structured and unstructured data.
- Optimized data ingestion pipelines using AWS Glue to automate data cataloging, ETL, and transformation processes, resulting in a 30% reduction in data processing time.
- Implemented data partitioning strategies and optimized query performance on AWS Athena, enabling data analysts to perform ad-hoc queries on large datasets.

Project 2: Multi-source Data Integration and Consolidation

- Led a project to integrate data from diverse sources, including relational databases, third-party APIs, and flat files, into a unified data warehouse.
- Designed a comprehensive ETL pipeline using Apache NiFi for data ingestion, Apache Spark for data transformation, and Snowflake as the data warehousing solution.
- Implemented complex data mappings, transformations, and data quality checks to ensure data consistency

and accuracy across sources.

Project 3: Data Warehouse Consolidation After Acquisition

- Led a critical data engineering project following the acquisition of a company, involving the consolidation of multiple data warehouses into a unified, scalable solution.
- Collaborated with cross-functional teams to define data integration strategies, including data mapping, transformation, and schema harmonization, ensuring seamless access to merged data.
- Implemented data deduplication and cleansing processes to eliminate redundancy and improve data quality.
- Utilized cloud-based data warehousing solutions using Snowflake, to achieve cost-efficiency and scalability in the merged environment.

Senior BI Developer, 06/2014 - 04/2018

Cognizant (Client: Keysight Technologies) – Santa Clara, CA

- Involved in Data Warehousing and worked extensively on RDBMS (Oracle, MS SQL) and ETL tools such as Informatica, Shell Scripting, and Python.
 - Facilitated knowledge transfer sessions for new resources.
 - Handled structured data projects ranging from 1-10 TB in size.
 - Defined Informatica job flows and processes.
- Gained expertise with Spark using Python.
- Managed loading data from UNIX file system to HDFS.

BI Consultant, 02/2011 - 05/2014

Cognizant (Client: Agilent Technologies) – Chennai, India

ETL & Data Management:

- Gathered client requirements for ETL objects and designed FTP data transfer jobs.
- Loaded data from LINUX file system to RDBMS using Informatica & Cron for automation.

Development & Scripting:

- Analyzed and transferred data to relational databases using Informatica & Python.
- Developed stored procedures and utilized BAPI and ABAP function calls in SAP.

Admin & Database Procedures:

- Administered repository with user logins and privilege assignments.
- Demonstrated expertise in writing and implementing PL/SQL procedures and functions.

SAP Integration:

- Employed mappings in Informatica leveraging BAPI and ABAP.
- Utilized Remote Functional Call (RFC) for inter-system communication.

Informatica Developer, 04/2010 - 01/2011

Cognizant (Client: EBay) – Chennai, India

Informatica Power Center Expertise:

- Utilized tools including Designer, Repository Manager, Workflow Manager, and Workflow Monitor.
- Developed robust mappings using transformations like Filter, Expression, Sequence Generator, and more.

ETL Development:

- Translated high-level design specifications into ETL coding and mapping standards.
- Extracted data from flat files/RDBMS into staging and further populated onto Data Warehouse.

Procedural & Scripting:

- Designed procedures for data extraction across systems into Data Warehousing system.
- Created UNIX shell scripts to automate Informatica sessions.

System Configuration & Data Movement:

- Worked on loading data from various flat files to XML Targets.
- Installed and set up Informatica Client tools, developing ETL mappings for heterogeneous data feeds.

EDUCATION

Bachelor of Science: Information Technology, 05/2009

Mount Zion College of Engineering & Technology - Tamilnadu, India

Aggregate: 80%

OPEN SOURCE PROJECTS

- Teckiy, Developer support platform. <https://www.teckiy.com>
- HelpU, India's first customer experience sharing portal., <https://www.helpu.in>