# NIRMALKUMAR ANGAMUTHU

# 📞 +1 (380) 238-9506|📧 clicknirmal@gmail.com|🌐 [LinkedIn](http://www.linkedin.com/in/nirmalkumar-angamuthu)|📍 Dublin, Ohio

**Professional Summary:**

Experienced Data Engineer Architect with 15+ years in IT and 10+ years in cloud data platforms. Expert in building scalable, secure solutions on **AWS, Azure, GCP, and Databricks.** Skilled in **Airflow, Cloud Composer, Step Functions, Kafka, Kinesis,** and **Dataflow** for real-timedata pipelines. Proficient in **Snowflake, Databricks, Snow pipe, RBAC, PySpark, Terraform, CloudFormation,** and **Cassandra**. Strong background in ETL/ELT, data lake architecture, streaming analytics, and cost optimization.

* Experienced in working with leading cloud platforms including **AWS, GCP**, and Azure for building scalable, secure, and cost-effective data solutions.
* Proficient in AWS services such as EMR, Glue, Lambda, S3, EC2, Redshift Spectrum, Kinesis, and Data Lake for big data and analytics workloads.
* Hands-on experience with **GCP components** like Big Query, Dataflow, Pub/Sub, and Cloud Composer for serverless and streaming data pipelines.Skilled in Azure Synapse, **Azure Data Factory**, and Azure Storage for hybrid cloud data warehousing and integration solutions.
* Extensive experience with **Databricks** for collaborative development, interactive analytics, **Delta Lake** management, Unity Catalog implementation, and optimizing performance across structured and semi-structured data.
* Proficient in **Talend** for end-to-end **ETL development**, including data ingestion, transformation, and job orchestration across heterogeneous systems.
* Expertise in PySpark, Spark SQL, and Scala for processing large-scale datasets in batch and streaming modes.
* Strong Python and SQL programming skills for data transformation, pipeline development, and advanced analytics.
* Experience with orchestration tools like Apache Airflow, AWS Step Functions, Oozie, Control-M, and Prefect for managing complex workflows.
* Built scalable data pipelines using Delta Lake, **Avro, Parquet, Kafka**, and Talend for structured and semi-structured data.Worked extensively with Snowflake and Redshift, including **Snow pipe** for automated ingestion, Streams & Tasks, and RBAC for access control.
* Deep knowledge of relational and **NoSQL** databases such as PostgreSQL, **MySQL, SQL Server, RDS**, Aurora, Teradata, Cassandra, and HBase.
* Skilled in implementing **CI/CD** pipelines using **Jenkins, GitHub Actions**, and Concourse for automated deployments across environments.
* Hands-on with Terraform and **AWS CloudFormation** for Infrastructure as Code, enabling repeatable, scalable infrastructure provisioning.
* Proficient in Docker and Kubernetes for containerization and orchestration of microservices in production environments.
* Experienced in implementing observability using CloudWatch, OpenSearch, Prometheus, **ELK Stack, and Splunk** for monitoring and logging.
* Integrated BI tools such as **Power BI, Tableau, Quick Sight, Looker, and Grafana** for advanced dashboards and real-time visualization.
* Exposure to machine learning techniques using tools like Vector Assembler, KMeans, and PySpark MLlib for data clustering and transformation.
* Well-versed in implementing enterprise-grade security and governance using IAM, **RBAC, SSO, KMS**, and encryption at rest/in transit.
* Developed and enforced metadata management and data quality frameworks using tools like iCEDQ and in-house validation tools.

**Education** **Technical Skills Summary:**

* Bachelor of Technology– Information Technology

**Technical Skills Summary:**

| Category | Tools & Technologies |
| --- | --- |
| Cloud & Big Data | AWS, GCP, Azure, EMR, Glue, S3, EC2, Lambda, Data Lake, Kinesis, Big Query, Azure Synapse, Redshift Spectrum, Snow pipe, Talend, Redis |
| Data Engineering | PySpark, Python, SQL (Advanced), Scala, Airflow, DBT, Spark SQL, Kafka, Delta Lake, Avro, Parquet |
| Data Platforms | Snowflake, Redshift, Databricks, RDS, Aurora DB, PostgreSQL, MySQL, Mongo DB, SQL Server, Teradata, HBase, Cassandra |
| ETL/ELT & Pipelines | AWS Glue, Snow pipe, Talend, DataStage, Dataflow,AutoSys , Fivetran, Cloud Composer |
| DevOps & Automation | Terraform, Git, Jenkins, GitHub Actions, Docker, Kubernetes, CloudFormation, CI/CD, Vault, Concourse |
| Monitoring & log | CloudWatch, Prometheus, OpenSearch, Splunk, ELK Stack  |
| BI & Visualization | Power BI, Tableau, Grafana, Looker, Quick Sight |
| ML & Analytics | Vector Assembler, KMeans |
| Workflow & Scheduling | Apache Airflow, Cloud Composer, Step Functions, Control-M, Oozie, Cron, Prefect |
| Security & Governance | IAM, RBAC, SSO, KMS, Encryption (at Rest & Transit), iCEDQ, Data Quality Frameworks, Metadata Management |

**PROFESSIONAL WORK EXPERIENCE:**

**Strategic System Inc. | Data Engineer | Client: Waste Management | May 2025 - Till Now**

**Project Name**: NDO
**Skills**:Azure, Python, PySpark, AutoSys, Apache Airflow, Databricks, Snowflake, Apache Spark, Hive, Talend, PyDeequ, Apache Nifi, Azure Synapse Analytics, Azure Databricks, Azure Data Factory, Azure Functions, Azure Data Lake Storage Gen2, Azure Event Grid, Azure Event Hubs, Azure Service Bus, Azure Key Vault, Azure Monitor, Azure Log Analytics, Azure API Management (APIM), Azure DevOps, MongoDB, Cassandra, PostgreSQL, MySQL, Azure SQL Database, Docker, Kubernetes, Terraform, Jenkins, Splunk, Prometheus, Grafana, Kafka, RabbitMQ, Delta Lake, Unity Catalog, CI/CD, GitHub Actions, Feature Store, Data Quality, Data Lineage, Data Governance, Streaming Pipelines, Batch Processing, Medallion Architecture, Slowly Changing Dimensions (SCD), CDC (Change Data Capture), ETL, ELT, Parquet, JSON Schema, Data Masking, RBAC, OAuth2, OAuth, Service Principal, ARM Templates, Cloud Security, Observability, Real-Time Analytics, Scalable Architecture, Cost Optimization, DevSecOps, Metadata Management

**Description:** The Mid Atlantic System (MAS) manages customer and ticket information, where tickets include waste type, disposal codes, and container cabin sizes, and are categorized into Day Optimization and Night Optimization. Customers are classified as either Special Access or Normal, with Special Access customers requiring driver permission to enter during designated time windows. Commercial and common data streams are ingested through Kinesis and stored in S3 buckets and a Cassandra database to provide real-time access to ticket and customer service information. Additional commercial, residential, and geocode data is retrieved from Snowflake, while Kronos integration provides driver schedules, actual clock-in/out times, and other driver details. Currently, all data processing and storage systems are being migrated to Azure Databricks, leveraging Delta Lake, Azure Data Lake Storage, and Unity Catalog to unify real-time and batch workloads within a scalable and secure cloud architecture.

**Responsibilities**

* Built and automated cloud-native infrastructure using Azure Resource Manager (ARM) templates, Azure Data Factory, and Azure Databricks Workflows, enabling scalable deployments and real-time ingestion pipelines such as Kronos schedule integration with Delta Lake as part of enterprise Azure cloud adoption.
* Designed and maintained ETL/ELT pipelines using Snowpark for Snowflake, PySpark, Azure Databricks, and DBT, implementing modular model layering (staging, intermediate, mart) and using Jinja macros and custom DBT packages to support reusable, scalable, and testable SQL transformations.
* Migrated legacy SQL logic to DBT, standardizing data transformations and improving code versioning, testing, and lineage visibility; integrated with enterprise data catalog tools to ensure governance and model transparency.
* Orchestrated end-to-end pipelines using MWAA (Managed Workflows for Apache Airflow), coordinating task dependencies across DBT Cloud, Azure Data Factory, and Databricks Jobs, while embedding alerting, retries, and SLA enforcement for mission-critical workflows.
* Led enterprise-scale data platform modernization to Snowflake and Azure Databricks, leveraging Snowpipe, Delta Lake time travel, and RBAC, with dimensional and normalized models governed by Unity Catalog, ERWin, and data dictionary standards.
* Enhanced cloud performance and cost efficiency through Databricks Job optimization, autoscaling compute clusters, and tag-based cost tracking, while strengthening monitoring and governance using Azure Monitor, Log Analytics, and audit trails.
* Supported collaborative analytics and BI environments by delivering production-grade, metadata-driven data pipelines integrated with lineage, glossary, and business metadata, ensuring discoverability and compliance with enterprise governance standards.
* Delivered ML-ready datasets and integrated anomaly detection pipelines by partnering with data science teams to expose clean, feature**-rich data**, supporting both training and inference in **production-grade AI workflows**.

**Infosys Ltd. | Data Engineer Architect | Client: Waste Management | May 2022 – April 2025**

**Project Name**: Build OD
**Skills**: AWS, Python, Airflow, EMR Serverless, DynamoDB, PySpark, Lambda, API Gateway, S3, EC2, Glue, Fivetran, CloudFormation, ECS, Secret Manager, RDS, Redshift, Athena, Kinesis, SNS, SQS, Step Functions, CloudWatch, CloudTrail, VPC, IAM, Route 53, Redis, CloudFront, EFS, EBS, ,CloudFormation, Cognito, Elastic Load Balancing, Auto Scaling, Snowflake, Jenkins, Terraform, Amazon Key space, Splunk, Kubernetes, Docker, Databricks, MongoDB and Cassandra.

**Description:** The Mid Atlantic System (MAS) manages customer and ticket information, where tickets detail waste type, disposal codes, and container cabin sizes. Tickets are categorized into Day Optimization and Night Optimization. Customer-level attributes include classifications such as Special Access or Normal customers; Special Access customers require driver permission to enter their premises, with specific time windows assigned for driver entry. Commercial and common data streams are ingested via Kinesis, then stored in S3 buckets and a Cassandra database for real-time access to customer service, and ticket information. Additional commercial, residential, and geocode data are retrieved from Snowflake. The system also integrates Kronos data to receive driver schedules, actual clock-in/out times, and other driver details from the Kronos application.

**Responsibilities**

* Built and automated infrastructure using AWS CloudFormation, Airflow, and Databricks Workflows, enabling scalable deployments and real-time pipelines (e.g., Kronos schedule ingestion with Delta Lake) as part of enterprise AWS migration strategies.
* Designed and maintained ETL/ELT pipelines using Snowpark for Snowflake, PySpark, and Databricks, integrating CDC strategies via Kafka, AWS DMS, and Snowflake Streams, while ensuring robust data lineage and auditability.
* Led large-scale data migration and modernization projects to Snowflake and Databricks, leveraging Snow pipe, Delta Lake time travel, RBAC, and advanced data modeling techniques across Snowflake, Redshift, Cassandra, and Unity Catalog.
* Improved cloud efficiency through optimized Databricks Jobs, PySpark, EC2, and tag-based cost controls, while enhancing observability using CloudWatch, OpenSearch, and audit logs to support data governance and compliance standards.
* Orchestrated enterprise-grade workflows using AWS Step Functions, Airflow, and Databricks Workflows, managing data across S3, Glue, DynamoDB, and Delta Lake with full IAM and governance policy enforcement.
* Supported data warehousing initiatives and enabled seamless collaboration between engineering and analytics teams by building scalable solutions with embedded data governance, metadata tracking, and compliance with enterprise lineage standards.
* Explored and integrated machine learning solutions for predictive analytics and anomaly detection, supporting data science teams with clean, feature-rich datasets optimized for training, inference, and downstream consumption.

**Cognizant Technology Solutions | Data Engineer | Client: Comcast | Jun 2019 – Apr 2022**

**Project Title**: Digital First
**Skills**: GCP (Cloud Storage, Cloud SQL, Compute Engine, Cloud Composer, API Gateway, Cloud Monitoring, Cloud CDN, IAM, Pub/Sub, Deployment Manager), Python, SQL, Vue.js, Webhooks, Jenkins, Concourse CI, PySpark, Dataflow (ETL), Spark SQL, Docker, Secret Manager (Vault equivalent), Snowflake, Grafana, Prometheus, MySQL, Agile (Scrum, Kanban)

**Description**: Developed a Security Vulnerability Dashboard leveraging GCP services, including API Gateway, Cloud Functions, and Cloud SQL, to monitor and track high, medium, and low-severity security patches with real-time data visualization in Grafana. Automated data ingestion from ServiceNow (SNOW) REST APIs using Python-based and built an end-to-end ETL pipeline using Cloud Scheduler, Pub/Sub, Dataflow (PySpark), and Spark SQL, seamlessly updating vulnerability data in BigQuery for advanced analytics. Led Agile ceremonies, including daily stand-ups.Top of Form

**Responsibilities**

Bottom of Form

* Built and optimized ETL pipelines for Cirrus, Catamaran, and Polaris using Google BigQuery, Cloud Composer (Airflow), and Spark, handling claims data from Teradata DWH and storing in ORC, Avro, and JSON.
* Designed and deployed Data Lakes and star schema models in BigQuery, with seamless integration to Snowflake for cross-platform analytics and reporting.
* Authored complex SQL for high-volume analysis in BigQuery and Snowflake, delivering real-time business reports and CSV outputs for downstream systems.
* Developed a Vulnerability Stats Dashboard in Grafana, sourcing data from Cloud SQL, and built Python/SQL-based ETL scripts to support business-critical insights.
* Streamlined IaC with Terraform, integrating with CI/CD tools like Jenkins, GitHub Actions, and Azure DevOps, while managing Terraform states in Git for infra consistency.

**Continuity2 | Tech Lead | Jul 2017 – Jun 2019**

**Project Title:** Business Continuity Management

**Skills:** AWS, Python, MySQL, Spark, Glue, Lambda, API Gateway, Aurora DB, IAM, EC2, Data Pipeline

**Description:** Developed Business Continuity Management (BCM) encompasses three key processes: risk identification, risk analysis, and risk evaluation. Risk identification involves discovering and describing risks that could impact the achievement of organizational objectives. Risk analysis examines the nature, sources, and causes of these risks, estimating their levels while assessing impacts, consequences, and existing controls. Risk evaluation compares the analysis results against established risk criteria to determine if the risk levels are acceptable or require mitigation. The Business Impact Analysis (BIA) within BCM identifies key products, services, and supporting activities, analysing them according to organizational impact criteria with a fully configurable system aligned to the organization's approach

**Responsibilities**

* Built vulnerability dashboards using Grafana from Aurora DB sources. Automated weekly ETL using Lambda, CloudWatch, and Glue Jobs.
* Created reusable Terraform modules for AWS infrastructure provisioning. Worked on ETL jobs for claims processing using Spark and DataStage.
* Designed UI apps to extract data from Teradata into SAS/Flat files. performance tuning for faster data delivery. Collaborated with QA for validation, unit testing.

**Cognizant Technology Solutions | Software Engineer | Client: Comcast | Feb 2014 – Jul 2017**

**Project Name:** Business Voice Edge
**Skills**: Azure Function, HDFS, MapReduce, Pig, Hive, Sqoop, MySQL, Azure Storage, Azure Data Factory, Azure Synapse Analytics, Azure Databricks, Azure Event Hubs, Azure Blob Storage, Azure Logic Apps, Azure Key Vault, Azure Monitor, Azure DevOps, Power BI, Data Lake Storage Gen2, PolyBase, Apache Oozie, Apache Flume, Apache Kafka, Spark SQL, ETL pipelines, Big Data Analytics, Data Warehousing, SQL Server, NoSQL, Containerization, Docker, Agile methodologies.

**Description**: The aim of this project is to provide Comcast business customers with enhanced digital self-service experience, focusing on self-service evolution, 360 customers view and product enhancements. Today customers have a decentralized portal experience to manage their account and subscribed services. They may have multiple portals with multiple services to manage their subscribed services. The user experience is difficult, and the self-service options are limited today. This project, through a new Business Class Portal, will extend tools to business class customers to manage their account and their subscribed services from within a single web portal with a single set of login credentials.

**Responsibilities**

* Build scalable and maintainable backend logic using Python, JavaScript (Node.js), or other languages.
* Connect backend code with cloud-native services like Azure Functions, S3, Blob, and RDS/MySQL. Trigger workflows based on events. Develop APIs that expose application functionality securely and efficiently.

**HCL Technologies | Software Engineer | Client: AIG | Jun 2010 – Feb 2014**

**Project Title:** Smart Pad

**Skills:** Python, HDFS, Map Reduce, Pig, Hive, Sqoop, MySQL, SSIS and SQL server 2008

**Description**: Smartpad System for AGLA Field Agents: Utilized Smartpad, a tablet-based system, as a primary tool for AGLA field agents (Marketing Team), uniquely identified by an Enterprise Code.

Managed business functions across multiple systems: POS (Point of Service): Facilitated customer transactions and service requests.CSF (Customer Service): Handled customer queries and support activities.NBA (New Business Application): Processed new insurance applications.BNA (Beneficiary Information). Ensured daily transaction data from all Smartpads was synchronized with the Excellent Scheduler by 11 PM, which transferred data to a secure mailbox. Configured the mailbox to create folders identified by Enterprise Code (e.g., F2RPOS.txt, F2RCSF.txt), with text files containing policyholder details delimited by a pipeline (|) symbol.

**Responsibilities**

* Developed and maintained data ingestion workflows using Sqoop, Hive, and Pig to process transactional data from Smartpad devices into HDFS for centralized analysis.
* Designed and executed MapReduce programs and Hive queries to transform, merge, and analyze policyholder data (age, salary, tobacco status) for accurate premium calculations.
* Automated synchronization of daily Smartpad transactions to the Excellent Scheduler, ensuring secure and timely data delivery to mailbox folders identified by Enterprise Code.
* Utilized Python scripts to parse pipe-delimited files (e.g., F2RPOS.txt, F2RCSF.txt), and integrated data from POS, CSF, NBA, and BNA systems into structured formats.
* Supported data persistence and validation across MySQL and SQL Server 2008, maintaining integrity for new insurance applications and beneficiary records

**OconnorComps | Software Engineer | Nov 2008 – May 2010**

**Project Title**: O’Connor Management Suite

**Skills**: ASP.NET 3.5, C#, JavaScript and SQL Server 2008.

**Description:** Developed a comprehensive marketing search engine for property tax consultants and real estate appraisers in the USA, covering 13 core counties and maintaining a database of property and appraisal data spanning over five years. Implemented core modules, including Comparable Property Analysis, Residential and Commercial Sales Study, Comparable Sales Analysis, Control Panel, and Multiple Listing Service (MLS) integration. Enabled users to search and filter properties using criteria like SQFT, Year Built, Grade, and CDU, allowing them to generate customized report documents for tax protests.

**Responsibilities**

* Creating study documents, gathering requirements, and performing in-depth analysis to understand business needs.
* Suggesting technical improvements and contributing to continuous enhancement of software development processes. Participating in code reviews to maintain coding standards and improve code Quality.

**CAP Digisoft Solutions | Software Engineer | Dec 2006 – Oct 2008**

**Project Title**: Cap Connect **Skills:** C# , SQL Server 2008, Ajax and Infragistics

**Description:** Developed a desktop application for viewing, annotating, and exporting files (image, native, and text formats) with advanced search capabilities. Enabled users to print files in HTML format and ensured seamless software maintenance through a one-touch deployment method for automatic updates. Designed and implemented Cap Connect, a secure platform for hosting and reviewing litigation documents, allowing users to search, organize, and review documents using custom criteria. Integrated an Uploader and Downloader module, a desktop engine that operates in the system tray to automatically upload and download files via FTP, based on server requests. Implemented pause and resume functionality for file transfers and established an integrity check process to verify that uploaded or downloaded files match the original files, ensuring reliable data transfer.

**Responsibilities**

* Creating study documents, gathering requirements, and performing in-depth analysis to understand business needs.
* Designing intuitive and efficient screen flows, UI layouts, and wireframes based on functional specifications.
* Coding, unit testing, and debugging applications to ensure high-quality deliverables.
* Participating in code reviews to maintain coding standards and improve code quality.