



ANIL PANTHAM

DATA ENGINEER

Email: panthamanil8@gmail.com

Contact: (937) 342-3893

LinkedIn: [Anil Pantham](#)

PROFESSIONAL SUMMARY:

- Senior Data Engineer with 4 years of experience in building on-premises and cloud-based data solutions using cutting edge tools like AWS, Azure Data Factory, Data bricks, Spark, SQL, and Python.
- working experience in AWS, Spark, Hadoop Ecosystem, Python for Data science, data pipeline and Tableau.
- Strong experience in AWS S3, EMR, EC2, Glue, Lambda, IAM, Kinesis, RDS, Route 53, VPC, Code build, Code pipeline, Cloud watch and Cloud formation.
- Adept at using Databricks for developing ETL and ELT processes using PySpark, Spark-SQL, and Scala
- Proficient in building and orchestrating ETL pipelines using Azure Data Factory and SQL Server Integration Services (SSIS).
- Experienced in implementation of Lakehouse architecture on Azure using Azure Data Lake, Delta Lake, Delta Tables, and Databricks.
- Experience in transferring the data using Informatica tool from AWS S3 to AWS Redshift.
- Hands on experience with Amazon Web Services along with provisioning and maintaining AWS resources such as EMR, S3 buckets, EC2instances, RDS and others.
- Worked in agile projects delivering end to end continuous integration/continuous delivery pipeline by Integration of tools like Jenkins and AWS for VM provisioning.
- In-depth understanding and usage of various Databricks optimization techniques like Partitioning, Bucketing, Adaptive Query Execution, DAG, Dynamic File Pruning, Data Skipping, Optimize, Bin-packing, Z-Order clustering on delta tables.
- Well versed in data warehousing and normalization techniques for architecting data models including star schema, snowflake schema, and slowly changing dimensions etc.
- Experience in provisioning access to Azure Data Lake from Databricks using Secret Scope, and mounting the storage account into Databricks.
- Experienced in large data migration projects from on-prem databases to Azure SQL and Azure Data warehouse.
- Experienced in designing and building data models for OLTP and OLAP applications for a variety of business domains.
- Worked on Amazon Web service (AWS) to integrate EMR wif Spark 2 and S3 storage and Snowflake.
- Expert in writing efficient SQL queries and optimizing poor performing queries and stored procedures.
- Solid experience in working as a data analyst to manage, document, analyze, and resolve data related issues.
- Pragmatic and proactive in performance tuning on OLTP and Data Warehouse systems (OLAP) for fast and efficient data load and data retrieval.



TECHNICAL SKILLS:

■ Azure Data Factory, SSIS	■ Databricks, PySpark	■ Azure Data Lake, Data center, Delta Lake, Snowflake
■ AWS, Azure	■ IAM, S3, EC2, VPC, EMR,	■ Cloud Formation, Lambda
■ SQL Server, Oracle ■ Spark-SQL	■ Python, Scala, PySpark	■ Azure Key Vault, Logic Apps

PROFESSIONAL EXPERIENCE:

CareSource, Dayton, OH


Aug 2023 – Current

AWS Data Engineer

Responsibilities:

- As a Data Engineer, assisted in leading the plan, building, and running states within the Enterprise Analytics Team.
- Worked in AWS environment for development and deployment of Custom Hadoop Applications.
- Design and Develop ETL Processes in AWS Glue to migrate Campaign data from external sources like S3, ORC/Parquet/Text Files into AWS Redshift
- Lead architecture and design of data processing, ware housing and analytics initiatives.
- Engaged in solving and supporting real business issues with your Hadoop distributed File systems and Open-Source framework knowledge.
- Responsible for data governance rules and standards to maintain the consistency of the business element names in the different data layers.
- Involved in various phases of development analyzed and developed the system going through Agile Scrum methodology.
- Performed detailed analysis of business problems and technical environments and use this data in designing the solution and maintaining data architecture.
- Build a program with Python and Apache beam and execute it in cloud Data flow to run Data validation between raw source file and Big query tables.
- Built the data pipelines that will enable faster, better, data-informed decision-making within the business.
- Used Rest API with Python to ingest Data from and some other site to Big Query.
- Loaded and transformed large sets of structured, semi structured and unstructured data using Hadoop/Big Data concepts.
- Performed Data transformations in Hive and used partitions, buckets for performance improvements.
- Optimized Hive queries to extract the customer information from HDFS.
- Strong experience in working with ELASTIC MAPREDUCE(EMR) and setting up environments on Amazon AWS EC2 instances
- Involved in scheduling Oozie workflow engine to run multiple Hive jobs.



- 
- Continuously monitor and manage data pipeline (CI/CD) performance alongside applications from a single console with GCP.
 - Developed Spark scripts by using python and bash Shell commands as per the requirement.
 - Compared Self hosted Hadoop with respect to GCPs Data Proc, and explored Big Table (managed HBase) use cases, performance evolution.
 - Write a Python program to maintain raw file archival in GCS bucket.
 - Implemented business logic by writing UDFs and configuring CRON Jobs.
 - Implemented scripts that load Google Big Query data and run queries to export data.

National Institute of Mental Health (NIMH), Bethesda, MD

Aug 2022 – Jul 2023

Azure Data Engineer

Responsibilities:

- Develop, design data models, data structures and ETL jobs for data acquisition and manipulation purposes.
- Develop deep understanding of the data sources, implement data standards, maintain data quality, and master data management.
- Worked with various file formats like CSV, JSON, Parquet, and Snappy Parquet.
- Used Python and Pyspark in Synapse notebooks to efficiently process huge amounts of data and load them into delta lake silver layer
- Utilized Delta caching for faster data processing using Databricks and PySpark.
- Worked on improving poorly performing PySpark code by using coalesce, repartition, query hints, broadcast joins, and delta caching.
- Involved in optimized Delta tables by using Optimize command and applying Z-Order Clustering.
- Transformed batch data from several tables containing tens of thousands of records from SQL Server, MySQL, PostgreSQL, and csv file datasets into data frames using PySpark
- Performed ETL operations in Azure Databricks by connecting to different relational database source systems using JDBC connectors.
- Developed Python scripts to do file validations in Databricks and automated the process using ADF.
- Developed an automated process in Azure cloud which can ingest data daily from web service and load into Azure Data Lake Gen2.
- Used Logic App to take decisional actions based on the workflow.
- Developed custom alerts using Azure Data Factory, SQLDB and Logic App.
- Developed Databricks ETL pipelines using notebooks, Spark Data frames, SPARK SQL, and python scripting.
- Developed complex SQL queries using stored procedures, common table expressions (CTEs), temporary table to support Power BI reports.
- Worked with enterprise Data Modeling team on creation of Logical models.
- Development level experience in Microsoft Azure providing data movement and scheduling functionality to cloud-based technologies such as Azure Blob Storage and Azure SQL Database.
- Worked on developing JSON Scripts for deploying the Pipeline in Azure Data Factory (ADF) that process the data.
- Independently manage development of ETL processes – development to delivery.

American Express, India


Jun 2020 – Dec 2021

Data Engineer

Responsibilities:

- Develop proof of concepts (POC) using Azure Cloud services such as Azure Data factory (ADF), Logic Apps, Data Bricks, Azure SQL DB, Power BI and Azure Data Lake storage (ADLS) to validate proposed solutions and seek feedback from stakeholders.



- 
- Working on getting the data from on prem SQL server database to a Raw Zone in Azure Data Lake by using Azure Data Factory.
 - Used Parquet file formats on Azure Blobs to store raw data.
 - Monitored the SQL scripts and modified them for improved performance using PySpark SQL
 - Develop Azure Databricks notebooks to apply business transformations and perform data cleansing operations. Used spark, python, delta while reading data from parquet files in ADLS to Data frame and writing to destination ADLS location. Partitioned data and implemented schema drift by using metadata JSON file before writing data to destination location in ADLS.
 - Implemented ETL platform using Azure Data Factory, Data Bricks, Data Lake, Azure SQL.
 - Create External tables in Azure SQL Database for data visualization and reporting purpose.
 - Performance improvement of existing views and procs and identify deadlocks, missing indexes and alter indexes where necessary
 - Created T-SQL Stored Procedures, User Defined Functions, Indexes, Views, Constraints, Triggers and wrote complex queries using Joins, Aggregate Functions, Sub-queries, Derived tables and CTEs
 - Implemented SCD type1, SCDType2 logic using Azure data bricks for Delta tables
 - Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Azure Databricks, and Azure Data Lake Analytics.
 - Created Pipelines in ADF using Linked Services/Datasets/Pipeline to Extract, Transform and load data from different sources like Azure SQL, Blob Storage, and Azure SQL Data warehouse.
 - Working on data Ingestion & transformation from different source systems using Blob Storage/ADLS into Data bricks and processing.
 - Develop and maintain application code in Databricks by using appropriate data structures and algorithms for optimal performance and better storage that increase the speed and consistency of the application.
 - Created CI/CD Pipelines in azure Devops to build and deploy data factory ARM templates
 - Loaded the data into facts and dimensions based on Data modeling using Azure Data factory
 - Scheduling ADF Pipelines as per the data dependency
 - Used Key Vault to maintain secrets in Azure environment.
 - Used Azure Devops to maintain Product Backlog Items, User Stories, Tasks and Bugs.

Goldman Sachs, India


Jan 2019 – Jun 2020

BI Developer

Responsibilities:

- Created Complex ETL Packages using SSIS to extract data from staging tables to base tables, performing sanity check for integrity of the data and removal of bad data.
- Designed and developed SSIS Packages to extract data from various data sources such as Access database, Excel spreadsheet and flat files into SQL server for further Data Analysis and Reporting by using multiple transformations provided by SSIS such as Data Conversion, Conditional Split, merge, union all and lookup transformation and send mail task.
- Analyze source system data, profile data, create metadata definitions; define data attributes for data flow & lineage and create source-to-target mapping.
- Use Power BI to pull Data from other OLTP AND OLAP systems and from flat files and Excel files to build dashboards and Reports for the business users as Required
- Document technical ETL specifications for the data warehouse. Perform periodic T-SQL code reviews and test plans to ensure high performance, data quality and integrity.
- Used T-SQL to develop and implement procedures and functions. Reviewed and interpreted ongoing business report requirements.
- Create database objects like Stored procedure, Functions, Views, indexes, and triggers to base on the business requirements documents and user requests



- 
- Collaborate with team members in the Dev environment to promote all T-SQL Code, SSIS Package, Reports and all other Database Object into source control using Team Foundation Server (TFS), and GitHub.
 - Designed cubes to satisfy the senior management requirement, deployed the cube into different environments and created KPI's and dashboard reports and extensively written MDX queries in report generation
 - Involved in Performance Tuning of Stored Procedures and SQL queries using SQL Profiler and Index Tuning Wizard in SSIS.
 - Generated functional and technical documentation including as-built documents, dashboard design specifications, operating manuals, release notes, and security requirements for reports generated and maintained.

Education:

Master of Information Technology from University of Cumberland's, Williamsburg, KY

Certifications:

- Microsoft Azure Fundamental
- Microsoft Azure Administrator: AZ104
- AWS Solution Architect – Associate

