# AVINASH

## Data Engineer| [avinashycsm@gmail.com](mailto:avinashycsm@gmail.com) | (571)-210-5323

**Professional Summary:**

* Around 8 years of professional experience with a proven expertise in cloud technologies such as Azure and AWS, specializing in designing and implementing data pipelines, data storage solutions, and data warehousing systems, while also excelling in real-time data processing, big data platforms, and business intelligence data architecture.
* Good knowledge on Azure Cloud, Azure App Services, Azure Data Factory, Azure Data Lake Storage, Azure Synapse Analytics, Azure Analytical services, Azure Cosmos NO SQL DB, Azure HDInsight Big Data Technologies (Hadoop and Apache Spark), Data bricks, Event Hub and Event grid.
* Experience working in reading Continuous Json data from different source system using Kafka into Databricks Delta and processing the files using Apache Structured streaming, Pyspark and creating the files in parquet format.
* Created data pipelines for both batch process, Micro-batch streaming and continuous streaming process in Databricks for high latency, low latency, and ultra-low latency of data accordingly by using inbuilt Apache spark modules.
* Well versed and experienced in creating pipelines in Azure Cloud ADFv2 using different activities like Move &Transform, Copy, filter, for each, Data bricks etc.
* Providing Azure technical expertise including strategic design and architectural mentorship, assessments, POCs, etc., in support of the overall sales lifecycle or consulting engagement process.
* Hands on experience in Hadoop Ecosystem components such as Hadoop, Spark, HDFS, YARN, TEZ, Hive, Sqoop, MapReduce, Pig, OOZIE, Kafka, Storm, HBASE.
* In-depth understanding of Spark Architecture including Spark Core, Spark SQL, Data Frames and Spark Streaming.
* Experience in designing and implementing data pipelines, data storage solutions, and data warehousing systems using AWS tools such as S3, RDS, DynamoDB, Redshift, and Athena.
* Experience in implementing data security and privacy policies to ensure the confidentiality and integrity of data using AWS tools such as IAM and VPC.
* Strong understanding of data architecture and design principles and the ability to develop and implement scalable data solutions using AWS services such as EC2, Glue, and Lambda.
* Ability to perform data analytics, predictive modeling, and data-driven decision-making using Azure tools such as HDInsight, Data Factory, and Synapse Analytics.
* Experience in working with data streams and real-time data processing systems using Azure tools such as Event Hubs, Stream Analytics, and Service Bus.
* Skilled in automating data migration processes using Azure Data Factory and scheduling pipelines for timely data updates.
* Designed and optimized data pipelines using Azure Data Factory, Azure Data Lake Storage, and Databricks Spark jobs with PySpark, specifically tailored to integrate diverse data sources with Adobe Experience Platform, ensuring data consistency and quality for personalized customer experiences..
* Strong knowledge of big data platforms, including Hadoop, Spark, and Cassandra, and their applications in data processing and analytics.
* Strong experience in writing applications using Python using different libraries like Pandas**,** NumPy, SciPy, Matplotlib etc.
* Good understanding of NoSQL databases and hands on work experience in writing applications on NoSQL databases like CosmosDB.
* Experience in performing Data Modelling by designing Conceptual, Logical data models and translating them to Physical data models for high volume datasets from various sources like Oracle, Teradata, Vertica, and SQL Server by using Erwin tool.
* Expert knowledge and experience in Business Intelligence Data Architecture, Data Management and Modeling to integrate multiple, complex data sources that are transactional and non-transactional, structured, and unstructured.
* Also, design and develop relational databases for collecting and storing data and build and design data input and data collection mechanisms.
* Well versed with Relational and Dimensional Modeling techniques like Star, Snowflake Schema, OLTP, OLAP, Normalization, Fact and Dimensional Tables.
* Good knowledge in creating SQL queries, collecting statistics and Teradata SQL query performance tuning techniques and Optimizer/explain plan.

**Technical Skills:**

|  |  |
| --- | --- |
| Azure Cloud Platform | ADFv2, BLOB Storage, ADLS2, Azure SQL DB, SQL server, Azure Synapse, Azure Analytic Services, Data bricks, Mapping Dataflow (MDF), Azure Cosmos DB, Azure Stream Analytics, Azure Event Hub, Azure Machine Learning, App Services, Logic Apps, Event Grid,  Service Bus, Azure DevOps, GIT Repository Management, ARM Templates |
| Reporting and BI Tools | Power BI, Tableau, Tableau Prep and Cognos |
| ETL Tools: | ADFV2, Informatica Power Center 10.x/9.x, DataStage 11.x/9.x, SSIS |
| Programming Languages | PySpark, Python, U-SQL, T-SQL, LINUX Shell Scripting, AZURE PowerShell, C#, Java |
| Big data Technologies | Hadoop, HDFS, Hive, Apache Spark, Apache Kafka, Pig, Zookeeper,  Sqoop, Oozie, HBASE, YARN |
| Databases | Azure SQL Warehouse, Azure SQL DB, Azure Cosmos No SQL DB, Oracle, Microsoft SQL Server |
| IDE and Tools | Code, Eclipse, NetBeans, SSMS, Maven, SBT, MS-Project, GitHub, Microsoft  Visual Studio |
| Cloud Stack | AWS, GCP, Azure, Snowflake |
| Methodologies | Waterfall, Agile/Scrum, SDLC |

**Professional Experience:**

**Client: CVS Health, RI Aug 2022 - present**

**Role: Azure Data Engineer**

**Responsibilities:**

* Collaborated with Business Analysts and Solution Architects to gather client requirements and translated them into Azure-based design architectures.
* Designed and maintained data pipelines using Azure Data Factory, achieving reduction in data processing time.
* Created High-Level Technical Design and Application Design documents, ensuring clear and comprehensive documentation for stakeholders.
* Developed Data Flow specifications and mapping for developers, facilitating efficient data processing.
* Collaborated closely with Business Analysts and Solution Architects to ensure seamless alignment of Azure-based design architectures with Adobe Experience Platform Integration patterns
* Designed and implemented interfaces using Azure Data Share for seamless file transfers.
* Engineered complex data transformations and manipulations using ADF and PySpark with Databricks.
* leveraged Spark and Scala for efficient data processing and analytics tasks. Spark was used to handle large-scale data transformations and computations, while Scala provided a robust and functional programming language for developing scalable data pipelines.
* Leveraged U-SQL scripts to ingest and transform data into Azure Data Warehouse.
* Designed and implemented real-time market data processing solution using Azure Stream Analytics, Azure Event Hub, and Service Bus Queue.
* Managed data ingestion using Azure Data Lake Storage and created pipelines with Azure Data Factory v2, extracting data from diverse sources.
* Created numerous Databricks Spark jobs with PySpark for various data operations.
* Implemented a Power BI integration module for canned reports from ADLS2.
* Proficiently used SQL Server Import and Export Data tool.
* Collaborated with cross-functional teams and assisted in troubleshooting, risk identification, and resolution.
* Led team members in technical issue resolution and mentored for skill development.
* Developed and maintained ETL processes using SQL Server Integration Services (SSIS) to extract, transform, and load data into a centralized data warehouse.
* Collaborated with stakeholders to gather data requirements and optimize SQL queries, resulting in improvement in query performance.
* Implemented data quality checks and validation procedures to ensure data accuracy.
* Managed database performance through index optimization and maintenance tasks.

**Environment:** Azure Cloud, Azure Data Factory (ADF v2), Azure functions Apps, Azure Data Lake, Blob Storage, SQL server, Teradata Utilities, Windows remote desktop, Unix, Azure PowerShell, Data bricks, Python, PySpark, Azure Cosmos DB, Azure Stream Analytics, Azure Event Hub, Power BI.

**Client: Wells Fargo, NC Jan 2021 – Jul 2022**

**Role: Azure Data Engineer**

**Responsibilities:**

* Ensured Wells Fargo's customer eligibility system complied with Securities and Exchange Commission (SEC) regulations for data collection, storage, and usage within the Azure environment.
* Spearheaded data privacy and security measures, maintaining full compliance with SEC guidelines for handling securities-related data.
* Designed and implemented data architectures for managing large volumes of home loans data while adhering to GDPR and CCPA data privacy and security regulations.
* Developed and implemented robust data governance policies and procedures within the Azure ecosystem, improving data quality, accuracy, and reliability.
* Established a culture of data stewardship, ensuring data usage remained aligned with SEC regulations.
* Led the development and implementation of a transformative customer eligibility project within the Azure platform.
* Utilized cutting-edge data analytics techniques to analyze customers' financial and personal data, revolutionizing the process of determining eligibility for auto financing.
* Designed and implemented an efficient Extract, Transform, Load (ETL) architecture using Azure services for seamless data transfer from source servers to the Data Warehouse.
* Implemented automated data cleansing and integration processes, resulting in improved data quality.
* Demonstrated a deep understanding of the Software Development Life Cycle (SDLC) principles, particularly in the context of campaign platforms. This included requirements gathering, design, development, testing, deployment, and maintenance phases, ensuring the successful delivery of campaign-related data solutions.
* Actively participated in designing and developing CI/CD pipelines for data engineering within the Azure ecosystem.
* Implemented automation from code commit to deployment using Azure DevOps.
* Managed a cloud data warehouse on Azure, facilitating batch processing and streaming.
* Enhanced data visualization for customer data using Azure Power BI.

**Environment:** Azure Data factory, Azure Databricks, Azure Event Hubs, Azure SQL   
Datawarehouse, Power BI.

**Client: Molina Healthcare, CA Jan 2017 – Jul 2019  
AWS Data Engineer**

**Responsibilities:**

* Examined claims and supporting documentation to ensure policy compliance before processing.
* Developed a strong understanding of claim processing from both client and service provider perspectives, identifying key metrics for each.
* led the migration of the data warehouse to the AWS cloud platform, leveraging services such as AWS Glue for ETL and data cataloging.
* Managed policy servicing and maintenance operations, including coverage changes, beneficiary data updates, and premium payments.
* Processed claims data efficiently through the system.
* Possess a good understanding of Electronic Health Record (EHR) systems, including their functionalities, data models, data elements, and data privacy and security regulations.
* Designed high-performance batch ETL pipelines using AWS cloud services.
* Extracted data from relational databases and APIs using AWS Glue to seamlessly store it in Amazon S3, which serves as AWS data lake storage.
* Utilized PySpark scripts in AWS Glue within the AWS Databricks environment for efficient data transformations and conversions.
* Designed data warehousing solutions leveraging AWS Redshift, a powerful service within AWS, for storing and analyzing transformed data.
* Implemented and designed Python microservices in the healthcare domain.
* Monitored productivity and resources efficiently by utilizing AWS CloudWatch Logs for log analytics.
* Implemented CI/CD pipelines with AWS DevOps for automated build, test, and deployment processes.
* Utilized AWS Event Hub to capture real-time data streams and route them to the appropriate data stores.
* Monitored data pipeline performance using AWS Monitoring and Analytics tools to ensure seamless data flow and identify potential bottlenecks.
* Played a critical role in a data migration project involving EHR, ensuring accurate, efficient, and secure data migration.
* Ensured data pipeline security using AWS security features, including role-based access control and encryption, to safeguard data privacy and confidentiality.
* Managed encryption keys and passwords through AWS Key Management Service.
* Utilized AWS Kinesis to orchestrate complex business processes and workflows.
* Implemented serverless computing solutions using AWS Lambda Functions for cost-effective and scalable data processing.
* Designed visualization dashboards for data analytics using Power BI.
* Proficient in practicing Agile methodology to update workflows and manage project lifecycles and sprints.

**Environment:** AWS Glue, Amazon S3, AWS Redshift, Python, AWS Kinesis, AWS Step Functions , Key Management Service AWS, Log Analytics, Scala, Power BI.

**Client:** **Deloitte, Hyderabad May 2015 – Dec 2016**

**Data Engineer Associate**

**Responsibilities:**

* Implemented data validation, cleansing, and enrichment processes to ensure compliance with regulatory requirements, including SEC and Finra rules.
* Developed and automated multiple ETL jobs using Amazon EMR, facilitating seamless data transfer from HDFS to S3.
* Created batch data pipelines for extracting data from S3 and loading it into RedShift using Glue jobs.
* Utilized PySpark and Scala to automate data ingestion from various sources, including APIs, AWS S3, and Redshift.
* Configured Spark streaming to store and process real-time data from Kafka.
* Leveraged AWS EMR to store structured data in Hive and unstructured data in HBase.
* Cleaned and transformed data in HDFS using MapReduce (YARN) programs for ingestion into Hive schemas.
* Developed and maintained data reporting and analytics solutions to support regulatory reporting and compliance monitoring.
* Created a data lake in Snowflake using Stitch, App Testing, and Production support.
* Managed S3 buckets, implemented policies, and utilized S3 and Glacier for storage and backup on AWS.
* Generated reports for the BI team by exporting analyzed data to relational databases for visualization using Sqoop.
* Created custom User Defined Functions (UDFs) to extend Hive and Pig core functionality.
* Enabled ODBC/JDBC data connectivity to Hive tables and worked with tools like Tableau and Flink

**Environment:** AWS S3, Glue, AWS EMR, Glacier, Redshift, Snowflake, Spark SQL, Sqoop, Flink,

YARN, Kafka, MapReduce, Hadoop, HDFS, Hive, Tableau, Spotfire, HBase.

**Education:  
University of Texas Arlington**  **Dec 2020**

Masters in Computer Science

**KL University** **May 2015**  
Bachelor’s in Computer Science