

~~ADITYA NITTALA~~

*~~Data Engineer~~*

~~Mail:~~ [*~~aditnit208@gmail.com~~*](mailto:aditnit208@gmail.com) *~~Phone: +1 972-665-7318~~*

**~~PROFESSIONAL SUMMARY~~**

* ~~9+ years experienced professional with a proven track record of successfully leveraging Azure Data~~ ~~Engineering and its associated technologies, such as~~ **~~Azure Data Factory~~**~~,~~ **~~Azure Databricks~~**~~,~~ **~~Logical~~ ~~Apps~~**~~,~~ **~~Azure Function Apps~~**~~,~~ **~~Snowflake~~**~~, and~~ **~~Azure DevOps~~** ~~services.~~
* Strong development proficiency in building scalable and efficient **ETL (Extract, Transform, Load) data ~~pipelines~~** ~~using~~ **~~PySpark~~** ~~and~~ **~~Spark SQL~~** ~~within~~ **~~Azure Databricks~~**.
* Competent in developing high-performance data pipelines using Scala, optimizing data processing, and ~~transforming complex data into valuable insights.~~
* Skilled in utilizing **Azure Logic App Integration tool** to seamlessly integrate diverse data sources and ~~streamline data workflows.~~
* ~~Strong understanding of data governance principles and best practices, ensuring data quality, security, and~~ ~~compliance.~~
* Strong implementation experience in fully implementing **Azure Functions**, **Azure Storage Services**, and

**~~Service Bus~~** ~~queries to seamlessly integrate a large-scale ERP system for enterprise-level integration.~~

* Hands-On experience in creating and managing **Azure DevOps** tools, enabling **efficient continuous ~~integration and deployment (CI/CD) pipelines~~** ~~for streamlined application delivery.~~
* ~~Developed robust data ingestion workflows, enabling the seamless extraction of data from diverse sources,~~ ~~and transforming it into Avro, Parquet, Sequence, JSON, and ORC file formats for optimized storage and~~ ~~rapid retrieval.~~
* ~~Extensive experience in Hadoop ecosystem technologies, including HDFS, Map-Reduce, Hive, Tez, and~~ ~~Sqoop~~.
* Proficient in developing large-scale data pipelines using Spark and Hive, optimizing performance through ~~query tuning, bucketing, and partitioning techniques.~~
* ~~Hands-on expertise in importing and exporting data between HDFS and Relational Database systems using~~

**~~Sqoop~~**.

* ~~Leveraged the powerful capabilities of~~ **~~Apache Oozie~~** ~~for efficient scheduling and comprehensive~~ ~~management of Hadoop jobs.~~
* Extensive hands-on experience in tuning Spark jobs and ETL workflows for optimal performance and ~~efficiency.~~
* Strong background in data ingestion, data modelling, and encryption techniques for secure and efficient ~~data processing~~
* Proficiency In **Apache Kafka**-driven **real-time streaming** analytics in Spark Streaming, enabling efficient ~~processing and analysis of high-velocity streaming data, while effectively utilizing Kafka as a~~ **~~fault-tolerant~~ data pipeline**.

**Commented [AN1]:** Add 2 more Kafka Points. Add more domain knowledge

* ~~Optimized Spark jobs and workflows through fine-tuning of~~ **~~Spark configurations~~**~~,~~ **~~strategic partitioning~~**, ~~and memory allocation, resulting in enhanced performance and resource utilization.~~
* ~~Demonstrated expertise in managing Snowflake's unique features such as~~ **~~Zero-Copy Cloning, Time~~ ~~Travel~~**~~, and~~ **~~Data Sharing~~**~~, for efficient data management~~.
* ~~Successfully~~ **~~integrated Snowflake with Azure Data Factory~~** ~~to orchestrate complex ETL pipelines,~~ ~~significantly optimizing data migration from diverse sources into Azure-based data warehouses.~~
* ~~Extensive experience in the development, maintenance, and implementation of Enterprise Data~~ ~~Warehouses (EDW), Data Marts, Operational Data Stores (ODS), and Data Warehouses, employing both~~ ~~Star schema and Snowflake schema designs.~~

**~~EDUCATION~~**



~~Azure data Factory, Azure Data~~ ~~Bricks, Logic Apps, Functiona~~l ~~App, Snowflake, Azure DevOps~~

**Details Links**

[LinkedIn](https://www.linkedin.com/in/aditya-nittala-b78b196a/)

**Skills**

AWS Services Python, R MySQL

Python Scripting Amazon AWS

NumPy, Pandas, Matplotlib Tableau Dashboards

Use Cases

Jupyter Notebook, PyCharm SDLC, Waterfall

* ~~Master of Science in Data Science~~.

**~~TECHNICAL SKILLS~~**

**~~Azure Services~~Aditya Nittala**

BUSINESS ANALYST

# Profile

I am a driven and diligent professional with exceptional communication skills that I have used to translate business requirements efficiently into finished solutions. I am capable of analyzing data to generate reports and dashboards highlighting areas of focus for future growth. Looking for a role that engages with businesses to develop robust and dynamic solutions.

# Employment History

## IT Systems Analyst, Atos India Pvt. Ltd.

JANUARY 2016 — DECEMBER 2020

Engage with Pharma Industry clients to transition databases and on- premises applications to Atos cloud solutions.

**Gathered and analyzed customer requirements** for Cloud Migration and Application Support. Drafted SOP documents working closely with Security, Cloud, and Database Administrator teams

**Developed Analytics application** in VMware to aggregate customer and team performance data to generate and display reports of Employee response time

**Presented client with Incident Management system Proof of Concept**

tailored to cloud migration and future application maintenance.

Compiled all available training documents and customer feedback as training manuals.

## Data Engineer, Hexaware Technologies

JUNE 2022 — AUGUST 2022

Demonstrated **Batch Data storage and processing** for large Datasets containing mixed, unclean data using **AWS S3 as a warehouse.**

Engineered ETL workflow for Kaggle Dataset stored in S3 bucket **using AWS Glue and triggers/events accessing data for Analysis or Visualization**.

Upgraded manual workflow to automated querying from Kaggle Data source to Target S3 using AWS CLI and Kaggle API

# Education

## Master of Science in Data Science, Stevens Institute of Technology

JANUARY 2021 — DECEMBER 2022

## B.Tech in Electronics and Communications Engineering, National Institute of Technology Trichy



JULY 2012 — MAY 2016

# Courses

**AWS Cloud Certified Practitioner Training, Udemy**

|  |  |
| --- | --- |
| ~~Big Data Technologies~~ | ~~MapReduce, Hive, Teg, Python, PySpark,~~ ~~Scala, Kafka, Spark streaming, Oozie~~, ~~Sqoop, Zookeeper~~ |
| ~~Hadoop Distribution~~ | ~~Cloudera, Horton Works~~ |
| ~~Languages~~ | ~~Java, SQL, PL/SQL, Python, HiveQL, Scala~~. |
| ~~Web Technologies~~ | ~~HTML, CSS, JavaScript, XML, JSP, Restful,~~ ~~SOAP~~ |
| ~~Operating Systems~~ | ~~Windows (XP/7/8/10), UNIX, LINUX,~~ ~~UBUNTU, CENTOS.~~ |
| ~~Build Automation tools~~ | ~~Ant, Maven~~ |
| ~~Version Contro~~l | ~~GIT, GitHub.~~ |
| ~~Methodology~~ | ~~Agile, Scrum.~~ |
| ~~IDE &Build Tools, Design~~ | ~~Eclipse, Visual Studio Code, PyCharm,~~ ~~Jupyter Notebook~~ |
| ~~Databases~~ | ~~Azure SQL DB, Azure Synapse, Cosmos DB,~~ ~~MS Excel, MS Access, Oracle 11g/12c, MS~~ ~~SQL Server 2016/2014/2012~~ |

**~~PROFESSIONAL EXPERIENCE~~**

**~~Client: C&S Wholesale Groceries, Edison, New Jersey~~.**

**Role: Data Engineer April 2022 -**

**~~Present~~ ~~Responsibilities:~~**

**Commented [AN2]:** Move this to the top

* Oversaw end-to-end operations of ETL data pipelines in Azure Data Factory, maintaining high-quality data

~~integration, transformation, and loading processes.~~

* ~~Implemented Logic Apps to streamline order processing workflows and enhance operational efficiency in~~ ~~the retail industry, resulting in improved order fulfilment and streamlined data exchange.~~
* ~~Leveraged Snowflake to optimize data warehousing and analytics capabilities in the retail industry, enabling~~ ~~valuable insights, data-driven decision-making, and improved targeted marketing campaigns.~~
* ~~Deployed Azure Event Hub for real-time data ingestion in the retail industry, facilitating real-time monitoring~~ ~~of foot traffic, enhanced fraud detection, and improved operational agility.~~
* ~~Managed the orchestration of Terraform for infrastructure automation in Azure deployments, streamlining~~ ~~provisioning and management of resources.~~
* ~~Utilized Hive within the Azure ecosystem to derive valuable insights, optimize queries, enable efficient data~~ ~~warehousing, and deliver comprehensive retail analytics as an Azure Data Engineer, enabling data-driven~~ ~~decision-making and enhancing business performance.~~
* ~~Employed PySpark and Scala in conjunction with Azure HDInsight, Kafka, and Spark Streaming to enable~~ ~~advanced data processing, real-time streaming analytics, and scalable data solutions within the Azure~~ ~~ecosystem, driving actionable insights and enhancing decision-making capabilities in the retail industry.~~
* Effectively utilized Oozie and ZooKeeper for orchestration and coordination of data workflows in Azure, ~~optimizing storage efficiency with ORC, Avro, Parquet, and delimited file formats as part of Azure data~~ ~~solutions, ensuring seamless data processing and efficient data storage in the retail data engineering~~ ~~landscape.~~



* Ensured data security through encryption while seamlessly handling data ingestion, protecting sensitive ~~data and maintaining data integrity in compliance with industry regulations in the retail industry.~~
* Supervised the development and deployment of SSIS/SSRS packages for improved data accuracy and ~~business insights in the retail industry, enabling data-driven decision-making, generating comprehensive~~ ~~reports, and facilitating data visualization.~~
* ~~Prepared technical specifications, data flow diagrams, and process documentation, fostering clear~~ ~~communication and effective collaboration within the team.~~

**~~Tools & Technologies~~**~~: Azure Databricks, Data Factory, Logic Apps, EventHub, Spark Streaming, Data pipeline,~~ ~~Terraform, Azure DevOps, Oracle, HDFS, MapReduce, YARN, Spark, Hive, SQL, Python, Scala, PySpark, GIT~~, ~~JIRA, Kafka, Power BI.~~

**~~Client: JP Morgan Chase, USA~~.**

**Role: Azure Data Engineer. Jul 2020 – Mar**

**~~2022~~**

**~~Responsibilities:~~**

* ~~Managed the development and execution of ETL data pipelines in Azure Data Factory, improving the~~ ~~processing of millions of financial transactions daily and facilitating faster financial reporting~~.
* Administered Azure Data Lake and Azure SQL Database for optimal data storage solutions, preserving ~~customer data for over 10 million accounts and ensuring high availability.~~
* Leveraged Azure Databricks and HDInsight for Big Data processing, enhancing credit risk modelling and ~~customer segmentation, and driving actionable insights for data-based decision making~~.
* ~~Implemented robust data security measures using Azure Security Center and Azure Key Vault, protecting~~ ~~sensitive financial information, and ensuring compliance with GDPR and other financial industry regulations.~~
* Orchestrated automated loan approval workflows using Azure Logic Apps, reducing processing time and ~~enhancing customer experience.~~
* ~~Worked with Azure DevOps to streamline CI/CD pipelines, enabling faster updates to customer-facing~~ ~~banking applications and reducing time-to-market for software releases~~.
* ~~Streamlined deployment and scalability of data processing environments using Docker containerization and~~ ~~Azure.~~
* Implemented real-time data processing using Spark Streaming and PySpark, enabling rapid insights from ~~transaction data and improving fraud detection and risk management.~~
* ~~Enhanced data quality and integrity by ingesting and transforming data from diverse sources, and~~ ~~developed ETL processes using SSIS and SSRS packages to facilitate data-driven decision-making.~~
* ~~Played a pivotal role in cross-functional teams to develop data models and solutions that drove a 20%~~ ~~improvement in meeting business needs.~~
* Implemented robust data governance and security measures, safeguarding sensitive customer data, and ~~ensuring compliance with banking industry regulations.~~
* Boosted system efficiency by optimizing data pipelines and queries, and improved speed of generating ~~banking reports using efficient Hive queries.~~
* Enhanced data processing efficiency by working extensively with RDDs, Data frames, and Hive for data ~~analysis and processing, and optimized batch processing of streaming data with Spark Streaming.~~
* ~~Managed project workflows using JIRA and maintained version control using Git, leading to improved~~ ~~project management and collaboration.~~

**~~Tools & Technologies~~**~~: Azure Databricks, Data Factory, Logic Apps, EventHub, Spark Streaming, Terraform~~, ~~Azure DevOps, YAML, Oracle, HDFS, MapReduce, YARN, Spark, Hive, SQL, Python, Scala, PySpark, GIT, JIRA,~~ ~~Kafka, Power BI.~~

**~~Client: FREDDIE MAC, VA~~**



**Role: Big Data Engineer. Feb 2019 – Jun**

**~~2020~~**

**~~Responsibilities:~~**

* Constructed and maintained end-to-end data pipelines with Sqoop, Kafka, and Spark, facilitating seamless ~~data transfer from MySQL and Oracle to HDFS for more precise mortgage risk assessments~~.
* ~~Worked extensively with Data Lakes and big data ecosystems like Hadoop, Spark, Hortonworks, and~~ ~~Cloudera, to process vast volumes of structured and unstructured data, aiding in loan performance analysis.~~
* Developed optimized Hive and Spark SQL queries for custom business requirements, enhancing efficiency ~~in data-driven decision-making processes.~~
* Built HBase tables with Hive integration for efficient storage and retrieval, facilitating quicker access to vital ~~market data.~~
* Deployed Kafka and Spark Streaming for real-time data analysis, contributing to immediate insights from ~~streaming data for credit risk evaluations.~~
* Leveraged Spark and PySpark for rapid data testing and processing, expediting the data validation for more ~~accurate analytics.~~
* ~~Maintained data accuracy and quality by developing custom automation scripts in Oracle's PL/SQL,~~ ~~resulting in improved data integrity.~~
* ~~Implemented CI/CD pipelines in the Hadoop environment, enhancing development and deployment~~ ~~efficiency for various data projects.~~
* ~~Utilized JIRA for effective project management, enhancing collaboration and efficiency within cross-~~ ~~functional teams.~~
* Managed clusters using Zookeeper and scheduled jobs with the Oozie workflow engine, optimizing data ~~processing workflows for mortgage market analyses.~~
* ~~Collaboratively resolved JVM-related issues, ensuring smooth system performance for data processing~~ ~~tasks.~~
* ~~Maintained code repositories with Git, bolstering version tracking, and efficient team collaboration~~.

**~~Tools & Technologies~~**~~: Sqoop, MYSQL, HDFS, Apache Spark Scala, Hadoop Hive, Cloudera, HBASE, Kafka,~~ ~~MapReduce, Zookeeper, Oozie, Python, PySpark, Ambari, JIRA~~.

**~~Client: WSO2, CA~~**

**Role: Hadoop Developer. Jul 2017 – Jan**

**~~2019~~**

**~~Responsibilities~~**

* ~~Designed and implemented ETL jobs using Spark-Scala to migrate extensive datasets from Oracle to~~ ~~MySQL, streamlining data accessibility for a mission-critical financial analysis project.~~
* Utilized Spark-Scala and Spark-Cassandra Connector APIs for tasks such as migrating customer data and ~~generating business reports, enabling the customer retention team to take data-driven decisions.~~
* Built a Spark Streaming application that tracked real-time sales, providing instant insights for the sales team ~~during peak retail periods like Black Friday and holiday sales.~~
* Analysed data types and source data, ensuring optimal data quality for PowerBI ad-hoc report generation ~~used in quarterly business review meetings.~~
* Converted complex SQL scripts into PySpark for an improved compatibility with the Hadoop ecosystem, ~~leading to a substantial increase in query performance for the data analytics team.~~
* Seamlessly extracted data from diverse data sources into HDFS using Sqoop, contributing to a multi-source ~~data integration project aiming to consolidate customer data.~~
* Managed data importing from various sources, performed transformations using Hive and MapReduce, and ~~loaded data into HDFS, supporting a large-scale marketing campaign analysis project~~.
* Implemented automation for deployments using YAML scripts, accelerating the build and release processes ~~for the new customer recommendation system.~~
* ~~Deployed key Hadoop ecosystem components such as Apache Hive, Apache Pig, HBase, Apache Spark,~~ ~~Zookeeper, Flume, Kafka, and Sqoop for a big data migration project.~~



* ~~Applied data classification algorithms using MapReduce design patterns for an AI-based fraud detection~~ ~~system.~~
* ~~Improved MapReduce job performance by creating combiners, partitioning, and implementing distributed~~ ~~cache for a project aimed at optimizing the organization's data processing pipeline.~~
* Managed source code in Git and GitHub repositories, enhancing team collaboration and productivity during ~~the development of a new real-time data analytics platform.~~

**~~Tools & Technologies~~**~~: Hadoop, Hive, spark, PySpark, Sqoop, Spark SQL, Cassandra, YAML, ETL.~~

**~~Client: Change Healthcare, Nashville, TN~~**

**Role: Data Warehouse Developer. Feb 2015 – Jun**

**~~2017~~**

**~~Responsibilities:~~**

* ~~Served as a SQL Server Analyst/Developer/DBA, using SQL Server 2012, 2015, and 2016 to optimize~~ ~~large-scale healthcare databases, enhancing data availability.~~
* ~~Designed and scheduled DTS/SSIS Packages, improving data transfer efficiency within healthcare data~~ ~~systems.~~
* Updated Erwin models for Consolidated Data Store (CDS), Actuarial Data Mart (ADM), and Reference DB, ~~aligning with evolving healthcare standards and user requirements.~~
* ~~Applied TFS for source control and deployment tracking, ensuring consistent script deployments~~.
* ~~Exported current Data Models as PDFs, sharing them via SharePoint for increased stakeholde~~r ~~accessibility.~~
* Managed the development, administration, and maintenance of key databases, enabling reliable healthcare ~~data warehousing~~.
* ~~Authored Triggers, Stored Procedures, and Functions using Transact-SQL (TSQL) to facilitate robus~~t ~~healthcare data operations.~~
* ~~Deployed scripts according to Configuration Management and Playbook requirements, maintaining~~ ~~seamless data operations.~~
* ~~Optimized data storage and access by managing Files/File groups and Table/Index associations.~~
* ~~Enhanced data processing speed and efficiency through query tuning and performance tuning.~~
* ~~Ensured data accuracy and integrity using Quality Center for defect tracking and resolution.~~
* ~~Safeguarded sensitive healthcare data and compliance by maintaining user roles and permissions.~~

**~~Tools & Technologies~~**~~: SQL Server 2008/2012 Enterprise Edition, SSRS, SSIS, T-SQL, Windows Server 2003,~~ ~~Performance Point Server 2007, Oracle 10g, visual Studio 2010.~~

**~~Client: Aetna Inc., Hartford, CT~~.**

**Role: SQL Engineer. Feb 2013 – Jan**

**~~2015~~**

**~~Responsibilities:~~**

* ~~Developed stored procedures, triggers, and functions, optimizing SQL Server performance through indexing~~ ~~and monitoring techniques.~~
* ~~Implemented ETL data flows using SSIS, facilitating data migration and transformation from various sources~~ ~~including SQL Server, Access, and Excel.~~
* Gained experience in dimensional data modelling and identified facts and dimensions for Data Mart design, ~~along with developing fact tables and dimension tables using Slowly Changing Dimensions (SCD)~~ ~~techniques.~~
*  ~~Managed error and event occurrences during ETL processes, employing techniques such as precedence~~ ~~constraints, breakpoints, checkpoints, and logging~~.
* ~~Acquired experience in building SSAS cubes, implementing aggregations, defining KPIs, partitioning cubes,~~ ~~and creating data mining models, contributing to deployment, and processing of SSAS objects.~~
* ~~Developed a range of reports, including parameterized, chart, graph, linked, dashboard, scorecards, and~~ ~~drill-down/drill-through reports on SSAS cubes using SSRS~~.

**~~Tools & Technologies~~**~~: MS SQL Server 2016, Visual Studio Legacy Versions, SSIS, Share point, MS Access, Git.~~