**NAME**  
Data Engineer XXXX@gmail.com | PHONE NUMBER | LinkedIn account

**Summary: -**

Data Engineer with 6 years of IT experience including 5 years of experience with analytical and intuitive skills in building successful Big Data and cloud projects. Expertise in designing scalable Big Data applications, migrating data warehouse models into cloud on large-scale distributed data. Expertise in all components of Big Data Ecosystem- Spark, Pyspark, data bricks, knowledge on data serialization techniques like Delta, Avro, excel, Sequence Files, Parquet, JSON, and ORC. Proficient in Azure data services such as Azure SQL Database, Azure Cosmos DB, Azure Data Lake Storage, and Azure Data Factory. Skilled in ETL/ELT processes, data modelling, and data analysis. Strong experience and knowledge of real-time data analytics using Spark Structured Streaming. Good experience developing applications using Pyspark and Scala. Extensively worked using Azure Databricks cloud by using (ADF, ADLS Gen2, Azure SQL, Blob storage, Azure Databricks).

**TECHNICAL SKILLS:**  
Languages: Python, SQL, Scala  
Big Data: Apache Spark, PySpark, Hadoop  
Cloud: Azure (ADLS, SQL, Data Factory), Azure Databricks  
Data Warehousing: Slowly Changing Dimensions  
Version Control: Git, Azure DevOps  
  
**PROFESSIONAL EXPERIENCE**  
  
COMPNY NAME | Senior Data Engineer | YEAR TO MENTION- Present  
  
**Project: WALMART Real-time Analytics Platform**

- Technologies: PySpark, Databricks, Azure Data Factory, ADLS Gen2, Azure SQL  
- Developed a real-time data ingestion pipeline processing 1TB+ daily using Spark Structured Streaming  
- Implemented Delta Lake for ACID transactions and time travel capabilities  
- Reduced data processing latency from hours to minutes, improving business decision-making

Responsibilities: -

* Worked on replacing existing Hive scripts with Spark Data-Frame transformation and actions for faster analysis of the data
* Developed PySpark scripts to Reduce costs of organization by 30% by migrating customers data in SQL Server to Hadoop
* Experience in handling JSON datasets and writing custom Python functions to parse through JSON data using Spark
* Worked on best buy applications using PySpark to integrate data coming from other sources like ftp, csv files processed using Azure Databricks
* Developed Spark applications for data extraction, transformation and aggregation from multiple systems and stored on Azure Data Lake Storage using Azure Databricks notebooks
* Created Pipelines in ADF to copy parquet files from ADLS Gen2 location then moved into Azure SQL server and finally historical data moved into to Azure Synapse Analytics Data Warehouse
* Generate weekly based reports and ops reports, customer goals reports, mobile scan and pay goals and usage in sales data by using power BI
* Environment: Azure ADF, Scala, Pyspark, Spark, SQL, Snowflake, Databricks, GitHub, Azure Git, , ADF Gen2, ADF Blog Storage

**Project: Data Warehouse Migration**  
- Technologies: Azure Synapse Analytics, ADLS Gen2, Azure Data Factory  
- Migrated on-premises data warehouse to cloud-based solution using Azure Synapse  
- Designed and implemented slowly changing dimensions (SCD) Type 2 for historical tracking  
- Achieved 40% cost reduction and 3x performance improvement in query response times

**Responsibilities**: -

* understanding of data profiling, cleansing, transformation, and migration techniques.
* documenting data migration processes at a functional and technical level, be detail oriented, highly organized, and understand data structures and business requirements.
* Identified the data that you are currently storing and processing.
* Map out data sources, consumers, and flows between systems. Document the format, frequency, and size of the data.
* Rehosting–also referred to as “Lift and Shift”– Involved migrating applications as-is into the cloud.
* Refactoring involved making modifications to your existing applications to maximize efficiency and make them compatible with Microsoft products in the cloud.
* Rearchitecting takes the application one step further by changing the application code base altogether in preparation for the cloud.

SEOCND COMPANY NAME | Data Engineer | from year to till year  
  
**Project: Commodity Data Platform**

**Project Description:** Compass ag solutions is a commodity risk management and consulting firm helping cattle feeders in North America measure, manage, and mitigate market risks. Managing profitable cattle feeding operation has never been more difficult. With increasing volatility, overwhelming data, and a global risk climate, this is a road best not traveled alone. While one can never control all the factors of cattle feeding, Compass Ag Solutions can partner with you on your journey to provide fresh ideas, strategic perspective, and peace of mind in a time of rising uncertainty.

- Technologies: PySpark, Databricks, Azure Data Lake Storage, Azure SQL  
- Built scalable ETL processes to integrate customer data from different source systems like Azure ADLS, AWS S3, REAT APIS.  
- Implemented data quality checks and data lineage tracking using Azure Purview  
- Reduced customer data consolidation time from 12 hours to 1 hours  
  
**Project: Automated Reporting System**  
- Technologies: Azure Data Factory, Azure SQL  
- Designed and implemented an end-to-end automated reporting solution  
- Reduced manual reporting effort by 80%, saving approximately 20 hours per week for the analytics team  
  
EDUCATION  
Bachelor of Science in XXXXXXXXXXX | XXXXXXXXXXXX University | 2000 - 2004  
  
CERTIFICATIONS  
- Microsoft Certified: IF ANY  
- Databricks Certified Associate Developer for Apache Spark IF YOU HAVE