**NIKHILESH SRINIVAS**

**SR.DATA ENGINEER**

**571-267-5928**

**MSNIKK4518@GMAIL.COM**

<https://www.linkedin.com/in/niksrinivas/>

**CARRER OBJECTIVE**

* Highly experienced Data Engineer with 10 years of experience and proven track record in ETL processes, Data Analysis, Data Architecture, Data Modeling, BI analytics, and dashboard development. Proficient in leading BI tools like Tableau, Cognos BI for producing insightful. Experience on Migrating SQL database to Azure Data Lake, Azure data Lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse and Controlling and granting database access and Migrating On premise databases to Azure Data lake store using Azure Data factory. Azure Data Engineer working with Kafka, you would be responsible for integrating Kafka, a distributed streaming platform, Working as a Data Engineer with PySpark involves using the Python API for Apache Spark to develop and maintain data pipelines for large-scale data processing. Extensive experiences in web based multi-tier enterprise applications; Client / Server Applications using **Java/J2EE** technologies and have good exposure in Model View Controller (**MVC**) Architecture based applications using Spring Framework 3.2.2 and Struts Framework 2.3.12 with expertise in presentation, business and database layers. Specialized in Core Java, Servlets, JSP, JPA, JDBC, JNDI, MVC, Struts, EJB, JMS, XML, JUnit, Custom Tags and Design Patterns. Azure Data Factory (ADF), Integration Run Time (IR), File System Data Ingestion, Relational Data Ingestion. Hands on Experience in Azure Synapse Analytics. Expertise extends to big data technologies, enabling effective data processing and analysis. Azure is a cloud computing platform provided by Microsoft, and PL/SQL (Procedural Language/Structured Query Language) is a programming language designed for managing data in Oracle databases.

|  |
| --- |
| **EXPERIENCE** |

**SR. DATA ENGINEER FEB2022-PRESENT**

MCLEAN, VA NAVY FEDERAL CREDIT UNION,

|  |
| --- |
| * Developed Kafka producer and consumers, HBase clients, Spark, and Hadoop, MapReduce jobs along with components on HDFS, Hive * Designed end-to-end ETL processes to extract data from various sources, perform complex transformations, and load it into destination databases or data warehouses. * Developed and managed data models using Cognos Framework Manager, ensuring optimized data access, security, and performance. * Worked extensively in Java, J2EE, XML, XSL, EJB, JSP, JSF, JDBC, MVC, Jakarta struts, JSTL, Spring2.0, Design Patterns and UML. * PySpark is the Python library for Apache Spark. Learn the fundamentals of Spark, RDDs (Resilient Distributed Datasets), * PL/SQL is closely associated with Oracle, it's worth noting that Azure supports various databases, * Integrate Apache Kafka with Azure Event Hubs to facilitate the ingestion of large volumes of streaming data. * Databricks is built on top of Apache Spark as a Data Engineer, * Analyze, design and build Modern data solutions using Azure PaaS service to support visualization of data. Understand current Production state of application and determine the impact of new implementation on existing business processes. * Azure cloud environment using Apache Cassandra as the underlying NoSQL database. * Managed Data Architecture and ETL Designing services for various institutions on behalf of an international IT, Digital Services, and Business Solutions firm. * Designed star schemas, snowflake schemas, and created calculations, measures, and dimensions to support reporting requirements. * Worked on migration of data from On-prem SQL server to Cloud databases (Azure Synapse Analytics (DW) Azure SQL DB). * Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL and U-SQL Azure Data Lake Analytics. Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in in Azure Databricks. * Leverage PySpark's powerful transformations and actions to transform and process data at scale. * Reviewed and approved hundreds of data modeling/database design documents. Helped increase the efficiency of data modeling procedures. Developed a statistical tool to optimize the design of large database tables and achieve data storage savings more than 70%. * Revised our Data Architecture standardization in per the new platform the Data Lake. * Worked on Power Shell scripts to automate the Azure cloud system creation of Resource groups, Web Applications, Azure Storage Blobs Tables, firewall rules. * Build and maintain optimized and highly available data pipelines that facilitate deeper analysis and reporting. * Review and analyze moderately complex business, operational or technical challenges that require an in-depth evaluation of variable factors. * Big Data ecosystems using Hadoop, Pig, Hive, HDFS, MapReduce, Sqoop, Storm, Spark, Airflow, Snowflake, Teradata, Flume, Kafka, Yarn, Oozie, and Zookeeper. * Automated ETL workflows using tools like Apache Airflow, Luigi, or custom scripts, improving efficiency and reducing manual intervention. * Design and develop ETL Data Pipelines using Spark, Spark Streaming, Sqoop and Scala for Open-source Hadoop applications to ingest, transform, analyze customer data and maintain ETL, technical documentation. * Ingest, process, and analyze real-time data streams using Kafka. * Experience in design, development, maintenance and support of Big Data Analytics using Java, Horton work, Cloudera Hadoop Ecosystem tools like HDFS, Hive, Sqoop, Pig, Spark, Kafka. * Extensive experience working in migration projects and moving thousands of reports from other tools like Cognos to Tableau. * Executed Data Analysis and Data Visualization on survey data using Tableau Desktop as well as Compared respondent’s demographics data with Univariate Analysis using Python (Pandas, NumPy, Sklearn, and Matplotlib) * Familiarize with the Databricks Workspace * Data Extraction and Transformation and Load (Databricks Hadoop) * Written python code for exploratory data analysis using Scikit-learn machine learning python packages- NumPy, Pandas, Matplotlib, Seaborn, stats models, pandas profiling. * Perform installation, configuration and of Oracle databases on Solaris, Linux, HP-UX and VMware. * Developed various machine learning models such as Logistic regression, KNN, and Gradient Boosting with Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn in Python. * Developed rest APIs using python with flask and Django framework and done the integration of various data sources including RDBMS, Shell Scripting, Spreadsheets, and Text files. * Lookup, Change Capture, Funnel, Peek, Row Generator, Filter stages in accomplishing the ETL Coding. * Designed ETL jobs incorporating complex transform methodologies using Data Stage tool resulting in development of efficient interfaces between source and target systems. * Understand the transformations and actions provided by PySpark for data processing. * Responsible for estimating the cluster size, monitoring, and troubleshooting of the Spark databricks cluster * Developed A.I machine learning algorithms like Classification, Regression, Deep Learning using python. * Implemented robust error handling mechanisms, logging, and notifications within ETL workflows to ensure smooth and reliable data processing. * Azure Data Factory (ADF) is a cloud-based data integration service that allows you to create, schedule, and manage data pipelines for moving and transforming data. * Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform and load data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards. * Used Python Scripting to test the resultant excel data of both Cognos and Tableau/ AG Grid saving hundreds of man hours with Manual Testing * Build data pipelines in airflow in GCP for ETL related jobs using different airflow operators. * Developed machine learning models using recurrent neural networks - LSTM for time series, predictive analytics. * Use PL/SQL to create stored procedures, functions, and triggers within Oracle databases. * Design, Architect and Support Hadoop cluster: Hadoop, Map Reduce, Hive, Sqoop, Ranger, Presto and high-performance SQL query engine and Druid for indexing. |

DATA ENGINEER, MAR 2020-JAN 2022

VERIZON, ASHBURN, VA

|  |
| --- |
| * Develop data pipelines to consume data from Enterprise Data Lake (MapReduce, Hadoop distribution - Hive Table/HDFS) for the analytics solution. * Used Map Reduce Programs that are running on the cluster. * Created Hive external tables to stage data and then move the data from Staging to main tables. * Implemented the Big Data solution using Hadoop, Hive, and Informatica to pull/load the data into the HDFS. * Used Hive to analyze the partitioned and bucketed data to compute various metrics for reporting on the dashboard using Power BI * Design and develop data pipelines using Azure Data Factory to orchestrate the movement and transformation of data. * Worked on Azure Services like IaaS, PaaS and worked on storages like Blob (Page and Block), SQL Azure. Well experienced in deployment configuration management and virtualization. * Spark using Scala on cluster for computational (analytics), installed it on top of Hadoop performed advanced analytical application, * Ingest data from various sources into PySpark DataFrames. * Use Databricks to import and export data from various sources. * Azure data engineer working with Cassandra would be involved in managing and optimizing data workflows, storage, * Ensure that PL/SQL code adheres to security and compliance standards. * Extensive experience in Object Oriented Programming, using Java &amp; J2EE (Servlets, JSP, Java Beans, EJB, JDBC, RMI, XML, JMS, Web Services, AJAX). * Created rich reports using tools like Cognos FM and Crystal Reports and worked with Business Users in presenting the data. * Distributed database Design, Data modeling, Development and Support in Data tax Cassandra distribution. * Enhanced the data architecture for the Aircraft-wing and control segments control system * Developed JSON Scripts for deploying the Pipeline in Azure Data Factory (ADF) that process the data using the Sql Activity. * Utilize Azure Stream Analytics to process and analyze streaming data from Kafka. * ETL Restarting capability for a date or date range or from point of failure or from beginning. * Create ETL scripts for the ad-hoc requests, requests to retrieve data from analytic sites. * Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL. * To analyze the data Vastly used Athena to run multiple queries on processed data from Glue ETL Jobs and then used Quick Sight to generate Reports for Business Intelligence. * Optimized report performance by tuning queries, implementing caching strategies, and utilizing aggregate awareness techniques in Cognos. * Have good experience working with Azure BLOB and Data lake storage and loading data into Azure SQL Synapse analytics (DW). * Architect implement medium to large scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Data bricks, NoSQL DB). * Streamed data from numerous sources using the Spark Streaming API and created Apache Spark apps to clean and validate the data. * Visionary Cloud, GCP, DevOps, and IT Infrastructure Specialist with career - long record of success leading the full project lifecycle from development * Worked on Python Open stack APIs and used Python scripts to update content in the database and manipulate files * Created Canonical model in Snowflake as a single source of truth by building data pipelines connecting and migrating data. * Design implement migration strategies for traditional systems on Azure (Lift and shift/Azure Migrate, other third-party tools. * Spark for ETL follower, Databricks Enthusiast, Cloud Adoption Data Engineering enthusiast in Open source community. * Involved and worked on Python Open stack APIs and used several python libraries such as wxPython, NumPy and matplotlib. * Utilized Oozieworkflow to run Pig and Hive Jobs Extracted files from MongoDBthrough Sqoopand placed them inHDFS and processed * loading and analyzing large datasets with Hadoop framework (MapReduce, HDFS, PIG, HIVE, Flume, Sqoop, SPARK, Impala, Scala), * Integrate data from various sources, both on-premises and in the cloud, using Azure Data Factory. * Create ETL scripts to retrieve data feeds, page metrics from Google analytic services (for Star Wars site). * Worked on Airflow 1.8(Python2) and Airflow 1.9(Python3) for orchestration and familiar with building custom Airflow operators and orchestration of workflows with dependencies involving multi-clouds. * Continuously tuned Hive UDF for faster queries by employing partitioning and bucketing * Enhanced report loading times by optimizing SQL queries, incorporating efficient filtering techniques, and leveraging in-memory caching in Cognos, * PySpark supports reading from and writing to various data formats and storage systems such as CSV, Parquet, Avro, JSON, * Used Data Stage Manager for importing metadata from repository, new job categories and creating new data elements * Written Map Reduce java programs to analyze teh log data for large-scale data sets. * Demonstrate expertise in automating and scaling technologies, like Docker, Kubernetes, Open Shift, and GCP platform * Databricks supports connectors for various data sources and storage systems such as Delta Lake, Azure Data Lake Storage, AWS S3, * PL/SQL skills based on the database systems used in your organization, whether it's Oracle or another supported database platform in Azure. * Used various sources to pull data into Power BI such as SQL Server, Excel, Oracle, SQL Azure etc. |

**BI/DATA ENGINEER, AUG 2017-FEB 2020**

WASHINGTON DC WORLD BANK,

|  |
| --- |
| * Responsible for optimizing, improving the performance of existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frames, and Pair RDDs * Designed and standardized a set of custom report templates in Cognos, reducing report creation time and maintaining consistent branding across the organization. * Performed data operations like Text Analytics and Data Processing, using the in-memory computing capabilities of Spark using Scala, * Design and implement ETL processes using Databricks notebooks. * Used various Java APIs like Apache POI, Java Email, I Text etc. as part of test automation. * Developed ETL jobs to automating the several student’s data from enrollment step to graduation process and service fee calculation for weekly finance payment to Confidential. * Use Azure Data Factory's data flow capabilities to transform and cleanse data as it moves through the pipeline. * Create and manage data pipelines for moving and transforming data between different sources and destinations, including Cassandra. * Updated the Engine-parts Warranty System data architecture * Design Setup maintain Administrator the Azure SQL Database, Azure Analysis Service, Azure SQL Data warehouse, Azure Data Factory, Azure SQL Data warehouse. * Resolved a critical data inconsistency issue in Cognos reports by identifying and rectifying discrepancies within underlying data sources, preventing misinformation. * Scala Test for writing test cases and coordinated with QA team on end to end testing. * Material for users on the use of the business applications. ther Junior Consultants on various modules and functional setups in Oracle ERP. * Data modeling, Design, implement, and deploy high-performance, custom applications at scaleon Hadoop/Spark. * Implement MapReduceusing Sqoopto import the data from various relational databases to filter out unstructured data, perform data cleaning and pre-process on HDFS * Worked with Data Stage Designer for importing metadata from repository, new job categories and creating new data elements * Extract real-time data feed using Kafka, process core job using Spark Streaming to Resilient Distributed Datasets * Implement ETL framework to provide features such as Master Data Management, ETL - restart capability, security model and version control. * Experience working with Windows Hyper-V Server, Azure, Windows Clustering including. * Conducted data preparation using SQL for Tableau consumption and publishing dashboard to Tableau Server * Developed Tableau data visualization using Cross tabs, Heat maps, Bar Charts, Gantt charts, Waterfall charts, Scatter Plots, Geographic Maps, Pie Charts and Donut Charts * migration part of crash logs from Fabric Crashlyticsto GCP Firebase Crashlytics and setup of roles and responsibilities on accessing Firebase in GCP, like firebase Viewer and Firebase Admin. * Performed impact analysis on Data Stage jobs, Oracle Stored Procedures and Packages * Use Base SAS for cleaning the data from IBM Maximo and analyzing the historical data of maintenance for dashboard analysis and indicators to drive the continuous process management * Used Routine activity to call any routine in the Data Stage repository * Design and implement ETL processes using PySpark transformations. Perform data cleaning, filtering, aggregations, * Developed, maintained multiple Kafka producers and consumers as per the business requirement * Involved in Cassandra Cluster's planning and had a good understanding of the Cassandra cluster mechanism * Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala. * Defined Spark Streaming APIs to build a learner data model that gathers live data from Kafka and connects it to the Cassandra sink * Worked with ETL tools Including Talend Data Integration, Talend Big Data, Pentaho Data Integration and Informatica. * Used databricks notebooks for interactive analytics using Spark APIs * Assisted in Tableau admin activities including granting access, managing extracts and installations, caching configurations to support large volume usage * Use Terraform to set up GCP infrastructure like VPC, subnets, GCP storage buckets, GCP Composer, Secret manager, GKE, etc. * Optimize PL/SQL queries for better performance. * Worked Azure SQL Database Environment. * Migrate the Data using Azure Database Migration Service (AMS). * Migrate SQL Server and Oracle database to Microsoft Azure Cloud. * Configured apache airflow for workflow management and created workflows in python. * Scheduled data refresh on Tableau Server for weekly and monthly increments based on business change to ensure that the views and dashboards were displaying the updated data accurately * Designing and implement a fully operational production grade large scale data solution on Snowflake Data Warehouse |

**SQL DEVELOPER, , SEP 2015-AUG 2017**

JERSEY, NJ DELOITTE

|  |
| --- |
| * Applied sophisticated techniques such as text analytics and processing to Spark's in-memory computing capabilities * Design and implement ETL processes for Cassandra data using tools like Azure Data Facto * Scala and Spark-SQL were used to create Spark code for quicker processing and testing * Spark RDD transformations were used to map business analysis and implement actions on top of the transformations * Created Spark tasks to do lightning-fast analytics across the Spark cluster * Data bricks notebooks using SQL, Python and automated notebooks using jobs. * Capturing functional/technical details of Oracle Applications R12.1.1 ERP (Enterprise Resource Planning) and documenting the same. * Used Execution Plan, SQL Profiler, Query Store and Database Engine Tuning Advisor to optimize queries and enhance the performance of databases * procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala. * Created monitors, alarms, notifications and logs for Lambda functions, Glue Jobs using Cloudwatch. * Work on Google cloud platform (GCP) services like compute engine, cloud load balancing, cloud storage, cloud SQL, stack driver monitoring and cloud deployment manager. * Worked with Spark using Spark Context, Spark-SQL, Java, Scala, PySpark, Data frame API, Pair RDD, Spark Streaming, ML libraries and User Defined Functions to apply transformations on the Data.   **SITACORP, INDIA MAR 2012 TO MAR 2014**  **SQL DEVELOPER**   * Created technical specifications, test plans and test data to support ETL data flows * Assist the Development staff with the testing and implementation of the application. * Used Toad Client for accessing Oracle databases that contained views of the domestic application. * Created SSIS Packages that mimics the nature of the domestic application. * Created complex stored procedures, triggers, cursors, tables, views and other database objects using T-SQL. * Troubleshoot SQL Server performance issues for internal and external customers. * Used T-SQL to create queries and stored procedures to produce reports for end-users * Configured Transactional Replication between the Production server and Reporting server and to move the data to servers in various locations. |

|  |
| --- |
| **Education** |

|  |  |  |
| --- | --- | --- |
| **Master Of Science (M.S.) In Computer Science,**   **Springfield, IL, US**  **University of Illinois Springfield.** |  |  |

|  |  |  |
| --- | --- | --- |
| **Bachelor of Science (B.S.) in Computer Science**  **SNIT.** |  |  |