|  |  |
| --- | --- |
| **HARSHAVARDHAN KONAKALLA** | (720) 382-0479 • konakallaharsha74@gmail.com  Home Address: :6052 E Hinsdale Avenue,Centennial,CO,80112  • [LinkedIn](https://www.linkedin.com/in/harshavardhan-konakalla-246aa4234/): www.linkedin.com/in/harshavardhan-konakalla-246aa4234 |

**Senior Data Engineer**

***Data-driven and incisive professional with hands-on experience in steering the development and management of big data infrastructure, tools, processes, and solutions for multi-national financial services and IT companies.***

Proficient in designing data pipelines with Azure Data Factory and Confluent Kafka for seamless data transfer. Experienced in analyzing and describing massive volumes of customer, product, and operational data sets (terabytes) to optimize key business/operational use cases, mitigate compliance risks, and create net new revenue streams. Background in accelerating the delivery of several concurrent data engineering projects focused on data lifecycle design, data ontology alignment, process automation, and cloud computing. Demonstrated expertise in creating data-driven reports, presentations, and powerful data visualization dashboards to enable management to make calculated business decisions while developing relationships with clients. Success in driving the full data lifecycle from data aggregation to cleansing, transformation, enrichment, and visualization while transforming data into meaningful information. Possess solid working knowledge of several AWS technologies, including AWS Lambda, Kinesis, and RedShift with experience in managing data-streaming frameworks such as Apache Kafka and expertise in Git.

**CORE COMPETENCIES**

|  |  |  |
| --- | --- | --- |
| * Data Pipeline Architecture Design * Data Integrity, Reliability, & Modeling * Business Requirements Gathering * Infrastructure Scalability & Automation * Extraction, Transformation, Loading (ETL) * Internal Process Improvements * Data Delivery Optimization * Statistical Models & Econometric Models * Design Sessions & Quality Assurance * Data Stewards & Data Modelling * Business Intuition & Data Solutions * Complex Business Systems | * Data Availability & Quality * Complex Data Sets & Data Ingestion * Data Build Tool (DBT ) * Tech Stack & Debugging * Service Level Agreements (SLAs) * Data Manipulation * Data Sources * Business Process / Business Analysis * Orchestration Tools * Transformation Pipelines * Data Accessibility & Documentation * Internal Teams Collaboration | * Data Pipelines Development * Data Needs, Metrics, & Analysis * Data Analytics & Data Science * Data Models & Large Datasets * Data Collection & Refinement * Production Engineering * Data Structures & Code Reviews * Subject Matter Expertise (SME) * Data Governance & Visualization Tools * Data Analytics Platform |

**TECHNICAL PROFICIENCIES**

|  |  |
| --- | --- |
| **Big Data Technologies**  **Amazon Web Services (AWS)**  **Azure Services** | HDFS, Map Reduce, Hive, Sqoop, Oozie, Spark, Kafka, DBT  AWS Lambda, Kinesis, Elasticsearch, Redshift, RDS, S3, and AWS Glue  Azure Data Factory , Azure Databricks, Azure SQL Database |
| **Hadoop Frameworks** | Cloudera CDHs, Horton works HDPs. |
| **Database** | Oracle 10g / 11g, PL / SQL, MySQL, MS SQL Server 2012, DB2 |
| **Language** | Python, Scala, C++, Object-oriented programming (OOP), and Java |
| **Data-Streaming** | Apache Kafka, Kinesis Data Streams, Apache Spark, Apache Beam, Apache Airflow, Apache Flink, Apache Storm, Apache NiFi, Azure Stream Analytics, and Confluent, |
| **Databases** | NO-SQL, HBase, Cassandra, MPP databases, BigQuery, MongoDB, DynamoDB, Snowflake |
| **IDE Tools** | Pycharm, Eclipse, IntelliJ, Visual Studio Codes |
| **Operating System** | Windows 7/8/10, Vista, UNIX, Linux, Ubuntu, Mac OS X |
| **Others** | Fivetran, Tableau, Qlik Sense |

**PROFESSIONAL EXPERIENCE**

**Master Data Engineer | Geico, Maryland Nov 2022 - Present**

***Technical scope:*** *Azure Data Factory, PySpark, Git, Spark, Python, DBT, Confluent Kafka, Snowflake*

* Design data pipelines utilizing Azure Data Factory and Confluent Kafka to transfer data from source to destination.
* Develop DBT scripts to transform data in Snowflake and generate filtered datasets using the Azure release pipeline.
* Create Python scripts leveraging pandas and PySpark data frames in Azure Databricks for applying transformations and exporting data to Snowflake.
* Implement SQL Window functions in Snowflake (e.g., row number, rank, dense rank) to filter/analyze data effectively.
* Handle the transition of 500 million data records within Snowflake, utilizing Snowflake functions to parse unstructured XML data during the process.
* Ingest relational data daily into various data repositories, including Azure Database and SQL Server Database.
* Work with Confluent Kafka to transfer relational data from Oracle Database to Snowflake, utilizing source connectors to ingest data to topics and sink connectors to transfer data to Snowflake.
* Generate Power BI reports, enabling analysis of trends in claims data and driving informed decisions.

**Big Data Engineer | Freddie Mac, Virginia Feb 2021 - Oct 2022**

*Technical scope: HDP 3.1.5, PySpark, Git, Spark, Python, Hive, Kafka, HBase, Sqoop, Tableau*

* Steer design and development of data transformation framework components to support the ETL process while creating data lakes by extracting customer data from various data sources.
* Foster conscious management decisions by generating/presenting meaningful reports and visualizing the HDFS data to the customer using Tableau.
* Develop ingestion modules to ingest data into HDFS from heterogeneous data sources and build distributed in-memory applications using Spark and Spark SQL.
* Utilize spark transformations and actions to build simple/quick and complex ETL applications while using Apache Oozie to manage and schedule Spark Jobs on a Hadoop Cluster.
* Design efficient queries to query data from Hbase and develop internal and external table schemas in Hive with appropriate static and dynamic partitions for efficiency.
* Ingest, transform, and analyze large data sets by developing a data pipeline using Kafka, Spark, and Hive.
* Transform and integrate various data into a scalable data repository, such as a data warehouse, and data lake.
* Engineer large-scale parallel relation-learning system using Spark.

**Big Data Developer | American Express, Arizona Feb 2020 - Jan 2021**

*Technical scope: Hadoop, Big Data, HDFS, Kafka, Sqoop, Oozie, Hive, LINUX, Eclipse, HDP, AWS Lambda, RedShift, S3, Kinesis, RDS, Airflow, Shell Scripting*

* Produced data processing pipelines using tools and frameworks in the Hadoop ecosystem while building ETL pipelines to automate the ingestion of structured and unstructured data.
* Oversaw and managed the process of designing, developing, and setting up AWS cloud solutions through AWS Lambda, Kinesis, Redshift, RDS, and S3.
* Designed the architecture of big data platforms, maintained data pipeline, and customized integration tools, databases, warehouses, and analytical systems to optimize business processes and efficiency.
* Developed diverse AWS services, including RDS instance, and Lambda to build serverless applications while producing reusable, efficient, and scalable programs.
* Implemented and configured big data technologies as well as tuned processes for scaled performance by implementing ELT processes and maintaining the production system.
* Devised multiple Spark jobs in Python for data cleaning, quality checking, and applying business transformations.
* Built data pipelines using Lambda functions, EMR, RedShift, and Airflow.
* Collaborated with development teams during the application design process for highly complex and critical data projects.
* Developed and built frameworks and prototypes to integrate big data and advanced analytics to drive business decisions.

**DATA ANALYST | TravelPort, Colorado Jul 2019 - Jan 2020**

*Technical scope: Tableau, JIRA, R language, Oracle, MySQL*

* Aided senior leadership in making informed decisions by gathering and analyzing business requirements and designing processes required for data analysis and decision-making.
* Utilized highly effective data modeling and statistical analysis tools and techniques to identify underlying trends and develop / present data-driven reports to senior management.
* Analyzed the data generated using R language and generated reports in bar graphs and pie charts by executing the visualization tool Tableau.
* Developed relational databases such as Oracle and MySQL to load, analyze, and retrieve data using SQL queries.
* Collaborated with data science teams to develop reports per requirements while actively participating in identifying and capitalizing on process improvement opportunities.

**EDUCATIONAL BACKGROUND**

**Master’s Degree – Management Information Systems|** University of Colorado, Denver | 2019

**Bachelor’s Degree – Electrical and Communications Engineering|** RMK Engineering College, Chennai | 2017

**Higher Secondary Education |** Sri Chaitanya College of Education | 2013