NARESH KHURIWAL **+91 – 9413875686 || naresh.khuriwal89@gmail.com**

Seeking senior level position as Data Engineer with an organization of repute.

PROFILE SUMMARY

- He is offering 11+ years of experience in Data Engineering, Data Analysis, Flowchart & Diagram Creation, Auditing & Reporting, Stakeholder Coordination, Client Relationship Management, and Liaison & Coordination.
- Presently associated with Recommendation AI Model, US as Data Scientist.
- A technically sound professional with expertise in Azure Synapse Analytics, Azure Data Factory, Azure Machine Learning Workspace, Azure Dedicated Spark & SQL Pool, Databricks Spark, Azure functions, Logic Apps, etc.
- Proficiency in Python, Django, Microservices and data science (Machine learning & Deep learning & ChatGPT), Microservices / RESTful Services /APIs (REST and SOAP) Extensive Knowledge in designing and developing scalable and responsive web development.
- Demonstrated excellence in conceptualizing & executing program vision from start to finish, managing complex milestones while adapting to changes & shifting priorities.
- A peoples' manager with exceptional abilities to lead international, cross-functional teams in delivering strategic initiatives proven to improve systems, processes, and bottom-line results.
- Interacts easily with people of diverse backgrounds, cultures, and professional levels. Effective communicator with good analytical, problem solving, and organizational abilities.

AREAS OF EXPERTISE

Functional & Managerial Skills:

Data Analysis || Database Management || Requirement Analysis || Design & Estimation || Data Visualization || Client Relationship management || Stakeholder Management || Team Management || Distribution, Predictive, & Hypothetical Modelling || Liaison & Coordination || Stakeholder Coordination

Technical Skills:

Languages:	Python, Java, R, Typescript, Ajax, JavaScript, Shell Script, Perl
Databases:	SQL server, oracle, snowflake, Cosmos, SQL, Casandra, Redis, S3, Mongo DB, Maria DB, HBase, MySQL, SQL lite, SQL, T- SQL, API Gateway, Lambda, Bash
Bigdata:	Kafka, Ni Fi, Apache spark, Spark, pig, Hive, Impala, Hadoop, Yarn, Flume, Sqoop, Oozie, Talend Map reduce, Zookeeper.
Hadoop:	Cloudera enterprise, Data bricks.
Frameworks:	Django, Spring Boot, Angular JS, Flask, Kubernetes
Python:	Python, SciPy, Pygit, Py spark, Spy spark, Pandas, Scikit, Beautiful soap, pickle, HTTPLib2, TensorFlow, Pytorch, Flask, Azure Machine learning Studio, MLOps, DevOps, Azure-SDK, Azureinterpret-community, Fair learn, raiwidgets, TensorFlow, Sklearn, Pandas, Matplotlib, Seaborn and NumPy
IDE's:	PyCharm, Spyder and NetBeans, Eclipse, Visual Studio, IntelliJ
Cloud:	AWS, GCP, Glue Athena, Kinesis, Redshift cluster, Azure, ADF, Cosmos, AWS EM, EMR, SNS, SQS, Dynamo DB, CloudWatch, ANF, NSG, Azure data factory, Azure data lake, ADO, ADLS, Synapse.
Operating:	LINUX Red Hat, UNIX, Ubuntu, Cent OSMAC, WINDOWS.
Servers:	Apache, SOAP, Tomcat, AWS, Rest API, Fast API, Amazon EC2, ECS, Rabbit MQ, WebSphere, JBOSS
Development:	Agile, Waterfall, Scrum, Big query, Sage maker.
Version:	GIT, GitHub, Gitlab, CVS, XML, SVN, JSON, Tableau, Bitbucket, Flyway, JUNIT.
Business:	SSRS, SSAS, OBIEE, OAS, Power BI Desktop/service
Deployment:	Heroku and Jenkins, JIRA, SonarQube, Terraform, CI/CD, SOLID
Network:	HTTP, TCP/IP, DHCP, SMPT, SNMP, HTTPS, CDN, JS
ETL:	SSIS, SSRS, SSAS, Power BI, PentaHo, ADF

Other Technical Skills:

- Expertise in data warehousing and ETL software, such as Informatica. Worked on ingesting data using Sqoop from many sources, including SQL servers and HDFS. Experience in using various tools like Sqoop, Flume, Kafka, and Pig to ingest structured, semi-structured, and unstructured data into the cluster.
- Proficient with Apache Spark ecosystems such as Spark and Spark Streaming using Python.
- Developed highly optimized Spark applications to perform various data cleansing, validation, transformation, and summarization activities according to the requirement.
- Strong experience in implementing data models and loading unstructured data using HBase, Dynamo DB, and Cassandra.
- Hands-on Experience in Spark architecture and its integrations like Spark SQL, Data Frames, and Datasets APIs. Extensive experience in migrating on-premises Hadoop platforms to cloud solutions using AWS. Hands-on experience working on statistical procedures in R like ggplot, shiny apps, and dashboard building.

- Expertise in creating and customizing Splunk applications, searches, and dashboards as desired by IT teams and businesses. Experience collecting log data from various sources and integrating it into HDFS using Flume, staging data in HDFS for further analysis.
- Designed and developed logical and physical data models that utilize concepts such as Star Schema, Snowflake Schema, and Slowly Changing Dimensions.
- Good understanding of Cloud Based technologies such as AWS and Azure. Experience in changing over existing AWS infrastructure to Serverless architecture (AWS Lambda, AWS Kinesis) through the creation of a Serverless Architecture using AWS Lambda, API gateway, Route 53, S3 buckets.
- Experience in writing python as an ETL framework and Pyspark to process huge amounts of data daily. Experienced in transporting and processing real-time event streaming using Kafka and Spark Streaming.
- Hands-on experience with importing and exporting data from Relational databases to HDFS, Hive, and HBase using Sqoop.
- Experienced in processing real-time data using Kafka 0.10.1 producers and stream processors and implemented stream processing using Kinesis and data landed into Data Lake S3.
- Vast expertise using Spark tools like Spark SQL and RDD transforms. Integrated AWS DynamoDB with AWS Lambda to backup DynamoDB streams and store value items.
- Adept in statistical programming languages like R and Python including Big Data Technologies like Hadoop 2, HAVE, HDFS, MapReduce, and Spark.
- Expertise in building CI/CD on AWS environment using AWS Code Commit, Code Build, Code Deploy, and Code Pipeline and experience in using AWS CloudFormation, API Gateway, and AWS Lambda in automation and securing the infrastructure on AWS.
- Created Automation to create infrastructure for Kafka clusters with different instances as per components in the cluster using Terraform for creating multiple EC2 instances & attaching ephemeral or EBS volumes as per instance type in different availability zones & multiple regions in AWS.
- Developed version control software such as Bit-Bucket, GIT, and SVN. Has used AWS Elastic Beanstalk for app deployments and have experience with AWS Lambda and Amazon Kinesis.

EDUCATION

- Pursuing Ph.D. in Computational Neuroscience from Poornima University Jaipur.
- M.Tech. in Breast Cancer Diagnosis Using Convolutional Neural Network Algorithm in 2018 from Poornima University.
- B.Tech. in Computer Engineering in 2013 from Rajasthan Technical University Kota.

Certifications:
- Microsoft Certified: Azure Data Scientist Associate.
- Microsoft Certified: Azure AI Fundamentals.
- Microsoft Certified: Power BI Data Analyst Associate.
- Microsoft Certified: Azure Data Engineer Associate.

EMPLOYMENT DETAILS

Since Feb'23 | Recommendation AI Model, US | Data Scientist - (LTIMindtree Limited)
Environment: Python, Azure Machine learning Studio, Azure ADF, Azure Storage, Azure DataLake, SQL, Spark, Azure cloud, Azure SQL DB, ETL, Teradata, MLIB, Power BI, MySQL, Azure, TensorFlow, Scikit-learn, Keras, Pandas, NumPy, Azure DevOps, Azure DevOps Pipeline, Azure MLOps, Azure-SDK, Azure interpret-community, Fairlearn, raiwidgets etc.

Responsibilities:
- Developed real-time data processing/recommendation applications using Python, leveraging Apache Spark Streaming with Kafka and JMS as streaming sources.
- Responsible for developing Spark applications in Python, optimizing code for performance, and loading large volumes of data into Hive and other data storage systems.
- Managed cloud infrastructure management, including virtual machines, networks, & storage resources on GCP and AWS.
- Implemented Infrastructure as Code (IaC) using tools like Terraform, ensuring automated infrastructure provisioning and management.
- Responsible for building & architecture end-to-end data pipelines, from data ingestion to transformation and analytics, using technologies such as PySpark, Kafka..
- Handled data modelling, designing efficient and scalable relational databases, and ensuring data governance, security, and privacy.
- Designed & implemented 3NF models to organize data into tables, eliminating redundancy and ensuring data integrity.
- Applied normalization techniques, such as 3NF, to reduce data redundancy, minimize update anomalies, and enhanced data consistency.
- Designed tables that adhere to 3NF standards, minimizing data duplication and improved database efficiency.
- Leveraged cloud-native services and APIs to integrate applications with GCP and AWS, enabling efficient data processing and service orchestration.

- Excellent problem-solving and troubleshooting skills, demonstrated through performance monitoring, optimization, and issue resolution in GCP and AWS environments.
- Collaborated with cross-functional teams, assess business requirements, and design customized cloud solutions aligned with organizational goals.
- Strong programming skills in Python, Java, and SQL, with experience in developing scripts and code for data processing, transformation, and analysis.
- Used visualization tools like Power BI and Notebooks, enabling data analysis, cleaning, transformation, and the creation of visually appealing reports and dashboards.
- Implemented Apache Airflow for authoring, scheduling, and monitoring data pipelines, ensuring efficient and reliable execution of ETL-related jobs.
- Developed custom aggregate functions using Spark SQL, enabling advanced data analysis and interactive querying. Utilized Python for data analysis tasks, including data cleaning, transformation, statistical analysis, and deriving meaningful insights.
- Knowledgeable about the trade-offs involved in database normalization, understanding when to apply 3NF and when to demoralize for specific performance or reporting requirements.
- Worked with various data storage and processing systems, including Teradata, Oracle, MongoDB, and Elasticsearch.
- Implemented security measures, including IAM roles, network access controls, and encryption, to ensure data privacy and compliance.
- Collaborated with Cloud Architects to design scalable and secure infrastructure solutions using Terraform, ensuring consistent provisioning across environments.
- Hands-on experience with CI/CD pipelines, utilizing tools like Jenkins and Maven for automated builds, deployments, and version control.
- Developed and maintained conceptual, logical, and physical data models based on business requirements and industry best practices.
- Optimized query performance by structuring data using 3NF models, reducing redundant data and improving database responsiveness.
- Responsible for streamlining data processing using technologies like Apache Kafka, Spark Streaming, and AWS Kinesis for real-time data ingestion and analysis.
- Worked with cloud data warehousing solutions, such as Redshift and Snowflake, for storing and querying large volumes of structured and unstructured data.

Jun'20 – Jan'23 | ATS Talent System, India | Data Analyst / Data Scientist - (Cresitatech IT Solutions Pvt. Ltd.)
Environment: Python, AWS SageMaker, AWS Glue, S3, MYSQL, AWS cloud, ETL, Teradata, MLIB, Power BI, TensorFlow, Scikit-learn, Keras, Pandas, NumPy, AWS CodeBuild, CodePipeline, MLOps, SageMaker-SDK, AWS Application Gateway. Route 53, Elastic IP, EC2 Instance, Load Balancer etc.

Responsibilities:
- Effectively used Python and experienced in web application development using frameworks such as Django and Flask.
- Responsible for front-end development using HTML, CSS, Typescript, angular, & jQuery to create dynamic & responsive webpages.
- Developed and maintained the entire frontend and backend modules of web applications using the MVC design pattern.
- Created Python/Flask modules to handle specific data formats and implemented data transformation and migration scripts.
- Designed & implemented object models, data models, tables, constraints, stored procedures, functions, triggers, and packages for Oracle Database.
- Contributed to the design and development of ETL flows using SQL Server Integration Services (SSIS) to change data structures.
- Utilized NumPy for numerical analysis and Spark for data cleansing, transformations, and aggregations.
- Implemented Django Database APIs for seamless access to database objects.
- Developed Flask applications to generate reports from Google Analytics on a daily, monthly, and weekly basis.
- Utilized Apache Spark SQL for fetching and generating reports in a reporting module.
- Employed AJAX and jQuery to transmit JSON data objects between the front-end and controllers.
- Worked on front-end frameworks like Bootstrap for creating responsive webpages.
- Utilized Celery as a task queue and RabbitMQ and Redis as messaging brokers for executing asynchronous tasks.
- Designed and developed stored procedure queries to support SSRS reports and Tableau visualization solutions.
- Proficient in SQL for data analysis, profiling, and optimizing data collection procedures.
- Created SQL queries and reports to support inventory control, sales profitability, and data analysis.
- Used Microsoft Excel for pivot tables, pivot reporting, and advanced functions like VLOOKUP.
- Familiar with statistical software such as Microsoft SPSS for tracking and analysing data.
- Ensured data compliance with Data Quality Centre initiatives and performed data analysis, gap analysis, and data mapping.
- Dealt with T-SQL and PL/SQL for developing queries, stored procedures, and data validation.
- Generated reports using SQL and UNIX for business analysis and fulfilled ad-hoc client requests using Excel and customized SQL queries.

Jun'18 – Jan'20 | Messaging System | Sr. Data Engineer - (Cresitatech IT Solutions Pvt. Ltd.)
Environment: Python, AWS SageMaker, AWS Glue, S3, MYSQL, AWS cloud, ETL, Teradata, MLIB, Power BI, TensorFlow, Scikit-learn, Keras, Pandas, NumPy, AWS CodeBuild, CodePipeline, MLOps, SageMaker-SDK, AWS Application Gateway. Route 53, Elastic IP, EC2 Instance, Load Balancer etc..

Responsibilities:
- Maintained Spark applications using Python, along with expertise in data development and software engineering.
- Promoted customer success by building and migrating applications and services to the AWS Cloud platform, utilizing various AWS services such as IAM, S3, EMR, Lambda, SNS, Data Lakes, Glue, Athena, and Redshift.
- Designed, built, implemented, and maintained AWS Cloud systems for clients, including the conversion of Ab-initio Jobs into Spark Applications using Python and running them on AWS EC2 machines.
- Responsible for automating AWS infrastructure through infrastructure as code using Terraform, creating various AWS resources, and managing IAM users, groups, roles, and policies.
- Designed secure access environments and architecture Amazon Glue/Spark pipelines with a high number of DPUs (Data Processing Units).
- Utilized Azure Data Factory as an orchestration tool for integrating data across upstream and downstream systems.
- Built data pipelines using Apache Airflow in GCP for ETL-related jobs, using different airflow operators. Skilled in analysing SQL scripts and designing solutions for implementation using Py Spark, with experience configuring GCP services such as Data Proc, Storage, and Big Query using cloud shell SDK.
- Responsible for building and architecture end-to-end data pipelines and ETL/ELT processes for data ingestion and transformation in GCP.
- Managed performance monitoring and tuning, identifying and resolving issues within the code realm.
- Led development and implementation of business intelligence architecture for delivering end-to-end BI solutions.
- Created Python scripts to automate AWS S3 data operations, control instance operations using AWS API, and utilize various AWS services such as EC2, ELB, S3, Glacier, CloudFormation, Elastic Beanstalk, Lambda, VPC, RDS, Trusted Advisor, and CloudWatch.
- Worked with databases, including Teradata, Oracle, and MongoDB, involving tasks such as creating databases, tables, triggers, stored procedures, and implementing ETL processes.
- Developed web applications using frameworks like Django and Flask, as well as creating and scheduling jobs using tools like Autosys and Arrow.
- Coordinated with AWS services such as Lambda, Data Pipeline, Route 53, EC2, ECS, and Docker for application deployment and maintenance.
- Developed and implemented scalable data pipelines using technologies like Spark, Databricks, Kafka, and AWS services for real-time data processing and analytics.
- Utilized Power BI for data analysis, integration, and decision-making, as well as developing code to read data streams from Kafka and perform transformations using Spark.
- Worked with Azure services such as Azure Data Lake, Azure Storage, Azure SQL, Azure DW, and utilizing Python and Spark/Py Spark for data processing and analysis.
- Designed ETL processes and data integration strategies, mapped & transformed data, and ensured data accuracy, consistency, and integrity.
- Used version control systems like Bitbucket/GitHub and utilizing continuous integration and deployment (CI/CD) tools like Bamboo and Jenkins.

Apr'16 – Jan'18 | Healthcare, New York, NY | Sr. Data Engineer / Data Scientist - (Cresitatech IT Solutions Pvt. Ltd.)
Environment: Java Spring Boot, EKS, ECS, AWS SageMaker, AWS Glue, S3, MYSQL Arora, AWS cloud, ETL, Teradata, MLIB, Power BI, TensorFlow, Scikit-learn, Keras, Pandas, NumPy, AWS CodeBuild, CodePipeline, MLOps, SageMaker-SDK, AWS Application Gateway. Route 53, Elastic IP, EC2 Instance, Load Balancer etc.

Responsibilities:
- Worked on Patient/Doctor/Physician Monitoring processes, validating and ensured adherence to HIPPA guidelines.
- Integrated diverse data sources, such as transaction data, customer information, and external data feeds, into a centralized data repository for AML purposes.
- Ensured data accuracy, consistency, and completeness through data validation checks and data cleansing techniques to support effective AML monitoring.
- Implemented robust data processing pipelines for handling large volumes of data in near-real-time or batch modes, using technologies like big data platforms, distributed computing frameworks, and data streaming tools.
- Implemented efficient workflows and automation scripts within PyCharm to streamline repetitive tasks, saving time and increasing productivity.
- Developed Light GBM regression models to predict associated risk ratings with locations.
- Proficient in Agile methodology, responsible for providing data science solutions from data gathering to deliverables.
- Researched and created AWS architecture in collaboration with data engineers and DevOps team to migrate models from on-premises to the AWS cloud.

- Developed and maintained ETL processes, including running SQL scripts, creating indexes, and stored procedures for data analysis. Skilled in designing various ETL strategies from heterogeneous sources and working with different data formats such as JSON and XML.
- Used Spark-SQL and Python on the Spark engine to develop end-to-end ETL pipelines.
- Worked with Apache Kafka for stream processing and utilizing Amazon S3 and CloudWatch for storage and monitoring purposes.
- Familiarity with AWS services and Blue Prism classification models for automating document classification. Strong mentorship skills, providing guidance and support to analysts across different teams on python libraries, frameworks, and AWS services.
- Responsible for using Python (NumPy, SciPy, pandas, scikit-learn, seaborn) and Spark (Py Spark, MLlib) for developing a variety of models and algorithms for analytics.
- Extensively used machine learning algorithms such as linear regression, classification, multivariate regression, Naive Bayes, Random Forests, K-means clustering, KNN, PCA, and regularization for data analysis.
- Used Apache Spark and Databricks for big data processing, utilizing Data Frame API, Spark SQL, and Spark Streaming.
- Handled data manipulation, transformation, and aggregation using Databricks' Data Frame API and Spark SQL.
- Integrated Databricks with cloud-based services i.e. AWS S3 or Azure Blob Storage for efficient data storage & processing.
- Hands-on experience in using natural language processing (NLP) techniques and visualization tools like Matplotlib for optimizing customer satisfaction and presenting findings effectively.

Jun'13 – Jun'16 | Salesforce Analytics, India | Data Analyst / Software Engineer - (Avyukta Intellicall Pvt. Ltd.)
Environment: OLAP, PL/SQL, SQL, OLTP, SSIS, SSAS, T-SQL, SSRS, UAT, Salesforce, Salesforce einstein, Python, Flask API, PHP Salesforce SDK etc.

Responsibilities:
- Designed ARM templates and used custom PowerShell scripts in Azure, saving 140 hours per new environment setup.
- Created Azure Data Factory pipelines for seamless data extraction, transformation, and loading across various storage systems.
- Migrated on-premises SQL databases to Azure SQL DB with SSIS packages for smooth data transfer to the cloud.
- Developed Azure notebooks for efficient ETL operations and data staging.
- Landed source datasets into Azure Data Lake Storage (ADLS) in Parquet file format for storage and retrieval. Implemented Agile methodology for data modelling projects.
- Automated regression testing with Python scripts to improve testing accuracy and efficiency.
- Developed Python programs/scripts for data transformation and harmonization.
- Managed Azure cluster maintenance, including node management.
- Administered user and group access in Databricks clusters for data security.
- Conducted data cleaning and transformation with SQL queries in Snowflake worksheets.
- Performed data quality analysis in Snowflake, building analytical warehouses to resolve data quality issues.
- Expertise in data warehouse architectures, ETL/ELT processes, and data security.
- Utilized advanced T-SQL features for database interactions and interfaces.
- Implemented conditional formatting in SSRS reports to enhance data presentation.
- Contributed in developing testing procedures, test cases, and User Acceptance Testing (UAT) for data-related projects.

References available on request