Farooq Ahmed Khan  
farooqahmedkhan240@gmail.com

Mobile: +1 (732) 832-3692

**Qualification Summary**

A self-motivated and diligent professional with over 11+years of progressive industry experience in data analytics, governance, business analysis, and engineering across various domains. Skilled in project management/process management, including Agile and Scrum methodologies, I excel in delivering data solutions utilizing the latest technologies, such as Enterprise Data Warehouse (cloud), Master Data Management, Big Data, and Business Intelligence BI. Known for meticulous attention to detail and a strong technical acumen, I effectively translate business needs into actionable analytics solutions. Adaptable and efficient, I thrive in fast-paced environments and work seamlessly both independently and within team settings.

**Skills Summary**

**Industry Experience:** Consumer/Retail, HR, Supply Chain, eCommerce, Financial/Banking, and GIS

**Competencies:** Big Data solutions, Data Engineering, Business Intelligence Solutions, Enterprise data warehouse design and development, Extract Transform & Load (ETL), Data Analysis, Data Integration and Extraction, Data quality, Machine learning, Master Data Management (MDM), Data Modeling, Data Governance, Visualization, and Reporting.

**Languages:** SQL, T-SQL, Python, Python shell scripting, Java, R, PHP, html and C++

**Software/Tools:** AWS Redshift, Glue, IAM roles, S3, AWS Lambda and Step Function, Glue Workflow (Orchestration), AWS RDS, Redshift Spectrum, AWS Kinesis, CI/CD, VPC, AWS Athena, AWS Data Catalog, Amazon EMR, Hadoop(HDFS), Hive, Teradata, Sqoop, Apache Spark, CloudWatch, EC2, Snowflake, G-suite, Open Refine, Collibra, Jupyter, Spyder, Oyster, Teradata, Databricks, Weka, MS Access, MS Office, Mircrosoft Excel, Cloud Coumputing, Matlab, JIRA, SAS, Power BI tool, Tableau, GitHub, Git, Looker, and Qlit sense, Azure, GCP.

**Databases:** MySQL, Microsoft SQL Server, Oracle, PostgreSQL, and AWS Dynamo DB (NoSql)

**Operating System:** Windows, Linux, Unix.

**Experience :**

**Client: CITI BANK, USA (Jan 2023 – May 2024)**

**Role: Data Analyst**

* Spearheaded data governance initiatives by identifying critical data quality issues and creating comprehensive data governance policies and standards.
* Developed and implemented data quality rules to assess critical data elements across multiple data quality dimensions, and designed data profiling processes by cleaning census data using Python libraries.
* Assisted in the delivery of operational reports by writing complex SQL and T-SQL queries to extract and manipulate data from databases, ensuring accurate and timely data reporting.
* Utilized PySpark to efficiently process and analyze large data, ensuring scalable data handling and improved performance in data processing tasks.
* Expertise in analyzing data warehouse and creating multidimensional cubes using SQL Server Analysis Services SSAS.
* Collaborated with risk management, finance, and IT teams to streamline reporting and ensure timely delivery of regulatory reports.
* Developed and maintained financial models to forecast business performance, improving predictive accuracy and helping reduce operational risks by 15%.
* Analyzed key financial risk metrics and trends, providing data-driven recommendations that improved the bank's credit risk assessment process.
* Automated routine data extraction and reporting tasks using Python and SQL, reducing manual efforts by 30% and increasing operational efficiency.
* Utilized SQL and MS Excel to extract, manipulate, and analyze complex data sets, providing insights that led to a 10% increase in operational efficiency.
* Implemented data quality checks and processes, ensuring the accuracy and consistency of data across various departments, resulting in a 20% reduction in data discrepancies.
* Worked with senior stakeholders to define and track key performance metrics, ensuring alignment with corporate objectives and enabling more effective resource allocation.
* Created insightful reports and dashboard to track work on census data for internal and external stakeholders using PowerBI.
* Documented procedures and protocols to standardize data quality measures and improve overall data integrity.

**Client: JP Morgan Chase (May 2020 – July 2022)**

**Role: Data Process Specialist/Subject Matter Expert**

* Successfully executed data export/import operations into AWS S3 and DynamoDB from various sources, including databases, market data providers, and banking APIs, increasing data accessibility and integration efficiency by 15%.
* Expertly designed, developed, and maintained scalable and high-performance big data ETL pipelines using AWS Glue, Python, Apache PySpark, and Hive, extracting data from diverse sources and loading it into Amazon Redshift, resulting in a 30% improvement in data processing speed and enabling more efficient analytics and decision-making across the organization.
* Efficiently ingested data from databases into Amazon Redshift and RDS, facilitating improved data accessibility and analysis, which cut down data retrieval time 20%.
* Developed and executed complex SQL queries across Amazon Athena, Redshift, and Redshift Spectrum to analyze and derive insights from large datasets, enhancing data-driven decision-making processes.
* Orchestrated complex data transformations using AWS Lambda, Step Function, and AWS Glue workflows, refining raw data from multiple sources into valuable data assets, improving data delivery efficiency by 30%.
* Efficiently processed, implemented and managed big data processing workflows on AWS EMR cluster consist of EC2 nodes, demonstrating expertise in configuring and optimizing PySpark script. Leading to an increase in processing efficiency and created auto scaling groups of EC2 instances for efficient scaling.
* Ensured data privacy & security and compliance by implementing measures to keep data separated and secure across national boundaries and multiple data centers, utilizing AWS IAM Roles and VPC to maintain stringent access controls.
* Implemented monitoring, forecasting models using advanced statistics, machine learning algorithms and statistical modeling to analyze complex historical data, driving data-driven decision-making and optimizing business processes. Also utilized data visualization techniques to analyze historical performance/descriptive analytics, correlations, and risk factors across diverse assets of capital market.
* Worked across various asset classes (Fixed Income, Equities and Commodities), financial instruments, benchmarks, and indices, complemented by proficiency in utilizing leading market data platforms such as Bloomberg, ICE, and Rimes to conduct comprehensive market analysis and support strategic decision-making processes.
* Conduct thorough risk assessments and scenario analyses using historical data and quantitative techniques to evaluate potential impacts on asset class performance, providing strategic recommendations to mitigate risks and enhance portfolio.
* Designed transaction dashboards that highlighted critical key performance indicators KPI such as revenue, total interest, card type, payment mode, transaction counts and so on, significantly improving strategic decision-making and operational efficiency.
* Built comprehensive credit card customer reports and dashboards, showcasing revenue insights by CSS, income, total interest, top-performing states, and other demographics, supporting more effective customer segmentation and marketing strategies.

**Client: APPLE (March 2017 – April 2020)**

**Role: Senior Data Analyst**

* Led end-to-end projects independently, from gathering requirements to on-time delivery, effectively communicating with the project manager, team members, sponsors/client and stakeholders throughout the process by utilizing Agile Scrum framework in JIRA.
* Performed EDA exploratory data analysis by utilizing statistical techniques and tools to identify trends, outliers, and correlations, and supporting data collection, evaluation, and model execution. Expert in collecting, organizing, and analyzing data to generate accurate reports for both internal and external stakeholders.
* Expertise in analyzing data warehouse and creating multidimensional cubes using SQL Server Analysis Services SSAS.
* Presented actionable insights derived from SQL & T-SQL queries and data visualization tools, MS Excel guiding product strategy and roadmap decisions and driving revenue growth.
* Implemented rigorous data quality checks, validation rules, and cleaning processes for data preparation using programming language like Python libraries, MS Excel, and Open Refine, which resulted in improvement in data accuracy and reliability.
* Conducted comprehensive audits of financial, customer, GIS and HR data, assessing data completeness, positional accuracy, and attribute consistency, and providing detailed feedback and recommendations for data quality improvements to enhance map accuracy and user satisfaction.
* Proficient in creating, managing, and optimizing Snowflake databases for efficient data storage and retrieval.
* Expertise in data ingestion, specifically adept utilizing the SnowSQL CLI to seamlessly transfer data from diverse sources into the Snowflake database
* Analyzed employee production and validation data to identify bottlenecks and inefficiencies, leading to the implementation of process improvements that boosted overall productivity by 30% and enhanced data quality by providing validation checks.
* Led the development of a comprehensive GIS data strategy for Apple Maps, outlining a roadmap for integrating and optimizing vendor-provided geospatial datasets.
* Developed a roadmap outlining both short-term and long-term strategies to enhance operational efficiency and data quality for Apple Maps, achieving immediate improvements in production and validation by 15%.
* Led RCA (Root Cause Analysis) Team to successfully retrieve the root cause of the problems using 5 whys
* methodology based on the feedback and problems reported by the business.
* Prepared Corrective and Preventive Action (CAPA) and report basis the root cause analysis/error trend with the higher management to help business/client/customer with the insights for project, product growth and process improvement.
* Led consumer-focused projects to boost sales performance by 25% and improve customer retention by 15%, leveraging advanced business analytical models developed and maintained based on comprehensive sales data utilizing Statistical Analysis, Co-relation Analysis. This initiative enhanced decision-making related to sales performance, customer behavior, market trends, and product success.
* Designed and implemented ETL or data pipelines to load over 25 million records of customer and shipment data from a data lake consisting of datasets from multiple sources (Salesforce, APIs, databases, SAP, CRM) into Hadoop Distributed File System (HDFS) and transformed using Hive to load over relational database (RDBMS) for analysis.
* Conducted quantitative analysis and applied machine learning algorithms, resulting in a 30% improvement in identifying performance issues and delivering actionable insights that optimized sales strategies and improved overall effectiveness.
* Designed and implemented new key performance metrics, dashboards, and scorecards, transforming raw sales into comprehensive reports that supported strategic decision-making and performance tracking.
* Cultivated strong cross-functional relationships with sales, finance, and operations teams, using sales data to streamline processes and resolve reporting issues, thereby improving data integrity and operational efficiency.
* Responsible for reviewing financial statements, completing cost and revenue analyses, and performing other basic finance reporting tasks, including P&L, gross margin, and profitability.
* Completed in-depth business valuations, monitored, and analyzed financial data trends, and prepared new financial models based on the analysis.
* Conducted in-depth interviews and surveys with production teams, validation teams, internal and external stakeholders to understand top business challenges.
* Conducted data governance maturity model assessments to evaluate and strengthen data governance practices, collaborate with stakeholders to identify gaps and implement targeted improvements.
* Created and maintained comprehensive documentation, including FAQs, training materials, presentations, and business definitions, business strategies supporting the seamless adoption and ongoing use of business intelligence solutions.
* Streamlined data processing using DAX functions and power query editor for ETL to convert data into insights. Also performed analysis to monitor KPI’s and trends.
* Collaborated with cross-functional teams (project managers, software engineers, data engineers, data scientists) to develop and maintain interactive dashboards for tracking operation’s production and quality, customer sales and loyalty program metrics using PowerBI/Tableau providing real-time insights with stakeholders to support decision making processes.

**Client: Google (May 2013 – February 2017)**

**Role: Data Analyst**

* Led end-to-end data analysis projects, including data extraction, profiling, cleaning, data modeling, transforming, analyzing and reporting based on the business problems and requirements.
* Proficient in leveraging Big Query to perform advanced data analysis and reporting, including querying data from diverse sources such as Data Lakes, APIs, and data warehouses.
* Examined datasets and determined the best analysis plan to address key business questions and create value. Owned all aspects of data during the implementation of changes, driving improvements in customer engagement and value metrics.
* Developed advanced SQL/PL-SQL queries encompassing data extraction, filtering, aggregation, and integration techniques, including expertise in Joins, Stored Procedures, Views, window functions and Indexes for comprehensive data manipulation and optimization.
* Engaged with stakeholders to gather and understand business requirements through active communication and collaboration, ensuring alignment between technical solutions and organizational goals.
* Conducted thorough correlation, regression analysis and hypothesis testing to predict and uncover patterns, trends, and anomalies, empowering the organization to make data-driven decisions and drive continuous improvement.
* Analyzed user interactions, including search behavior, map usage, and engagement with Google Zagat Website to help business to refine their services and offerings to better meet customer needs, enhancing overall satisfaction and loyalty.
* Tracked and reported KPIs such as user engagement, click-through rates CTR, CMP and conversion metrics for Google Zagat by retrieving insights from the user usage, feedback, reviews, and rating data aiding in business and marketing strategies, allowing them to focus on what customers value most.
* Conducted funnel analysis on Google Zagat Website to identify user drop-off points, understand customer journey stages, and optimize the review process for enhanced user engagement and improved business insights.
* Executed comprehensive data wrangling processes, ensuring data integrity and consistency across multiple sources. Designed and conducted A/B tests to evaluate the effectiveness of marketing strategies, leading to a 15% improvement in customer engagement.
* Prepared and provided the final datasets to be integrated into Google Maps by extracting, cleaning, and merging data from multiple sources, ensuring accuracy and completeness for optimal mapping functionality.
* Performed behavioral analysis of user interactions with maps, such as search behavior and route preferences, to understand customer habits and preferences.
* Facilitated seamless data discovery and improved data governance by managing the organization's data catalog, including metadata, data schema, and data lineage, along with maintaining an up-to-date data dictionary.
* Ensured the accuracy and reliability of critical organizational data by effectively managing master data, incorporating feedback from internal teams to maintain high data quality standards.
* Conducted workshops and training sessions on data quality and governance, empowering users to take ownership of data assets and contribute to overall data integrity.
* Assisted Service Manager and Director in providing management status reports and designed intuitive dashboards, charts, and reports using MS Excel (charts, Pivot and Lookup), PowerBI that facilitate informed decision-making and enhance data-driven presentations for stakeholders.

**Recognitions and Certifications**

* Digital First Certification Cognizant (Aug 2021)
* ISOO 8000 Master Data Quality Manager (MDM) (Sep 2023)
* Data Analytics and Visualization Job Simulation Accenture (January 2024)
* Data Science Foundations by IBM (2024)
* Data Analysis with Python by IBM (2024)
* SQL and Relational Database Management 101 by IBM (2024)

**Education**

* Master’s in computer science and information science -University of Arkansas at Little Rock, Little Rock, Arkansas
* Bachelor of Engineering in Information Technology - Osmania University, Hyderabad, India (2012)