

Professional Summary:

A dedicated and data-driven professional with over 9 years of experience in data engineering, data analysis, and quality assurance. Proven proficiency in designing and maintaining scalable data pipelines, utilizing cutting-edge technologies and tools such as Spark, Python, AWS, PowerBI, and SQL. Experienced in leading cross-functional teams, conducting data-driven gap analysis, and optimizing data quality and systems. Excels in communicating complex data insights to non-technical stakeholders. Adept at executing both manual and automated testing, and implementing effective data governance strategies. Seeking a challenging Data Engineer role that allows me to leverage my extensive experience in data analysis, data quality assurance, and data engineering. Aiming to contribute to a forward-thinking team where I can utilize my skills in big data technologies, cloud services, and machine learning to drive data-informed business decision making and foster data-driven innovation.

Technical Skills:

Programming Languages:	Java, Python, SQL, NoSQL
Big Data Technologies:	Apache Spark, Hadoop, Hive, DataBricks
Data Warehousing/ETL Tools:	Informatica, Apache Spark ETL, SSIS, ETL Process Validation
Database Systems:	DB2, Teradata, SQL Server, Oracle
Data Modeling:	Logical Data Modeling, Entity-Relation Diagramming
Data Visualization Tools:	PowerBI, QlikView, Tableau, iRise
Data Analytics:	Predictive Modeling, Statistical Analysis, MLlib for Machine Learning
Project Management Methodologies:	Agile/Scrum, SDLC, JAD Sessions, Use Case Development, Gap Analysis
Software Development Tools:	Rational Application Developer (RAD), Eclipse, Github, PyCharm, Jupyter, Visual Studio code
Testing Tools:	SoapUI, Fitnesse, Postman, ALM/Quality Center
Reporting Tools:	Crystal Report, SSRS, Advance Excel
Data Governance Tools:	Informatica Master Data Management (MDM), IBM MDM
Cloud Services:	AWS, GCP, Azure
Defect Management Tools:	TFS Azure, Rally

Professional Experience:

CVS Health. Wellesley, MA (Remote)	May 2022 – Present
------------------------------------	--------------------

Big Data Engineer

Responsibilities:

- Develop and implement data processing algorithms, ensuring optimal data delivery architecture is consistent throughout Pharmacy Benefits Management (PBM) services projects.
- Design, build, and maintain efficient, reusable, and reliable data pipelines using pySpark, DataBricks and other new technologies.
- Leverage new technologies for enhancing the performance, scalability, and reliability of the ETL/ ELT development process.
- Collaborate with data scientists and architects on several projects to ensure data solutions are aligned with company initiatives.
- Supported existing cloud Data Management implementations, addressing performance issues, troubleshooting and resolving technical challenges, and implementing necessary enhancements.
- Automate tasks through appropriate tools and scripting.
- Designed data analysis pipelines with Python, leveraging cloud services for efficient processing and storage.
- Validate and maintain software functionality and security.

- Participate in the code review process, ensuring code quality and sharing knowledge with the team.
- Analyze raw data to draw insights and develop strategies.
- Extracted real-time data feeds with Spark Streaming, converting them to RDDs and processing data as Data Frames in HDFS.

General Motor- Austin, TX

Oct 2020 - May 2022

Data Engineer / Data Analyst

Responsibilities:

- Collaborate with cross functional team of 10 GM EPP (Error Proofing Platform) and Torque Strategy team member to improve the improvement the vehicle manufacturing validation process.
- Created Features, User Stories, and use case scenarios for the EPP system working with business SMEs, and technical engineers.
- Worked on Entity Relation, workflow diagram, dimensional model and mockup for ease of understanding of the business process using BPMN.
- Validated and logged defects/bugs in the TFS Azure server.
- Create Business reports/ dataset in PowerBI and python importing from excel/ multiple data source.
- Optimized report metric performance by using Power Query and DAX Query function.
- Create and maintain optimal CI/CD data pipeline architecture, transform raw data into a suitable format for analysis.
- Ensure data is compliant with all regulations and maintain data privacy. Develop and implement data standards, quality rules.
- Used cloud services (like AWS, GCP, Azure), databases, ETL tools, and DataBricks solutions for optimal extraction, transformation, and loading of data from a wide variety of data sources
- Analysed the data to find trends and patterns applying statistical methods to derive insights using SQL and big data technologies

Ford Motor Company – Dearborn, MI

Oct 2019 – Jun 2020

Data Engineer / Data Scientist

Responsibilities:

- Profiled technical and business metadata, context curation, lineage and relationship, data quality, and profiling usages.
- Led Informatica data catalog to capture technical metadata & lineage enabling project teams to make better decisions reducing project time by 10% & increasing the accuracy of data models by 5%-25%
- Developed ETL pipelines using Apache Spark/DataBricks solutions for efficient data processing and analytics, resulting in enhanced data quality and 15% improvement in processing speed.
- Applied knowledge of data science to improve data dimensional modeling techniques, leading to more insightful business intelligence reports and predictive models.
- Used pySpark's MLlib library for machine learning tasks on large datasets, boosting the predictive accuracy of data models by 20%.
- Implemented plans for extracting and loading HIVE metadata in a metadata repository within the Big data lake- Hadoop ecosystem.
- Created business process documentation and diagrams for users – to populate data assets in the Enterprise Data Catalog (EDC) environment.
- Curated Data profiling, Created dataset in QlikView dashboard, Business terms/ Data domain, and association with metadata through Machine Learning process.
- Successfully preserved the 7-point standard for data quality and governance, strengthening the organization's data integrity.

Xylem Inc. – Rye Brook, NY

Jul 2019- Oct 2019

Business Data Analyst

Responsibilities:

- Mastered customer and vendor information in Informatica Master Data Management (MDM). Cleansed data, ensured data standardization, stored golden records, and constructed MDM metric reports in the BI reporting tool.
- Initiated meetings with Data Governance/ Data Quality SMEs, and IT team members to discuss organizational requirements; held daily meetings with onshore and offshore teams to provide status updates
- Identified and obtained requirements critical to the business process flow, creating MDM metric Dataset as needed; prepared report through Tableau reporting.
- Participated in the Extract/Transform/Load (ETL) process by configuring various sources to target databases; supported the Secure Shell File Transfer Protocol (SFTP)/Amazon (AWS)-S3 process for transferring report files to the Tableau BI
- Developed, updated, and maintained report designs and interactive dashboards with the Tableau BI vendor.

Ford Motor Company – Dearborn, MI

Jan 2018 – Jun 2019

Data Quality Engineer/ Analyst

Responsibilities:

- Acted as the Quality Assurance (QA) Engineer for a data-centric project that entailed managing and improving the quality of Service Component Architecture (SCA-C) data, refining IBM MDM Framework (GMF) processes, and standardizing data quality interfaces.
- Developed and updated Use Case documentation to bolster existing data services and establish Proof of Concept (POC) solutions for new data-oriented services.
- Analyzed and improved Java code pertaining to data processing and manipulation; migrated corresponding Java files from AccuRev/Github into a Rational Application Developer (RAD)/Eclipse workspace.
- Facilitated daily SCRUM meetings, centering discussion on data quality and assurance, to address concerns and issues from the IT, product management, and business departments.
- Executed Gap Analysis on the Data warehouse to uncover existing issues; maintained a record of data defects in an issue log using a defect management tool (ALM/Quality Center) and updated tasks in Rally based on data quality findings.
- Created and executed unit and functional data test cases using SoapUI, Fitnesse, and Postman for both manual and automated testing; implemented XML request and response validation and collaborated with end-users to conduct User Acceptance Testing (UAT).
- Devised and deployed data validation mechanisms and data set reports using DB2/Teradata/SQL, ensuring data integrity and reliability.

United States Army Reserves – Austin, TX

Aug 2015 – Sept 2022

CBRN Specialist

Responsibilities:

- Fulfilled ordnance, maintenance, decontamination, and disposal responsibilities in support of domestic and global Chemical, Biological, Radiological, and Nuclear (CBRN) operations.
- Provided gas chamber training and awareness to military personnel, ensuring the proper use of the M50 protective mask
- Advised, managed, and performed environmental protection programs, coordinating the relevant laws and regulations accordingly
- Gathered and analyzed information, identified areas for improvement, drafted findings and recommendations, and prepared reports in Excel, Powerpoint, Visio accordance with executive guidance and professional standards

Department of Health and Environmental Control – Columbia, SC

Mar 2014 – Nov 2015

Technical Data Analyst

Responsibilities:

- Implemented a Data Governance-Master Data Management (IBM MDM) project to design one master record from different systems including Healthcare, Medicare/Medicaid, and finance operations.
- Communicated with data stewards and users to better understand needs and areas for improvement, providing solutions utilizing a structured Software Development Life Cycle (SDLC) approach; verified and validated associated Medicare/Medicaid eligibility requirements.
- Developed a MDM- data warehouse system metric, document survivorship process, data mapping, logical data modeling, and sharing documents
- Developed business process diagrams and Entity-Relationship (ER) diagrams, creating Pseudo code and SQL queries to filter data within the Oracle/SQL database
- Established and maintained a traceability matrix (RTM) to ensure the completeness of requirements during various SDLC stages.
- Defined User stories using Case Scenarios and narratives. Wrote Test Cases and Test Scenarios. Tracked UAT issues/bugs and prepared reports.
- Produced sketches, wireframes, and clickable prototypes for testing and illustration purposes.
- Effectively communicated UAT results between Users and the Development team, and provided recommendations for Change Control Requests (CCR).

Pima County- Tucson, SC**Oct 2011- Mar 2013****Business Systems Analyst****Responsibilities:**

- Developed CRM (Advantage and Maximo) application and Crystal Reporting with a focus on translating business needs into solutions.
- Translated business requirements into Technical Design specifications and Mockups for implementation of the “Overtime Reporting” system for the Finance and HR group.
- Facilitated Joint Application Development (JAD) sessions, and workshops with the development team, business stakeholders, and end-users.
- Extracted data from Database/ Data Warehouse using SQL server for data mapping and mart, FMS to AMS, Maximo.
- Planned Test cases, Test scenarios, and implemented them on the interfaces in the test environment.
- Executed System Test and worked with the customer on a manual workflow for UAT ensuring reports confirm the specifications of business requirements and performance criteria.

Education:

Wilmington University

May 2015

Master of Science | Information System Technology | Concentration in Internet/Web Design

University of Nebraska at Omaha

Dec 2010

Bachelor of Science | Pre-Actuarial Mathematics

REFERENCES AVAILABLE UPON REQUEST