Mate Yilma Lead Data Engineer

ThreeFourSeven ThreeFiveThree EightTwoFiveFive

Summary:

- Lead Data Engineer with over 8 years of experience in designing, developing, and optimizing complex SQL databases, proficient in crafting efficient and high-performance SQL queries, stored procedures, and triggers.
- Proficient in managing and optimizing NoSQL databases, including MongoDB, Cassandra, and DynamoDB, with a track record of successfully implementing scalable and flexible data solutions.
- Expertise in utilizing SSAS to design and build multidimensional data models, OLAP cubes, and data mining structures.
- Led and managed end-to-end SQL development projects and migrations, ensuring successful delivery within specified timelines and meeting project objectives.
- Experience in Extracting, Transforming and Loading (ETL) data from Excel, Flat file, Oracle to MS SQL Server by using BCP utility, DTS and SSIS services.
- Hands on experience on various SQL Server Services like Integration Services (SSIS), and Reporting Services (SSRS).
- Extensive knowledge of Normalization and Relational Database Management Systems (RDBMS).
- Proficient in the entire software development life cycle (SDLC) and various project methodologies.
- Extensive experience implementing ETL Packages using Data Transformation Services (DTS) and SSIS across multiple versions.
- Successfully migrated DTS packages to SSIS and adeptly modified using DTS editor wizard.
- Created dashboard solutions for analyzing STD statistics using SSIS cubes and Tableau.
- Highly skilled in SSIS tools, including Import and Export Wizard, Package Installation, and SSIS Package Designer.
- Worked with Version Control Systems such as Team Foundation Server (TFS) and Git.
- Proficient in using SSIS Control Flow items (For Loop, execute package/SQL tasks, Script task, send mail task) and SSIS Data Flow items (Conditional Split, Data Conversion, Fuzzy lookup, Fuzzy Grouping, Pivot).
- Trained and experienced in Splunk, developing dashboards on Splunk server.
- Extensively worked on Erwin Data Model Tool with strong knowledge of Relational Database Modeling/Loading and Dimensional database modeling concepts.
- Expertise in using other ETL and Reporting tools, specifically Informatica and Cognos.
- Proficient in Microsoft toolsets, including Cognos and Power BI.
- Experience in utilizing SAS for extracting data from diverse operational sources on multiple platforms, building data warehouses, and integrating extracted data into data marts.
- Extensive experience in Hadoop (HDFS) to store and process large volumes of structured and unstructured data.
- Proficient in Python and Shell scripting to automate data processing tasks, improve data quality, and enhance ETL workflows.
- Skilled in leveraging Azure cloud services to design and implement data solutions, including data lakes, data warehousing, and real-time data processing.
- Expertise in designing and building OLAP Cubes, providing multidimensional data analysis capabilities to business stakeholders and supporting advanced data modeling and querying for complex analytical needs.
- Proficient in creating comprehensive Crystal Reports, transforming raw data into visually appealing and insightful reports.
- Proficient in developing (complex) stored procedures, optimizing data retrieval and manipulation for critical business applications, and ensuring efficient database performance and data integrity.
- Experienced in using Scala programming language and Apache Spark, enabling real-time and batch data analytics for large volumes of structured and unstructured data.
- Expert in data cleansing and data mining techniques utilizing SQL queries and data profiling tools to identify and rectify data quality issues.
- Experienced in working with JSON data formats, extracting, parsing, and transforming JSON data using SQL and other programming languages to support data integration and interoperability between systems.
- Extensive experience in working with relational database management systems (RDBMS), including SQL Server, Oracle, NoSQL, and MySQL, developing complex queries, stored procedures, and database design.
- Proficient in using MS Visio to create data flow diagrams and visual representations of database structures.
- Skilled in developing MDX scripts for querying and analyzing multidimensional data in OLAP cubes, providing valuable insights.
- Experience managing job scheduling and automation using Control M, ensuring efficient and timely execution of SQL processes and data-related tasks.

- Experience using IBM Information Server for data integration and governance, successfully designing and implementing data pipelines to ensure data quality and compliance with data standards.
- Skilled in creating ad-hoc reports using SQL queries, enabling on-demand data analysis and reporting for stakeholders.
- Expert using Hive to handle large-scale, distributed data processing and analysis.
- Skilled in implementing QlikView BI Tool to develop interactive and dynamic data visualizations, empowering stakeholders with data-driven insights for decision-making.

Experience:

Bank of America - NYC, NY 11/2020 - Present Lead Data Engineer

- Developed and implemented complex SQL queries, stored procedures, and functions to extract, transform, and load (ETL) data from diverse sources into target databases for migration or integration purposes.
- Designed and optimized database schemas, table structures, and indexes (SQL/NoSQL/PLSQL) to enhance query performance and improve overall system scalability for large-scale data migrations.
- Developed comprehensive data migration strategies and plans, including data mapping, data validation, and data cleansing activities, to ensure accurate and reliable data transfer between source and target systems.
- Utilized NoSQL-specific query languages, such as MongoDB Query Language (MQL) and Cassandra Query Language (CQL), to perform data retrieval, filtering, and aggregation for analysis and reporting purposes.
- Developed and optimized complex NoSQL queries and data manipulation operations to extract, transform, and load (ETL) data from diverse sources into target NoSQL databases.
- Utilized Tableau's data blending and data connection capabilities to integrate data from multiple sources, including relational databases, flat files, and APIs.
- Implemented ETL (Extract, Transform, Load) workflows using SQL and Hadoop tools like Sqoop and Apache Hive.
- Developed and optimized complex ETL processes using SSIS, facilitating efficient data integration and migration from various sources into SQL Server databases.
- Designed and implemented data warehouse solutions using SSIS, including data modeling, dimensional modeling, and ETL workflows, to support analytical reporting and business intelligence requirements.
- Developed and maintained SSRS reports, leveraging SQL queries and data visualization techniques, to present meaningful insights and key performance indicators (KPIs).
- Used DevOps methodologies to implement automated performance testing and load testing for data engineering processes, ensuring system scalability and stability.
- Developed and maintained Splunk apps and add-ons, extending the capabilities of Splunk through custom SQL queries and configurations, enabling integration with external systems and data sources.
- Developed and maintained automated data pipelines using Snowflake with DevOps, orchestrating complex data workflows and scheduling data processing jobs.
- Utilized AWS Glue for data extraction, transformation, and loading (ETL), and automating data processing workflows.
- Designed and implemented data pipelines using AWS services, including AWS Lambda, AWS Step Functions, and AWS Batch, to orchestrate data processing tasks and workflows.
- Implemented version control and change management processes for Power BI artifacts, utilizing tools like Git and SVN, ensuring proper tracking and traceability of changes made during the project lifecycle.
- Implemented data lineage and metadata management practices within Power BI, ensuring transparent and traceable data flows and promoting data governance and compliance.
- Actively monitored and resolved data integration and reporting issues within Power BI, conducting root cause analysis, troubleshooting, and performance tuning.
- Utilized Snowflake's semi-structured data support for handling and querying JSON, Avro, and other data formats.
- Developed and implemented normalization strategies to decompose database tables into smaller, more manageable entities, ensuring data integrity, eliminating data redundancy, and minimizing update anomalies.
- Conducted data analysis and business requirements gathering to identify data dependencies, functional dependencies, and normalization opportunities within the database schema.
- Implemented Azure SQL Data Warehouse (now Azure Synapse Analytics) for scalable and distributed data processing, leveraging SQL-based querying and data manipulation techniques to optimize performance and ensure efficient data retrieval.
- Designed and implemented database schemas and table structures within Azure SQL Database, leveraging SQL skills to ensure efficient data storage and retrieval, and optimizing performance for cloud-based applications.

- Conducted performance tuning and optimization of OLAP cubes, analyzing cube processing times, query performance, and usage patterns to optimize cube design, aggregation levels, and indexing for improved performance.
- Conducted data modeling workshops and training sessions, utilizing MS Visio to educate stakeholders on data modeling concepts, ER diagrams, and best practices.
- Implemented forward engineering using Erwin to generate SQL scripts and DDL statements, automating the creation of database objects and tables based on the logical and physical data models.
- Developed and maintained dynamic Cognos reports and interactive dashboards, providing vital data insights to key stakeholders.
- Deployed Control M to meticulously schedule and automate ETL processes, ensuring timely and accurate execution of critical tasks.
- Managed and administered IBM Information Server, including DataStage and QualityStage components, to create robust data integration and data quality processes.
- Expertise in Hive to efficiently process and analyze large-scale datasets within Hadoop environments for seamless data querying and advanced analysis for large-scale big data projects.
- Created MDX scripts to effectively query and analyze multidimensional data in OLAP cubes.
- Used Scala's ability to handle JSON data formats to enable seamless data integration and interoperability between various systems.
- Developed and maintained data transformation workflows using Scala and Apache Spark.
- Conducted data profiling and data quality assessments using Informatica Data Quality (IDQ).
- Implemented Informatica mappings, sessions, and workflows to orchestrate data movements, transformations, and data cleansing tasks.
- Led a team of developers and analysts in designing interactive and dynamic Cognos reports, visualizations, and data models.
- Utilized Python libraries for data analysis and data transformation, such as Pandas and NumPy, to enable advanced data processing and data enrichment for business intelligence.
- Conducted thorough testing and validation of Python and Shell scripting-based solutions, including functional testing, data validation, and performance testing.
- Developed and optimized DTS packages to efficiently transfer and transform data between databases, ensuring seamless integration and data synchronization.
- Conducted performance tuning and optimization of stored procedures, analyzing execution plans, optimizing SQL code, and leveraging indexing strategies to enhance system performance and responsiveness.
- Utilized stored procedures to encapsulate business logic and data processing rules, ensuring data integrity and consistency across applications and databases.
- Utilized data cleansing methodologies to identify and rectify data quality issues, improving data integrity and reliability within databases.
- Utilized JSON data for seamless integration with NoSQL databases, enabling hybrid data environments and supporting flexible data storage and retrieval.
- Utilized data visualization tools and techniques to present ad-hoc reports in a clear and concise manner.

Verizon - NYC, NY 11/2020

Senior Data Engineer/Lead

 Conducted data profiling and analysis to identify data quality issues and implemented data cleansing and transformation techniques using SQL to improve data integrity and consistency during migration processes.

04/2018 -

- Implemented data extraction methodologies from various sources, such as flat files, databases, and APIs, using SQL scripts and tools, ensuring seamless data integration into target databases.
- Utilized SQL performance tuning techniques, including query optimization, index management, and query plan analysis, to enhance database performance and reduce query execution time for critical migration projects.
- Developed and maintained Tableau workbooks and dashboards, implementing best practices for data visualization, user interactivity, and performance optimization.
- Developed and optimized complex SQL queries to perform data transformations and aggregations on HDFS data, leveraging Apache Hive for querying and analyzing large datasets.
- Collaborated closely with business stakeholders to understand their reporting needs and translate them into visually impactful SSRS reports and dashboards.
- Developed and maintained Snowflake data warehouse automation scripts, facilitating efficient deployment and version management.
- Utilized SSAS multidimensional and tabular models to design and implement OLAP cubes, enabling fast and efficient data analysis and multidimensional reporting.

- Conducted data profiling and analysis to identify data quality issues and implemented data cleansing and transformation techniques within SSIS to ensure high-quality data for reporting and analysis purposes.
- Implemented correlation searches and alerts within Splunk, leveraging SQL and Splunk query language to proactively identify and notify stakeholders of critical events, anomalies, and security threats.
- Conducted data parsing, extraction, and enrichment within Splunk using SQL-based transformation techniques.
- Developed and maintained Power BI dashboards, reports, and visualizations using SQL queries and Power Query, enabling users to explore and analyze data through interactive charts, graphs, and tables.
- Implemented data transformation services and cleansing techniques within Power BI using SQL-based data manipulation queries.
- Utilized DAX (Data Analysis Expressions) language within Power BI to create calculated columns, measures, and custom aggregations, providing advanced data calculations and analysis capabilities.
- Implemented security measures within Snowflake, defining user access controls and permissions, to protect sensitive data and prevent unauthorized access.
- Designed and implemented data migration strategies, utilizing AWS services such as AWS Database Migration Service (DMS) and AWS Snowball, to seamlessly transfer and consolidate data between on-premises and cloud environments.
- Implemented DevOps automation using AWS OpsWorks and AWS Elastic Beanstalk, simplifying the deployment and scaling of data engineering applications.
- Developed and implemented normalization strategies to decompose database tables into smaller, more manageable entities, ensuring data integrity, eliminating data redundancy, and minimizing update anomalies.
- Developed and maintained Azure Analysis Services models, utilizing SQL-based data modeling techniques to create multidimensional and tabular models for efficient data analysis and reporting.
- Implemented Azure Data Lake Storage as a data repository for structured and unstructured data, utilizing SQL skills to guery and analyze data stored in Azure Data Lake Store.
- Developed and maintained data connections and data sources within OLAP cubes, leveraging SQL skills to extract and transform data from various NoSOL databases.
- Collaborated with data architects and analysts to ensure proper integration of data sources and alignment with dimensional models within Cognos, utilizing SQL skills to extract, transform, and load (ETL) data into Cognos.
- Implemented version control and change management processes for MS Visio diagrams, utilizing tools like Git and SVN.
- Conducted reverse engineering of existing databases using Erwin, extracting metadata and generating visual representations of database structures.
- Utilized Python and Shell scripting to automate data processing tasks, improve data quality, and enhance ETL workflows.
- Developed and optimized ETL processes using Informatica PowerCenter, extracting, transforming, and loading data from various sources into relational databases and NoSQL systems.
- Implemented security and access control measures within DevOps processes, ensuring data protection and compliance with data privacy regulations.
- Utilized MDX scripts to query and analyze multidimensional data in OLAP cubes, providing valuable insights to business stakeholders.
- Experienced in using Control M for job scheduling and automation, ensuring timely execution of ETL processes and data-related tasks, while maintaining data consistency and accuracy.
- Conducted performance tuning and optimization of data engineering processes using DevOps tools and practices, ensuring efficient data processing and system scalability.
- Proficient in IBM Information Server, including DataStage and QualityStage, to develop and maintain data integration and data quality processes.
- Generated ad-hoc reports and dashboards using SQL and business intelligence tools, presenting data insights to key stakeholders for data-driven decision-making.
- Handled JSON data formats, extracting, parsing, and transforming JSON data to support data integration and interoperability between various systems.
- Designed, implemented, and maintained complex stored procedures in SQL Server and Oracle databases
- Utilized SQL and Python to build data models and data transformations, supporting data integration and reporting requirements.
- Utilized Scala and Spark to develop and optimize data mining processes, extracting valuable insights from vast datasets, and uncovering trends and patterns

- Implemented version control and change management processes for SQL scripts and database objects using tools like Git and SVN, ensuring proper tracking and traceability of changes made during the project lifecycle.
- Conducted performance tuning and optimization of Tableau dashboards and visualizations, analyzing query performance, implementing data extracts and aggregation techniques, and leveraging Tableau's performance monitoring tools to ensure optimal performance.
- Implemented SQL queries, scripts, and data ingestion processes to extract data from various sources, including relational databases, flat files, and APIs, and load it into HDFS for processing and analysis.
- Developed and optimized SQL queries and stored procedures to support SSRS and SSAS reporting, ensuring efficient data retrieval and calculation performance.
- Implemented parameterization and interactivity features within SSRS reports, allowing users to dynamically explore and filter data for personalized insights.
- Implemented data security measures within SSIS, SSRS, and SSAS, including user access controls, row-level security, and encryption, to protect sensitive data and ensure compliance with data privacy regulations.
- Created and customized Splunk dashboards, reports, and visualizations using SQL queries, charts, graphs, and tables, providing actionable intelligence and real-time monitoring of key performance indicators (KPIs) and system metrics.
- Developed and implemented normalization strategies to decompose database tables into smaller, more manageable entities, ensuring data integrity, eliminating data redundancy, and minimizing update anomalies.
- Implemented complex SQL queries, aggregations, and calculations within OLAP cubes for data summarization, slicing, and dicing.
- Utilized SQL Server Management Studio (SSMS) to design, deploy, and manage OLAP cubes.
- Developed and optimized complex SQL queries and data extraction scripts to retrieve and transform data from various sources for integration into Cognos.
- Developed and maintained documentation for Crystal Reports, including report specifications, data source connections, and SQL queries.
- Conducted performance tuning and optimization of Crystal Reports, analyzing data retrieval and processing times, optimizing SQL queries, and implementing caching strategies.
- Utilized MS Visio to design and document database schemas, including tables, relationships, primary and foreign keys, and constraints.
- Developed error handling and exception management mechanisms within Python scripts, improving data consistency and ensuring graceful handling of data-related issues.
- Conducted performance tuning and optimization of Python scripts, analyzing code execution, implementing caching mechanisms, and employing parallel processing techniques to enhance data processing efficiency.
- Conducted impact analysis of proposed changes to data models using Erwin, assessing the impact on downstream systems, applications, and reporting, to make informed decisions during project planning and migrations.
- Implemented SharePoint (MOSS) content types and metadata to organize and categorize project-related documents and data.
- Developed and maintained MDX scripts to query and analyze multidimensional data in OLAP cubes.
- Utilized Control M for job scheduling and automation, ensuring timely execution of ETL processes and datarelated tasks.
- Developed and maintained documentation for SQL queries, ETL workflows, and data mining processes, ensuring knowledge transfer and compliance with best practices.
- Generated ad-hoc reports using SQL and QlikView BI tool, presenting data insights to key stakeholders for data-driven decision-making and strategic planning.
- Utilized Python and Shell scripting to automate data processing tasks, improving data quality and efficiency of ETL workflows.
- Designed and implemented data transformation services (DTS) to facilitate seamless data flow and integration between heterogeneous systems.
- Leveraged JSON data format for data exchange and storage, developing JSON-based APIs and data schemas, enabling interoperability and data interchange between systems.
- Conducted performance tuning and optimization of SQL queries and stored procedures, analyzing query execution plans, indexing strategies, and SQL Server query hints to enhance system performance and reduce query execution time.
- Proficient in NoSQL data modeling, designing schema-less data structures and document-oriented databases, ensuring efficient data storage and retrieval for unstructured and semi-structured data.
- Utilized Informatica PowerCenter to develop and maintain data integration workflows, facilitating efficient data extraction, transformation, and loading (ETL) processes between diverse data sources and data destinations.
- Utilized JSON data formats to handle and process semi-structured data, integrating JSON data seamlessly with relational databases, enabling flexible and dynamic data handling.

• Designed and implemented NoSQL data models, defining collections, documents, and key-value pairs to efficiently store and retrieve data in NoSQL databases, ensuring optimal data organization.

Wells Fargo - San Francisco, CA 09/2014 - 08/2016 SQL Developer

- Developed and maintained data integration processes and data pipelines using SQL and ETL tools, ensuring smooth and automated data flows between systems and data sources.
- Implemented data security measures, such as data encryption, role-based access controls, and data masking, using SQL techniques and database security features, to protect sensitive data during migrations and integrations.
- Conducted performance benchmarking and optimization exercises, analyzing SQL execution plans and query performance metrics, and making recommendations to improve SQL code efficiency and database performance.
- Implemented data security measures within Tableau, including user access controls, row-level security, and data encryption, to protect sensitive data and ensure compliance with data privacy regulations.
- Conducted data profiling and analysis on HDFS data, identifying data quality issues, anomalies, and patterns using SQL queries, and implementing data cleansing and normalization techniques to improve data integrity.
- Conducted performance tuning and optimization of SSIS packages, SSRS reports, and SSAS cubes, analyzing query execution plans, optimizing SQL code, and fine-tuning ETL processes to enhance overall system performance.
- Implemented data security measures within SSIS, SSRS, and SSAS, including user access controls, row-level security, and encryption, to protect sensitive data and ensure compliance with data privacy regulations.
- Designed and implemented database tables, ensuring adherence to normalization principles (1NF, 2NF, 3NF) to eliminate data duplication and maintain data consistency.
- Utilized SQL data definition language (DDL) statements, such as CREATE TABLE and ALTER TABLE, to implement normalization principles and enforce referential integrity constraints.
- Designed and developed OLAP cubes using SSAS, utilizing SQL skills to define multidimensional or tabular models that enable fast and interactive data analysis.
- Collaborated closely with stakeholders to understand their data analysis and reporting requirements, translating them into effective OLAP cube designs that provide actionable insights and support decisionmaking processes.
- Collaborated with report consumers to validate and verify Crystal Reports, ensuring accuracy and alignment with business requirements.
- Developed and maintained documentation for Crystal Reports, including report specifications, data source connections, and SQL queries.
- Developed and maintained detailed entity-relationship (ER) diagrams, data flow diagrams (DFDs), and system architecture diagrams using MS Visio.
- Developed and maintained documentation for data models using Erwin, including data dictionaries, metadata, and data flow diagrams.
- Created and maintained project-specific SharePoint sites and portals, providing a centralized location for project documentation, data, and communication.
- Collaborated with cross-functional teams, utilizing Team Foundation Server (TFS) and Git for version control and collaboration.
- Developed and maintained complex MDX scripts for multidimensional data analysis, enabling powerful querying and reporting capabilities in OLAP cubes.
- Created ad-hoc reports using SQL queries and QlikView BI tool, presenting data insights to key stakeholders for data-driven decision-making.
- Leveraged Data Transformation Services (DTS) in SQL Server to automate ETL processes, ensuring seamless data integration and synchronization between various data sources and target databases.
- Utilized .NET & C# to build custom data-driven applications, integrating SQL databases with web-based solutions and providing user-friendly interfaces for data access and reporting.
- Developed and optimized complex SQL queries, stored procedures, and data extraction scripts to retrieve and transform data from various sources for integration into NoSQL databases, facilitating seamless data migration and consolidation.
- Utilized NoSQL databases to store and manage large volumes of unstructured and semi-structured data, providing flexible and schema-less data storage solutions for diverse data types and formats.
- Developed error handling and exception management mechanisms within stored procedures.
- Utilized stored procedures to optimize data retrieval and manipulation for reporting purposes, enabling efficient generation of ad-hoc reports and data-driven dashboards.

- Developed and optimized complex C# applications to interact with SQL databases, implementing data retrieval, manipulation, and analysis functionalities for critical business applications.
- Utilized C# and ADO.NET to build custom data-driven applications, enabling seamless data access, reporting, and analysis for end-users and stakeholders.

Education: Bachelors in Computer Science, University of Virginia, Class of 2014

Mate Yilma

• Queens, NY, US

Contact Information

- trd-1fj-2dg@mail.dice.com
- 3473538255

Work History

Total Work Experience: 10 years

- Lead Data Engineer Bank Of America Nov 01, 2020
- Senior Data Engineer/Lead Verizon Apr 01, 2018
- Data Engineer/SQL Developer California Department Of Public Health Oct 01, 2016
- SQL Developer Wells Fargo Sep 01, 2014

Skills

- business intelligence 8 years
- data integration 8 years
- data integrity 8 years
- database 8 years
- etl 8 years
- integration 8 years
- migration 8 years
- normalization 8 years
- nosql 8 years

- optimization 8 years
- performance tuning 8 years
- qa 8 years
- reporting 8 years
- scripting 8 years
- sql 8 years
- olap 7 years
- data modeling 6 years
- erwin 6 years
- meta-data management 6 years
- powerbi 6 years
- azure 5 years
- tableau 5 years
- aws 4 years
- snowflake 4 years

Work Preferences

- Desired Work Settings: Remote or On-Site or Hybrid
- Likely to Switch: False
- Willing to Relocate: True
- Travel Preference: 75%
- Work Authorization:
 - o US
- Work Documents:
 - o US Citizen
- Security Clearance: False
- Third Party: False
- Employment Type:
 - o Full-time
 - o Part-time
 - Contract Corp-to-Corp
 - o Contract Independent
 - o Contract W2
 - Contract to Hire Corp-to-Corp
 - o Contract to Hire Independent
 - o Contract to Hire W2

Profile Sources

Dice: https://www.dice.com/employer/talent/profile/2ab97fc8134a60d0800527bab76b7701	