

Personal Profile

Data Software Engineer with strong experience in Big Data solutions, SQL/NoSQL databases, and building ETL/ELT pipelines using cloud services. Proficient in AWS cloud technologies, dbt for analytics engineering, Docker for containerization, and Apache Airflow for workflow orchestration. Skilled in Python, Java, SQL, and PowerShell, with a solid foundation in Computer Science concepts and software engineering best practices. Experienced in requirement gathering and SDLC, with a focus on scalable, efficient data solutions.

As a Green Card holder. I do not need any sponsorship in the future.

Core Skills

- Python/SQL
- PySpark/Snowflake/
- DBT/Docker/Airflow
- Azure/AWS
- DataLake/DeltaLake/ETL/ELT
- DBT/Docker/Airflow
- Data Visualizations

Career Summary

2021 Nov – 2024 Oct

Data Engineer – Capital Group

Role outline

Serve in the Investment Group. Implement ETL logic to meet business requirements from stakeholders. Technologies include but are not limited to AWS S3, Lambda, Glue, Athena, Postgres RDS and DBT.

Achievements

- Generated Glue jobs by analyzing data from different sources. Such as files in S3, CloudWatch/CloudTrail logs and other third-party APIs.
- Utilized Glue jobs and Lambda functions to handle data processing logics with different kinds of triggers such as SQS and Glue catalog events.
- Automated ETL process with Glue jobs by using AWS CDK which extract CloudTrail log into parquet files and dump them into Athena DB.
- Initiated, Developed and managed an ETL pipeline. Converted IRR, MoM, FundReturnRatio and other Excel functions in Pyspark. More than 20 Lambda / Glue jobs were included for data staging, logic processing, and datasets validation. All permissions and resources access are defined in AWS CDK.
- Persistent the ETL results into Postgre DB via AWS Glue jobs. Partitioned and indexed monthly coming datasets with pre-defined store procedures to improve performance.
- Managed and automated the Glue jobs in DBT. Including the data validations, SCD checks and documentation generation. Modularized and materialized the middle steps into tables. Other features such as seeds for ad-hoc analysis. Macro for user defined functions.
- Rendered Postgre datasets into Tableau visuals and published into different workspaces. Implemented Role-Level-Security to limit different data access requests.
- Read and backfill data from Dremio data source in the AWS Glue job with rotated PAT(private access token) enabled. With the optimized the indexes, the API provide the best performance when compare with APIs from by other teams.
- Modified the Sproc and tuned the data archive sproc by modifying indexes and job schedules. Made the front-end query returned in milliseconds rather than mins than before in Postgres DB.

2019 May – 2021 May

Software Data Engineer – Microsoft

Role outline

Work as a software data engineer under Office IP Antispam team. Response for maintaining and upgrading the ETL/ELT data pipeline, Lens/ADF jobs, and stream dataset in Cosmos and ADLS with Scope script and Kusto queries. As

well as generate Ad-hoc visuals with various tools based on different requirements.

Achievements

- Migrated Cosmos uploader/downloader with Scope SDK, T-SQL, and PowerShell from old JScript logic. Equipped with alert and pipeline monitor to enhance durability.
- Troubleshoot issues, errors in ETL services. Generated probes, monitors, alerts for different components and services with C#, Scope language, and PowerShell.
- Built and maintained a big data platform (ADF/ADLS) in Cosmos Scope script as per industry standards and best practices.
- Performed DB admin activities on the multiple DB Servers. Optimized data pipeline by rewriting new logic and SQL schema, gained both performance and service durability. Reduced up to 20% storage cost with archiving and re-indexing stale data as well as increase the query performance.
- Optimized slow running scope queries by using scope hints. Generate TB level data daily bases.
- Implemented Ad-hoc data status monitors and dashboards with PowerBI and Lens platform. Delivered reports to end users weekly and monthly.

2018 Nov – 2019 May

Data Engineer – Conagra Brands Inc.

Role outline

Serve under Cost to Serve team. Generated annual reports from the sketch. Implemented data modeling, user validation and dynamic filters based on stakeholders' requirements.

Achievements

- Designed and implemented the whole ETL (Extract, Transform, Load) data process with various transactions within SSIS packages to implement business rules and logic.
- Generated PowerBI reports, dashboards with Star schema. Implemented complex requirements in PowerBI by using DAX and M language.
- Reduced refresh waiting time by using PowerBI API in PowerShell scripts, deployed, automated the scripts on CtrlM Servers.
- Created customized visuals with R libraries. Embedded the PowerBI visuals within Sales-force homepage with PowerBI-JavaScript libraries.
- Accumulated agile Product Management experience - a full gamut of creating stories, generated PowerBI reports and dashboards, working through user acceptance testing (UAT).
- Hold scrum meetings as a scrum master. Wrote various documents such as data mapping document, work detail descriptions for KT sessions, and future development.

2013 Sep – 2015 May

SQL BI developer – KSDHC Inc.

Role outline

As a SQL BI engineer, utilize the MS BI stack (SSIS/SSRS/SSAS) to perform ETL processes with a big scale of data. Includes data cleansing, data labeling and aggregation, etc. Generate different visuals with Microsoft SSRS.

Achievements

- Created logical and physical data models based on the requirements utilizing Erwin and Navicat.
- Improved the performance of the official website by optimizing slowly running queries, utilized indexes, partitions, and deadlock monitor via SQL Profiler and DTA.
- Normalized database tables to avoid redundancy and DML anomalies.
- Developed complex SQL scripts and procedures for data profiling and auditing purposes. Combined C# logic in SSIS packages.
- Utilized Microsoft SSIS/SSAS/SSRS to handle ETL logic, create ad-hoc reports with both tabular model and cube model.
- Developed various Dashboards and Stories using advanced Tableau features including calculated fields, parameters, table calculations, row-level security, R integration, and dashboard actions while dealing with Cloud data in SSRS.

Education

2015 - 2018	Oregon State University Master degree in computer science
2010 - 2014	Liaoning Normal University Bachelor degree in computer science and technology

Certificates

Microsoft:	Azure Database Administrator Associate Since 2019
Microsoft:	Data Analyst Associate Since 2019