**Bryan Eaton**

**Data Engineer**

**Professional Summary**

* Over 10+ years of experience in Data Engineering, Data Pipeline Design, Development, and Implementation as a Data Engineer/SAP Developer.
* Seasoned Python developer with over 19 years of experience crafting robust CLI tools, web services, and open-source contributions.
* Proven expertise in designing and implementing data marts and data warehouses by extracting, transforming, and loading data from OLTP systems. Skilled in creating scalable, high-performance data architectures to support business intelligence and analytics
* Proficient in developing efficient pipelines using RabbitMQ and Apache Kafka, ensuring seamless data flow and real-time processing for data-driven solutions
* Leveraged PySpark against Spark clusters to accelerate ETL transformations, efficiently extracting and processing large-scale OLTP data into optimized data marts and warehouses for Shell Energy.
* Engineered robust Airflow pipelines to seamlessly transfer data from vendor SQL Server databases to Amazon S3 in Parquet format, enabling optimized data access and consumption through a Dremio Lakehouse.
* Strong experience and knowledge in Data Visualization with SAP Business Objects and creating interactive dashboards using SAP’s Design Studio.
* Proficient in optimizing Airflow DAGs, integrating with cloud-native tools, and ensuring high availability and performance in Kubernetes clusters
* Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL and U-SQL Azure Data Lake Analytics. Data Ingestion to one or more Azure Services – Azure Data Lake, Azure Storage, Azure SQL, Azure DW and processing the data in Azure

**Technical skills**

|  |  |
| --- | --- |
| ETL Tools | Airflow, Spark, Apache Kafka, RabbitMQ, |
| NoSQL Databases | MongoDB, Cassandra, Amazon DynamoDB, HBase |
| Data Warehouse | Google Cloud Storage, Snowflake, Teradata, Azure Synapse |
| Databases | PostgreSQL, Microsoft SQL Server,SSIS, SSRS, IBM DB2, PostgreSQL, Amazon RDS, ElasticSearch, MongoDB |
| Web Development | HTML, XML, JSON, CSS, JQUERY, JavaScript |
| Monitoring Tools | Splunk, Chef, Nagios, ELK |
| SourceCode Management | J Frog Artifactory, GitHub, GitLab, |
| Containerization | Docker & Docker Hub, Kubernetes, OpenShift, Podman |
| Programming  and Scripting | Python, Java, C#, GoLang, Lua, Kotlin, PostgreSQL, SQLServer, MySQL, Shell Scripting |
| Build & Development Tools | GitHub Actions, GitLab CI/CD, Maven, Gradle |
| Methodologies | Agile/Scrum, Waterfall |

**PROFESSIONAL EXPERIENCE**

**Data Engineer**

**RegScale, Knoxville, TN July 2022 to August 2025**

**Responsibilities:**

* A principal developer for the RegScale CLI application (https://pypi.org/project/regscale-cli/)
* Developed official integrations for RegScale with Wiz and Tenable, streamlining data connectivity and provided automated compliance updates to the RegScale platform.
* Built integrations for many other cybersecurity tools like Tenable Nessus, BurpSuite, Nexpose, and Prisma
* Built robust parsers handling diverse data schemas and formats from various compliance tools and asset management systems
* Implemented comprehensive validation logic to ensure data integrity and compliance with RegScale's data model requirements
* Designed fault-tolerant processing with detailed logging and graceful error recovery for production environments
* Designed and deployed a disk caching solution to optimize performance when processing massive volumes of API data, reducing API call overhead and improving local system responsiveness by 90%.
* Engineered robust Airflow DAGs to automate and schedule command-line interface (CLI) jobs within the RegScale compliance automation platform.
* Implemented efficient pickle and JSONL serialization strategies with compression (gzip/lz4) to minimize disk I/O and storage footprint.
* Built comprehensive reporting solutions from the ground up using Python and the Rich library to generate visually compelling, interactive terminal-based reports for data analysis and operational monitoring.
* Created extensible Python mixin classes that could be composed with different data models, promoting code reusability and maintainability
* Developed C# services to export RegScale System Security Plans (SSPs) to OSCAL XML format, ensuring prevalidation and full compliance with FedRAMP's OSCAL validation rules .
* Designed and implemented containerized development environments using Docker and Docker Compose to streamline the design, testing, and deployment of RegScale platform integrations and job schedulers, reducing development cycle time and ensuring consistent testing across environments.
* Performed a full data migration from ZenGRC to RegScale using native APIs and custom mappers built in Python.
* Built CI/CD pipelines for the RegScale CLI to push to PyPi test and prod when dev and main branches were updated .
* Developed C# models, services and controllers to support customer needs when ingesting Cybersecurity data.
* Work closely with sales engineers, platform engineers, customer contacts, and the CISO to translate business and technical strategies into data-driven solutions for the clients of RegScale.
* Created and maintained technical documentation for all the workflows for RegScale (https://regscale.readme.io/docs/overview)

**Environment**: Python, C#, Linux, Angular, SQL Server, Azure SaaS deployments, GitLab, Airflow, Docker

**Data Engineer**

**Shell Energy, Houston, TX July 2021 to July 2022**

**Responsibilities:**

* Built the Dremio lakehouse primarily against our retail and revenue management systems.
* Load and Transform huge datasets of structured data from OLTP (Online Transaction Processing) systems to a Dremio lakehouse using PySpark.
* Configured Apache Airflow as the orchestration engine to automate and schedule ETL pipeline jobs that extract data from our OLTP systems and load it into Dremio for analytical workloads.
* Used custom PySpark routines to load parquet files into Dremio for more complicated pipelines.
* Migrated a legacy MS Access ETL VBA scripts to Python enabling much easier code maintenance.
* Configured Apache Airflow as the orchestration engine to automate and schedule ETL pipeline jobs that extract data from our OLTP systems and load it into Dremio for analytical workloads.
* Designed and developed Power BI reports to present data clearly and concisely, using charts, graphs, and tables. Created report models and deployed them to the report server.
* Utilized legacy SSIS packages for SQL Server jobs.
* Developed data-driven subscriptions to automate report generation and delivery.
* Involved in Analyzing system failures, identifying root causes, and recommended course of actions
* Created and maintained technical documentation for all the workflows.
* Created database access layer using JDBC and SQL stored procedures.
* Wrote custom database functions to use in our ETL stack.
* Collaborated with the infrastructure, network, database, application, and BI teams to ensure data quality and availability.

**Environment:** Python, Airflow, Dremio, SSIS, SSRS, PowerBI, SQL Server, PySpark.

**Data Engineer**

**Oak Ridge National Laboratory, Oak Ridge, TN Oct 2018 to July 2021**

**Responsibilities:**

* Designed and developed RESTful APIs using Spring Boot for application backend services.
* Managed database interactions using Hibernate and wrote complex PostGIS SQL queries (e.g., ST\_Intersects, ST\_Distance) for the PlanetSense project.
* Integrated secure RabbitMQ messaging with Java Spring within SIPRNet for reliable communication.
* Configured Hazelcast as an in-memory distributed cache to store image data, utilizing eviction policies (LRU, LFU) for scalability.
* Supported the WorldStamp project by using R for data fetching and transformation, orchestrated via Airflow DAGs.
* Wrote unit and integration tests using JUnit and monitored application metrics with Spring Actuator.

**Environment**: Java, PosgreSQL, Elasticsearch, Python, PostGIS, Linux