

Leo Qin - Principal Data Engineer

About Me

I am a product-focused data engineer with strong proficiency in systems architecture, infrastructure, and R&D. I have substantial subject matter experience in identity resolution, payments, lending, and fraud detection. Prior experience/background in product management, data science, and applied machine learning.

History

- Staff Data Engineer at Wisetack (Aug 2022 - Present)
 - Design, implement, and maintain a platform for ingesting, transforming, and delivering data to various stakeholders, including product, engineering, analytics, and other business users
 - Design, implement, and maintain a platform for designing, training, and deploying machine learning models on AWS Sagemaker
 - Onboard over 30 users to the platform, including ~10 non-engineering (credit/analytics team) developers requiring development environments, local setup, and mentorship
 - Architecture: Airbyte for data ingestion, dbt for data transformation, Dagster Cloud for orchestration, AWS Athena for compute
- Data Science Manager at Spokeo (Dec 2018 - Aug 2022)
 - Tech lead, product owner, and software architect. Primary focus is on complete rebuild of multi-source entity resolution system.
 - Secondary emphasis on engineering and architecture for multi-TB-scale production ETL system and any greenfield R&D efforts (“is this possible”-type questions). People manager and scrum team product owner.
 - Architecture: shell scripting for batch data ingestion, PySpark and Pentaho for transformation, Airbyte for orchestration, AWS EMR for compute
- Predictive Analytics Manager at Expedia (Nov 2017 - Dec 2018)
 - Define, design, develop, and deliver data science solutions primarily supporting Payment Operations and Strategy/Innovation teams. Collaborate with BI tech team to define future architecture for data science modeling and analytics environment. Serve as scrum master for own team and data science product manager for others.
 - Developed novel non-parametric time series forecasting algorithm and accompanying library including features for cross validation, visualization, and speculative regression. Performance comparable with FBProphet with 6-20x faster convergence, richer functionality.
 - Oversee and facilitate annual planning process from ideation through scoping and dependency resolution
- TransUnion (Sep 2015 - Nov 2017)
 - Develop data analytics capabilities within Fraud and Identity Solutions team. Primary emphasis on development of fraud detection products (scores/models) using advanced analytics and rapid/iterative prototyping. Secondary emphasis on data development strategy and executing go-to-market product launch initiatives with internal and external stakeholders.
 - Lead product R&D analytics for new online fraud detection capabilities from business case development through to implementation and validation
 - Develop and maintain ETL procedures for application-specific data extracts within IBM Netezza and Hadoop/Hive-based data warehousing environments
 - Manage team of data scientists in prioritizing and producing analytics efforts for internal and external customers, interfacing with product managers and pre-sales consulting staff.
- PreCash (Jan 2013 - Sep 2015)
 - Primary business data analyst for senior management and c-suite for building business cases and determining product development strategy, concurrently serving as industry subject matter expert, business systems analyst, and data integration specialist.
 - Provides ad-hoc and structured analysis for senior management and c-suite users; operationalizing generalized business questions into specific findings and technical recommendations to influence product strategy and direction

- Customer education during integration phase through test design, API request/response analysis, and implementation of reporting solutions out of Oracle data warehouse environment to track future outcomes, metrics, and KPIs
- Serve as key contact within the escalations path for technical and business issues, ongoing subject matter expertise in business processes and industry best practices regarding such topics as cybersecurity, user experience, and data aggregation and reporting

Tools and Languages

- Languages: Python, SQL, bash/shell script
- Data Transformation: dbt, SparkSQL/pyspark
- Data Orchestration: Dagster, Airflow
- Data Ingestion: AWS Glue, Airbyte
- Data Compute: AWS EMR, AWS Athena (Trino/Presto)
- Data Storage: AWS S3, AWS DynamoDB, HDFS
- Data Permissions: AWS Lakeformation, AWS IAM, AWS SSO
- Infrastructure: Terraform, Pulumi, Docker (Compose), Kubernetes