

Adrij Shikhar

Personal Info

Email
adrijshikhar85@gmail.com

Phone
+918218058928

Github
<https://github.com/adrijshikhar>

LinkedIn
<https://www.linkedin.com/in/adrij-shikhar/>

Website
<https://adrijshikhar.github.io/>

Education

July 2018 - May 2022

Indian Institute of Technology, Roorkee

Chemical B.Tech.

Work History

Senior Software Development Engineer

Jun 2022 - Current

Hevo Data, Bangalore

Data Pipeline & CDC Framework Development

- *CDC Framework Ownership:* Led the end-to-end implementation of a real-time Change Data Capture (CDC) framework using the **Debezium Engine**, significantly enhancing data synchronization capabilities.
- *Performance Optimization:* Achieved a **5x** increase in data ingestion speed for the MySQL connector through targeted optimizations, including advancements in binlog processing.
- *Schema Catalog Service:* Designed and built a generic schema service from scratch to manage, validate, and version data schemas within ETL pipelines. This service supported source-specific features like PostgreSQL's unchanged toast datum and ensured data quality through compatibility checks.
- *Connector Ecosystem Expansion:* Spearheaded the development and launch of the SurveyMonkey connector, broadening the platform's integration capabilities.

System Architecture & Optimization

- *ETL Platform Scalability:* Engineered core database interaction improvements by implementing a metadata system that scaled the ETL platform by **15x**.
- *Microservices & Orchestration:* Integrated a microservices architecture on AWS Fargate and implemented a Temporal-based task execution system to improve reliability and streamline complex, hierarchical DAG processes.
- *Process Optimization:* Optimized the source object generation flow, resulting in a **10x improvement** in pipeline performance and enabling support for significantly larger datasets.
- *System Reliability:* Designed and implemented a standardized error-handling framework for both HTTP and gRPC services, creating consistent error classification and improving user experience.
- *Advanced Scheduling:* Developed an advanced scheduler with support for complex timing configurations, enhancing the flexibility of data pipeline execution.

Performance Engineering & Scalability

- *Large-Scale Object Support:* Architected and delivered system-wide enhancements to successfully support up to 25,000 objects in a single pipeline, a key scalability milestone.
- *Load Testing & JVM Tuning:* Conducted extensive performance and scale testing to identify and resolve system bottlenecks. Subtly improved system efficiency through deep-dive analysis, including JVM analysis and performance tuning.
- *Service Scaling:* Proactively scaled core services, including the Hermes API and Session Logs, to handle increased load and ensure high availability.
- *Data Footprint Reduction:* Optimized the metadata storage by leading the Catalog Document Size Reduction initiative, improving data retrieval times and reducing storage costs.

Data Governance & Security

- *Real-time PII Redaction:* Implemented a data governance tool for PII redaction in near real-time using AWS S3 and AWS Comprehend, ensuring data privacy and compliance.
- *API Security:* Enhanced the REST API connector by implementing OAuth 2.0 authorization, strengthening security and improving user access management.

Customer Impact & Collaboration

- *Critical Issue Resolution:* Successfully managed and resolved several complex, high-priority issues for key enterprise clients, significantly improving customer satisfaction and retention.
- *Knowledge Sharing:* Regularly contributed to improving internal documentation and conducted knowledge transfer sessions, fostering team collaboration and upskilling.

References

Tariq Iqbal | Senior Engineering Manager - Atlassian

contact info: [linkedin.com/in/ta-riq](https://www.linkedin.com/in/ta-riq)

Shubham Goyal | CTO - Powerplay

contact info: [linkedin.com/in/shubham7stark](https://www.linkedin.com/in/shubham7stark)