Junaid Rahim

+91 9673184875 • junaidrahim5a@gmail.com • junaid.foo

Professional Experience

Atlan Pte. Ltd. **Remote**

Software Engineer II *July 2023 - Present*

- Architected and led the implementation of the central publishing service with parquet-based WAL architecture processing **600M+ entity changes daily**. Achieved **50% runtime improvement** through resumable workflows, time-travel and rollback capabilities.
- Core contributor to the Atlan App SDK powering **30+ partners** and all internal teams. Reduced app development-to-deployment cycle to **1 week**, with partners launching **10+ marketplace apps**.
- Designed universal MCP integration for App SDK making all Temporal-based apps MCP servers with single config change. Enabled **20+ apps** to expose activities as MCP tools without code modifications.
- Accelerated entity diff-ing by **5x** for **100GB+ scale JSON datasets** using Merkle tree comparisons and hash bucket partitioning removing the single-pod bottleneck enabling unlimited parallel processing.
- Implemented ring-based release controller with GitHub Actions enabling cohort deployments reducing production regressions by **90%** across **4000+ releases** while enabling confident, gradual rollouts.
- Built the internal agent platform using LangGraph and LangMem improving SDLC process with automated release control, code quality and quality assurance agents.

Software Engineer I *Jan 2022 - July 2023*

- Led the optimization effort for the search library used in pipelines achieving **500x improvement in P99 query latency**. Implemented domain-specific inverted indexes and intelligent disk spilling with SQLite.
- Improved internal spill-to-disk library performance by **75%** across multiple connectors through SQLite parameter tuning, LRU caching, and optimized indexes.
- Designed metadata schemas for **12+ BI connectors** outlining attributes, relationships, and lineage patterns which became standard for all **300+ enterprise customers**.
- Implemented end-to-end column-level lineage from warehouse tables to dashboard widgets for Tableau, Looker, PowerBI, dbt, and Sigma. Achieved widget-level granularity tracking data flow across **50M+ dashboard elements**.
- Designed, implemented and advocated the lineage graph framework for BI connectors leveraging tree pruning, lazy evaluation, and spill-to-disk patterns. Improved runtime by **90%** and reduced missing lineage support tickets by **50%**.
- Implemented the schema drift detection system to safeguard against flaky upstream metadata APIs. Automated circuit breakers prevented **100+ potential data loss incidents** from harmful delete operations.
- Migrated analytical pipelines from JSONL to Parquet using DuckDB. Reduced processing time by **80%** for pipelines handling **100M+ records**.
- Designed and implemented the Fivetran connector supporting **200+ upstream data source types** with automated schema detection, eventually improving orphaned asset detection.
- Built multiple dbt connector versions generating lineage for both model-level and warehouse-level assets powering an improved experience for analytics engineers using Atlan.
- Built PowerBI M query lineage extractor leveraging open-source parsers to traverse PowerQuery ASTs. Increased lineage coverage by **15%** for **32M+ assets**, surfacing complex data lineage for **50+ customers**.
- Reduced connector development time by **70%** (from **30 to 10 days**) through internal tooling, reusable libraries, and GitHub Actions automation. This acceleration enabled the team to ship **20+ additional connectors** within the first year.
- Resolved **100s of initial connector support tickets** post-launch. Established first QA and support functions with internal tooling, documentation, and observability frameworks.
- Implemented the data quality offering using Soda with SQL pushdown to warehouses, this enabled automated quality checks across data assets without data movement.
- Created company's internal documentation platform (Kryptonite) inspired by g3doc for docs-as-code use-cases, serving **8k+ weekly page hits** across all engineering repositories with automated API and test coverage documentation.

Software Engineering Intern *July 2021 - Dec 2022*

- Designed and shipped **5 foundational metadata connectors** (Snowflake, Tableau, PowerBI, PostgreSQL and Looker) using python and argo workflows. Presently deployed across **98% of enterprise customer base** and process **800M+ metadata assets daily** across **3.7k customer data sources**.
- Implemented end-to-end lineage generation capabilities in BI connectors achieving **70%+ lineage coverage** across metadata assets providing critical data flow visibility for governance and impact analysis.
- Established core marketplace platform components including composable UI configs, central registry, and standardized workflow templates. These foundations now support the entire connector marketplace used by **30+ engineers**.
- Designed and implemented Atlan Playbooks, a drag-and-drop feature to bulk-edit metadata. Replaced manual CSM scripts with **4000+ playbooks**, eliminating hundreds of hours of repetitive work.

International Conference on Learning Representations **Remote**
Guide: Prof. Alexander Rush, Cornell NLP, Cornell University *March 2020 - April 2020*

- Worked with Prof. Rush's team to build the open source virtual conference portal for ICLR 2020. Implemented backend to manage all the submitted papers via the OpenReview API. (Python, Flask) [website] [github]
- The portal gathered **1M+ page views**, **100k+ video watches** and **80k+ chat messages**

Open-source Projects

Atlan App SDK
Python SDK to build apps on the Atlan platform *Jan 2025*

- The SDK offers a complete Platform-as-a-Service (PaaS) toolkit that enables developers to create integrations, data processing workflows, and custom applications that seamlessly extend the Atlan experience.

argopm
CLI to ship Argo Workflows in packages *Nov 2022*

- Package manager for Argo Workflows. It enables developers to distribute and consume argo workflow templates as reusable modules.

HalideOS
An Experimental OS written entirely from scratch. *Sept 2020*

- A bare bones x86 operating system written in C++. Implemented the framebuffer, shell and a mini standard library for console programs.

Publications and Talks

1. **Scalable Data Pipelines with DuckDB at DuckCon 2024**
   Presented in-person at DuckCon 2024 in Seattle.
2. **Shipping Argo Workflows in Packages at KubeCon 2023**
   Presented in-person at ArgoCon 2023 hosted w/ KubeCon Europe 2023 in Amsterdam.
3. **Predicting Semen Motility using three-dimensional Convolutional Neural Networks** Priyansi, Biswaroop Bhattacharjee, and Junaid Rahim. In *17th International Conference On Distributed Computing And Internet Technology.*

Education

Bachelor of Technology (B. Tech) 9.04/10.0
Computer Science and Engineering, Kalinga Institute of Industrial Technology *2019-2023*

Technical Strengths

- **Languages:** Python (primary), Go, Rust, Typescript and SQL.
- **LLMs & Agents:** LangChain & LangGraph, MCP and PyTorch.
- **Data & Workflow Orchestration:** Argo Workflows, Temporal, Daft, Apache Spark, dbt and OpenLineage.
- **Databases & Storage:** SQLite, DuckDB, PostgreSQL, Apache Iceberg, Snowflake and BigQuery.
- **Infrastructure:** Kubernetes, vCluster, Dapr, Github Actions, AWS (EC2, S3, EKS) and Victoria Metrics.

Last Updated on November 3, 2025. References available upon request.