

Maggie Cunningham

📞 (617) 555-7314 ✉️ maggie.cunningham@example.com

🌐 [linkedin.com/example/maggiecunningham](https://www.linkedin.com/example/maggiecunningham) 📍 Boston, MA 02118



SUMMARY

Software engineer with six years building and operating production data infrastructure, distributed processing systems, and query engine capabilities. Work spans query representation, planning, optimization, execution, storage, and table management, with close attention to correctness, compatibility, and predictable performance. Proficient in Java, Rust, Scala, and Python, with hands on delivery across Apache Spark, **Apache DataFusion**, **Apache Iceberg**, Apache Arrow, and Apache Parquet. Designs optimizer rules, execution operators, **incremental processing** paths, metadata integrations, and concurrency controls. Uses profiling, **representative workloads**, benchmarks, and regression tests to diagnose performance and correctness issues. Partners with product, application, security, build, and data management teams to turn operational needs into reusable infrastructure that supports reliable analytical workloads.

EXPERIENCE

Senior Software Engineer, Query Infrastructure

May 2024 - Present

Harborline Data Systems

Boston, MA

Owns query infrastructure for a production analytics platform, guiding planning, execution, **storage**, and table management improvements across interactive and scheduled workloads.

- Implemented optimizer rules in **Java** and **Rust**, improving plan stability across representative customer workloads.
- Extended DataFusion operators with parallel execution and Arrow movement, protecting throughput through benchmark coverage.
- Built Iceberg **catalog integrations** for metadata evolution, **transactions**, snapshots, and efficient table updates.
- Diagnosed planner, operator, and Parquet defects, then added regression tests that prevented recurring failures.
- Partnered with product and platform teams on reusable interfaces supporting **security**, build, and data services.

Software Engineer, Distributed Data Systems

January 2022 - April 2024

Meridian Compute Labs

Cambridge, MA

Built distributed processing capabilities for batch and analytical workloads, combining Spark development, storage optimization, incremental computation, and production **performance** analysis.

- Built Spark processing capabilities for recurring transformations and exploratory analytical workloads.
- Refined incremental computation logic, reusing valid outputs while preserving deterministic results for changed inputs.
- Profiled CPU, memory, shuffle, and I/O behavior, then tuned operators against representative workload bottlenecks.
- Developed Iceberg and Parquet reads with metadata pruning, reducing unnecessary movement across analytical tables.
- Maintained **Scala** and Java services with Gradle and GitHub workflows, strengthening **concurrency** and recovery tests.
- Evaluated upstream data system changes through reproducible **benchmarks** and compatibility findings before production adoption.

Software Engineer

October 2020 - December 2021

BluePeak Analytics

Providence, RI

Developed production data services and shared transformation infrastructure, improving distributed execution, columnar interoperability, concurrency safety, and release confidence for application teams.

- Developed Java, Scala, and **Python** services supporting distributed transformations for application teams.
- Improved Spark joins and aggregations through execution plan analysis and runtime metrics.
- Built Parquet and Arrow ingestion paths, improving **interoperability** between analytical services and consumers.
- Investigated race conditions and resource contention, then verified fixes with focused load tests.
- Created representative benchmark suites and **GitHub** regression checks for engine upgrade safety.
- Defined stable transformation semantics and service contracts, reducing workload specific logic in shared infrastructure.

Associate Software Engineer

October 2019 - September 2020

Quarry Software Group

Worcester, MA

Built backend components for a distributed analytics service, supporting transformation, scheduling, execution, incident diagnosis, testing, and safe resource access under supervision.

- Built Java and Python components for transformation scheduling and distributed execution support.
- Analyzed query plans and runtime traces, documenting reproducible production failures across service boundaries.
- Implemented **data structures** and concurrency controls for deterministic shared processing behavior.
- Added tests for serialization, partition handling, and recovery scenarios, strengthening release confidence.
- Used **Gradle**, GitHub, and code review feedback to ship incremental backend improvements.

- Supported **profiling** and workload replay during performance investigations across storage and processing paths.

PROJECTS

Incremental Query Processing Framework 📅 2024

Created reusable incremental execution semantics for changed inputs, valid prior results, deterministic output handling, and representative workload validation across distributed analytical pipelines.

Columnar Storage Interoperability Layer 📅 2022

Connected Iceberg metadata with Parquet and Arrow data paths, supporting selective reads, table updates, compatibility checks, and efficient data movement.

LEADERSHIP & AWARDS

- Received senior engineering recognition for query infrastructure improvements that protected correctness and predictable production performance.
- Led peer learning sessions on workload profiling, benchmark design, concurrency diagnosis, and regression prevention.
- Earned team recognition for clear incident analysis and collaborative delivery across product and platform groups.

EDUCATION

Bachelor of Science in Computer Science

2019

University of Massachusetts Lowell GPA: 3.7

Lowell, MA

Coursework: *Algorithms, Data Structures, Distributed Systems, Database Systems, Operating Systems*

CERTIFICATIONS

- Oracle Certified Professional, Java SE 📅 2023
- Databricks Certified Associate Developer for Apache Spark 📅 2022

TECHNICAL SKILLS

- Programming Languages:** Java, Rust, Scala, Python
- Query Engines:** Apache DataFusion, Apache Spark, Velox
- Table Formats:** Apache Iceberg, Apache Parquet, Apache Arrow
- Query Optimization:** Physical Planning, Optimizer Rules, Execution Plans
- Distributed Systems:** Parallel Execution, Data Movement, Incremental Processing
- Performance Tools:** Profiling, Workload Replay, Runtime Metrics
- Build Systems:** Gradle, Cargo, GitHub Actions
- Testing Practices:** Regression Testing, Load Testing, Compatibility Testing
- Storage Systems:** Columnar Storage, Metadata Pruning, Snapshot Reads
- Concurrency Engineering:** Data Structures, Memory Management, Failure Recovery
- Development Workflows:** GitHub, Code Review, Automated Testing

SKILLS

- | | | | |
|----------|--------------------------|--------------------------|-------------------|
| • Java | • Query Planning | • Storage Systems | • Concurrency |
| • Rust | • Query Optimization | • Incremental Processing | • Algorithms |
| • Scala | • Execution Engines | • Parallelism | • Data Structures |
| • Python | • Distributed Processing | • Memory Management | |

PROFESSIONAL AFFILIATIONS

- Member, Apache Software Foundation community discussions focused on Spark, Iceberg, Arrow, and data systems.
- Participant, Boston data engineering community, sharing practical lessons on distributed processing and query performance.

LANGUAGES

- English (Native)
- Spanish (Intermediate)

ADDITIONAL INFORMATION

Work Status : Authorized to work in United States. No sponsorship required.

REFERENCES

AVAILABLE ON REQUEST