Culley Harrelson, PhD

Email ~ LinkedIn ~ GitHub

Skill Index

Data Architecture:
- Data Lakes/Warehouses
- Data Modeling
- Data Catalogs & Data Governance
- FinOps, Cloud Tooling

Data Engineering:
- Python, SQL, Bash, Pandas, GitHub, Docker
- ETL pipelines (DAGs), DevSecOps
- JSON, NoSQL, ProtoBuf
- Databricks & BigQuery

Data Analytics:
- Looker, Tableau, Panel Dashboards
- Jupyter Notebooks, pandas, numPy
- Plotly, Vega, Seaborn, HvPlot

Leadership: • data lifecycle management • engineering management • technical program management •

Experience

February 2021 - Current

HCL America, Inc., Sunnyvale, CA – *Principal Data Architect*
Client: Google (Business Intelligence Group, Global Affairs)

Data Architecture:

- Technical lead for three compliance reporting data models. Engineering management for the teams supporting the ETL pipelines, dashboards and reports based on these models
- Played a critical role in developing a compliance data platform to serve as the single source of truth for compliance workflows at Google. Helped define a framework-agnostic data model for abstracting GRC (governance, risk, compliance) data, driving automation and collaboration across teams
- Launched a project to refactor materialized tables into logical views on the SOX pipeline, and performed code reviews for the data engineers doing the work
- Data security evangelist and primary point of contact for Access Control Lists (ACL) updates to sensitive data and dashboards, including annotating metadata entries in the corporate data catalog.
- Maintained awareness of corporate policies in addition to privacy & security regulations including GDPR, CCPA, EU DMA, EU DSA, and COPPA
- Technical lead for the data quality team on an executive-level reporting pipeline, ensuring the accuracy, completeness and timeliness of compliance data by the product areas teams
- Launched a project to refactor a critical legacy production pipeline for engineering compliance reporting. Created the SQL testing framework, and documented issues/bugs for tracking the project. Performed code reviews for the data engineers performing the work. Facilitated weekly meetings with the stakeholders to track progress and handle issues
- Stayed current with Google technical documentation for data management and compliance. Tools included: plx (similar to Bigquery, Snowflake and Databricks SQL), in-house unix tools, GoogleSQL, data materialization, Looker Data Studio, GSuite, and much more

Data Engineering:

- Author/reviewer of over 700 change lists (CLs) in the Google code base for data engineering projects.
- Orchestrated compliance reporting pipelines for Engineering Compliance, SOX, Cloud, and RegReady teams/programs to meet critical regulatory deadlines. Using Google's in-house version of GCP Dataflow / Cloud Composer (built on Apache Airflow) to construct Directed Acyclic Graphs (DAGs)

- Developed multiple ETL pipelines to move data between [Protocol Buffer](#) output (similar to [JSON](#) and [Thrift](#)) from [Cloud Spanner](#) production systems into Google's in-house version of [BigQuery](#) for analytics and reporting
- Developed a greenfield reporting pipeline for a [Risk-Control Self-Assessment](#) module in the Enterprise Risk Management System to calculate residual risk ratings for Google products after the application of known compliance controls.
- **Secure CI/CD data pipelines:** delivered [binary authorization of continuous build processes](#) (CI/CD) including service-based identity checks, containerized deployments, and scheduled, verifiable builds of ETL Pipelines for programmatic tracking of compliance requirements of the data. Built software packages on gLinux with Bash, Blaze (build automation like [Make](#) and [Bazel](#)), Rapid (similar to [AWS CodeDeploy](#)), Piper (similar to [Git](#)) and Borg (predecessor to [Kubernetes](#)).
- Implemented process discipline and engineering best practices for engineering team members, including code review workflows. Adopted internal SQL readability style guidelines to ensure alignment with SQL formatting and documentation standards
- Moved the pipelines and front-end visualizations for two work streams into the corporate version control system, and brought code reviews into the development processes
- Migrated SQL scripts between database engines, changing syntax, built-in functions and metadata structures. This was required to stop using deprecated software
- Automated SQL tests and alerts to notify engineers and stakeholders of anomalous data injected into the system. Checked ingestion data for missing values and validated account details against HR datasets

Data Analysis & Dashboards:

- Expert in exploratory data analysis, demonstrated through the design and development of compliance-related dashboards using Looker Data Studio and Plx (a Google-internal data visualization tool similar to Tableau). These dashboards tracked the status of compliance regulations, risk assessments, remediation plans, assigned teams, and compliance control relationships to systems, products, and various types of data. Dashboards were converted to text format and checked into the corporate version control system, undergirding Google's [second line of defense](#) for regulatory readiness.
- Helped develop KPIs in alignment with program management dashboards to track progress and drive operational insight in weekly status meetings for DevOps, Tooling and other teams
- Primary point of contact for analysis on multiple data pipelines. Worked with analysts and stakeholders to validate and address data anomalies
- Successfully refactored the user interface for over 130 SOX critical dashboards, including developing prototypes, getting approvals, and managing a team of front-end engineers to ensure timely delivery. Facilitated the review and acceptance of the updates by stakeholders

Leadership:

- Tech lead for the launch of new data pipelines and analytic efforts in collaboration with cross-functional team members and stakeholders
- Facilitated daily scrums for data engineers (I'm a [certified scrum product owner](#)). Flexed time to be available to teams in India and Ireland.
- Backlog grooming/prioritization and issue management
- Contributed to the development of program documents, including project communications, business requirements documents, product requirements documents, solution design proposals, proof-of-concept prototypes, newsletters, and slide decks
- Managed and contributed to end user and technical documentation. Led a team of technical writers and business analysts in revising the help center documentation for multiple applications
- Managed technical projects using in-house Google project management software and custom dashboards tailored to each team

- Provided training and mentorship to new team members on data models and internal data engineering tools. Facilitated knowledge sharing and changes to the team structure
- Trained advanced stakeholders on domain-specific data models, Google SQL functions, spreadsheet data feeds, and other data visualization tools
- Maintained up-to-date knowledge of rapidly-evolving corporate privacy, security, and data governance policies, and applied these policies in making design decisions for data engineering tasks
- Gathered requirements for reporting and dashboards, defined acceptance criteria, and proposed potential positive outcomes. Developed solutions based on consensus
- Conducted retrospectives and post-mortems to review and analyze project processes and outcomes

2001 - 2019
Tuolumne Technology Group, Sonora, CA – *President*
Role: Founder, Product Owner, Lead Engineer

Founder / inventor of FlashcardExchange (now called cram.com), a disruptive technology startup that was first to market in 2001 for collaborative online study

- Devised and executed a business strategy that scaled the online user base from 0 to 2M users
- Sold the company for ~$1 million. cram.com is still a market leader with 3.3+ million customers and 195 Million+ flashcards
- Over time, developed and refactored the web and back-end application architecture as needed:
    - *Servers*: Built & managed unix servers and software. Package management, performance tuning, email server management, log file processing. FreeBSD, Varnish, Apache, Postfix, Memcache. Over 10 years on-call for impeccable server uptime.
    - *Design*: data modeling, UX design, application and infrastructure design
    - *Data*: PostgreSQL and MySQL database engineering and administration. Logical/conceptual schema development, ETL, database performance tuning for highly-available web applications, offsite backups, etc.
    - Early adopter of AWS S3 and SES

Certifications & Badges

Databricks Lakehouse Fundamentals – March 2023
Professional Scrum Product Owner – September 2019

Education

2021 – Meridian University, Petaluma, CA – *PhD in Psychology*
1995 – UC Santa Cruz, Santa Cruz, CA – *BA in Philosophy*

References on LinkedIn