Chris Blue

Data Engineer | Cloud, Automation, & Data Pipelines

hellochrisblue@gmail.com | GitHub |+1 (323) 240-5603

Professional Summary

Highly accomplished Senior Data Engineer with 7 years of experience designing, implementing, and optimizing scalable, high-performance data solutions that unlock insights across the content lifecycle. Proven expertise in building robust ETL pipelines, managing large-scale data migrations (100+ TBs), and leveraging AWS cloud services. Adept at automating complex workflows and ensuring data quality and consistency to support critical reporting, self-service analytics capabilities, and ML/AI systems in high-throughput environments, driving business decisions.

Technical Skills

- **Programming & Scripting:** Python (Advanced), SQL, Bash/Shell Scripting.
- **Cloud Platforms (AWS):** S3, AWS CLI, Glue, EMR, Redshift, Kinesis, Lambda, Athena, CloudFormation, EC2.
- **Big Data & Data Warehousing:** ETL/ELT Pipelines, Schema Design, Relational & Dimensional Data Modeling, OLAP Concepts.
- **Data Management:** Data Integrity, Data Quality, Data Governance, Performance Tuning, Cost Optimization.
- **Automation & Workflow:** Data Validation, Monitoring Solutions, REST API Design & Usage, CI/CD Concepts.
- **Tools & Methodologies:** Streamlit, Grafana, Agile/Scrum Framework, Git/GitHub.

Professional Experience

BuzzFeed Inc.   **Los Angeles, CA**
Data Engineer   Jul 2023 – Present
- Lead the development and automation of 10+ ETL workflows using Python, implementing robust validation checks to ensure 100% data reliability and consistency.
- Lead the migration of over 100 TB of critical media content across multiple ecosystems (BuzzFeed, Meta, Adobe) using AWS S3 batch operations, AWS SDK, and Python scripting, handling complex datasets and ensuring seamless data transfers.
- Collaborated with internal IT teams to define requirements and empower them to independently perform S3 operations in bulk, reducing data migration preparation time by 50% through the design and deployment of an internal Streamlit web tool.
- Architect and maintain scalable AWS-based archiving and backup solutions, increasing weekly backup capacity 5x to 100 TB and streamlining monthly archiving of 100 TB of media assets.
- Proactively manage disk usage across cloud and on-prem storage pools, optimizing resource allocation and reducing storage costs by 50%, enabling reinvestment in content initiatives.
- Orchestrate large-scale data restores for critical projects and integrate a Restore Costs Calculator Webhook to improve operational insights and cost forecasting.

7till8 Wetsuits   **Los Angeles, CA**
Operations Manager   May 2021 – May 2022
- Managed a 15k+ record dataset of customer measurements, ensuring data integrity and accuracy to improve downstream tailoring workflows and pattern generation efficiency.
- Optimized customer data management workflows by streamlining input and storage processes, enhancing data consistency and reducing manual errors.
- Generate reports and analyze customer measurement data to surface insights that inform product design improvements and drive strategic tailoring decisions.

Osk Studio   **Los Angeles, CA**
Associate Data Engineer (Freelance)   Oct 2019 – May 2021
- Improved data accessibility by 30% for lead artists and developers by designing and implementing custom programming logic to support ML pipelines.
- Built ETL processes to collect, clean, and transform multi-source datasets for ML model training, ensuring high-quality, analysis-ready data for critical initiatives.
- Used Python (pandas, numpy) to perform advanced data cleaning and manipulation, resolving inconsistencies and ensuring data quality for ML model inputs.

Education

New York University, New York, USA
Bachelor of Fine Arts   Aug 2011 – May 2015

Scrum Alliance, Inc., Colorado, USA
Certified Scrum Product Owner®   Dec 2024