

Chris Blue

Data Engineer (L5) | Cloud & Big Data Solutions | Automation & ML Data Pipelines

+1 323 240 5603

hellochrisblue@gmail.com

[GitHub](#)

Professional Summary

Highly accomplished **Senior Data Engineer** with 7 years of experience designing, implementing, and optimizing scalable, high-performance data solutions. Proven expertise in building robust data pipelines, managing large-scale data migrations (100+ TBs), and leveraging **AWS cloud services** (S3, CLI, Glue, Redshift, Kinesis, Lambda, Athena). Adept at automating complex workflows and ensuring data quality to support critical reporting, analytics, and ML/AI systems in high-throughput environments.

Technical Skills

- **Programming & Scripting:** Python (Advanced), SQL (Advanced: complex queries, optimization, performance tuning), Bash/Shell Scripting.
- **Cloud Platforms (AWS):** S3, AWS CLI, Glue, EMR, Redshift, Kinesis, Lambda, Athena, CloudFormation, EC2, Iconik/AWS and Qumulo/AWS S3 configurations.
- **Big Data & Data Warehousing:** ETL/ELT Processes, Data Pipeline Development, Data Warehousing, Data Lakes, Data Modeling (Dimensional, Star/Snowflake), Distributed Systems, Real-time Data Processing.
- **Data Management:** Data Migration, Data Archiving, Data Quality, Data Governance, Data Integrity, Performance Tuning, Cost Optimization.
- **Automation & Workflow:** Python Scripting, Workflow Automation, REST API Design & Usage, CI/CD Concepts.
- **ML/AI Data Support:** ML/AI Data Pipelines, Feature Engineering Support, ML/GPU Concepts, Data Preparation for ML Models.
- **Tools & Methodologies:** Git/GitHub, Streamlit, FFMPEG, Agile/Scrum Framework.

Professional Experience

BuzzFeed Inc.

Data Engineer

Jul 2023 – Present

- Led the **migration of over 100+ TBs of critical media content data** across diverse ecosystems (BuzzFeed, Meta, Adobe, First We Feast), including **complex Adobe and Lostin/BringMe data migrations**, using AWS S3 batch operations, AWS CLI, and Python scripting, **improving transfer efficiency by 25% and ensuring 100% data reliability**.
- **Streamlined data migration processes for ML content delivery** by designing and building an **internal web tool (Streamlit, Python)** to automate manifest generation, **reducing preparation time by 50%** and empowering internal IT teams to independently perform bulk S3 actions.

- **Designed and implemented a self-service archiving workflow for the Post Production department**, which streamlined BuzzFeed's archiving process, enabling the archiving of 100 TBs of data monthly.
- **Architected and deployed a scalable weekly data backup system on AWS**, increasing capacity **5x from 20 TB/week to 100 TB/week**, ensuring high data durability and availability for critical assets and contributing to operational excellence.
- Proactively **managed disk usage across multiple cloud (AWS) and NAS/SAN storage pools**, optimizing resource allocation, reducing storage costs by 50%, and preventing service disruptions.
- Configured **Iconik/AWS and Qumulo/AWS S3 copy solutions**, enhancing data accessibility and storage efficiency for large-scale media assets.
- **Orchestrated and delivered large-scale data restores** for critical projects (e.g., Filmrise, Hot Ones/Meta), ensuring timely content availability and **minimizing business continuity risks**.
- Integrated a **Restore Costs Calculator Webhook** to enhance operational insights and cost management for data restoration processes.
- Developed and implemented automated data workflows using Python scripting to streamline post-production processes, **improving efficiency and data handling for over 10 critical workflows**.

7till8 Wetsuits

Operations Manager

May 2021 – May 2022

- Managed the collection, organization, and integrity of a comprehensive dataset of over **15k customer measurements**, ensuring data accuracy and accessibility for pattern makers and seamstresses, **directly contributing to improved tailoring accuracy and efficiency**.
- Developed and implemented streamlined processes for efficient data input and storage of customer measurement data, **optimizing data management workflows** for the tailoring department.
- Utilized customer measurement data to generate reports and insights, **identifying trends for design improvements and better tailoring strategies**.

Osk Studio

Associate Data Engineer (Freelance)

Oct 2019 – May 2021

- Assisted in the design and implementation of programming logic for machine learning projects, **improving data accessibility by 30%** for data scientists and developers.
- Developed and executed robust processes to collect, clean, and transform datasets from various sources for ML model training and analysis, **ensuring high-quality data for critical ML initiatives**.
- Utilized Python libraries for efficient data manipulation and analysis in support of machine learning models, implementing **advanced data cleaning techniques** to handle missing values and inconsistencies.

Education

New York University, New York, USA

Bachelor of Fine Arts

Aug 2011 – May 2015