

Allen Clement

607 454-1328 | aclemen4@binghamton.edu | Binghamton, NY
<https://github.com/allenclement7> | www.linkedin.com/in/allen-clement

EDUCATION

Binghamton University, SUNY, Thomas J. Watson College of Engineering and Applied Science

Master of Science in Computer Science

Aug 2022 - Dec 2023

Visvesvaraya Technological University (VTU), Bangalore, India

Bachelor of Engineering in Computer Science

Aug 2015 - May 2019

Cumulative GPA: 3.72/4.00

TECHNICAL SKILLS

Languages Known: Python, Java, C#, C, HTML, CSS, XML

Business Software and Databases: Riversand, **SQL**, PostgreSQL, Hadoop, Hive, Git, GitHub, Bitbucket, Amazon Sagemaker

AWS: S3, Lambda, Amazon Redshift, Sagemaker, Glue, Athena, DynamoDB, Quicksight, CodeBuild, CodeDeploy, CodePipeline

GCP: Storage, Cloud Functions, App Engine, BigQuery, Secret Manager

Certifications: Microsoft Azure Fundamental AZ-900, AWS, Data Engineering, Python, SQL, Data Camp PySpark

Additional Tools: GCP Cloud DevOps, GDSN, Kubernetes, Jenkins, Docker, Linux, Apache Spark, Argo CD, Azure Blob Storage, Storage Explorer, Retool, Service Now, Airflow, Spark, Oracle, Anaconda, PyCharm, Power BI, Tableau, JIRA, Zapier, PIM/MDM

PROFESSIONAL EXPERIENCE

Data Science Intern, Colaberry Inc. | Texas, United States

May 2023 - Aug 2023

- Engineered a **Python** script, slashing monthly GCP costs from **\$7000** to just **\$15**, by **automating** the removal of inactive users' GCP disks and PVCs and achieved cost reduction of over **95%**, showcasing strong cost optimization skills
- Collaborated with a 3-person team, deploying a Python Django website, driving a substantial **35%** boost in user participation
- Applied AWS services, including DynamoDB for data storage and AWS Lambda for seamless function execution
- Orchestrated an efficient AWS-based CI/CD pipeline, integrating **AWS CodePipeline, AWS CodeBuild, and AWS CodeDeploy** to achieve a **50%** reduction in deployment time, ensuring rapid content updates
- Demonstrated expertise in Git repository management and code maintenance

Senior Systems Engineer, Infosys Limited | Bangalore, India

May 2019 - Aug 2022

- Played a crucial role as a **Data Engineer** in a cloud migration project, utilizing Pyspark for ETL batch processing to migrate OnPrem data to AWS, including a **50%** reduction in data transfer time and cost savings of **\$100,000**, while effectively leveraging technologies such as **Python, Snowflake DB, AWS services (S3, EMR, Lambda, SageMaker, DynamoDB, CloudWatch), Airflow, and Sonarqube**
- Led migration efforts to GCP cloud platform in an agile environment, showcasing expertise in cloud architecture and infrastructure designs, resulting in a **20%** reduction in infrastructure costs
- Spearheaded a cross-functional team of Cloud Engineers for Google Cloud Migration, reducing deployment time by **45%**, leading to faster project delivery
- Designed and managed CI/CD pipelines (Git Actions, Docker, Argo CD, Kubernetes), enabling streamlining deployment processes
- Orchestrated application migrations to GCP, utilizing Google Secret Manager for enhanced security and cost-efficiency, which facilitated in a **90%** lower operational costs
- Utilized Google Cloud services such as Cloud Storage, Compute Engine, Kubernetes Engine for resource optimization, saving \$50,000 in annual infrastructure costs
- Implemented Riversand MDM product, enhancing master data management with refined data quality and governance. Configured and maintained the Riversand **data model** for various domains, established GDSN rules for compliance, and exploited Azure Blob Storage and Event Hub for data management and processing, cutting down data receiving time by **75%**

PROJECT EXPERIENCE

Real-time Data Orchestration, ETL Pipeline using AWS and Netflix Conductor, Independent

Jun 2023 - Present

- Engineered and deployed a streamlined real-time data orchestration and **ETL pipeline** using AWS services such as **Amazon Kinesis, AWS Glue, Amazon S3, Amazon Redshift, Amazon Athena, and Amazon CloudWatch**, with Netflix Conductor
- Delivered actionable insights and elevated user experiences by slashing data processing time by a remarkable **50%**

Interview Preparation System, Colaberry Inc., Texas

May 2023 - Aug 2023

- Spearheaded the development of a comprehensive interview preparation system, leveraging expertise in technologies such as Django, DynamoDB with OpenAI integration and a robust system architecture, implementing video recording, evaluation, and analytics using Python, Pandas, and SciKit-learn, resulting in a **15%** improvement in candidate performance

ETL on YouTube trending video data using AWS Services, Binghamton, New York

Feb 2023 – Mar 2023

- Employed AWS services, including S3, Glue Catalog, Lambda, Athena, and **AWS Glue ETL**, to streamline the ingestion and transformation of Kaggle's Trending YouTube Video Statistics data, **achieving a 40% improvement in data processing time**.
- Generated a clean dataset in Parquet format, facilitating efficient BI analysis

Real-Time Data Streaming using Apache Nifi, AWS, Snowpipe, Stream & Task

Jan 2023 – Feb 2023

- Independently designed and implemented a real-time data streaming and **data warehousing** solution utilizing **Apache NiFi, Snowflake, and AWS technologies**, leading to a 30% acceleration in data processing efficiency