

Arthur Vargas

Senior Software Engineer

Next-Gen Sequencing Data Pipelines

Location: Austin, Texas, US

Mobile: +1 (336) 416-7584

Email: ahvargas92@gmail.com

LinkedIn: www.linkedin.com/in/arthur-vargas-swe

PROFESSIONAL SUMMARY

Senior software engineer with 6 years developing bioinformatics data pipelines and Good Clinical Practice (GCP) software compliant with 21 CFR requirements. Advanced Python programming, Nextflow development, and modern engineering best practices. Knowledge of Illumina NGS data, sequencer types, and software tools. Skilled in containerizing pipeline components and configuring continuous integration/continuous deployment (CI/CD) pipelines in Azure DevOps (ADO). Experience using AWS services, including Simple Storage Service (S3), Elastic Compute Cloud (EC2), Lambda, and ECS (Elastic Container Service), for automation, parallelization, and monitoring. Biology bachelor's degree, passionate about life sciences and advancing translational biomarker research technologies.

WORK EXPERIENCE

IQVIA | Senior Software Engineer

NOV 2021 – AUG 2024 | Durham, NC

- Guided a remote team of engineers by delegating technical tasks, conducting code reviews, promoting object-oriented software designs, setting code coverage thresholds for unit testing, automating software builds in ADO, and adopting latest AI tools for development, like Github Copilot.
- Developed a primary analysis Nextflow pipeline performing demultiplexing with Illumina software tools like bcl2fastq and bclconvert, leveraged Illumina InterOp libraries to extract and report sequencing quality control metrics like read length, percentage of bases with Phred score ≥ 30 . Pipeline was relied on as a critical first step for Illumina assays targeting cancer relevant gene variants.
- Wrote modular software in Python, Groovy, and BASH to automate Illumina samplesheet creation for demultiplexing, verify FASTQ file integrity, and generate run reports. Containerized software components with Docker and orchestrated with Nextflow.
- Developed and deployed a Python-based CLI using the Typer library, featuring five commands with minimal arguments. This streamlined application replaced manual processes for evaluating Ion Torrent Genexus Sequencer run quality. It combined project management and genomic data into comprehensive quality control reports to facilitate client data delivery.

- Developed first iteration of a testing framework for Nextflow code using BASH Automated Testing System (BATS) to execute functional tests, resulting in more reliable pipeline orchestration.
- Configured a memory allocation strategy in a primary analysis Nextflow pipeline for resource-intensive components, incorporating dynamic retries and exponential backoff to support high-performance computing (HPC). This configuration minimized the impact such components had on shared resources.
- Collaborated with software architects and designers to ensure that pipelines adhered to corporate IT infrastructure standards and software designs were extensible, decoupled, and maintainable. Focus on architecture and design enhanced future project scalability.
- Documented all software, crafting clear and concise docstrings for functions, classes, and modules. Maintaining Confluence pages with build details and instructions for user acceptance testing. Wrote the official product documentation including design and architecture specifications, and user manuals

IQVIA | Software Engineer III

NOV 2019 – NOV 2021 | Durham, NC

- Utilized the dacite Python module to convert HTTP payloads into Python dataclasses, enabling efficient, type-safe data transformation and clean access via dot notation. This approach streamlined data handling and supported post-initialization actions with `__post_init__()`
- Containerized pipeline software components by defining base images, environment variables, entry points, dependencies, and project setup in Dockerfiles, ensuring a consistent, isolated runtime environment deployable across diverse systems.
- Configured CI/CD pipelines by defining steps and jobs in build scripts (azure-pipelines.yml) that used other templates, external code libraries, and docker images. This infrastructure enhanced company processes, security, and automated the end-to-end build and deployment process
- Collaborated with project managers to understand internal customer needs and translate them into actionable development tasks in an Agile methodology.

IQVIA | Software Engineer II

OCT 2018 – NOV 2019 | Durham, NC

- Wrote a command-line application for migrating large files from local network to AWS S3 where they could be input into a genomic profiling assay
- Assisted development of Python components that monitor S3 buckets allowing for immediate initialization of AWS EC2 instances when inputs were available. This parallelization significantly reduced overall runtime of the assay

PROJECTS

[Gene Annotator](#)

- Identifies, logs, and removes duplicate entries in gene annotation data. Performs data transformation steps and saves results to file
- SEP 2023

SKILLS

Programming Languages and Frameworks

- Proficient in: Python
- Working knowledge of: Groovy, Java, BASH, Nextflow, C/C++, JavaScript

Tools and Technologies

- Cloud Platforms: AWS, Azure
- API: FastAPI, Flask
- Code Management: Docker, GIT, ADO, AWS, Maven, Gradle
- API Development: Postman
- OS: Linux

EDUCATION

B.S. Biology

North Carolina State University, AUG 2002 – DEC 2006