Bijan Vakili

Senior ML Ops / Data Engineer

Washington, DC • bvakili.com • resume-w@mail.bvakili.com • (202) 699-8519

Summary

Experienced ML Ops and Data Engineering professional with 15+ years of expertise in cloud infrastructure, scientific workflow optimization, and healthcare data integration. Proven track record of implementing ML models, automating data pipelines, and optimizing research workflows in life sciences and healthcare environments.

Skills

ML Ops & AI

AWS SageMaker, LangChain, LLMs, scikit-learn, model validation, inference pipelines, Pandas, Dash, Python scipy, OpenAI, AWS Bedrock, Ollama, Multi-modal AI

Data Engineering

ETL processes, data mapping, transformation strategies, SQL/NoSQL, Redshift, data integrity validation, Dask, pandas, data transformation, data validation

Cloud Infrastructure

AWS (Lambda, S3, SQS, RedShift, EC2, RDS, ECS Fargate, EFS, EBS, DocumentDB, Aurora, DynamoDB, IAM, VPC, API Gateway, ALBs, CloudFront, Route53), IaC with Terraform, scalable ML infrastructure

Programming

Python, Java, JavaScript, R, Bash, Node.js, HTML/CSS, data structures, algorithms, API development

DevOps / CI/CD

GitHub Actions, Jenkins, GitLab CI, Docker, Kubernetes, automated testing, monitoring, Terraform, Atlantis, tfsec, S3 PyPi, S3-backed npm, pytest, tox, black

Certifications

AWS Developer Associate (DVWFK6C2PEBE19SV), AWS Solutions Architect Professional (FG18BPPLE1EE1F3X), Certified Java SCJA (259274625), Kubernetes Certified Administrator (LF-ih10o4kbs4)

Professional Experience

Major Biotechnology Company

Solutions Architect / Dedicated Support for Data Scientists
June 2022 – March 2024
Location: Washington, DC; West Coast; & Remote

Description:
In a transformative period for IT within the organization, I spearheaded the integration of bioinformatics research into the production environment, marking a pioneering initiative to accelerate scientific discovery. My role was pivotal in developing and deploying innovative software solutions that directly impacted decision-making in drug development.

I was embedded in the bioinformatics team that collaborated with senior oncology data scientists on several key initiatives. This collaboration provided valuable insights into the integration of complex biological data with clinical applications, particularly in the oncology space.

I led the implementation of cBioPortal, customizing it to meet the bioinformatics team's needs. This initiative was later showcased to the Chief Science Officer, establishing it as the organization's primary data-sharing mechanism, in line with FAIR principles.

As the steward of the AWS account, I managed cloud resources exceeding $100k/year across multiple environments, balancing cost-efficiency with the support of the bioinformatics team's internal workflows and the facilitation of company-wide data sharing.

Key Achievements:

- Helped bioinformatics team hit yearly target identification goals in first quarter, accelerating drug discovery timeline and improving research efficiency
- Developed automated data pipelines that reduced data update cycle from a month to under two days for critical drug target data, achieving a 10x efficiency gain
- Established a virtuous cycle between engineering/IT and scientists, bridging technical implementation with scientific research needs
- Helped lead scientist transition from SQLite to Aurora PostgreSQL with individualized schemas and custom permissions tailored to support their teams' scientific workflows
- Identified critical inefficiencies and security vulnerabilities in data management systems, spearheading the migration from a non-compliant, self-managed MongoDB cluster to Amazon DocumentDB for the R-based bioinformatics group
- Introduced the "restaurant model" for prioritizing research code's transition to production, requiring a code piece to be requested three times before being moved, significantly optimizing resource allocation
- Enhanced an innovative asynchronous system to overcome Python's single-threaded limitations, decoupling compute tasks into a scalable, queue-based architecture using Dask clusters for parallel computing
- Rapidly developed UI web smoke tests using AWS DeviceFarm to verify system functionality across all major browsers and platforms
- Proposed and developed a secure research preview environment for Dash apps, expediting the bioinformaticians' research cycle in a cost-efficient manner

Technologies Used:

- **AI/ML:** AWS SageMaker, JupyterLab on EC2, Python scipy, scikit-learn (SimpleImputer, RandomForestClassifier), OpenAI, AWS Bedrock, Ollama, LangChain, Multi-modal
- **Cloud/Infrastructure:** AWS ECS Fargate, EC2, Lambda, S3, EFS, EBS, DocumentDB, Aurora RDS (PostgreSQL + MySQL), DynamoDB, Redshift, IAM, VPC, Security Groups, Subnets, SQS, API Gateway, ALBs, CloudFront, Route53
- **Programming:** Python, R, Java, Bash, Node.js, HTML/JavaScript/CSS
- **Data Engineering:** Dask, pandas, ETL pipelines, data transformation, data validation, data integrity checks

- **DevOps/CI/CD:** GitHub Actions, Terraform, Atlantis, tfsec, S3 PyPi, S3-backed npm, pytest, tox, black, Docker
- **Visualization:** Plotly, Dash, ReactJS, AG Grid

Healthcare Industry Association

Portal Team Lead / Data Engineer
November 2013 – September 2014
Location: Midwest / Remote

Description:
A major healthcare industry association's marketplace website is a healthcare B2B platform. I was brought in to resolve critical issues and subsequently restructured their codebase to improve maintainability and performance. I rewrote their models, persistence layer, and fully implemented the UI using Liferay Service Builder, Spring Portlet MVC, and JSPs. Due to my competence in troubleshooting issues in a timely fashion, I was chosen to be the support point of contact for the 60-day warranty phase.

Key Achievements:

- Advised on the Roles Implementation; mapping Liferay functions to business needs; such as leveraging Org / Site Roles, Teams
- Wrote Site Map document for the portal; including details about the pages; such as themes, urls, permissions, layouts, and layout settings
- Restructured 50+ portlet plugins to make projects easier to maintain and follow best development practices
- Rescued project from performance issues by replacing custom code with Service Builder (SB) and removing custom and buggy caching code
- Resolved clustering replication bugs in Amazon EC2 environment
- Combined all portlet contexts into one; thus reducing the memory footprint
- Refactored hundreds of Java classes; over 50 Spring Portlet controllers
- Improved build time, startup time, and page load; at least 2x improvement on these points
- Extended the Liferay user profile to include dozens of new attributes and sync with external system (third-party web service)

Technologies Used:

- **Cloud:** AWS EC2 cloud environment
- **Backend:** J2EE (servlets/JSPs), Tomcat 7, Portlets, Liferay plugins (hooks, themes, ext), Liferay 6.1 GA2, Velocity, Spring Portlet MVC, Web Services, JAX-RS
- **Frontend:** jQuery, various jQuery plugins, Liferay Service Builder
- **Security:** OpenAM
- **Search:** Solr
- **Testing:** iMacros for web test automation

Environmental Research Organization

Portal Team Lead / Data Integration Specialist
August 2012 – September 2014
Location: Pacific Northwest / Remote

Description:
I developed a portal for a major environmental research organization to help their scientists with their assessment effort for environmental damage caused by a major industrial incident. This involved heavy customization of Liferay to support unique requirements regarding file storage, calendar sharing, and field operation workflows. The project required sophisticated data transformation strategies for scientific research data.

Key Achievements:

- Integrated Liferay with Kaltura using Web Content Portlet, Flash, and HTML5 video
- Managed multiple Amazon EC2 environments: DEV, SYSTEST, and PERFORMANCE
- Enhanced Document Library portlet to meet document management goals, ensuring WebDAV compatibility across all operating systems
- Heavily customized portal Calendar to allow sharing calendars among sites, pushing calendar events based on permissions, and implementing advanced filters
- Wrote custom forms portlet for scientists using Spring Portlet MVC for dispatching, with workflow handler for complex review processes
- Installed and configured Hudson CI box, restructuring plugins to be modular and to only build and deploy changed components
- Profiled, debugged, and optimized plugins, resolving a long-standing server degradation problem using JVisualVM, JStat, Grinder, and JMeter

Technologies Used:

- **Cloud:** AWS EC2 virtual environment
- **Backend:** J2EE (servlets/JSPs), Tomcat 7, Portlets, Liferay plugins (hooks, themes, ext), Liferay 6.1 GA2, Velocity, Spring Portlet MVC, Web Services, JAX-RS
- **Frontend:** jQuery, various jQuery plugins, Sencha Ext JS
- **Data:** Liferay Service Builder, WebDAV
- **Security:** OpenAM
- **Search:** Solr
- **CI/CD:** Hudson CI, Git and Github
- **Monitoring:** JStat, JVisualVM, Grinder, JMeter
- **Media:** Kaltura

Large Healthcare Benefits Organization

Portal Team Lead / Healthcare Data Engineer
May 2011 – September 2012
Location: Mid-Atlantic Region

Description:
I led a major technology consulting firm's healthcare benefits processing portal team from requirements through go-live, with a system serving millions of users. This healthcare data integration aspect of the project involved complex mapping of medical forms and secure data exchange between multiple systems using standards, such as ICD 10 codes, EDI 837, and HL7.

Key Achievements:

- Led 10+ senior portlet developers in custom development of plugins, including IPAC Ext Plugin (Inter Page Communication) and KFI Portlet Plugin (Key From Image) with mapping of over 1500 fields

- Designed data mapping strategies for medical bill processing system, handling complex healthcare data formats and ensuring data integrity
- Clustered Liferay with failover on 7 environments, including specialized testing environments, administering all portal, reporting, SSO, and LDAP related servers
- Integrated ActiveVOS Web Services, Actuate BIRT Interactive Viewer, and OpenAM/DJ SSO, creating CXF web service clients and defining WSDL contracts for the workflow
- Utilized RabbitMQ web service for message handling with graceful exception management
- Conducted performance testing of the infrastructure, especially portal and SSO, using Grinder and JMeter, with JVisualVM deep Memory/CPU monitoring and profiling
- Wrote a white paper to size the production hardware, taking responsibility for portal, SSO, and LDAP servers

Technologies Used:

- **Backend:** J2EE (servlets/JSPs), Tomcat 6, Portlets, Liferay plugins (hooks, themes, ext), Liferay 6.0.6, Velocity, Spring Portlet MVC, Web Services, JAX-RS, JAX-WS
- **Frontend:** jQuery, various jQuery plugins, Liferay Service Builder
- **Security:** OpenAM
- **Search:** Solr
- **Integration:** WebDAV, Hudson CI, ActiveVOS, Actuate BIRT
- **Monitoring:** JStat, JVisualVM, Grinder, JMeter, Cacti

Education

B.Sc. in Computer Science
University of Virginia, Charlottesville, VA
Graduated 2008

Study Abroad
Tsinghua University, Beijing, P.R.C.
Fall 2007

Awards

- MLOps Marketplace Competition (2025)
- IAM Excellence Award (2024)
- Top Value Consultant (2022-2024)
- Enterprise System Excellence Award (2018)
- Liferay Contributor of Year (2014-2015)
- Technology Solutions Delivery Award (2009)