Chris Dagdigian

Principal Consultant, BioTeam · Co-founder
Boston, MA and Western Maine · dag@bioteam.net · linkedin.com/in/chrisdag

SUMMARY

Co-founder of BioTeam and a hands-on infrastructure consultant for life-science research computing since 2002. Designs, builds, and troubleshoots HPC clusters, petascale storage, high-speed data networks, and AWS cloud environments for pharmaceutical, biotech, academic, and government clients. Works comfortably with bench scientists, system administrators, and executives, and still does the hands-on infrastructure work himself.

Current focus: cloud-native HPC (AWS ParallelCluster, Slurm), GPU scheduling for computational chemistry and cryo-EM, and infrastructure for AI and machine-learning workloads. AWS user since September 2007, during the invitation-only EC2 beta and more than a year before EC2 reached full production status (October 2008). Building production scientific computing environments on IaaS cloud platforms since 2010. Co-founder of the BioPerl project and a founding officer of the Open Bioinformatics Foundation.

CERTIFICATIONS

Seven current AWS certifications. AWS certified since 2014.

Certification	Valid through
AWS Certified Solutions Architect – Professional	Nov 28, 2028
AWS Certified Solutions Architect – Associate	Nov 28, 2028
AWS Certified Cloud Practitioner	Nov 28, 2028
AWS Certified Security – Specialty	May 7, 2028
AWS Certified AI Practitioner	Mar 8, 2028
AWS Certified SysOps Administrator – Associate	Mar 9, 2028
AWS Certified Developer – Associate	Mar 6, 2028

EXPERIENCE

BioTeam — Middleton, MA · Feb 2002 – Present

Principal Consultant (co-founder; previously Co-Founder and Senior Director, Infrastructure)

Principal consultant in BioTeam's HPC and infrastructure practice and a full-time, customer-facing consultant. Clients range from stealth-mode startups to global pharmaceutical companies, plus academic institutions and U.S. government agencies. Most work concentrates on HPC design and operations, petascale storage, high-speed networking for large scientific data sets, and AWS cloud architecture. Earlier government work includes several years as an HPC specialist supporting CDC scientific computing.

Recent work (2023 – 2026):

- Led a multi-deliverable HPC assessment for a biotech client: baseline of the current environment, requirements and gap analysis, and a strategic roadmap.
- Designed AI and HPC infrastructure for a research university, covering scope, deliverables, and the on-premises versus cloud decision.
- Diagnosed intermittent GPU job failures on a pharmaceutical computational-chemistry Slurm cluster. Root cause was a containerized ML stack claiming a GPU outside Slurm's control. Moved users to containers launched under Slurm (Docker and Apptainer) and kept the older V100 nodes in service alongside Blackwell-generation GPUs.
- Integrated Posit Workbench with AWS ParallelCluster through the Slurm job launcher, so user sessions start on auto-scaling compute nodes and large-memory or multi-GPU servers are billed only while in use.

- Designing and validating custom IAM roles and permission boundaries that let AWS ParallelCluster and its image-build pipeline deploy into a shared pharmaceutical production account under least-privilege guardrails. Separate boundaries for the image builder and the cluster operator, resource scoping by name and tag to protect other teams' resources, and checks with the IAM policy simulator and IAM Access Analyzer.
- Built and validated a closed-network proof of concept for SageMaker HyperPod on EKS: four GPU nodes, FSx for Lustre linked to S3, GROMACS scaling runs (81.7 ns/day on one node, 47.3 ns/day on two), and Kueue-based task governance with borrow-and-preempt behavior.
- Developed retrieval-augmented generation (RAG) pipelines and Model Context Protocol (MCP) servers for BioTeam that give AI assistants searchable access to two collections: HPC and scientific-software documentation (Slurm, Schrödinger, CryoSPARC, RELION, GROMACS, AMBER, Apptainer, and others) and BioTeam's consulting project archive. Both run behind federated single sign-on.
- Creates HPC training for life-science users and administrators. Began with Grid Engine (SGE) user and administrator classes taught across the US; the current Slurm curriculum has 45 modules for users, power users, administrators, and leadership across on-premises clusters, AWS ParallelCluster, and AWS PCS. It includes SGE, PBS, and LSF migration guides, Schrödinger and CryoSPARC integration modules, and hands-on exercises validated on a live test cluster. Also developed AWS ParallelCluster training sessions.

Earlier highlights: ParallelCluster deployments for pharmaceutical and agricultural-science clients, Schrödinger Suite and cryo-EM (RELION, CryoSPARC) cluster integrations, and storage and networking assessments for government and academic research organizations.

Blackstone Computing — Boston, MA · 2000 – 2001

Principal Consultant, Bioinformatics and Infrastructure Technology

Member of a consulting team that turned early Beowulf-style HPC designs into production clusters for biotech and pharmaceutical clients, including systems on the Top500 list. Trainer and integrator for Platform LSF and Sun Grid Engine.

Genetics Institute / Wyeth Research — Cambridge, MA · 1995 – 2000

Systems Manager, Department of Genomics (2000). Ran the datacenter and infrastructure for the genomics and discovery groups.

Systems Manager, Bioinformatics Hardware Group (1998 – 1999). Designed, purchased, deployed, and supported the scientific computing infrastructure for the computational biology groups, including the company's first purpose-built research datacenter.

Bioinformatics Programmer/Analyst (1996 – 1998). Wrote software and pipelines for automated, high-throughput cDNA analysis. Co-founded the BioPerl project.

Research Associate, Bioinformatics (1995 – 1996). First professional role, supporting an emerging cDNA sequencing project.

OPEN-SOURCE AND COMMUNITY LEADERSHIP

- Co-founder, BioPerl project.
- Helped found the Open Bioinformatics Foundation, the nonprofit umbrella for BioPerl and related toolkits and host of the Bioinformatics Open Source Conference (BOSC); served as Treasurer and board member for 10+ years.
- Founder of the Bioclusters mailing list and operator of the gridengine.info blog and wiki.

PUBLICATIONS

Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., … Birney, E. (2002). The Bioperl Toolkit: Perl Modules for the Life Sciences. *Genome Research*, 12(10), 1611–1618. doi:10.1101/gr.361602

SPEAKING AND WRITING

- Annual "Trends from the Trenches" address at Bio-IT World Conference & Expo, from the 2009 keynote onward; ten consecutive years as of 2020, and again on the 2022 keynote program.
- Frequent invited speaker and trainer at industry, academic, and government events.
- Technical blog at bioteam.net, 2008 – present.

EDUCATION

B.S., Biotechnology (concentration in Computer Science), Worcester Polytechnic Institute, 1995.

TECHNICAL SKILLS

- **Programming and infrastructure-as-code (IaC) automation:** Perl, Python, Ruby, shell scripting; Ansible, Terraform, CloudFormation, Packer.
- **HPC and schedulers:** Slurm, AWS ParallelCluster (formerly CfnCluster), Grid Engine, LSF, PBS Pro, Torque/Maui; SageMaker HyperPod; Kueue.
- **Cloud:** AWS (EC2, S3, FSx for Lustre, EKS, SageMaker, IAM and organization-level design); Kubernetes and OpenShift.
- **Containers and software delivery:** Docker, Apptainer, Spack, EasyBuild, Environment Modules and Lmod.
- **Scientific applications:** Schrödinger Suite, RELION, CryoSPARC, GROMACS, AMBER, NAMD, GATK, Posit Workbench (RStudio), Nextflow.
- **Infrastructure:** Linux, macOS, Windows; petascale storage (BeeGFS, Lustre, NFS, Isilon, Avere, Panasas, Qumulo); high-speed networking; datacenter design; Bright Cluster Manager, Rocks.