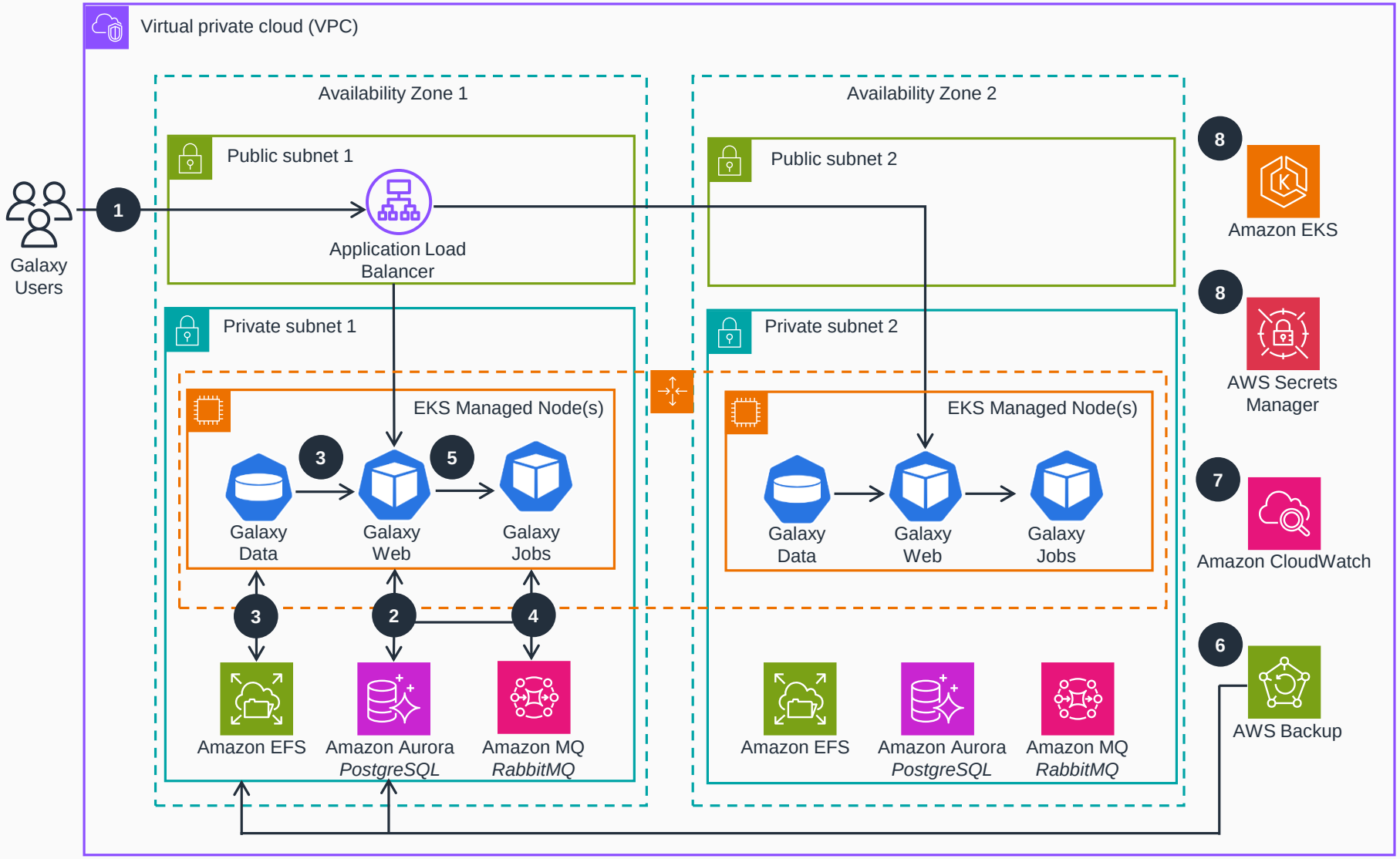
Guidance for Galaxy Deployment on AWS

This architecture diagram shows best practice recommendations on how to run Galaxy software on AWS at scale.



1. Galaxy users access the Galaxy Web application by the public endpoint of the **Application Load Balancer**.
2. Galaxy stores user metadata and history in the PostgreSQL database hosted in **Amazon Aurora Serverless**, and Galaxy users have access through the Galaxy Web server. The credentials for this access are stored in **AWS Secrets Manager**.
3. The Galaxy Data volume stores user data, including input data for processing and processed data. **Amazon Elastic File System (Amazon EFS)** provides the storage capacity.
4. Galaxy leverages a message queue for communication between internal processes. This Guidance hosts the message queue on **Amazon MQ** with the RabbitMQ message broker. Credentials for the broker are stored in **Secrets Manager**.
5. Users can review, schedule, and manage bioinformatics jobs in the Galaxy Jobs pod through the Galaxy Web server.
6. **AWS Backup** takes regular backups of **Amazon EFS** file systems and PostgreSQL databases.
7. The monitoring and log collection of both the Galaxy components and the AWS infrastructure are centralized in **Amazon CloudWatch**.
8. **Amazon Elastic Kubernetes Service (Amazon EKS)** provides the control plane, manages both the networking and the nodes for the Kubernetes pods, and horizontally scales by adding or removing nodes. Additional pods are deployed through **Amazon EKS** to synchronize secrets with **Secrets Manager** and to publish logs to **CloudWatch**.

Reviewed for technical accuracy January 17, 2024

AWS Reference Architecture