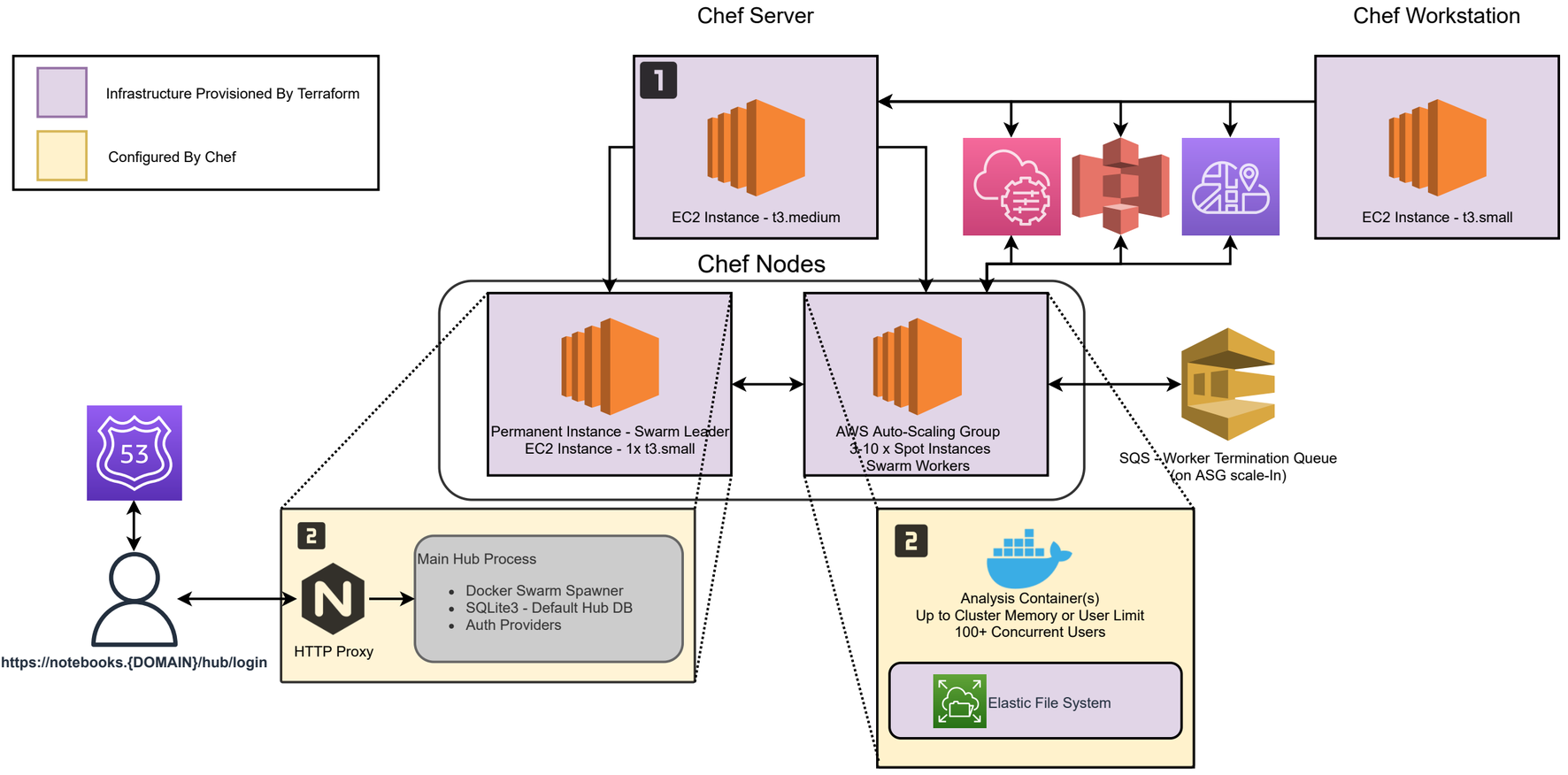

1

Ops Engineer with access to an AWS Admin Role runs a `terraform apply`. This creates several dozen resources (Purple), most important of which are:

- A VPC for all resources
- An EC2 instance that configures itself as a Chef Server.
- An EC2 autoscaling group; nodes in this group register with the Chef server on init
- An EC2 instance to act as a workstation
- An Elastic Filesystem (EFS) for sharing data between nodes

Chef Server instance's user data installs Chef and reads/writes SSM, S3, and CloudMap to configure the server.

The Chef Workstation waits on creation of the Chef Server before initializing and creating default roles for the hub deployment.

The Hub master node and ASG instances wait on the Workstation, and read from S3, CloudMap, and SSM on launch, finally, they register as Chef nodes

2

Cookbooks are pushed to AWS S3, consumed by the Workstation and added to the appropriate Node runlists.

Once nodes are registered, "Leader" cookbooks are run on a permanent node that registers the instance as the swarm leader. Includes:

- Docker - Launch nginx as a reverse proxy for the Hub.
- Docker Swarm - Register as Swarm leader
- JupyterHub Server w. Custom Docker-Swarm Spawner

These cookbooks start a Docker Swarm, publish the swarm token, and advertise the IP of this instance for ASG nodes to join.

"Worker" cookbooks run on each ASG instance

- Docker - Install Docker s.t. container can be spawned on instance
- Docker Swarm - Register as Swarm Node
- NFS Client - Share datasets between nodes

3

The main hub runs at **https://notebooks.{DOMAIN}/hub/login** and uses Nginx for SSL termination. Nginx forwards requests to the hub server running on the same instance.

The hub server is responsible for launching a new Python (i.e. Jupyter Notebook) container onto one of the ASG nodes.

Because this deployment uses NFS, users can safely terminate and restart their notebook servers (including on different machines) and have access to the same files.

When the resource utilization is too high, additional ASG nodes will spin up, up to the maxium cluster memory, concurrent users, or disk utilization parameters passed in the jupyterhub config