How can you make sure this application does not go out of hand and consume all the cluster resources?

Setting resources requests and limits.

How would you collect logs from this app for later analysis? Where and how would you save it?

The app logs might be shipped to CloudWatch. Log data might be exported to S3. Monitoring can be done with Loki, Promtail and Grafana as a visualization layer.