CS 4287/5287: Principles of Cloud Computing

Programming Assignment #3
Handed out 10/17/2024; Due 11/08/2024 (11:59 pm CDT on Brightspace)
Theme: Cloud-native Deployment of IoT Data Analytics Pipeline using Kubernetes and Docker

Team 14 Members:

- **Maddox**
- **Emma**
- **Abhay**

Project Components and Team Responsibilities

VM Termination and Ansible Playbook Setup

VM Termination (Completed by: Abhay)
Tasks Completed:

- Terminated all VMs from PA2, preserving the automation setup for future use.

Ansible Playbooks (Completed by: Maddox)

- Extended the master playbook with child playbooks to:
    - Install Kubernetes and required Python packages.
    - Configure and run a private Docker registry on VM1.
    - Modify `config.toml` to add the private registry mirror.
    - Restart containerd and Docker daemons.
    - Update firewall rules and disable swap for Kubernetes.

Deliverables:

- Documented Ansible playbook structure and VM termination steps.
- Verified package installations and Docker registry configuration.

Kubernetes Cluster Configuration

Cluster Setup (Completed by: Abhay)
Tasks Completed:

- Manually configured a 4-node Kubernetes cluster, designating VM1 as the master and tainting it to also function as a worker.
- Connected the remaining VMs as worker nodes.

Deliverables:

- Documented the manual steps for Kubernetes cluster setup and configuration.
- Verified cluster functionality across nodes.

Pipeline Deployment and Testing

Container Deployment and Private Registry Configuration

Docker and Registry Configuration (Completed by: Emma) Tasks Completed:

- Built Docker images for the producer, consumer, and ML inference server using Dockerfiles from PA2.
- Pushed images to the private Docker registry and ensured reachability across worker nodes.

Deliverables:

- Documented Dockerfile modifications and registry setup.
- Verified successful image pulls from all VMs.

Pod and Pipeline Deployment

Kubernetes YAML Configurations (Completed by: Emma) Tasks Completed:

- Created deployment and service YAML files for Kafka, ML inference, and the database pods.
- Configured Job YAML files for the producer and consumer, deploying components across the Kubernetes cluster.

Deliverables:

- Provided YAML configurations and tested deployment across Kubernetes.
- Verified pod communication and data relay functionality within the pipeline.

Pipeline Execution and Testing (Completed by: Maddox)

Tasks Completed:

- Executed the full data pipeline, ensuring each component's successful communication and data relay.
- Ran the producer at a frequency of one sample per second, capturing latency results for analysis.

Deliverables:

- Documented latency results and pipeline functionality.
- Reported any issues encountered and corresponding resolutions.

Workload Variation and Performance Analysis

Producer Scaling (Completed by: Abhay) Tasks Completed:

- Introduced multiple producers in Kubernetes (1 to 5) using the replica feature.
- Collected and recorded end-to-end latency data to analyze performance impact at each scale.

Deliverables:

- Collected data and plotted CDF curves for latency with varying producer counts.
- Analyzed 90th and 95th percentile response times for performance impact reporting.

General Collaboration and Integration

- **Documentation and Reporting (Completed by: Maddox)**: Compiled a final report detailing the project setup, testing processes, and analysis of performance outcomes.
- **Video Demonstration (Maddox)**: Created a video demonstration of the project, outlining each component and its functionality.

Milestones and Deliverables

1. **Milestone 1**: VM termination, Ansible playbooks extension, Kubernetes package installation (Status Report on Brightspace)
2. **Milestone 2**: Kubernetes cluster creation, single producer pipeline testing (Status Report on Brightspace)
3. **Final Submission**: Full pipeline testing, producer scaling, CDF plots, and complete project documentation and video demonstration