Multi-Agent System for Load Balancing in Cloud Computing

HAMMOUDI Sarra

1 Deployment Diagram Description

The **Deployment Diagram** illustrates the physical deployment of the system components across servers and clusters, emphasizing the dynamic assignment of requests based on their type (video, image, or text). Key elements include:

- **Resource Manager Server**: Hosts the **Principal-Agent**, which analyzes the type of client requests and delegates them to the appropriate cluster head:
  - *Video requests* are sent to the **Video Cluster Head**.
  - *Image requests* are sent to the **Image Cluster Head**.
  - *Text requests* are sent to the **Text Cluster Head**.
- **Cluster Heads**:
  - **Video Cluster Head**: Contains the **Agent-Video-Supervisor** and **Agent-Video-Annuaire**, which check the state of all servers and assign video requests to the *underloaded server*.
  - **Image Cluster Head**: Contains the **Agent-Image-Supervisor** and **Agent-Image-Annuaire**, which handle image requests and assign them to the *underloaded server*.
  - **Text Cluster Head**: Contains the **Agent-Text-Supervisor** and **Agent-Text-Annuaire**, which manage text requests and assign them to the *underloaded server*.
- **Servers (Server 1, Server 2, Server 3)**: Each server hosts the **Agent-Monitor**, **Agent-Surveillance**, and **Agent-Supervisor**, which manage database updates, monitor server states, and send alerts to the system administrator if issues are detected.
- **System Admin**: Receives alerts from the servers in case of failures or anomalies.

The diagram highlights the interaction between the **Resource Manager**, **Cluster Heads**, and **Servers**, emphasizing the dynamic assignment of requests to the *underloaded server* after checking the states of all servers.

2 Sequence Diagram Description

The **Sequence Diagram** depicts the sequence of interactions between the system components during a client request, emphasizing the analysis of request types and their assignment to the appropriate cluster head. The key steps are:

- The **Client** sends a request to the **Principal-Agent**.
- The **Principal-Agent** analyzes the request type and delegates it to the appropriate cluster head:
  - *Video requests* are sent to the **Agent-Video-Supervisor**.
  - *Image requests* are sent to the **Agent-Image-Supervisor**.
  - *Text requests* are sent to the **Agent-Text-Supervisor**.
- The cluster head forwards the request to its respective **Annuaire** agent:
  - **Agent-Video-Annuaire** for video requests.
  - **Agent-Image-Annuaire** for image requests.
  - **Agent-Text-Annuaire** for text requests.
- The **Annuaire** agent retrieves the CPU, storage, and database load information from all servers and assigns the request to the *underloaded server*:
  - *Video requests* are assigned to **Server 2**.
  - *Image requests* are assigned to **Server 1**.
  - *Text requests* are assigned to **Server 3**.
- The **Agent-Monitor** in the selected server detects database updates and notifies the **Agent-Surveillance** and **Agent-Supervisor**.
- The **Agent-Surveillance** increments a sequence number and sends it to the **Agent-Supervisor**, which updates the directory in the respective **Annuaire** agent.
- The **Agent-Surveillance** periodically checks the status of the **Agent-Monitor** and sends an alert to the **System Admin** if the monitor fails.

The sequence diagram emphasizes the dynamic and collaborative nature of the system, ensuring efficient load balancing and minimal latency.