**Requirements -**

**1. Compute**

* **Number of VMs:** Minimum **2**, Maximum **10** (depending on workload demand).
* **VM Sizes:**
* Baseline: **2 vCPUs / 4 GB RAM** (e.g., AWS t3.medium, Azure B2ms, GCP e2-standard-2).
* Comparative Option: **4 vCPUs / 8 GB RAM** (e.g., AWS t3.xlarge, Azure D4s\_v3, GCP e2-standard-4).
* **Operating System:** Ubuntu 22.04 LTS (preferred) or equivalent.

**2. Storage**

* **Type:** SSD (gp2/gp3 in AWS, Premium SSD in Azure, Persistent Disk in GCP).
* **Size:** 20 GB per VM (expandable if required).
* **Performance:** Sufficient baseline IOPS (e.g., ≥3000 IOPS for gp3 or equivalent).

**3. Cost Optimization**

* Support for **Reserved Instances / Committed Use Discounts** for long-term usage.
* Option to leverage **Spot / Preemptible Instances** for non-critical workloads.
* Cloud-native budget alerts & cost dashboards.

**4. Scalability & High Availability**

* **Multi-AZ Deployment:** VMs should be deployable across multiple availability zones for resilience.
* **Future Readiness:** Support for **Auto Scaling Groups (or equivalent)** to dynamically adjust VM count based on demand.
* **Disaster Recovery:** Ability to configure scheduled backups and VM snapshots.

**5. Concurrency & Workload Distribution**

* Each service will be deployed on a **dedicated VM**, ensuring:
  + **Isolation** – faults in one service do not affect others.
  + **Reliability** – independent performance and stability.
  + **Natural Concurrency** – services can operate in parallel without contention.

**5. Monitoring & Management**

* **Monitoring:** Cloud-native monitoring of CPU, memory, disk, and network metrics**.**
* **Logging:** Centralized log collection and integration with native log analytics tools.
* **Automation:** API/CLI support for provisioning, scaling, and lifecycle management

**Other Services (Future Consideration)**

* No additional services required at present.
* In the future, we may implement **Auto Scaling Groups** (or equivalent in other cloud providers) to dynamically adjust the number of VMs based on workload demand.