

Vorta Load Balancing - Fundamentals

Objective

Load balancing is done to minimize downtime, and maximize uptime. This is achieved by making all services available on all 6 nodes.

Each node has unique **IP**, and each module has unique **Port**, the combination of **IP + Port** forms a **Service**. Then **backend** service determines which **Service** is accessed by distinguishing the **IP + Port** combination.

In general, the main two services are **API (6010)** and **Dashboard (6011)**, through these services, the **module** can be accessed.

All **modules** are present across all **nodes**, the **modules** are the following:

- SDMS
- JDAA
- E-Form
- Iskrim
- Dokrim

The **Load Balancer** works by pointing access to available **nodes**, this is done by **redundancy**, making all **services** and **modules** available for each **node**.

Therefore, if a **node** is down, the **Load Balancer** points the request to other, available **nodes**.

Round Robin Load Balancing Concept

Vorta does this by using **Balance + Round Robin** load balancing concept, **Balance** distributes services equally to all nodes, while **Round Robin** listens for disturbances in each node, then picks the best node to be accessed by the user.

What happens in an event of LB downtime?

User has to manually access the service by using **origin IP and Port**

The Elephant in the Room: Load Balancer

Load Balancer itself has a unique **IP + Port**, therefore, it is also a **service**. **LB** eliminates the need to access individual **nodes** manually, this is possible because each **node's** IP address is registered and recognized by the **LB** but not to the **client**.

When **client** requests access to a **service**, the **LB** processes the request, and points the **client** to an available **node**.