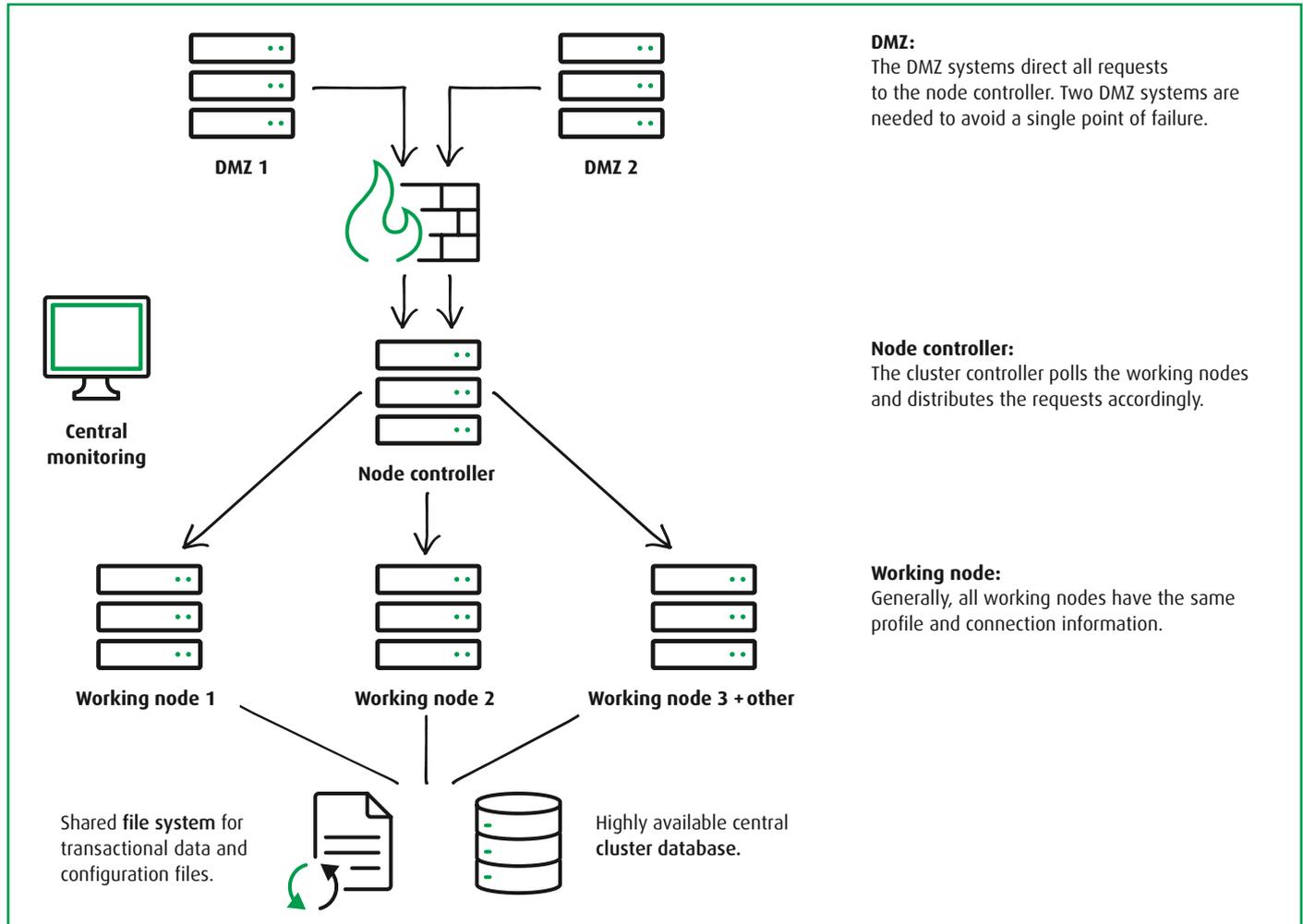# Lobster

OVERLOADED? AS IF!

LOAD BALANCING WITH LOBSTER.
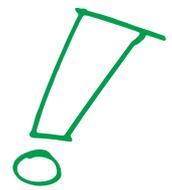
ADVANTAGES OF A LOAD-BALANCING SYSTEM

High availability: Working with a load-balancing active-active set-up increases reliability. Should problems occur in one system, then any outstanding jobs can simply be taken over by another system – thus preventing delays.

Scalability: A load-balancing system is able to distribute the load between different systems when processing very large volumes of data. However, it's worth noting that Lobster's software is so powerful that this is rarely necessary. For example: 2TB of data volume and 22 million data requests generally mean that a server with 16 cores and 48 GB of main memory shouldn't go above 30% utilisation.

INFRASTRUCTURE SET-UP

DMZ 1

DMZ 2

Central monitoring

Node controller

Working node 1

Working node 2

Working node 3 + other

Shared **file system** for transactional data and configuration files.

Highly available central **cluster database.**

DMZ:
The DMZ systems direct all requests to the node controller. Two DMZ systems are needed to avoid a single point of failure.

Node controller:
The cluster controller polls the working nodes and distributes the requests accordingly.

Working node:
Generally, all working nodes have the same profile and connection information.

LOAD BALANCING – YOUR QUESTIONS ANSWERED.

HOW DOES AUTOMATIC FAILOVER WORK IN AN ACTIVE-ACTIVE SET-UP?

If one of the working nodes fails, then the remaining working nodes take over the load.

If the node controller fails, then one of the working nodes takes on the role of the distributing node controller.

WHAT IS THE MINIMUM NECESSARY HA SET-UP?

In principle, you can start with one DMZ, one node controller and one working node. However, two DMZ systems are needed to avoid a single point of failure and two working nodes also make sense to ensure that after the failure of one node, there are still one node controller and one working node remaining. So the ideal minimum set-up includes two DMZs, one node controller and two working nodes.

HOW DO I INCORPORATE MY CURRENT LOBSTER SYSTEM INTO A HA SET-UP?

You can transfer your current licence to the HA set-up. Essentially, this means you've already licensed a node (and a DMZ). It's usually a good idea to use your current productive system as the node controller within a HA set-up. You don't need a load balancer on your test system, so you'll be able to set up a three-tiered test system (test, QA, production).

WHEN DO I NEED A TCP/IP BALANCER?

In case a working node assumes the role of the node controller, the IP address doesn't move over as well. If you're working with direct (internal) connections that rely on an IP address, with Lobster_data acting as a server, you will need a TCP/IP balancer. The TCP/IP balancer forwards requests to the DMZ. As the working node logs on to the Lobster DMZ systems as the new master, these systems can then simply forward any requests. This is irrelevant in the case of external connections, however, as the DMZ is responsible for forwarding the data.

WHAT DO I HAVE TO CONSIDER WHEN CONNECTING INDIVIDUAL SYSTEMS WITHIN A HA SET-UP?

TCP/IP communication should be enabled between all instances and it should be a high-speed, efficient connection. The working node ascertains the operational readiness of the node controller in a matter of milli-seconds.

The performance of the entire system is wholly dependent on its I/O performance. Additional times can be recorded and analysed using the system property hub. datawizard.traceIO=true.

WHAT ARE THE HARDWARE SPECIFICATIONS FOR THE NODE CONTROLLER?

If the node controller is only needed to forward requests, then a basic hardware set-up (comparable to your current DMZ) is sufficient. However, if the node controller is used to process a portion of the jobs directly (e.g. the node controller processes all documents up to 10KB), then we would recommend a hardware set-up that is similar to that of the working nodes (e.g. 4 CPU cores, 8GB RAM).

DO ALL SYSTEMS HAVE TO BE ON THE SAME RELEASE LEVEL?

Although the working nodes should ideally all be on the same patch level, it is absolutely essential that they be on the same release level.

WHAT DO I NEED TO KNOW WHEN OPERATING A HA SYSTEM?

Profiles are always transported from the test system to the productive node controller. So, in order to monitor all processes comprehensively and independently of the responsible working nodes, the user also needs to log on to the control centre of the node controller.

HOW DO I KNOW IF MY DATABASE/ MY DISK SPACE IS OVERLOADED?

Lobster can send warning messages, should this happen. If necessary, the system can also be configured ahead of time so the Lobster_data Server purposely shuts down in order to avoid a system overload.

HOW ARE THE SYSTEMS MONITORED?

Either via HTTP requests (e.g. with Nagios) or via SNMP Traps. Simple Network Management Protocol (SNMP) enables a central station to monitor network elements selectively. In this way, error messages from Lobster _data jobs can be sent via SNMP Traps or the status of services can be updated via SNMP Traps.

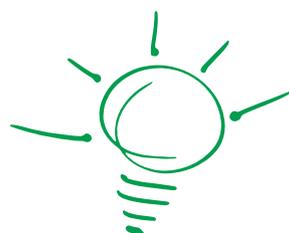## DOES LOBSTER OFFER OTHER SOLUTIONS FOR OPTIMISING MY PROCESS LOAD?

Every job/profile run is associated with a certain overhead (about 120 ms per profile run): address memory, open thread, close thread, free memory, etc. Although this overhead may be negligible when processing smaller job volumes, the Lobster Process Load Optimisation Module (PLO) can be extremely helpful when dealing with, say, ten million jobs per day – as it reduces this overhead to about 3 ms per profile run. For a personalised consultation, please send us your Lobster Statistics Module analysis.

Lobster