



Importance of High Availability in ESB

High Availability for the ESB

The enterprise service bus (ESB) has a foundational position within the organization's application architecture, and as such has both functional requirements (e.g., message routing, data transformation, orchestration) and non-functional. Often overlooked during the buying process, the non-functional requirements such as manageability, reliability, and availability, usually end up defining the effectiveness of the ESB and architecture as a whole.

High availability (HA) is paramount to an ESB because of its critical position and open design. A key strength of an ESB is the ability to interact with multiple cross-format applications; however, this strength is also creates multiple touch-points that can affect availability. A mission critical ESB requires a well-designed HA solution that can protect and maintain uptime. In this article, we examine the importance of high availability in the enterprise software stack and the benefits of HA configuration in Mule ESB.

Cost of Dismissing HA

Few circumstances have a broader impact on the enterprise than unplanned downtime. In early 2012, Emerson Network Power released research that reported businesses lose an average of \$5,000 per minute of downtime (\$300,000/hr.). This report is one of the first to include both direct costs (lost customer sales) as well as indirect costs, such as missed opportunities and decreased confidence of key stakeholders. This study illuminates the high cost of even brief system failures, and serves as a reminder that financial loss is not the only cost of downtime.

Without a business continuity and disaster recovery plan, downtime can be unapologetic, causing everything from minor hiccups to devastating crashes across critical applications. In one instance, Virgin Blue Airlines' (later renamed Virgin Australia) online booking system went down – and stayed down for 11 days. It cost the company \$20 million, in addition to negative press and a fury of protests from customers via social media.

Without a proper failover system, the cost of downtime can be crippling – It doesn't take 11 days to cause serious damage. High availability is no longer a perk reserved for financial institutions; it is now a fundamental requirement for any system architecture.

Establishing High Availability by Clustering

High availability is established in an ESB by deploying

clusters: components responsible for taking over the responsibilities of an application if one or more fail. Instead of waiting for a crash to resolve, current transactions are automatically transitioned from the crashed component to the backup (often referred to as a failover node) until the crash is resolved.

The simplest cluster is a single additional node that is responsible for supporting the entire suite of applications that interface with the ESB. A good starting point, this minimalist approach is often insufficient for established enterprises. These organizations will need to undergo a cost-benefit analysis to determine the optimal number of nodes for their architecture.

For companies that require help with this process, Mule has award-winning professional services and support that can work hand-in-hand with development and architecture teams to evaluate business needs and integration challenges, and recommend an efficient form of clustering.

High Availability Configuration

HA nodes are commonly configured in one of two ways, each method providing their own unique strengths and weaknesses.

Active/passive – Each node has a redundant instance that is only brought online when the primary node fails. This 1:1 configuration ensures each application has a dedicated backup, but requires extensive hardware.

Active/active – Traffic intended for a failed node is either passed onto an existing node or load balanced across the remaining nodes. This group-oriented approach provides enhanced reliability, better load balancing, and scalability.

Most enterprises find active/active configuration ideal for creating an efficient and cost-effective HA environment, as it makes better use of hardware. However, both configuration approaches have traditionally been cumbersome and complex to set up, demanding valuable time and developer resources.

Mule has simplified this process, providing preconfigured active/active nodes to ensure quick and painless deployment. This enables developers to spend their valuable time improving business processes, instead of diverting attention to failover nodes.

In-memory Data Grid for HA

Aside from cluster configuration, nodes can be organized in

different data models that affect how information is accessed and stored. An in-memory data grid distributes information across multiple servers (or nodes), to increase availability and reliability of data. This model provides key benefits for high availability clusters:

Performance – Data resides in-memory, making access quick and always available. Latency periods are lower when transitioning from a failed node to a backup.

Scalability – Easily scale your nodes up or down without disruption to the data grid.

Reliability – Information is distributed across the grid, making it both more reliable and safer in times of failure.

Mule ESB's HA uses an in-memory data grid that is deployed by companies with transaction rates, such as Facebook and Amazon. Unable to afford even seconds of downtime, these companies trust Mule to provide instant information replication. This same data model is the foundation for all HA deployments, ensuring every organization a highly reliable service.

Advanced High Availability, Simplified

Mule ESB Enterprise provides reliable, scalable and easy to manage clustering to ensure high availability for all mission critical applications. Using an active/active model, it ensures continual service without lose of scalability. (Note: Mule High Availability is available only with Platinum Support subscriptions).

Mule clusters are built on Mule's in-memory data grid, enabling all instances to run the same applications and load balance between them while maintaining transient state information. Should any Mule instance experience an outage, the other nodes immediately pick up the load with no service interruption.

Mule clusters are easy to manage for IT administrators, who can treat them as logical units from Mule's web-based management console. This familiar interface enables clusters to be managed as a single unit or per node, much like server groups.

Lastly, with Mule you are in good company. Mule is trusted to keep applications available in thousands of production deployments by leading organizations such as Walmart.com, Nokia, Nestlé, Honeywell and DHL, as well as 85 in the Global 500 and 5 of the world's top 10 banks. It powers mission-critical applications responsible for massive revenue streams for organizations ranging from major airlines to global banks.