



**MetalSoft Whitepapers -
RDMA Accelerated
Infrastructures**



MetalSoft Whitepapers - RDMA Accelerated Infrastructures

MetalSoft enables NVIDIA to offer a compelling end-to-end integrated, servers, storage and network solution to datacenter operators. MetalSoft is a bare metal management and provisioning solution that leverages the performance of bare-metal accelerators to build an easy to use, highly secure, multi-tenant, multi-datacenter bare metal private cloud.

Contents

MetalSoft Whitepapers - RDMA Accelerated Infrastructures	1
Background	2
RoCE-capable SDN	2
RDMA-ready diskless operation.....	2
RDMA-enabled Kubernetes clusters.....	3
GPU support.....	3
Enabling as-a-service consumption of RDMA-based solutions.....	3

Background

While RDMA has been around for many years especially in the HPC field, it has never achieved large scale acceptance for general purpose datacenter workflows due to a number of reasons including lack of support for Ethernet and sensibility to bursty environments.

For over 10 years now we have seen an steady increase in horizontally scaling of general purpose applications. Microservices orientated architectures, NoSQL databases and the increasing use of AI and Kubernetes all promote the use of many small components that talk to each other as opposed to a few big ones for a more resilient infrastructure, more efficient use of resources and easier management. Yet another nascent trend is serverless with even more disaggregated components called “functions”.

RDMA is clearly a solution to this need for lower latency but, in order to function properly in a general purpose datacenter it requires the the NIC, the switches and the operating system need to collaborate and require special configurations making RDMA difficult to use for regular enterprises or service providers.

RoCE-capable SDN

MetalSoft’s solution automates bare metal compute network and storage enabling seamless, cloud-like on demand, self-service consumption completely hiding the complexity of RDMA but offering it’s benefits to end-users. The key component to this is our Bare Metal SDN which supports secure multi-tenancy by using switch-supported and hardware accelerated VLAN, VPLS, eVPN based tenant isolation technologies. These capabilities allow MetalSoft to build end-to-end RoCE networks:

- **multi-vendor support** including Cumulus Linux enable heterogenous environments
- **RoCE QoS** support via **DSCP** configuration
- **Automatic configuration o NICs** (incl. smartNICs) via BMC, (EFI)BIOS or special tooling to enable virtual functions (SR-IOV) or other advanced functionalities depending on the end-user’s use case.
- **Policy based automatic firmware upgrade** enable clients to stay-up-to-date with the latest software

RDMA-ready diskless operation

MetalSoft's **diskless operation capabilities** allow instant provisioning, fast node replacement, stop-and-resume capabilities, full OS-and-Data snapshotting and more. This feature, coupled with an RoCE based NVMe-over-Fabrics enabled storage can provide unparalleled performance and flexibility to both regular hosts and also kubernetes persistent volumes via a MetalSoft Kubernetes CSI plugin.

RDMA-enabled Kubernetes clusters

MetalSoft's built-in integration with Kubernetes enables special provisioning of RDMA capabilities that are then used by the users without having to worry about complex configurations but simply using the cluster as any other kubernetes cluster.

- **Automatic configuration of RoCE SR-IOV** virtual functions in NICs
- **Automatic configuration of kubernetes** to use the SR-IOV CNI plugin
- **MetalSoft CSI** provider enables persistent volumes that can also utilize NVMe-over-Fabrics over RoCE or accelerated iSCSI.

GPU support

Native **GPU detection** enables a few hosts to be provisioned with GPUs and then allocated dynamically among users as needed. For example one of the most used use cases for this feature is ML training. The GPU enabled servers are allocated to a certain user for a number of hours until training is complete afterwhich the nodes get back to the pool.

Enabling as-a-service consumption of RDMA-based solutions

In summary, by using MetalSoft, end-users of Enterprises, or even service providers can benefit from ultra-high performance networking and storage enabled by RDMA without having to have the resources to build and manage this complex technology.