Leading 5G Innovations

ZENIC vDC Controller, Navigator on Cloud Networks

——A Perfect Combination of Elasticity and Intelligence

ZTE

ZENIC vDC Controller

—— A perfect combination of elasticity and intelligence

Being open programmable network system architecture powered by separated user/control architecture and centralized management and control, software defined networks (SDNs) nowadays have become more and more popular among data center networks. The ZENIC vDC controller released by ZTE is a carrier-class controller concentrating on the SDN demands of cloud data centers. Orienting to multiple application scenarios such as private cloud, public cloud NFVI and hybrid cloud, the ZENIC vDC controller designed with a service orchestrator and cloud platform provides carrier-class vDC networks with end-to-end SDN solutions.

The ZENIC vDC controller supports distributed controller clusters. The system is actually a 2+N multi-chassis system consisting of two active controller nodes and N (1-128) service controller nodes. The active controller working in an active/standby mode allows better reliability. Southbound interfaces are deployed on the service controller.

The basic function of the ZENIC vDC controller is as show in the following：

Architecture	**Dynamic component uploading and unloading based upon dynamic libraries;** **Support distributed controllers. Maximally 128 controller nodes are supported.** **Support open APIs. C++/Java and RESTful interfaces can be used for programming.** **Support Java programming interfaces which are compatible with ODL.**
Kernel & applications	Support switch access, topology computing and ECMP. Support address learning, ARP/ND reply and multiple QoS services such as Queuing and Meter services. Support Overlay and non-Overlay networking modes. Support basic forwarding decision services. Support embedded DHCP services. Support embedded ACL and traffic engineering. Support VPC applications and migration of virtual devices. Support interconnecting DCIs.

ZTE

Architecture	**Dynamic component uploading and unloading based upon dynamic libraries; Support distributed controllers. Maximally 128 controller nodes are supported. Support open APIs. C++/Java and RESTful interfaces can be used for programming. Support Java programming interfaces which are compatible with ODL.**
Operation & management	Support Web-based network management interfaces. Support NML configuration parameters. Support visible topology and traffic. Support failure observation, alarms and end-to-end traffic diagnosis.

Customer value

Rapid deployment of tenant service

Service on-line time:

week—>hour

Simplification of O&M

Fault location time:

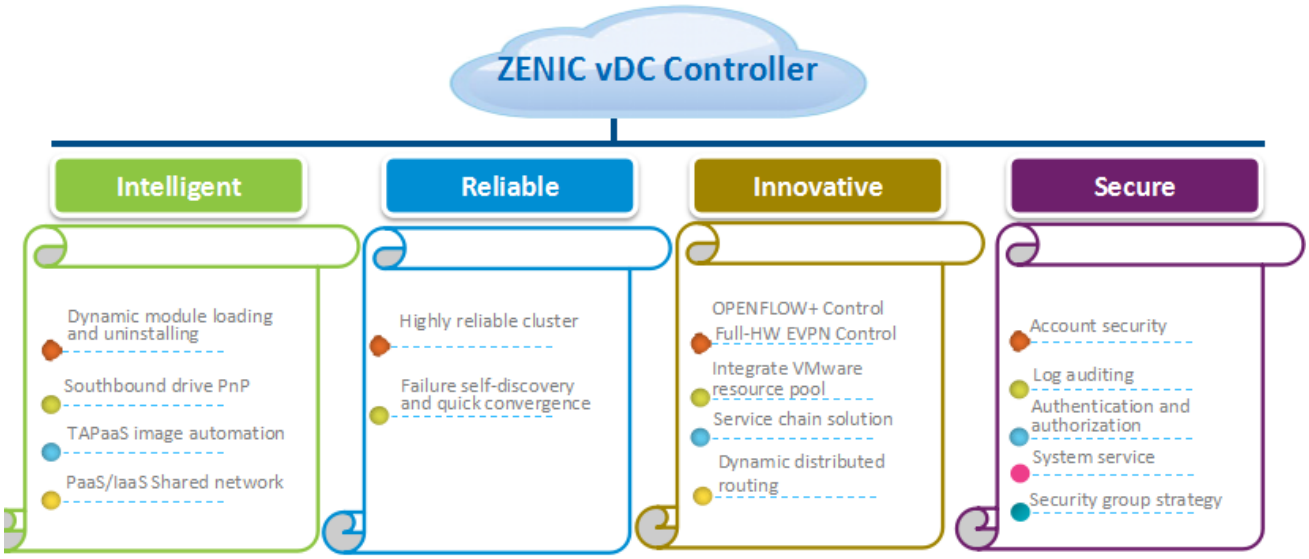
hour—>minute

Dynamic Adjustment of Massive Strategies

System integration time:

month—>week

Key Features Table



Modular components

Modular components are provided. System uploading and unloading are implemented based on the service modules. Instead of changing source codes, users can tailor configuration as per their demands.

Plug-and-play southbound drivers

Multiple southbound drivers can be uploaded. Accessible to multiple types of switches, the controller allows dynamic uploading and unloading of different switch drivers to offer real plug-and-play services.

TAPaaS Image Automation, Flow Accurate Analysis

Provide TAAS-API interface based on open source, can dock with different manufacturers.Specify the specific traffic collection rules for a particular port from the choreographer by VNF, SDN Network side automatic location flow collection point and distributed flow collection and distribution strategy.When the virtual machine is migrated, created, deleted, and bounced, it does not need to manually modify the traffic acquisition strategy, and the SDN network automatic adjustment strategy is fitted.

Can be based on the five-tuple selective collection of traffic, and in accordance with the actual needs of different characteristics of the flow can be sent to different purposes to receive equipment.

Highly Reliable Distributed clusters

Distributed controller clusters are supported. The controller cluster consists of a pair of active controller and 1-128 service controllers. Dynamic node expansion is supported. The active controller in an active/standby mode ensures proved reliability. The southbound interfaces are distributed on the service controller.

All configuration data and physical topology data are distributed to all the nodes. After being transferred to the active node via the service nodes under inspection, the physical topology will be sent to all the service nodes. Flow tables, host status and virtual network data are all under distributed processing via distributed hash tables (DHTs).

End-to-end fault diagnosis

Link failure discovery and path adjustment can be implemented in seconds according to timeout information generated by link layer discovery protocol (LLDP) and port status reported by the switch. The OpenFlow Management and Configuration Protocol (OF-Config) is used to enable collaboration of

topology discovery and bidirectional forwarding detection (BFD). The expansion of the OpenFlow resulting in faster failure detection finally enables fault convergence within 100ms.

The network management is capable of composing messages. Together with the capabilities of the OpenFlow in controlling flow tables, hop-by-hop message processing can be implemented to locate the failed nodes rapidly.

ZTE "Openflow+" Control Surface

The controller is completely offline and forwarding post remains.ARP and ICMP protocols are unloaded to the forwarding surface.Reduce the cost of the controller.Underlay and overlay Complete separation.

Controller does not participate in fault convergence of underlay network.Different types of SDN switches have different TTP models.Each flow table is mapped to a multistage ASIC assembly line.

Full-Hardware EVPN Control Surface

Cloud/Network separation: physical network control and virtual network separation, network side scheme is very simple, network departments to promote network automation configuration;

SDN only controls the hardware switch, not need to take over vSwitch, not need to install any software on the server;

The cloud department is responsible for computing virtualization, and the OpenStack cloud platform is responsible for VM-level policy and migration policy follow-up;

Dynamic Distributed Routing

Global Unified Control Surface—Based on the SDN controller, provides the overall unified management, the control plane.

Business boards have forwarding capabilities—Through the OpenFlow multistage flow table, each vswitch/hardware switch has the service forwarding ability.

Sync All Business Board key table entries—A global unified controller Mac is used to answer the Gateway ARP, and the controller guarantees the global synchronization of the table items such as Routing and ARP.

Optimize Traffic Path—Forwarding with the switching, traffic is done within the switching, forwarding across the switching, and traffic is done through the VxLAN Fabric network.

Integrated VMware Resource Pool

vSwitch as VTEP: A DVS is deployed on each ESXi, and the flow of this node is DVS and sent to SDN network.Each VM is assigned a separate VLAN and only communicates with the DVS.DVS take over external NIC.

TOR as VTEP: ESXi policy settings Bypass enables the VSS VM to connect to the 5960 switching, using VLANs to differentiate the VM and the controller to distribute the flow table to make the bypass work.Better performance than a pure software solution that hosts vswitch in a VM.

Dynamic insertion of service chains

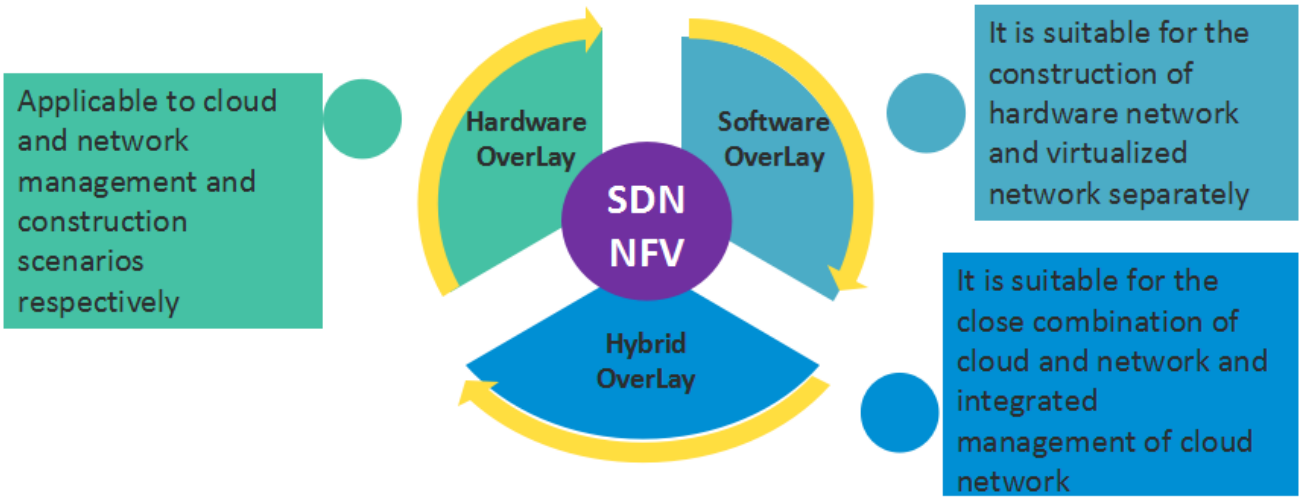
Dynamic service insertion and service bundles made up by L4-L7 services on the service chain can be provided via the traffic steering. Also network elements such as firewalls, IDS/IPS and Cache that offers transparent services can be added to the network. Moreover, the sequence of implementing multiple L4-L7 services can be changed to eliminate conflicting and co-dependent services.

Security Group Policy Controls Traffic Exchange

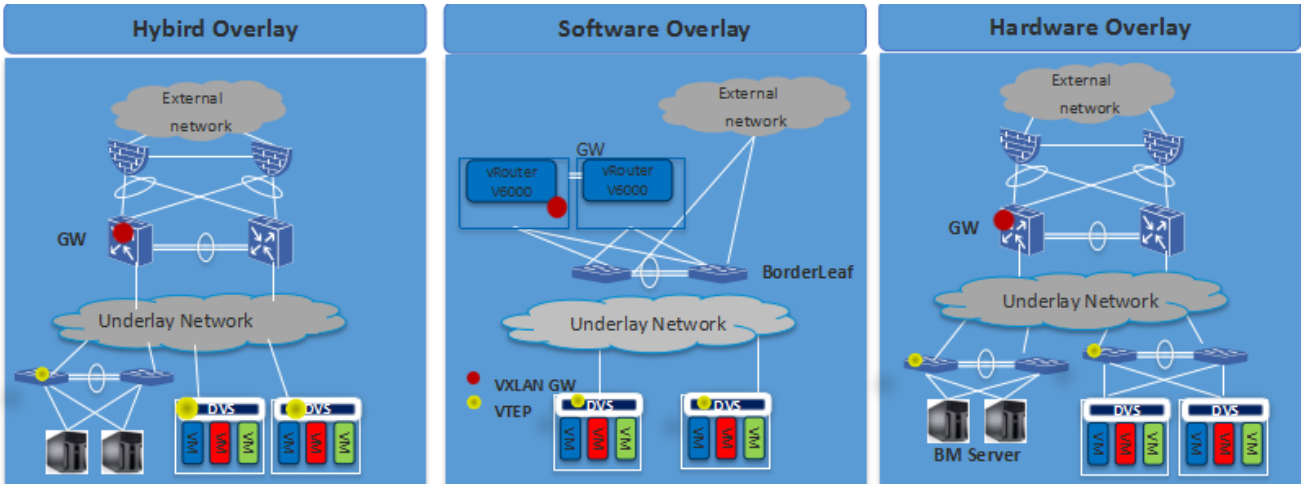
East-West Safety: Micro-segmentation, stateful security group based on Connection track.Cross-tenant, cross-router traffic is isolated by east-west firewall.

North-South Security: North-to-sorth level, each virtual router is bound to a virtual firewall.The external router is connected to the firewall, and the NAT is also implemented by the firewall.

Application Scene



Multiple OverLay Networking Scenarios：



Hyhird Overlay : For KVM compute nodes, DVS acts as VTEP, for bare metal servers and SR-IOV, TOR switches act as VTEP, and hardware devices act as VxLAN gateways.

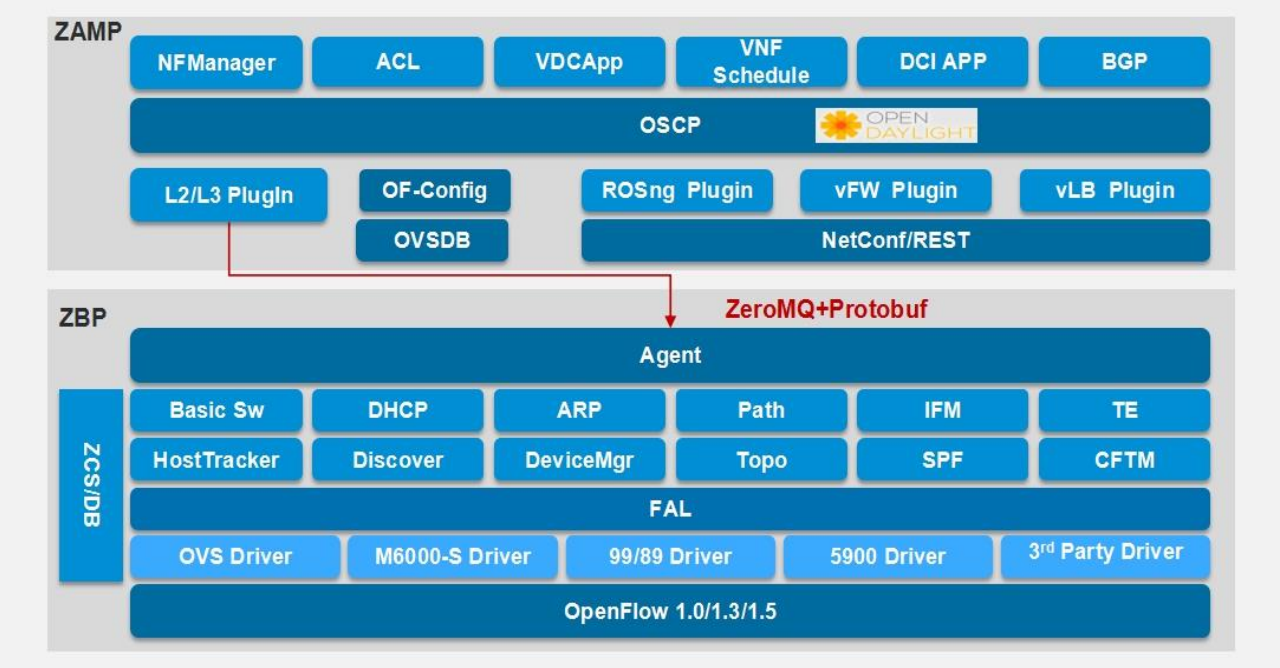
Software Overlay : For KVM compute nodes, DVS ACTS as VTEP, and vRouter ACTS as a software-like VxLAN GW&NAT.

Hardware Overlay : TOR switches act as VTEP for KVM nodes and bare metal servers, and hardware devices (routers, switches) act as VxLAN gateways.

The SDN controller only controls the hardware switch; When there are only bare metal servers in the cloud, the interface of bare metal server has one-to-one correspondence with the hardware network. If it is virtualized networking scenario, between the vSwitch and ToR switch is VLAN networking, between hardware switches is VXLAN networking, which is called hierarchical-port binding. The tunnel is established between the VTEP devices or between the VTEP and the GW devices。The control plane of VXLAN networking for DVS/OVS is OpenFlow protocol, the control plane of VXLAN networking for hardware switch is EVPN protocol ; and VNI, VRF, MAC, and route forwarding are controlled by the SDN controller. The plug-in creates virtual instances for devices such as firewall, load balance, and IPSecVPN, and uses the interface of SDN controller to connect network.

software architecture

The modular software architecture of the ZENIC vDC controller is as shown in the following figure:



The software of the ZENIC vDC controller mainly consists of two parts : ZBP and ZAMP. The specific descriptions of different modules are as follows:

Name		Description
ZBP	ZCS/DB	ZCS: It is a cluster management functional module providing cluster control and data synchronization services based on single-hop DHT.

Name		Description
		DB: It refers to in-memory database
	Driver	Driver here refers to switch drivers adaptive to different hardware and protocols. It contains the drivers of open vSwitches (OVS), ZXR10 M6000, ZXR10 9900/8900 and ZXR10 5900.
	FAL	The forwarding abstract layer provides unified abstract programming models for devices and networks.
	Core Applications	Basic network core services are provided as follows: BasicSW: Implement basic L2 switching services. DHCP: Implement DHCP Relay and reply services ARP: It refers to the host ARP address, location learning and ARP reply. Path: The path management discomposes forwarding decision requests and sends them to topology paths, so that path switchover can be implemented. IFM: It refers to interface management, including the management of physical interfaces, VLAN sub-interfaces and L3 interfaces. TE: It refers to traffic engineering modules. Discover: It is responsible for discovering the host. DeviceMgr: It refers to the management of plug-and-play southbound drivers as well as the management of the resources with switches and interfaces included. Topo: It refers to topology management, including reception of the events obtained through switch uplink and downlink detection and the establishment of the system topology. SPF: It is abbreviated from shortest path first. CFTM: It is abbreviated from flow table management
	OpenFlow	Openflow protocol stacks
	Agent	Interconnect the basic L2 and L3 services between the java domain and the C domain. ZMQ channel encapsulation and decapsulation are provided.
ZAMP	vDCApp	Multi-tenant-based virtual data center is adaptive to OpenStack interfaces.
	DCIAPP	One resource pool consisting of multiple dispersed DCs is adaptive to the OpenStack interface.
	ACL	Implement configuration, generation and distribution of policy routing tables.
	NFManager	It is a network management system responsible for user interface operation and information display.

Name		Description
	BGP	Implement BGP stacking services. Get routes from the peer end or send routes to the peer end.
	VNF Schedule	It enables scheduling of virtual network functions
	OSCP	It refers to the operation and control platform on the java domain embedded in the controller.
	Plugin	The plugin is mainly used to implement the following adaptive management features: L2/L3 plugin: It enables the interconnection of the basic L2 and L3 services between the java domain and C domain. Of-config/Netconf: It enables the management and configuration of physical OFS via the controller. OVSD: It enables the management and configuration of the virtual OFS via the controller. RosNG plugin: It enables the management and configuration of the rosng system of ZTE switch platforms via the controller. vFW/vLB plugin: It enables the management and control of virtual firewalls and load balancing devices via the controller.

Specification

Specifications	Items	Descriptions
Physical Specifications	PC server	Requirements for systems: X86 systems, dual-line 14-core systems (proposed) and E5-2680 or more advanced systems. Hard disk space shall be bigger than 600GB.
	Memory size	Support 32 switches and the memory size≥32G in the Tiny mode, Support 64 switches and the memory size≥64G in the Small mode Support 2048 switches and the memory size≥256G in the Medium mode.
	CPU	Tiny mode，≥4vCPU； Small mode，≥8vCPU；

Specifications	Items	Descriptions
		Medium mode , ≥16vCPU
Performance Specifications	Maximum number of controller clusters	128
	Maximum number of the switches accessing one single controller node	2048
	Traffic distribution rate per controller (PPS)	80K
	Single controller flow table size	12M
	Controller responding rate	<2ms (Time for message transport and flow table distribution)
External Interfaces	Northbound interfaces	RESTful/RESTCONF Universal northbound interfaces VDC APP application interfaces DCI APP application interfaces
	East-west interfaces	Routing protocol interfaces EVPN/MP-BGP interfaces
	Southbound interfaces	OpenFlow1.0/1.3/1.5 NETCONF SNMP CLI RESTful/RESTCONF
Reliability	Mean time to failure (MTTF)	99.999%
	Mean time between failure (MTBF) (Man-made interruptions and system upgrade-based interruptions are excluded.)	<5 minutes

Leading 5G Innovations

ZTE CORPORATION

NO. 55, Hi-tech Road South,ShenZhen,P. R. China
PostCode: 5180057
Web: www.zte.com.cn
Tel: +86-755-26770000
Fax: +86-755-26771999