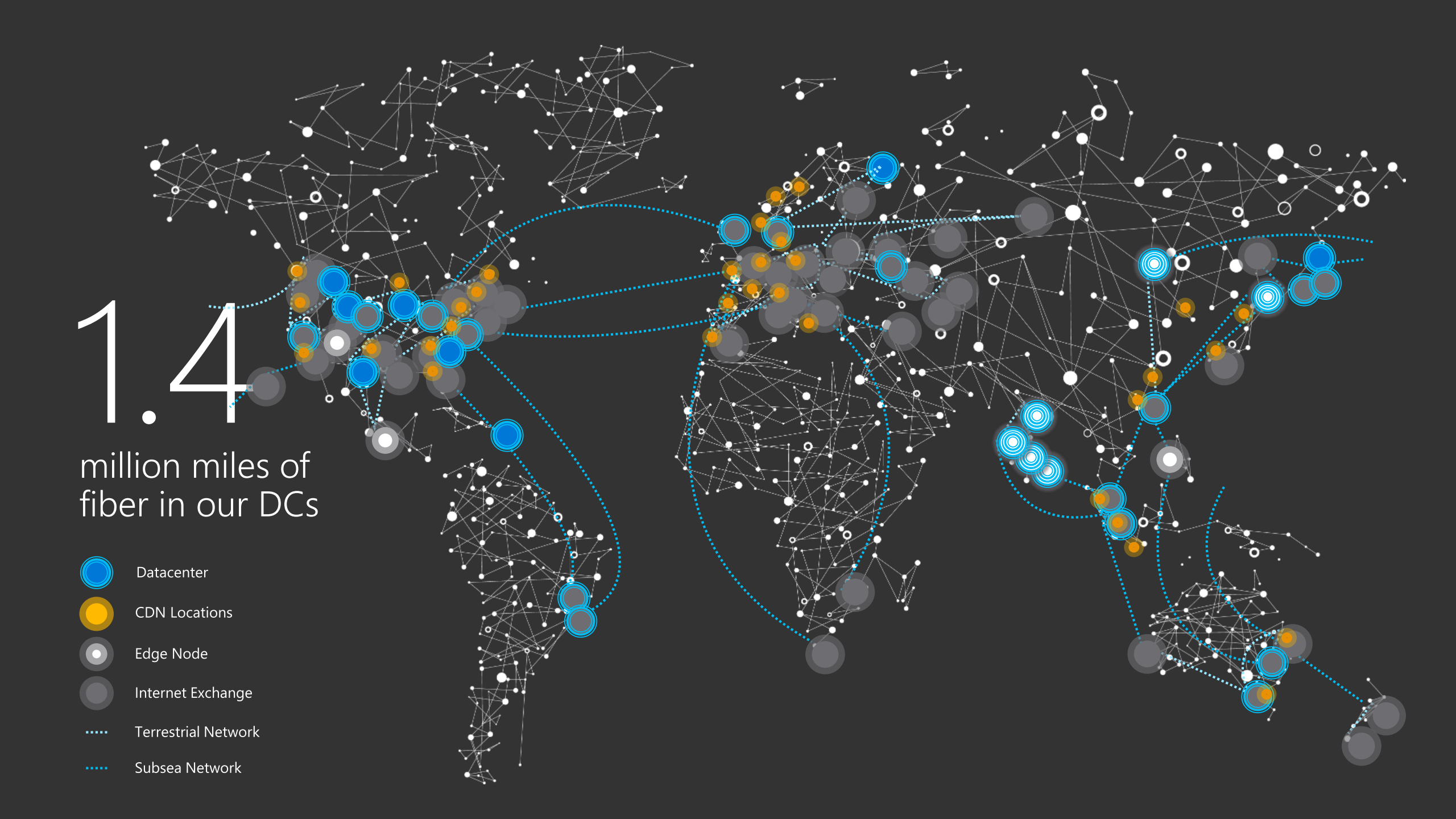
Engineering SDN for Scale

Deepak Bansal
Partner Development Manager
Microsoft Azure
dbansal@Microsoft.com

Azure momentum

>90,000 New Azure customer subscriptions/month	>1.5 Million SQL Databases running on Azure	>5 Billion Authentications per week using Azure Active Directory
2 Trillion Messages per week processed by Azure IoT	90 Trillion Storage Objects in Azure	>40% Revenue from Start-ups and ISVs


1.4
million miles of
fiber in our DCs
Datacenter
CDN Locations
Edge Node
Internet Exchange
Terrestrial Network
Subsea Network

>80%

of Fortune 500 use the Microsoft Cloud

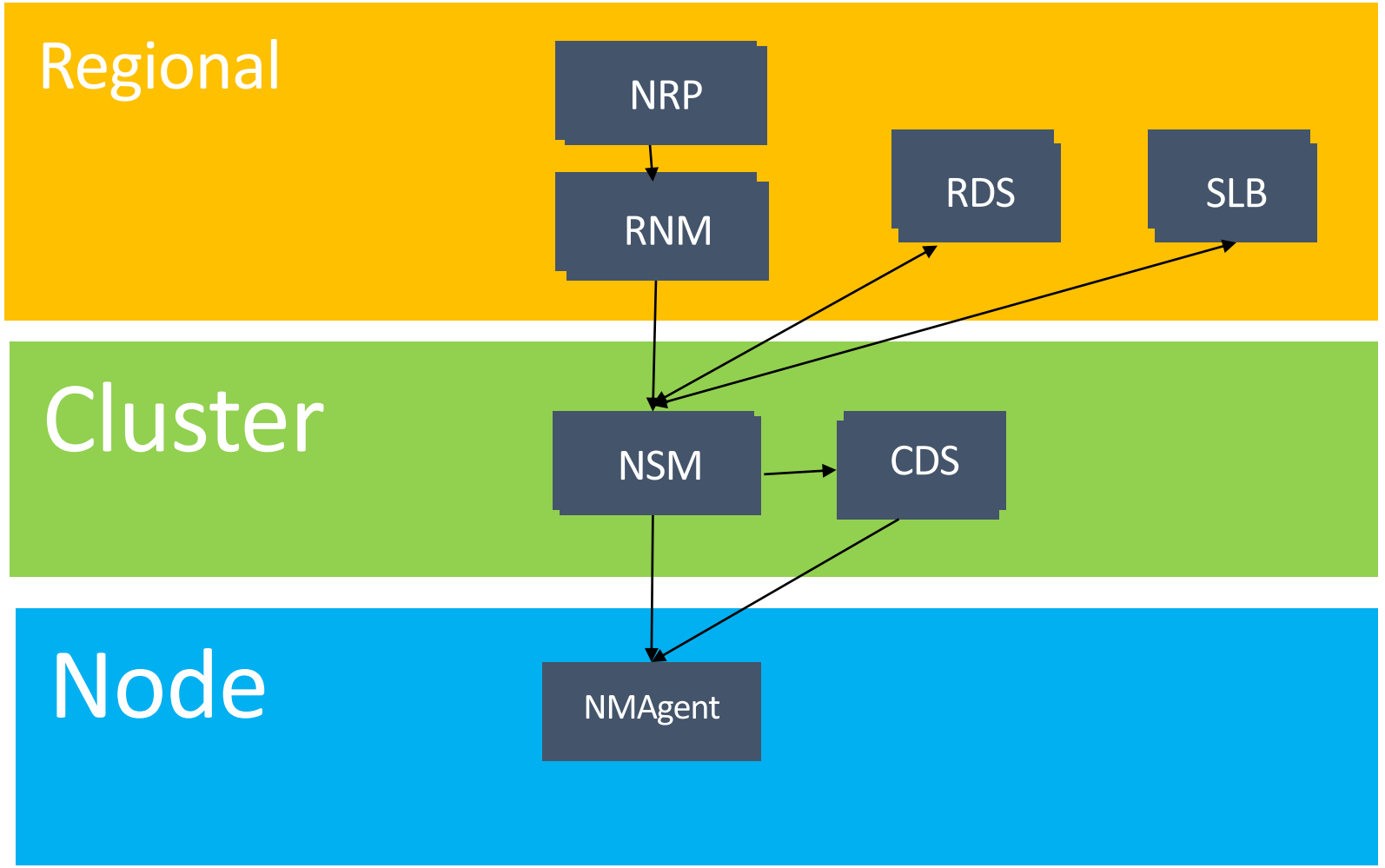
SDN Motivation and Challenges

- **Enable customers to run the cloud services with similar or better network than on-premise**
  - Per customer network, with richness, flexibility, control, isolation and programmability of on-premise network (firewalls, routes, load balancing, DMZ, VPN etc)
  - Virtual appliances in the “virtual network” in the cloud
- **Deploying and managing complex policies on physical devices does not scale**
- **Challenge: How do we deliver and scale virtual networks across millions of servers?**
- **Solution:**
  - Network Controller: Centralized, scalable, highly available, goal state based management of the customer’s virtual networks
  - Host SDN: scalable virtual network data plane via host agents and host drivers
  - Virtual network functions: Provide virtual network services like load balancing and address resolution implemented as services on Azure

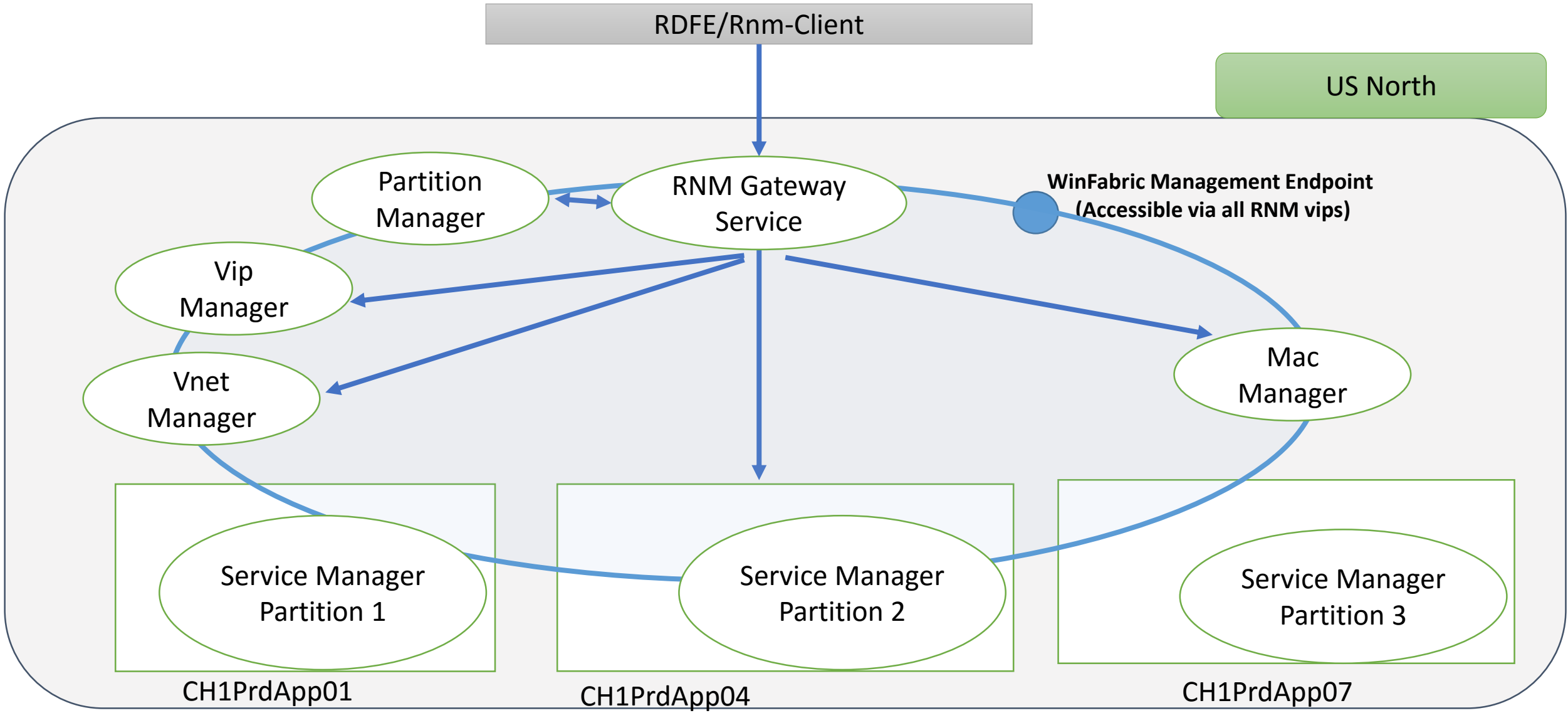
Network Controller

- Challenges:
  - Scale
    - 1DC/1 region -> ~50DCs/17 regions
    - 1000s of virtual networks and network endpoints -> 100,000s -> millions
  - Rate of virtual network provisioning
    - 30min provisioning time -> <5s provisioning time with containers
  - Scope of virtual network
    - Within cluster -> region -> global
- Solution: Scalable and highly available network controller that is
  - Hierarchical
  - Partitioned and regional
  - Micro-services based

Hierarchical Network Controller

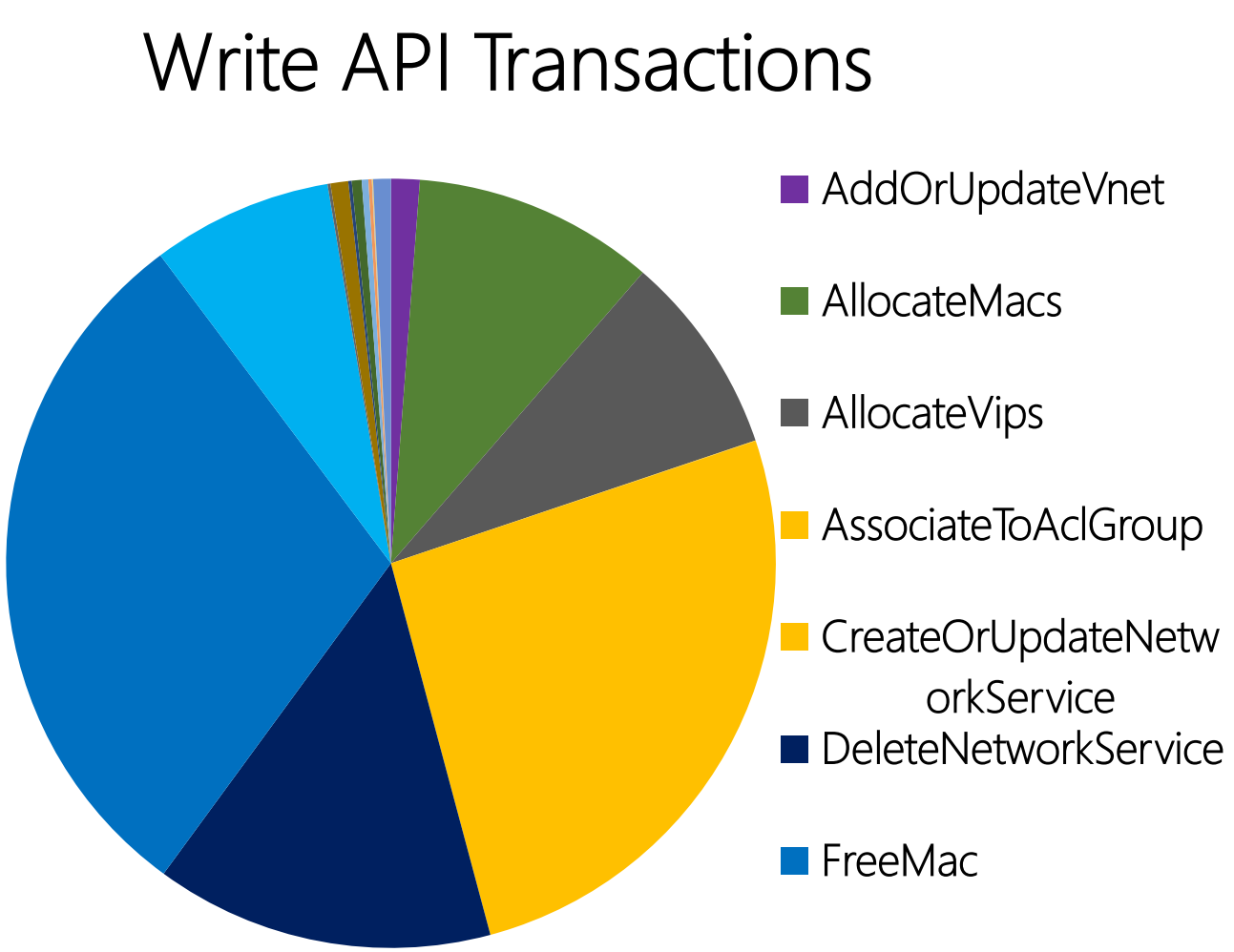


RNM Micro-Service Architecture

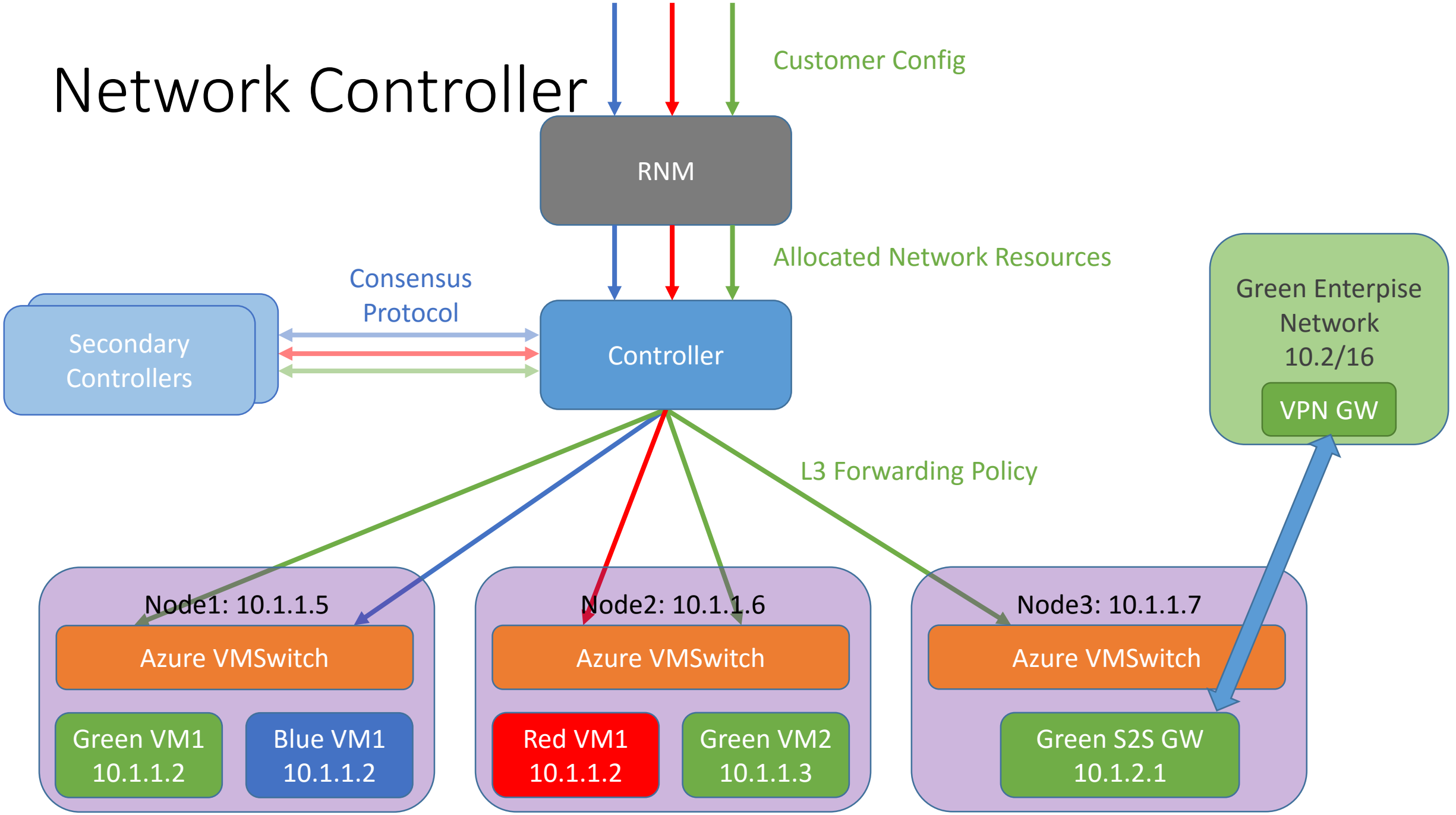


Regional Network Controller Stats

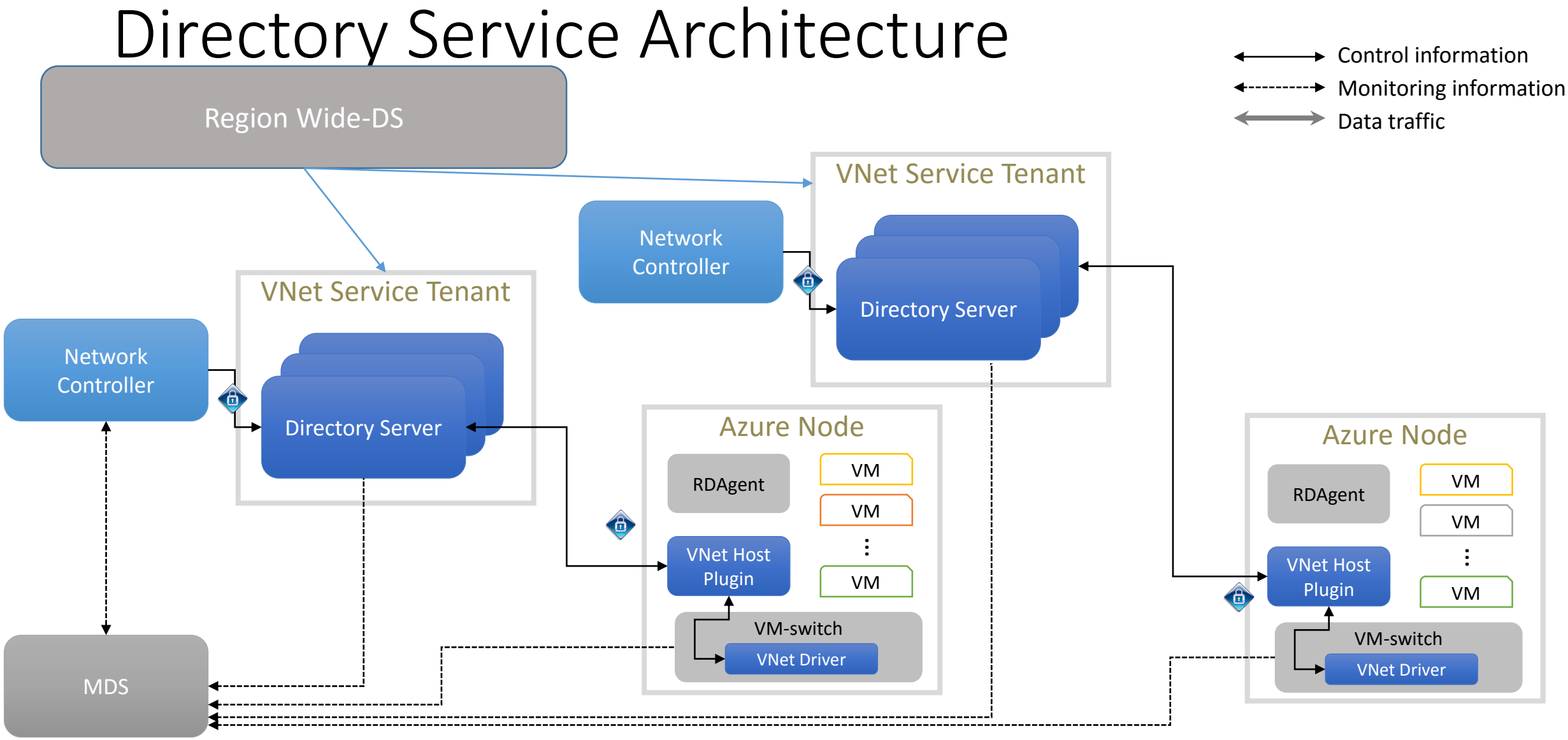
- 10s of millions of API calls per day
- API execution time
  - Read : <50 milliseconds
  - Write : <150 milliseconds
- Varying deployment footprint
  - Smallest : <10 Hosts
  - Largest : >100 Hosts



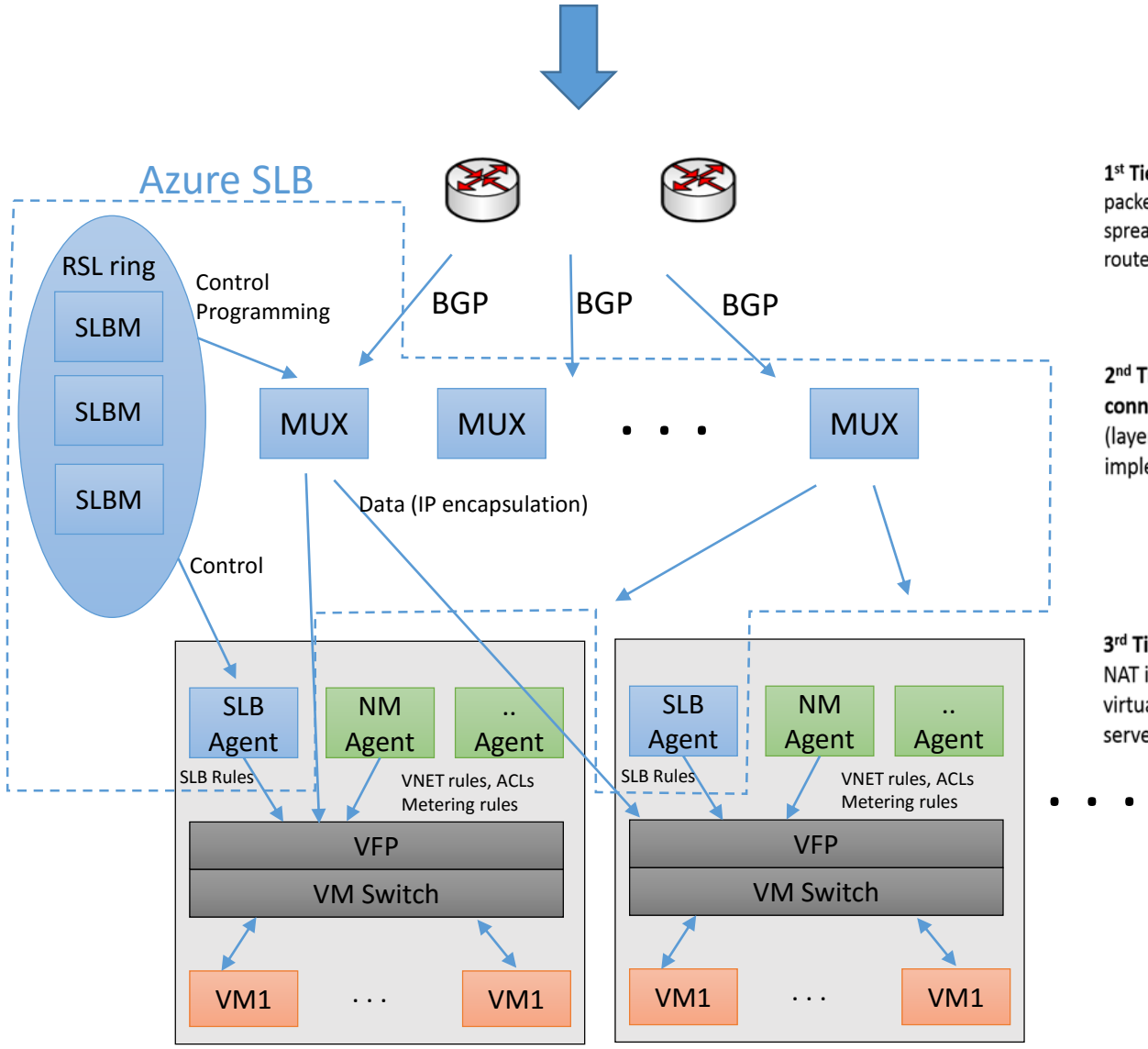
Network Controller



Directory Service Architecture



SLB high level architecture



1st Tier: Provides packet-level (layer-3) load spreading, implemented in routers via ECMP.

2nd Tier: Provides **connection-level** (layer-4) load spreading, implemented in servers.

3rd Tier: Provides stateful NAT implemented in the virtual switch in every server.

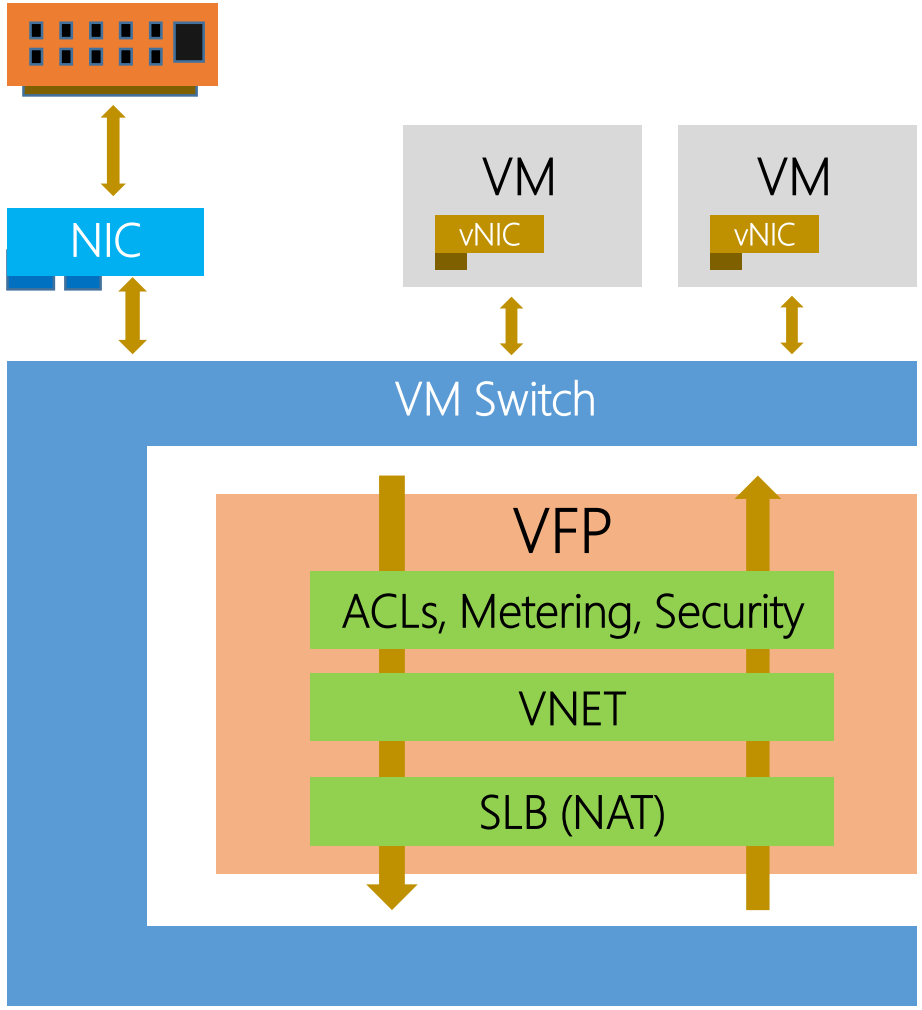
SLB Components:

- SLBM (SLB manager): Control plane that tracks vip dip mappings, health and programs Muxes and hosts
- Mux: Data plane that forwards traffic as well as announce routes
- SLB Host agent: Does NATting

SLB Scale

- 600+ SLBs in production
- 1 Tbps (aggregated) of data
- 17 million of SNAT slot requests
- 1 million total endpoints

Virtual Filtering Platform (VFP)



- Acts as a virtual switch inside Hyper-V VMSwitch
- Provides core SDN functionality for Azure networking services, including:
  - Address Virtualization for VNET
  - VIP -> DIP Translation for SLB
  - ACLs, Metering, and Security Guards
- Uses programmable rule/flow tables to perform per-packet actions
- Supports all Azure data plane policy at 40GbE+ with offloads
- Coming to private cloud in Windows Server 2016

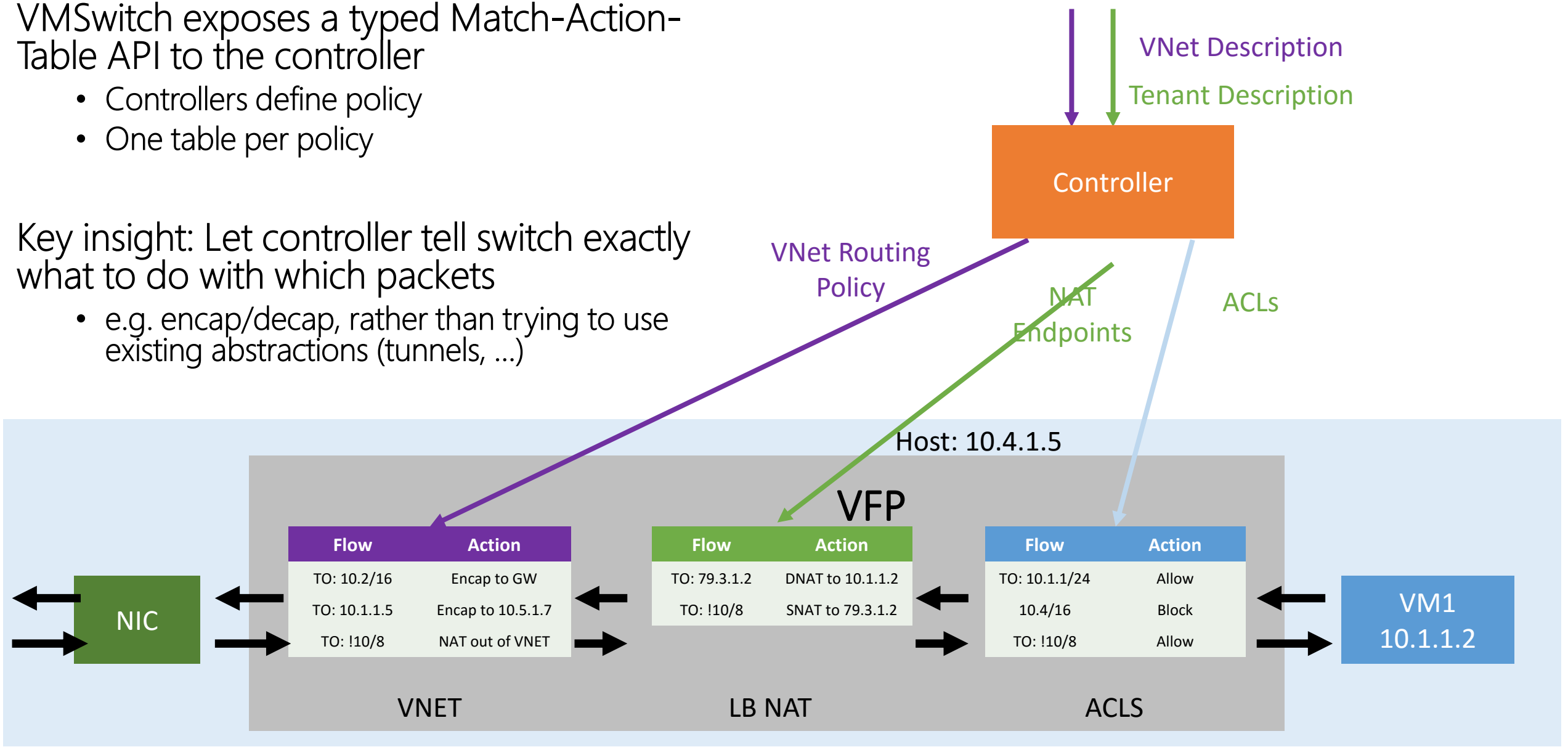
Flow Tables: the Right Abstraction for the Host

VMSwitch exposes a typed Match-Action-Table API to the controller

- Controllers define policy
- One table per policy

Key insight: Let controller tell switch exactly what to do with which packets

- e.g. encap/decap, rather than trying to use existing abstractions (tunnels, …)



We are Hiring!

- BA/BS, MS, PhD
- Interns, Full time positions

{Dbansal, parveenp,umkrishn}@microsoft.com