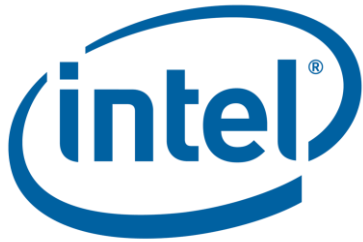


DPDK, Collectd and Ceilometer

The missing link between my telco cloud and the NFV infrastructure



Maryam Tahhan, Intel
Emma Foley, Intel
Carlos Gonçalves, NEC
Ryota Mibu, NEC



Do you know what's happening in your cloud?



Do you know what's happening in your cloud?



Do you know what's happening in your cloud?



Introduction

Medium-/large-scale cloud environments account for between **hundreds and hundreds of thousands of infrastructure systems**.

As the size of infrastructure, traffic and virtual resources grow, so does the effort of monitoring back-ends.

Data growth is exploding across the network. There are more users and devices are connected to the network than ever before. **The underlying expectation remains** that services are available whenever and wherever they are needed; and that those services meet a level of quality that is acceptable to the end user.

It is **vital to monitor systems for malfunctions** that could lead to users' application service disruption and promptly react to these fault events to facilitate improving overall system performance.

A key part of this includes **expanding the amount of data available about the system** (e.g. DPDK statistics), and **improving alarming functionality** in OpenStack Aodh.

The Corner Stone

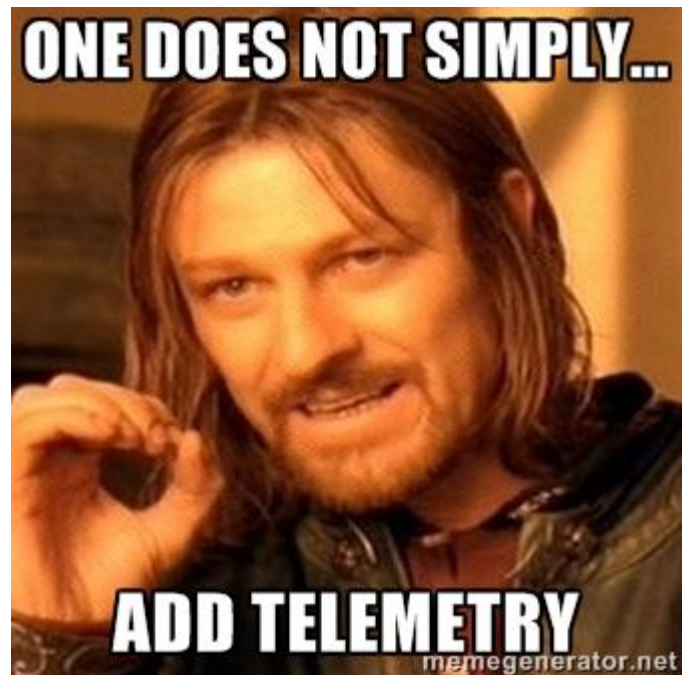
Telemetry is the cornerstone for:

Billing

Benchmarking

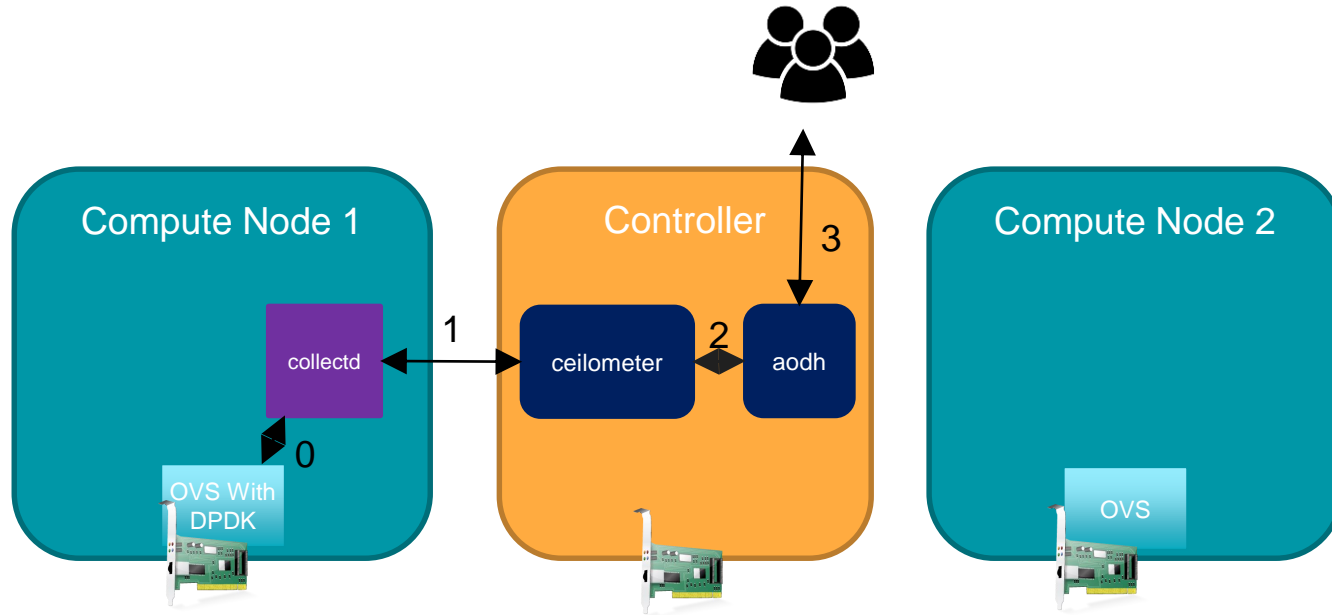
Intelligent orchestration

Fault management



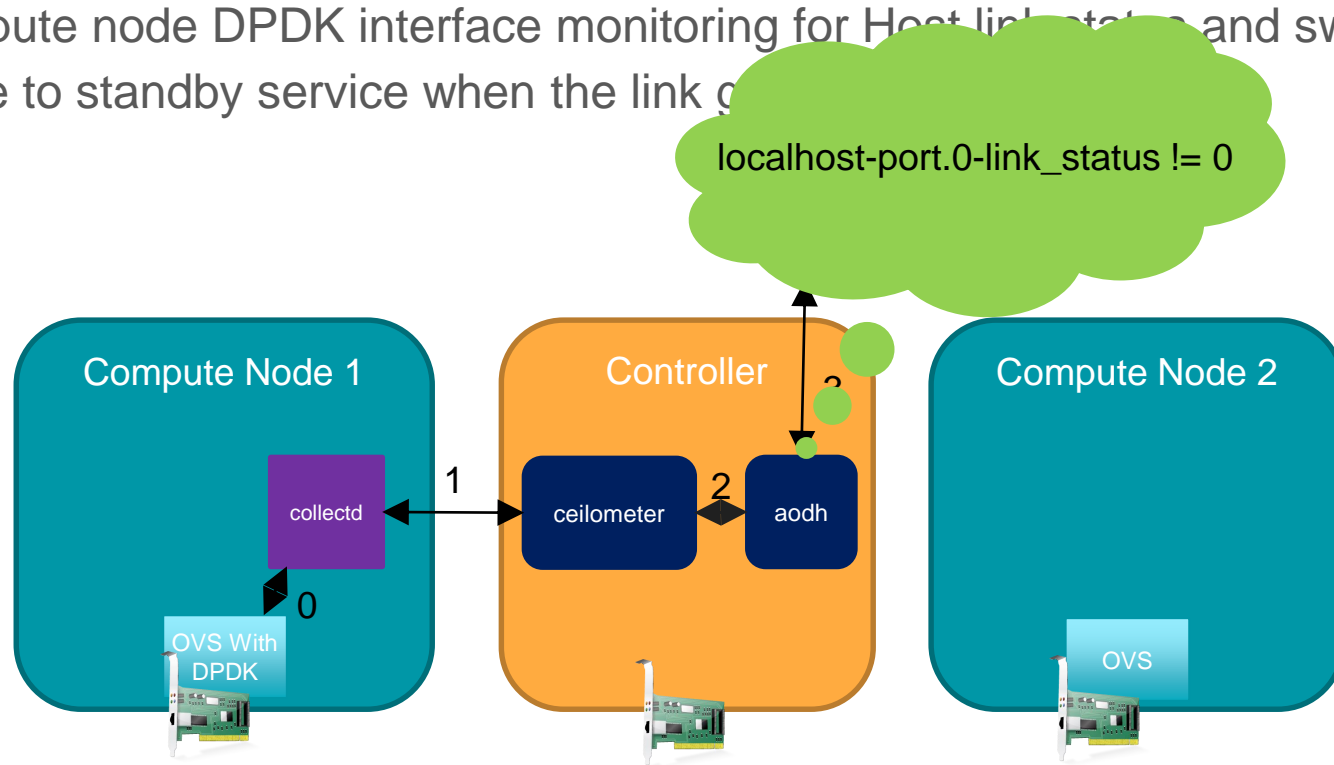
Use case Example

Compute node DPDK interface monitoring for Host link status and switches from active to standby service when the link goes down.



Use case Example

Compute node DPDK interface monitoring for Host link status and switches from active to standby service when the link goes down



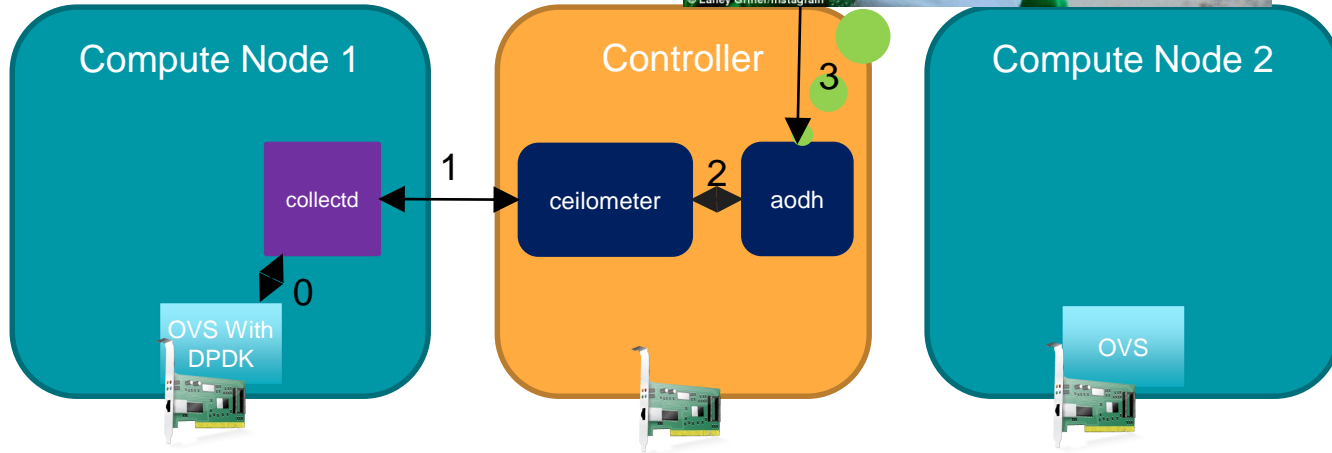
Use case Example

Compute node DPDK interface monitor

active to standby service when the link

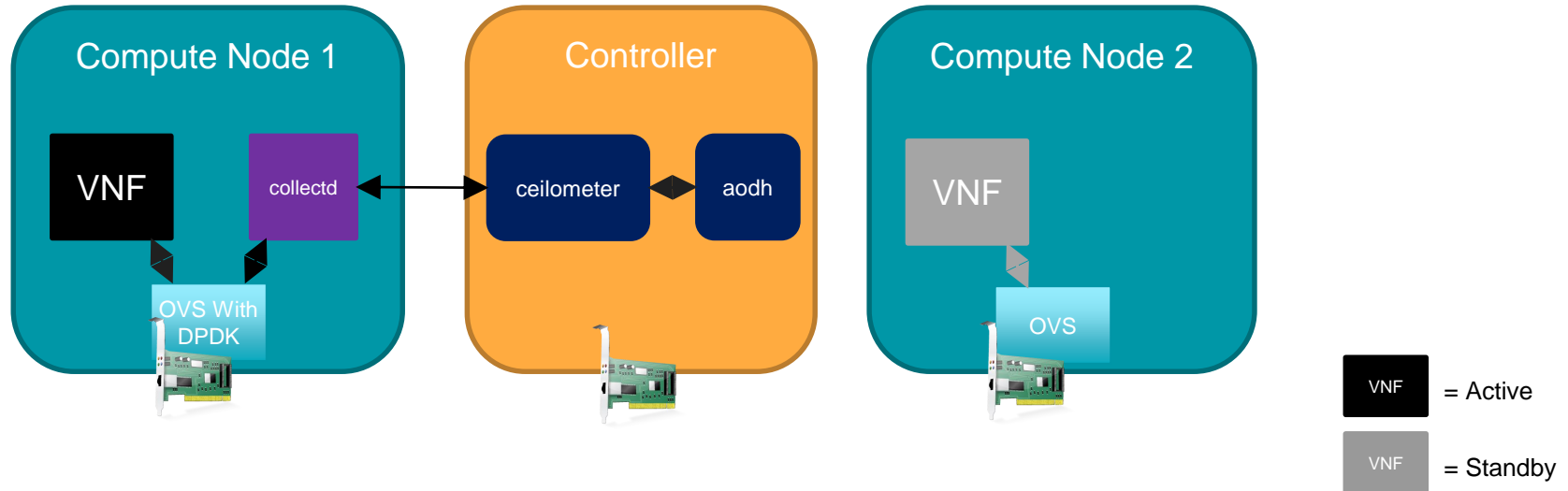


switches from



Use case Example

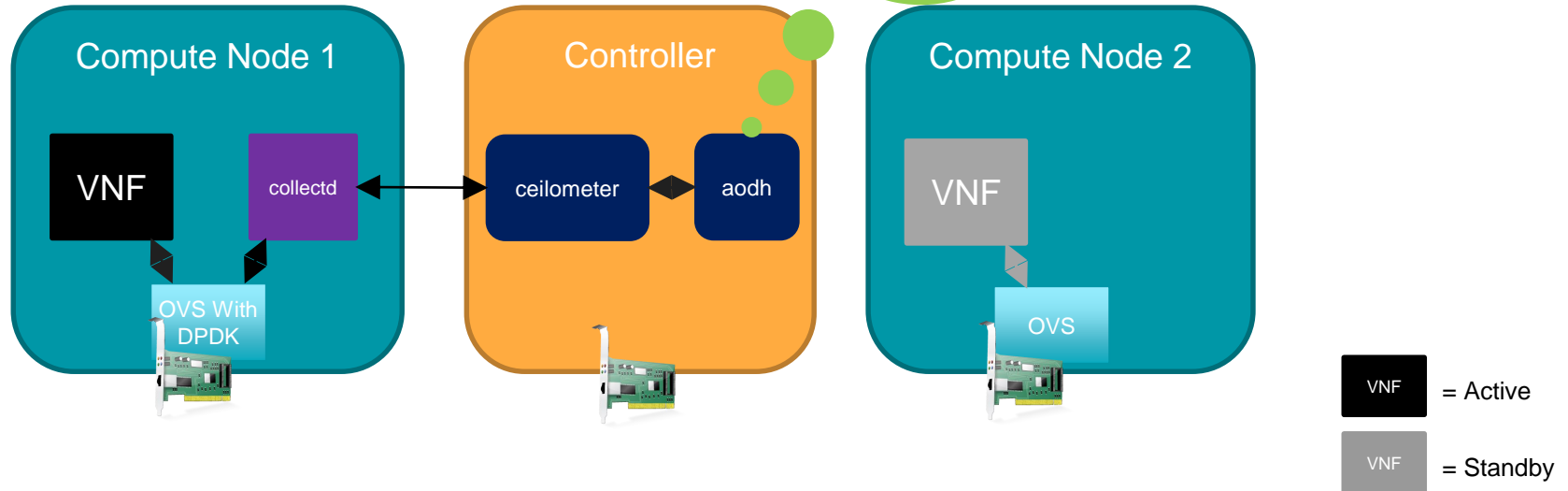
Compute node DPDK interface monitoring for Host link status and switches from active to standby service when the link goes down.



Use case Example

Compute node DPDK interface monitoring for Host link status and switches from active to standby service when the link goes down

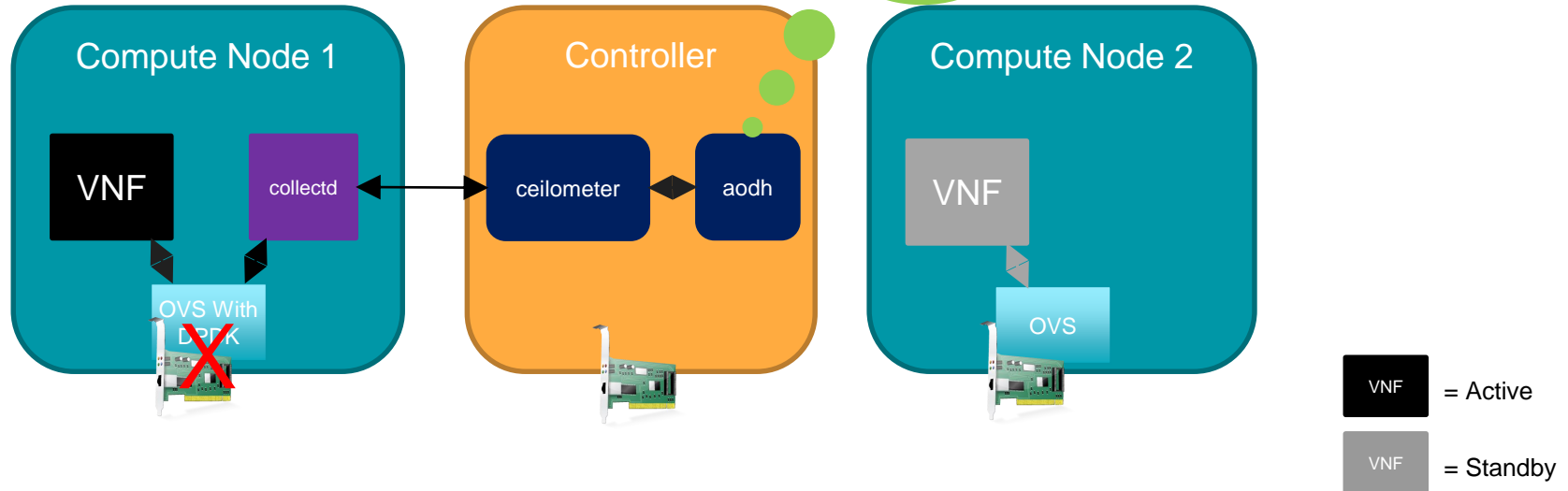
`localhost-port.0-link_status != 0`



Use case Example

Compute node DPDK interface monitoring for Host link status and switches from active to standby service when the link goes down

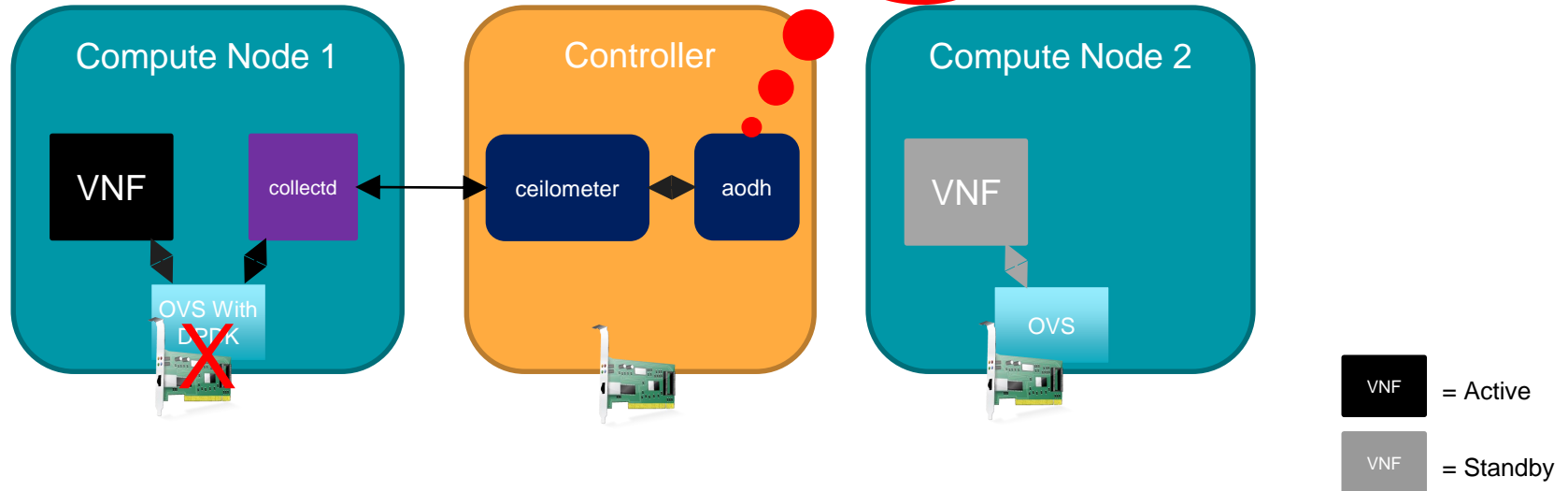
localhost-port.0-link_status != 0



Use case Example

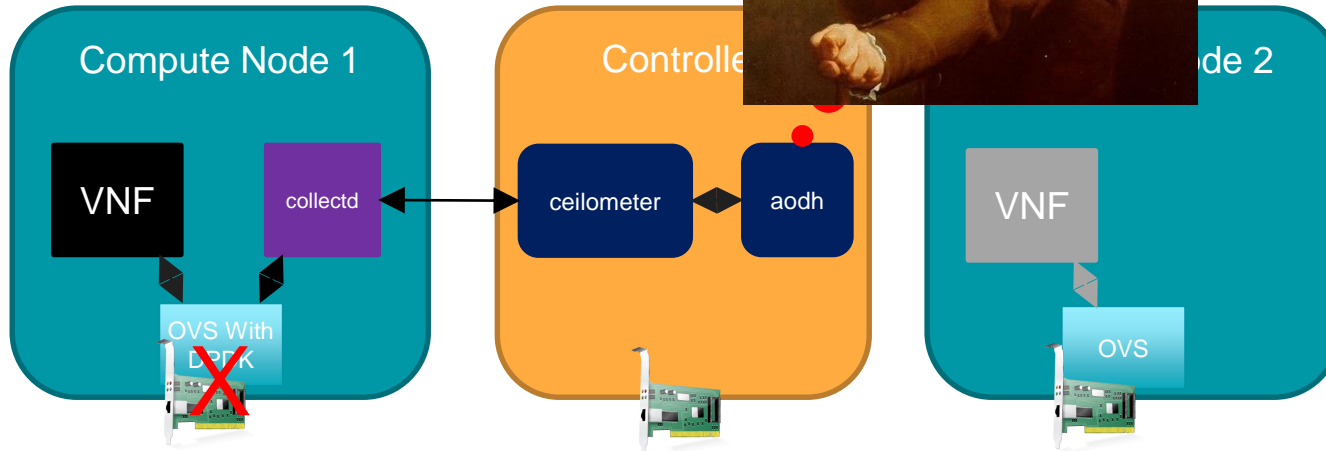
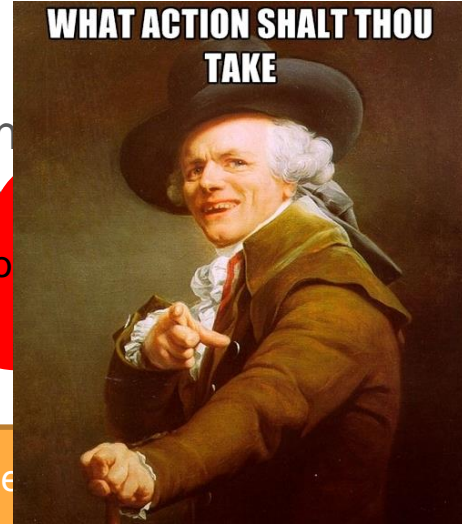
Compute node DPDK interface monitoring for Host link status and switches from active to standby service when the link goes down



`localhost-port.0-link_status == 0`



Use case Example

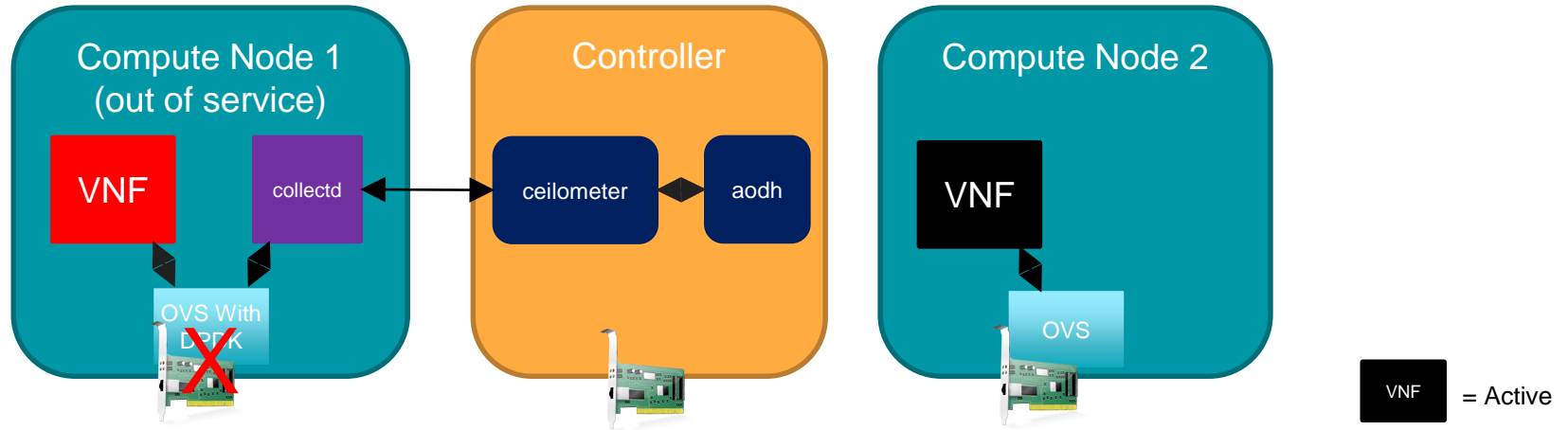
Compute node DPDK interface monitoring and switches from active to standby service when the link of



 = Active
 = Standby

Use case Example

Compute node DPDK interface monitoring for Host link status and switches from active to standby service when the link goes down.



FYI



SFQM Overview



OPNFV



DPDK
DATA PLANE DEVELOPMENT KIT



- Develop the utilities and libraries in DPDK to support:
 - Measuring Telco Traffic and Performance KPIs. Including:
 - Packet Delay Variation.
 - Packet loss.
 - Monitoring the performance + status of the DPDK interfaces.
 - Detecting and reporting violations that can be consumed by VNFs and higher level fault management systems.

The ability to measure and enforce Telco KPIs in the data-plane will be mandatory for any Telco grade NFVI implementation.

SFQM Overview



OPNFV



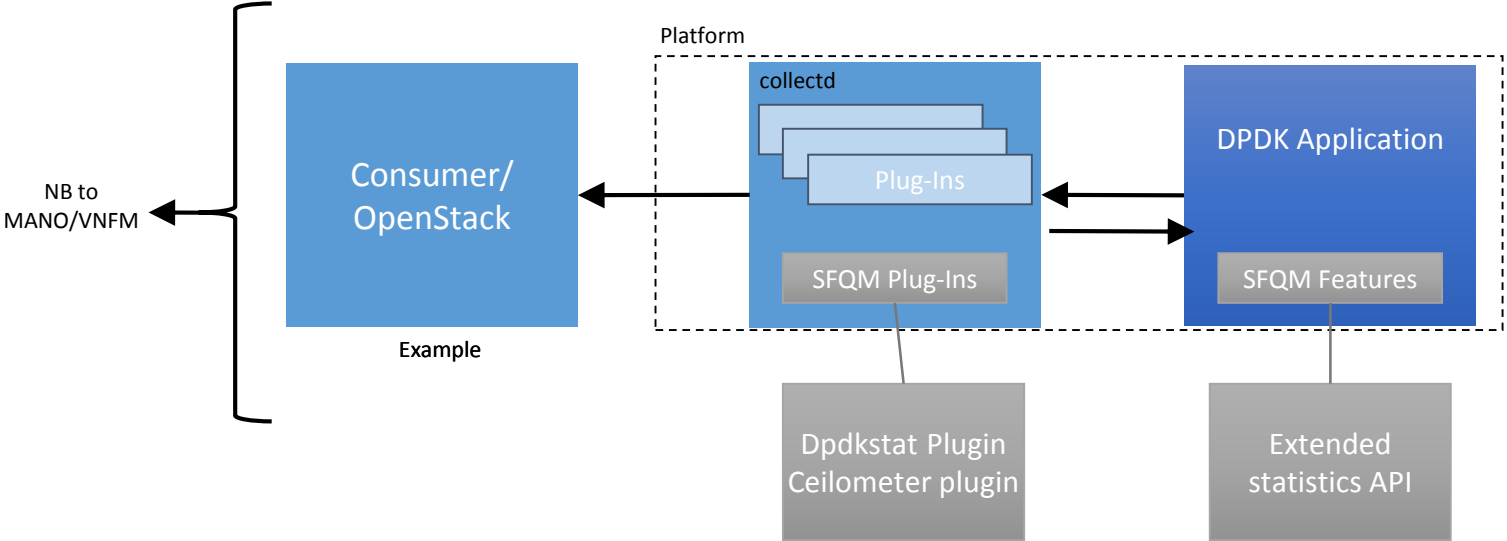
DPDK
DATA PLANE DEVELOPMENT KIT



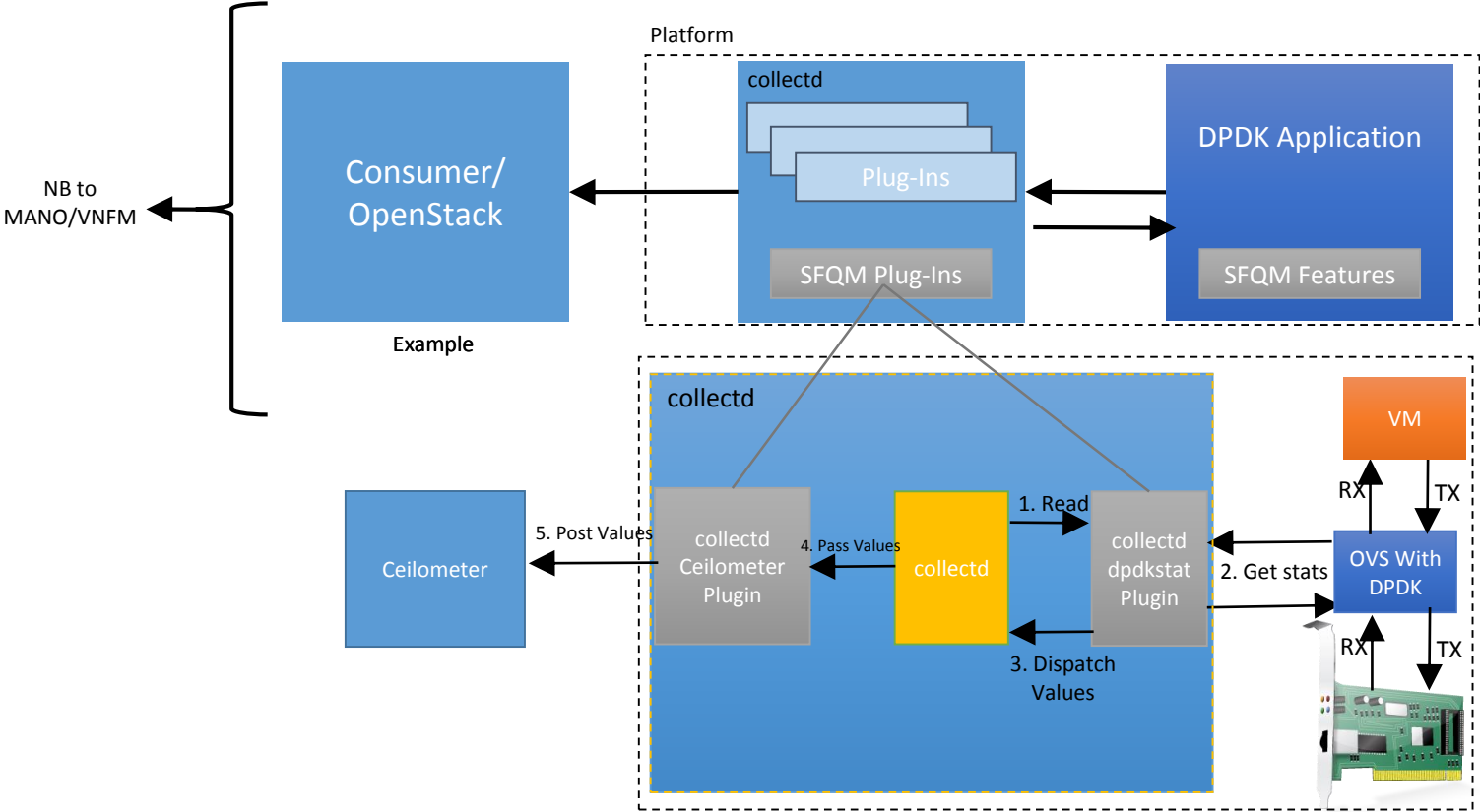
- Develop the utilities and libraries in DPDK to support:
 - Measuring Telco Traffic and Performance KPIs. Including:
 - Packet Delay Variation.
 - Packet loss.
 - Monitoring the performance + status of the DPDK interfaces.
 - Detecting and reporting violations that can be consumed by VNFs and higher level fault management systems.

The ability to measure and enforce Telco KPIs in the data-plane will be mandatory for any Telco grade NFVI implementation.

Collecting DPDK Interface Statistics with collectd



Collecting DPDK Interface Statistics with collectd



Collectd - dpdkstat

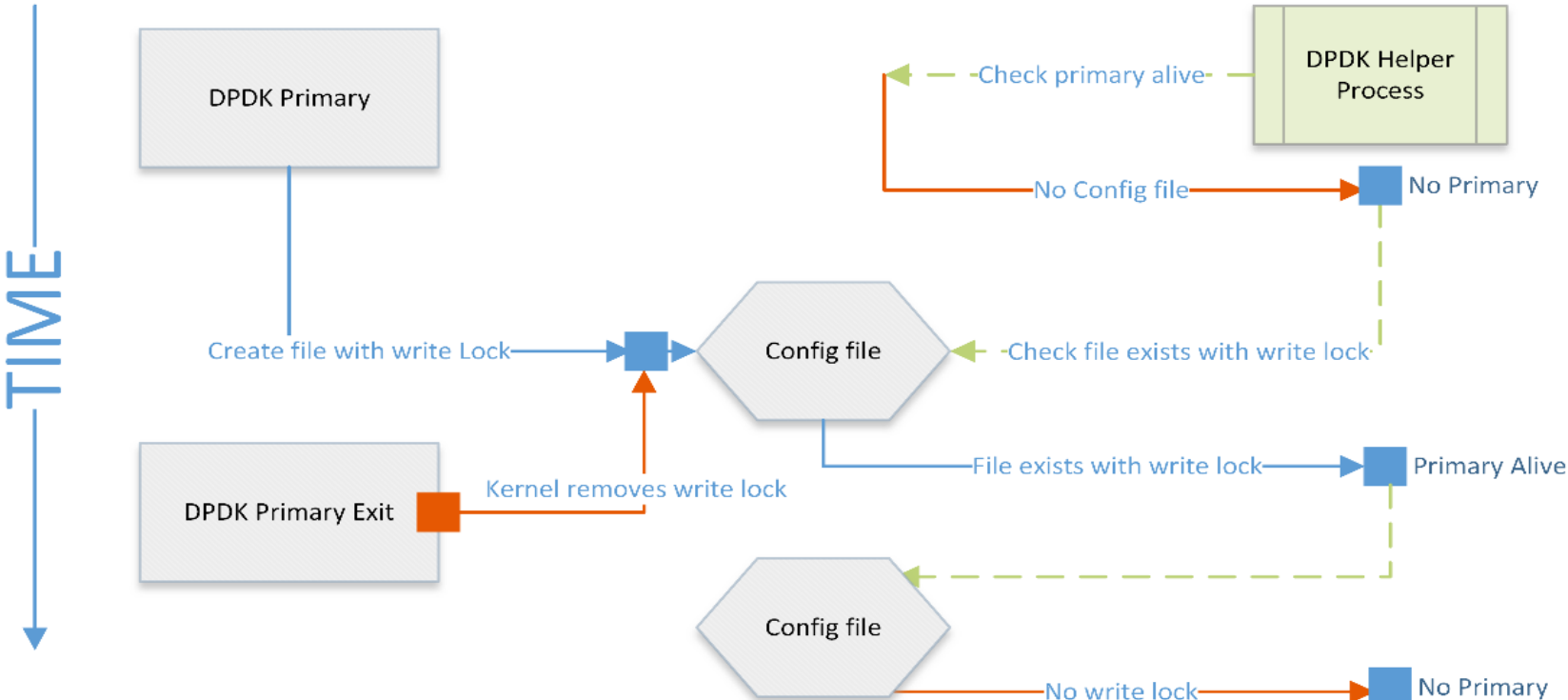


Diagram courtesy of Harry Van Haaren

Collectd - dpdkstat

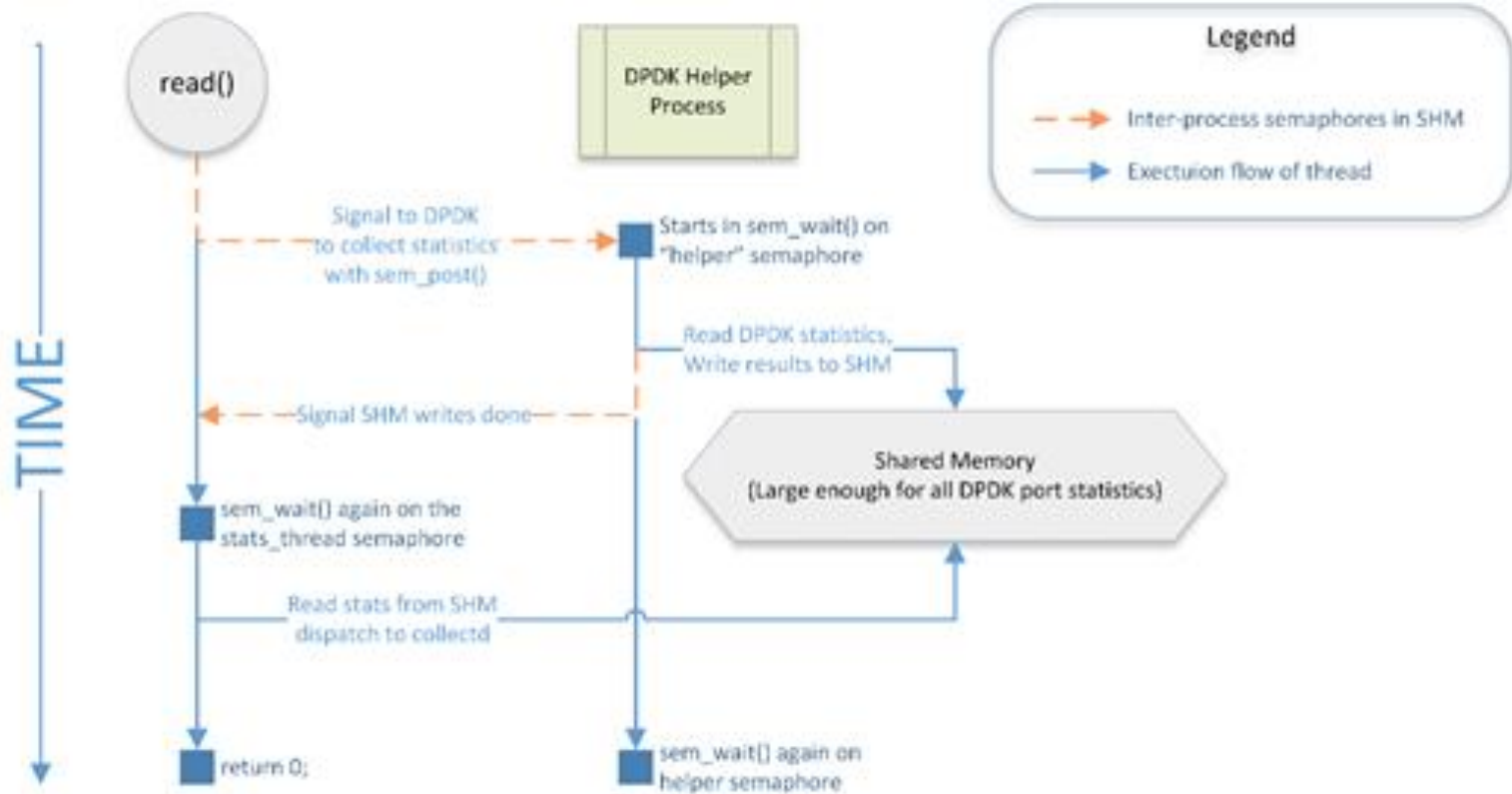
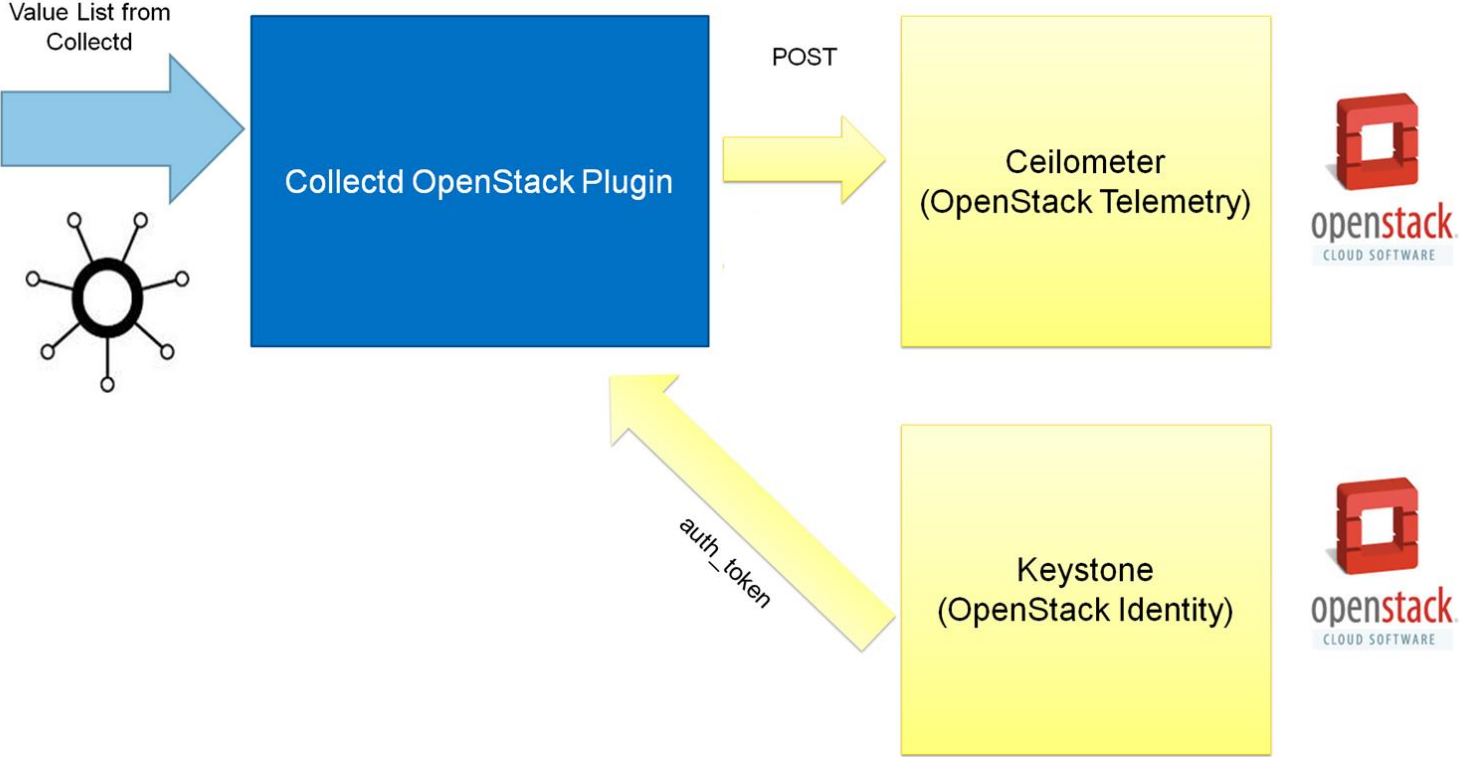


Diagram courtesy of Harry Van Haaren

Ceilometer collectd plugin



Future work

Taking advantage of the notification plugin architecture in collectd to post an event (like link status failure or application thread failure) directly to the notification bus for immediate alarming in Aodh.

Integrating DPDK Keep Alive with collectd and enabling it to take advantage of the notification plugin to Aodh.

Performance, scalability and aggregation analysis.

Gnocchi integration

Open vSwitch (with DPDK) flow and port stats collectd plugin.

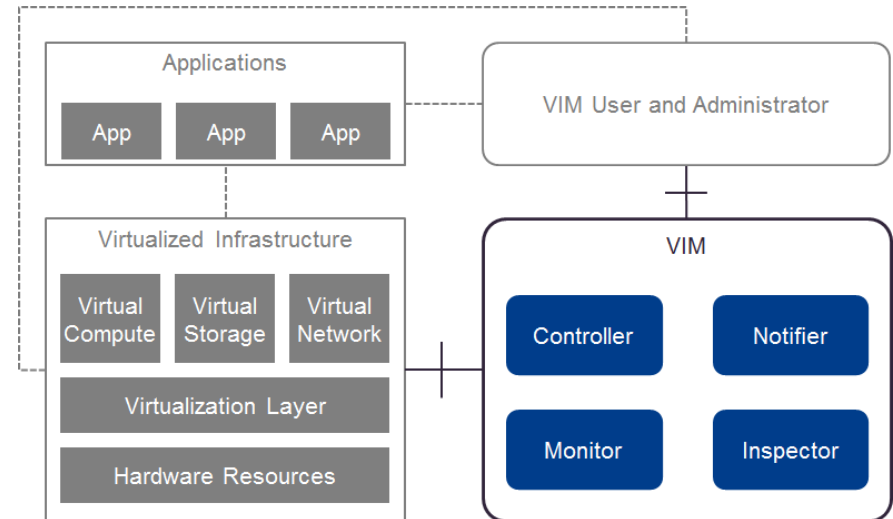
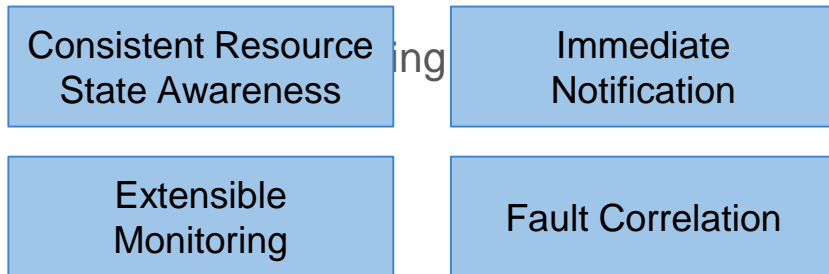
Doctor

Project in OPNFV working on building an **open-source NFVI fault management and maintenance** framework to ensure **Telco VNFs availability** in fault and maintenance events

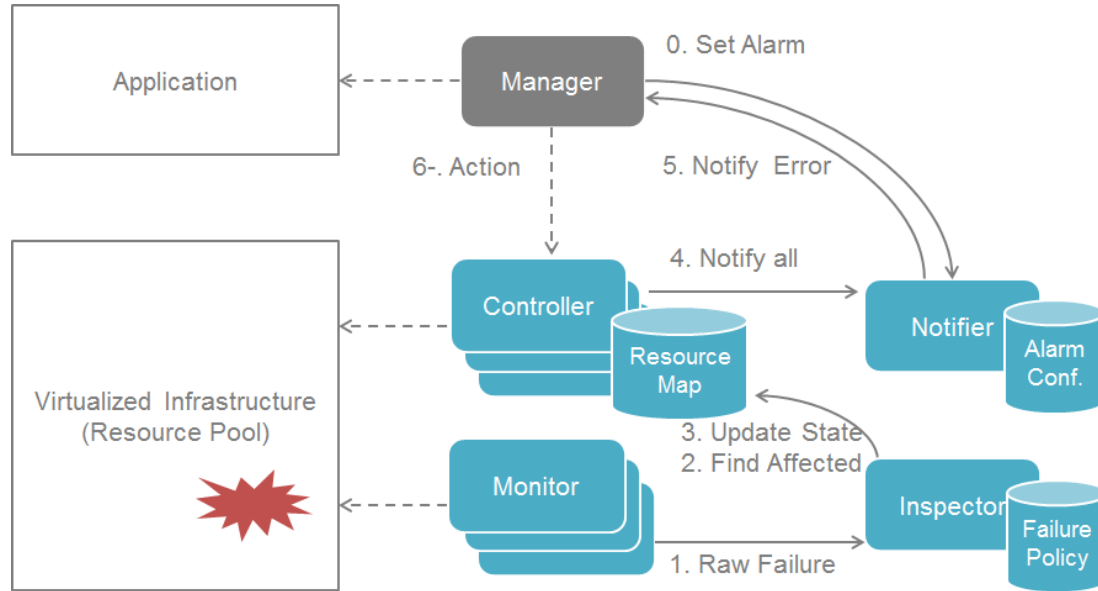
Identify requirements

Gap analysis

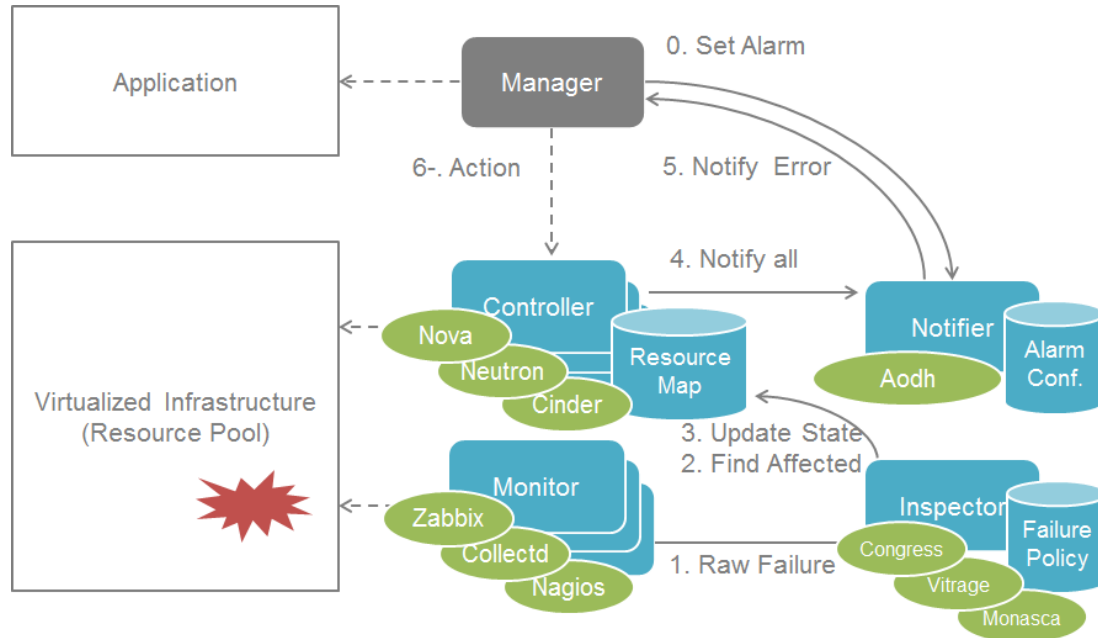
Implementation work in upstream



Doctor: fault management use case



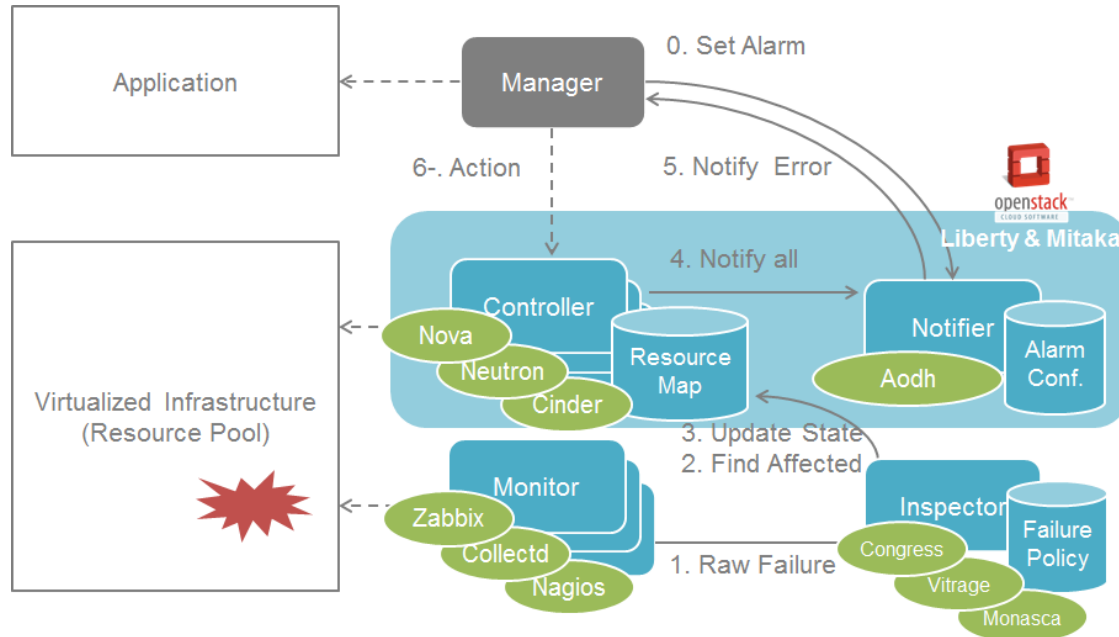
Doctor: mapping to the OpenStack ecosystem



Doctor: focus of initial contributions

Consistent Resource State Awareness

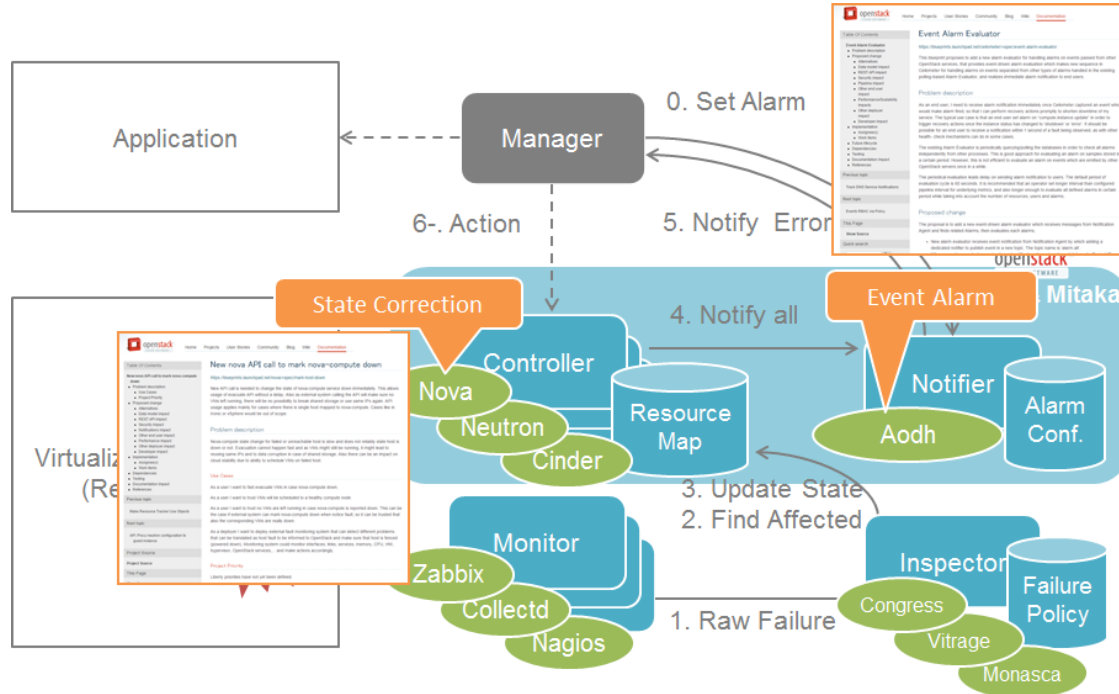
Immediate Notification



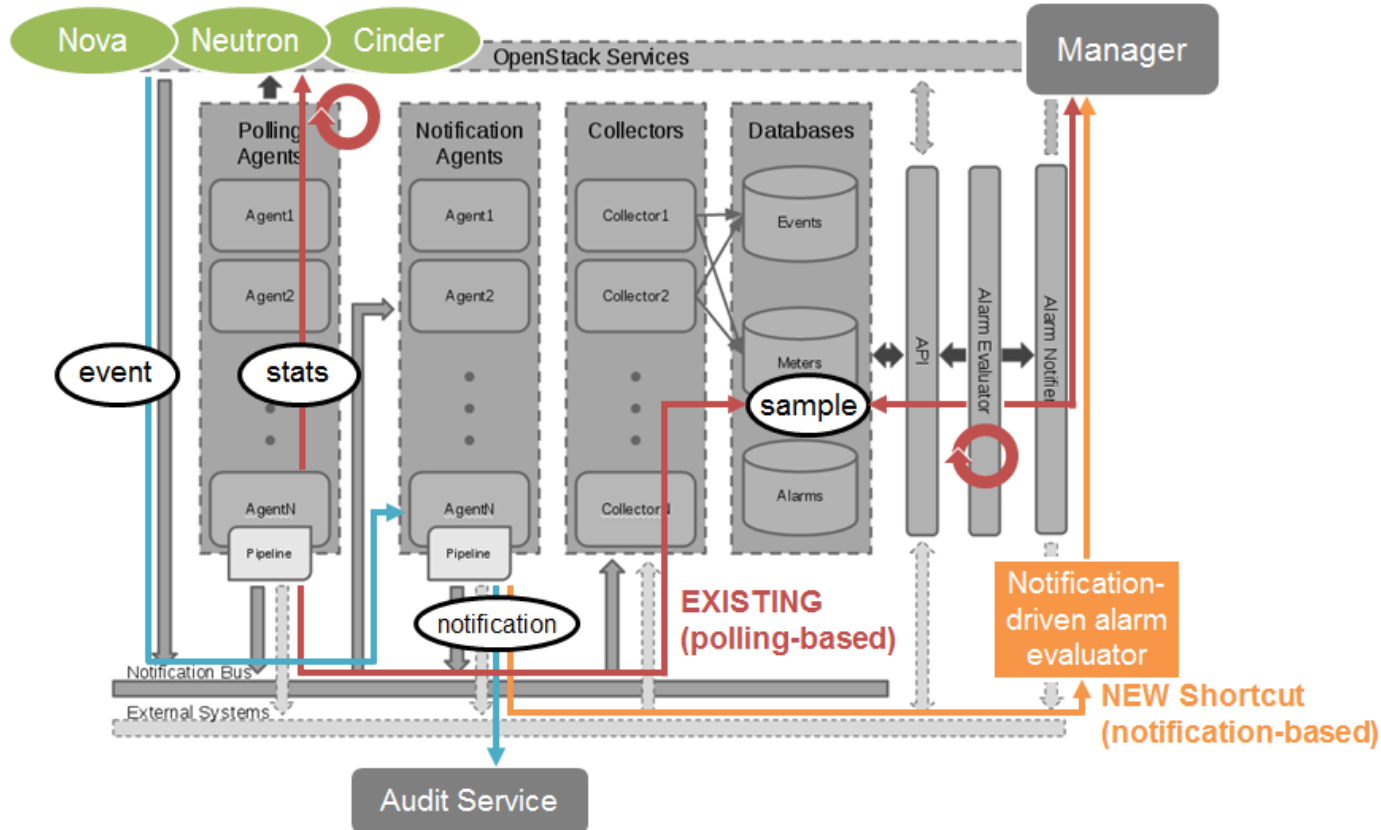
Doctor: focus of initial contributions

Consistent Resource State Awareness

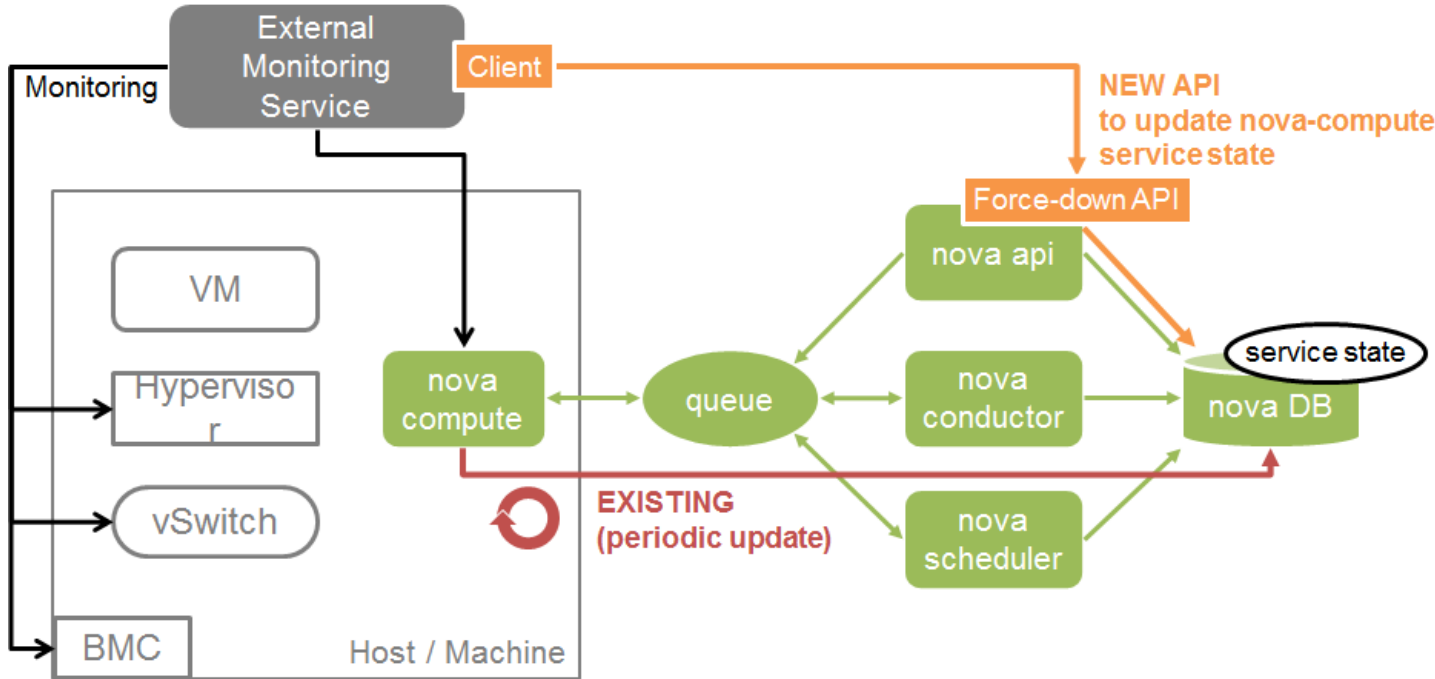
Immediate Notification



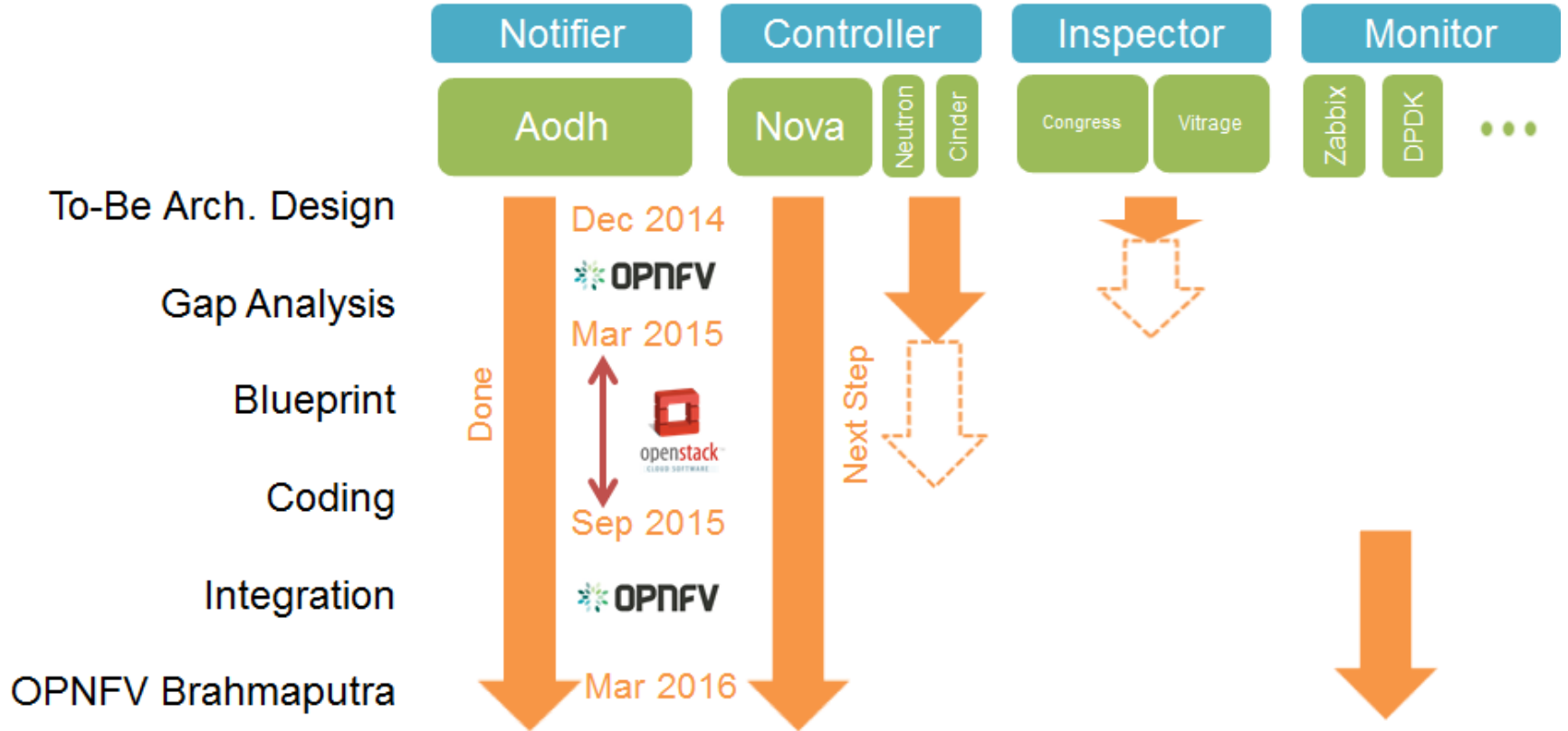
Immediate event alarming



State correction



From project creation to Brahma Putra release



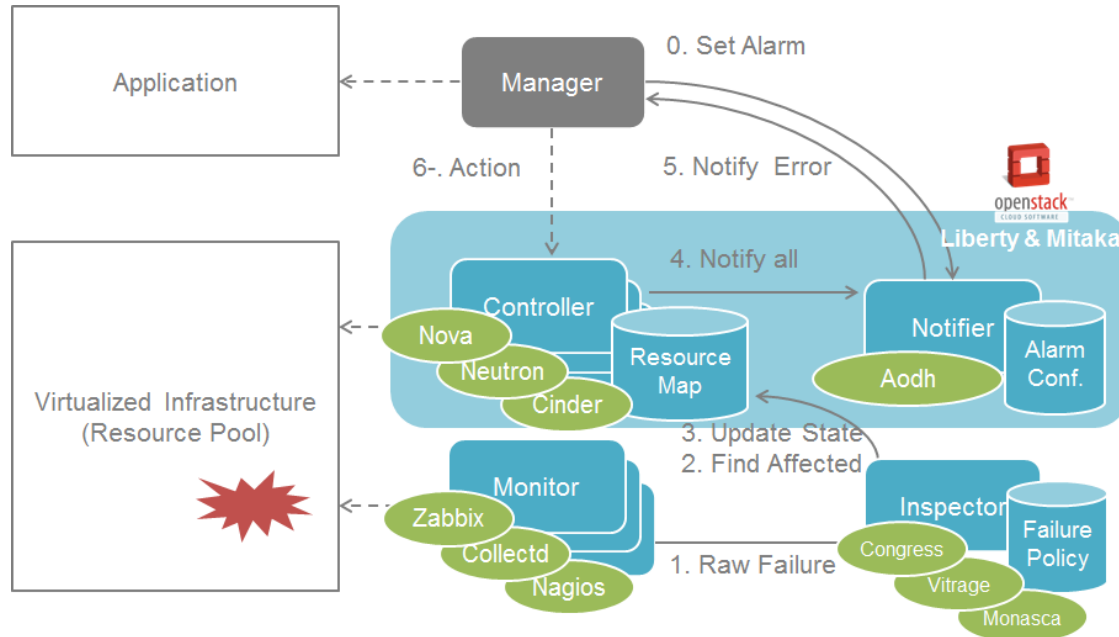
Doctor blueprints in OpenStack Liberty

Project	Blueprint	Spec Drafter	Developer	Status
Aodh	Event Alarm Evaluator	Ryota Mibu (NEC)	Ryota Mibu (NEC)	Completed (Liberty)
Nova	New nova API call to mark nova-compute down	Tomi Juvonen (Nokia)	Roman Dobosz (Intel)	Completed (Liberty)
	Support forcing service down	Tomi Juvonen (Nokia)	Carlos Goncalves (NEC)	Completed (Liberty)

Doctor: focus of initial contributions

Consistent Resource State Awareness

Immediate Notification



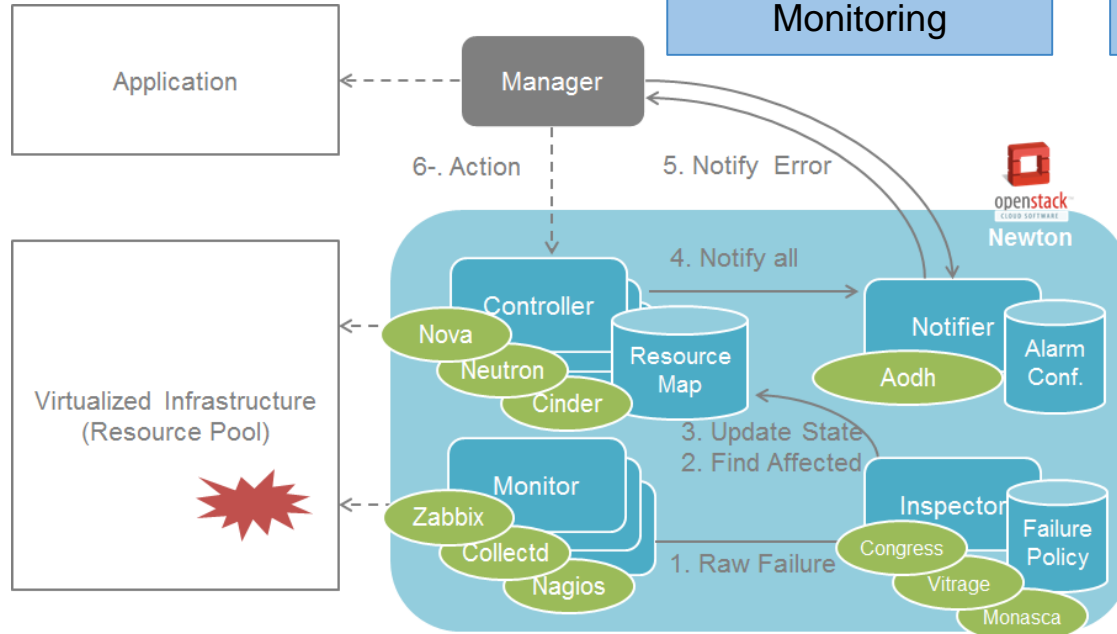
Doctor: extending contribution focus

Consistent Resource State Awareness

Immediate Notification

Extensible Monitoring

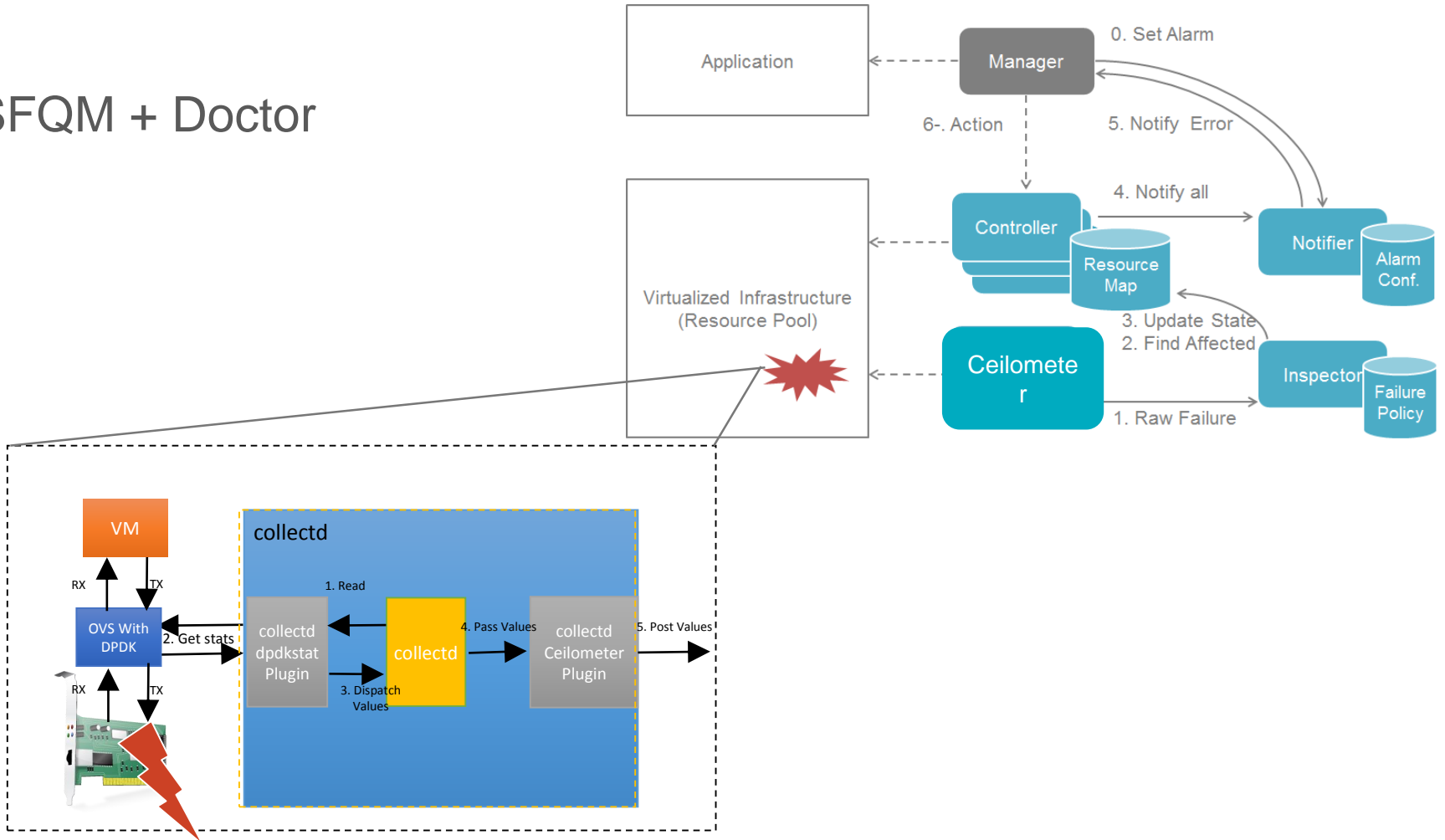
Fault Correlation



Doctor blueprints in OpenStack

Project	Blueprint	Spec Drafter	Developer	Status
Aodh	Event Alarm Evaluator	Ryota Mibu (NEC)	Ryota Mibu (NEC)	Completed (Liberty)
Nova	New nova API call to mark nova-compute down	Tomi Juvonen (Nokia)	Roman Dobosz (Intel)	Completed (Liberty)
	Support forcing service down	Tomi Juvonen (Nokia)	Carlos Goncalves (NEC)	Completed (Liberty)
	Get valid server state	Tomi Juvonen (Nokia)	Tomi Juvonen (Nokia)	Completed (Mitaka)
	Add notification for service status change	Balazs Gibizer (Ericsson)	Balazs Gibizer (Ericsson)	Completed (Mitaka)
Congress	Push Type Datasource Driver	Masahito Muroi (NTT)	Masahito Muroi (NTT)	Completed (Mitaka)

SFQM + Doctor



Demo

What will be demonstrated

- Link status check (DPDK)
- Fault detection propagation (collectd-ceilometer-plugin)
- Resource state correction (Nova)
- Alarm the OpenStack admin/user (Aodh)
- Active-Standby service switching (User)

```
every 0.1s: cellmeter sample-list --meter dpdkstat.gauge --limit 1 Thu Apr 28 17:47:41 2016
```

Resource ID	Name	Type	Volume	Unit	Timestamp
localhost-port_0-link_status	dpdkstat.gauge	gauge	1.0	none	2016-04-28T15:47:37



```
every 0.1s: nova service-list Thu Apr 28 17:47:38 2016
```

ID	Strategy	Host	Zone	Status	State	Updated_At	Disabled_Reason
0	nova-conductor	openv	indiana	enabled	up	2016-04-28T15:47:33.000000	-
14	nova-api	openv	indiana	enabled	up	2016-04-28T15:47:33.000000	-
25	nova-scheduler	openv	indiana	enabled	up	2016-04-28T15:47:33.000000	-
10	nova-compute	openv	indiana	enabled	up	2016-04-28T15:47:33.000000	-
12	nova-compute	compute1	nova	enabled	no	2016-04-28T15:47:33.000000	-
18	nova-compute	compute1	nova	enabled	no	2016-04-28T15:47:33.000000	-

```
every 0.1s: nova list --name server --fields ... Thu Apr 28 17:48:04 2016
```

ID	Name	Status	Power State
x320b0e-Pl00-4174-9017-8c19146cc6d	server2	ACTIVE	Running
Td13k7Tf-8802-479a-80ac-c807c931948a	server2	ACTIVE	Running

Application Manager Log

Event at	Resource	Status	Message
----------	----------	--------	---------



Summary

“Trying to manage a complex cloud solution without a proper telemetry infrastructure in place is like trying to walk across a busy highway with blind eyes and deaf ears. You have little to no idea of where the issues can come from, and no chances to take any smart move without getting in trouble”. [1]



Painting the pedestrian crossing

References

[1] <https://azure.microsoft.com/en-us/blog/cloud-service-fundamentals-telemetry-basics-and-troubleshooting/>



Thank You

Backup

SFQM Overview



- Develop the utilities and libraries in DPDK to support:
 - Measuring Telco Traffic and Performance KPIs. Including:
 - Packet Delay Variation (by enabling TX and RX time stamping).
 - Packet loss (by exposing extended NIC stats).
 - Performance + status Monitoring of the DPDK interfaces (by exposing extended NIC stats + collectd Plugin).
 - Detecting and reporting violations that can be consumed by VNFs and higher level fault management systems (through DPDK Keep Alive).

The ability to measure and enforce Telco KPIs in the data-plane will be mandatory for any Telco grade NFVI implementation.

SFQM Features

Collected

DPDK 2.0

DPDK 2.1

DPDK 2.2

DPDK 16.04

- Callback API
- RX/TX timestamping sample app.

- Extended NIC stats for ixgbe.
- proc_info

- Extended NIC stats for igb, i40e and VFs.
- Extended NIC API Alignment across drivers.
- DPDK KeepAlive

- Primary process liveness check.
- [dpdkstat plugin – pull request](#)
- [Ceilometer plugin.](#)

- Measure packet latency through DPDK

- Retrieve statistics directly from NIC registers.
- Detect DPDK application thread failure.

- Detect DPDK primary process liveness.
- Relay DPDK interface stats and status back to ceilometer.

Statistics we can relay to Ceilometer

collectd-ceilometer plugin

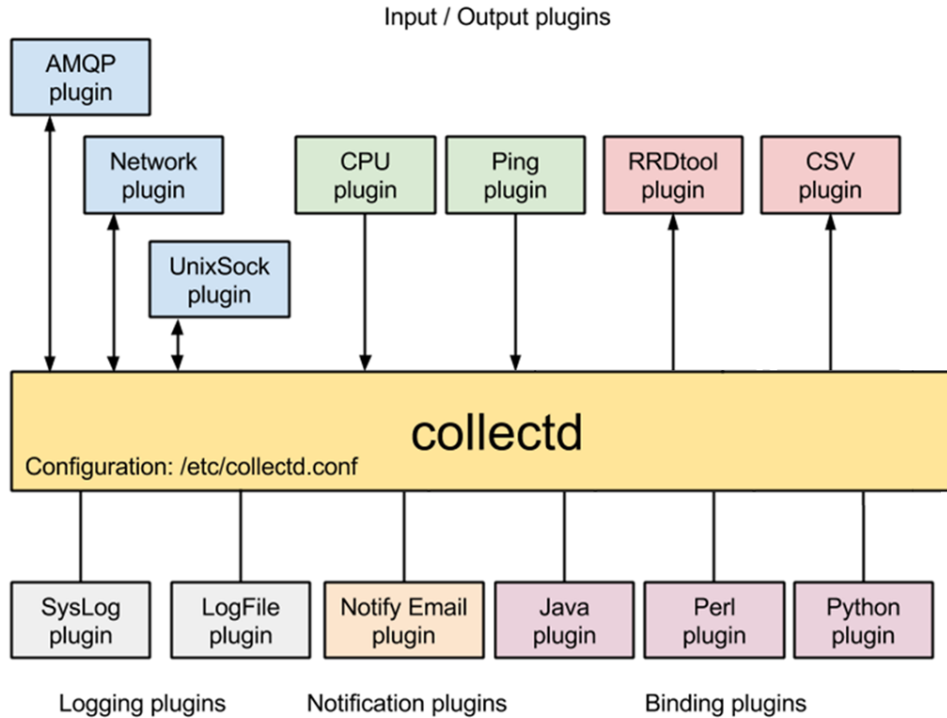
Any of the existing read plugin stats:

- Cache_size, ceph_latency, ceph_rate, cpu count, cpufreq, fan speed, fork_rate, if_collisions, if_dropped, if_errors, virt_cpu_total ...
- Full list of plugins can be found [here](#)
- Full data set spec can be found [here](#)

collectd-dpdkstat plugin

- Any of the hardware NIC statistics registers for drivers that support the extended NIC stats API.
- Any of the generic SW statistics registers
- DPDK Interface Link Status.
- rx_good_packets, rx_crc_errors, tx_errors, mac_local_errors, mac_remote_errors, rx_priority0_dropped....

Collectd architecture



Demo

```

every 0.1s: cellmeter sample-list --meter dpdkstat.gauge --limit 1 Thu Apr 28 17:47:41 2016

```

Resource ID	Name	Type	Volume	Unit	Timestamp
localhost-port-0-link_status	dpdkstat.gauge	gauge	1.0	none	2016-04-28T15:47:37



```

every 0.1s: nova service-list Thu Apr 28 17:47:38 2016

```

ID	Strategy	Host	Zone	Status	State	Updated_At	Disabled_Reason
0	nova-conductor	openv	ind.arena	enabled	up	2016-04-28T15:47:33.000000	-
14	nova-api	openv	ind.arena	enabled	up	2016-04-28T15:47:33.000000	-
25	nova-scheduler	openv	ind.arena	enabled	up	2016-04-28T15:47:33.000000	-
16	nova-compute	openv	ind.arena	enabled	up	2016-04-28T15:47:33.000000	-
17	nova-compute	compute1	nova	enabled	no	2016-04-28T15:47:33.000000	-
18	nova-compute	compute1	nova	enabled	no	2016-04-28T15:47:33.000000	-

```

every 0.1s: nova list --name server --fields ... Thu Apr 28 17:40:34 2016

```

ID	Name	Status	Power State
x320b0e-Pl00-4174-9017-8c70149cc6d	server2	ACTIVE	Running
Td13k77f-8801-479a-80ac-c807c931948a	server2	ACTIVE	Running

Application Manager Log

Event at	Resource	Status	Message

