

# Profiles and Flavours

June 2021 LFN Developer & Testing Forum

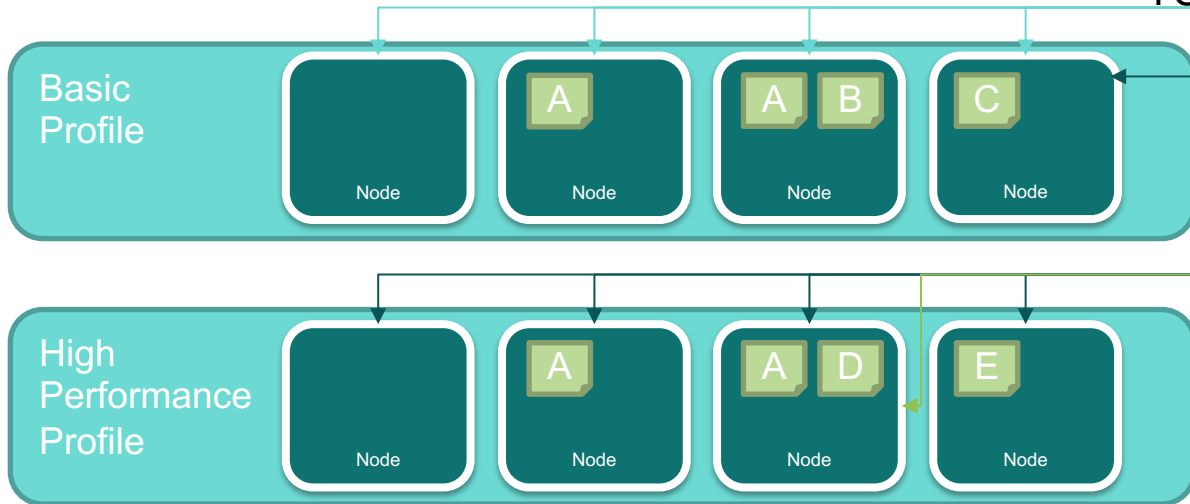


- **Network Functions** and Telco Applications expect certain **capabilities** and **performance** from the infrastructure where they run on.
- Creating workload-specific silos makes resources unmanageable and un-shareable.
- Anuket RM has defined a number of common **Profiles** and **Flavours** to address this issue and streamline Telco Cloud infrastructure management and consumption.

# Profiles, Extensions and Flavours

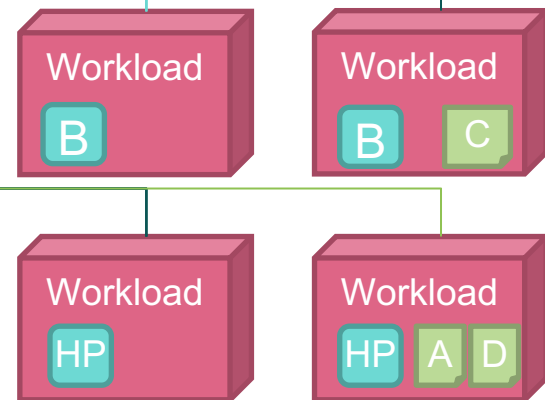
## Infrastructure Profiles and Extensions

- Profiles - Top level categorisation
  - Partition the infrastructure
- Extensions
  - Second level categories



## Workload Flavours

- Capability requirements for workloads to run
- Used to filter infrastructure resources based on capabilities



# Basic Profile vs High Performance



## Basic Profile

Only suited for workloads that tolerate variable performance, including latency, and resource over-subscription.

- **Simultaneous Multi-Threading (SMT).**
- **No NUMA** alignment,
- Supports **over-subscription** (using CPU Allocation Ratio) which is specified as part of sizing information in the workload profiles

## High Performance Profile

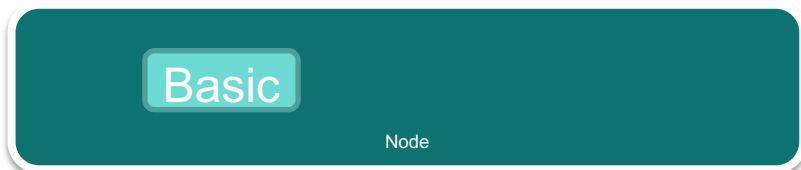
Used for workloads that require predictable performance, high network throughput requirements and/or low network latency.

- **NUMA alignment**
- **CPU pinning**
- **Huge pages**
- **Doesn't support over-subscription.**

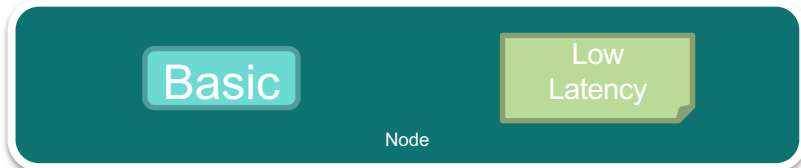
# Adding Extensions



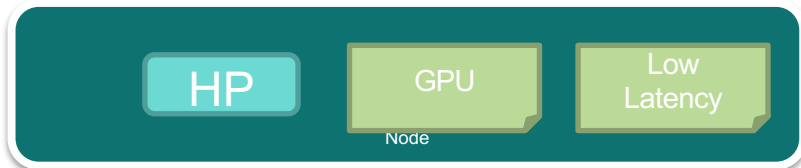
Profile **Extensions** are meant to be used as **labels** for infrastructure, identifying the nodes that implement **special capabilities** that go **beyond** the profile baseline.



Label: *B*



Label: *B.low-latency*



Label: *H.compute-gpu.low-latency*

# Extension Examples - 1



Profile Extension Name	Description	B	HP
Compute Intensive High-performance CPU	Predictable computing performance and higher clock speeds	✗	✓
Storage Intensive High-performance storage	Low storage latency and/or high storage IOPS	✗	✓
Compute Intensive High memory	> X amount of RAM	✗	✓
Compute Intensive GPU	GPU compute resource on the node	✗	✓
Network Intensive High speed network (25G)	25G NICs	✗	✓
Network Intensive Very High speed network (100G)	100G NICs	✗	✓

# Extension Examples - 2



Profile Extension Name	Description	B	HP
Low Latency - Edge Sites	Node located in an Edge site, < X ms latency to final users or geographical distribution (TBD)	✓	✓
Very Low Latency - Edge Sites		✓	✓
Ultra Low Latency - Edge Sites		✓	✓
Fixed function accelerator	fixed function accelerator (non programmable, eg Crypto, vRAN adapter)	✗	✓
Firmware-programmable adapter	consumable Firmware-programmable adapter (eg Network/storage adapter).	✗	✓
SmartNIC enabled	Programmable accelerator for vSwitch/vRouter, Network Function and/or Hardware Infrastructure	✗	✓
SmartSwitch enabled		✗	✓

- ***Infra Profiles & Extensions – qualitative/quantitative?***
- ***More Workload requirements?***
- ***Different types of compute hw?***
- ***Accelerator cards?***





**Anuket**

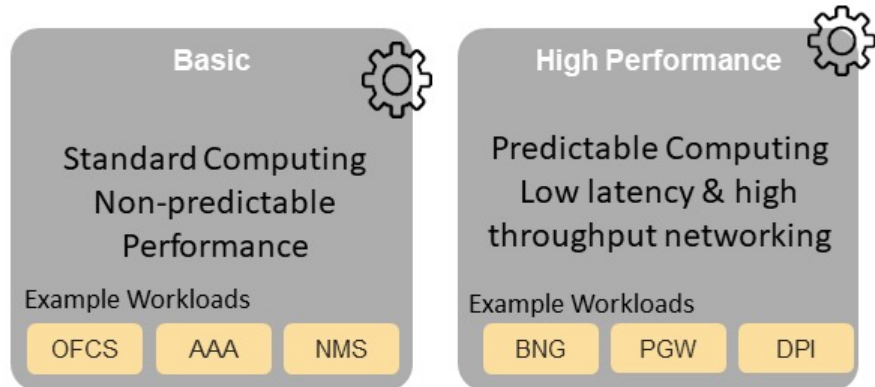
# Introduction to Profiles

## Host (Node) Profiles

- › Mechanism to manage configurations or request for host with certain configurations
- › Define the configuration that can then be applied to any number of hosts
  - › Results in configuration consistency and correctness
- › Can be used as a label identifying the hosts (servers) that are configured as per the host profile
  - › can apply a profile label to a host that has more hardware configurations, but not fewer, than what has been specified in the host profile.
- › Variability in hardware configurations captured in “Profile Extensions”

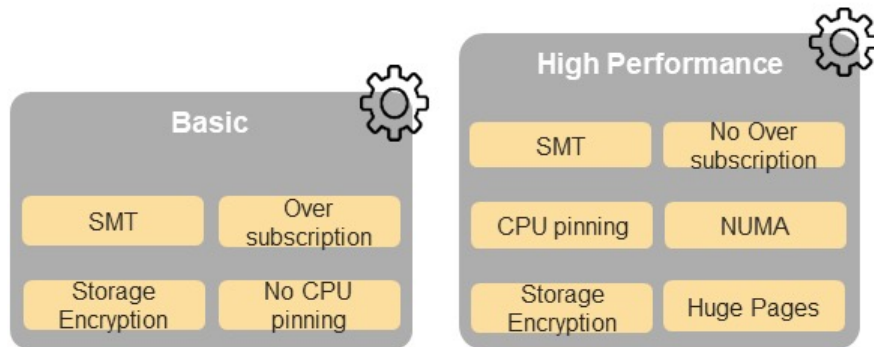
# Number of profiles

- › Profiles partition infrastructure into separate pools
  - › Can lead to surplus and shortage
- › Analysis of the workloads and with abstractions we can establish
- › Two (2) top-level profiles (RM [Section 2.4.1](#)):
  - › Basic (no performance expectations) and
  - › High Performance (predictable performance)



# Profiles

- › RM [Section 4.1](#) specifies the node capabilities
- › RM [Section 5.4](#) specifies the profile capabilities



- › These top-level profiles can be further specialized – Profile Extensions
  - › The label of the parent profile can be applied in addition to the Profile Extension label.
  - › Example, High Performance – Compute Intensive: spread workload across all NUMA nodes to overcome any NUMA memory access bandwidth limitations

## Profile Extensions (RM [Section 2.4.2](#))

- › Profile Extensions are used as labels for infrastructure, identifying the nodes that implement *special capabilities* that go beyond the profile baseline.
- › Example abstracted Profile Extensions for High Performance nodes:
  - › **Network Intensive:** *for Workloads that require higher networking speeds (e.g., using DPDK or SR-IOV).*
  - › **Enhanced Network Intensive:** *for Workloads that require specific network resources (e.g., crypto acceleration, SmartNIC).*
  - › **Storage Intensive:** *for Workloads that require low storage latency and/or high storage IOPS.*

# Introduction to Flavours

Virtual Server (container or VM) Flavours:

- › Mechanism to specify size, geometry and other capabilities of the requested virtual resources
- › Can be pre-defined or defined at virtual resource request time
  - › Virtual resources can be requested consistently