



**KubeCon**



**CloudNativeCon**

**Europe 2023**





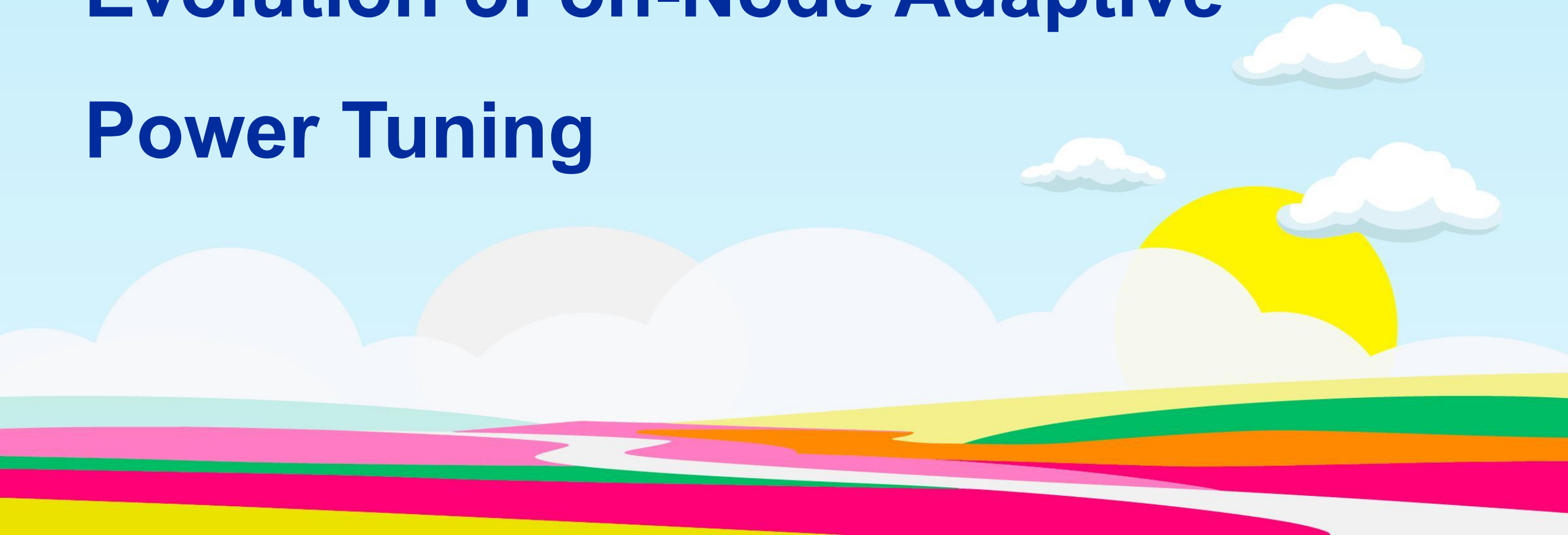
KubeCon



CloudNativeCon

Europe 2023

# Evolution of on-Node Adaptive Power Tuning



# Who are we?



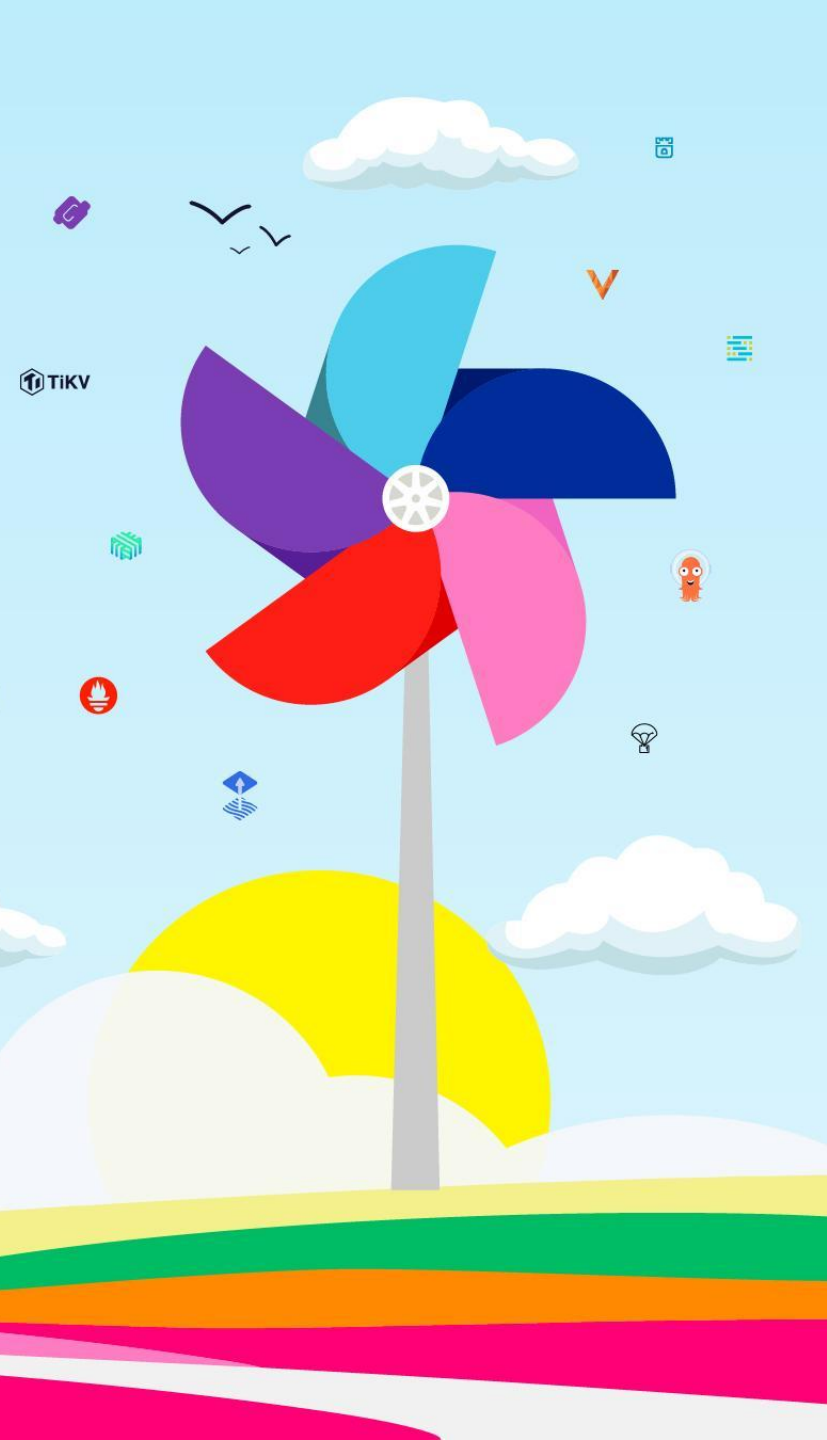
**Dr. Atanas Atanasov**  
Senior Cloud-Native Dev.  
*Intel*

*atanas.atanasov@intel.com*



**Rimma Iontel**  
Chief Architect  
*Red Hat*

*riontel@redhat.com*



# Power Controls

# Node-Level Optimizations

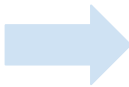
- **Hardware optimization**

- Granular control of unused components
  - Cores, NICs, disks, hardware accelerators, etc.
- Turning off unused components
  - NIC, disks, cores, hardware accelerators, etc.
- Alternate hardware architectures
  - SmartNICs, IPU, ARM CPUs, etc.

- **OS optimizations**

- Power efficient profiles
  - Tuned, Performance Profiles
- Granular selection of settings for energy efficiency
  - CPU governors, p/c-states, uncore frequencies

**TuneD** is a service that monitors the system and optimizes its performance based on ***use case*** specific ***profiles***

- High throughput
  - Low latency
  - **Power Savings**
- 
- throughput-performance
  - latency-performance
  - network-latency
  - network-throughput
  - **balanced**
  - **powersave**

## **P-States:** voltage-frequency control states

- CPUfreq in the Linux kernel enables scaling the CPU frequency
- Governor controls allowed settings

```
# cpupower frequency-info
analyzing CPU 0:
  driver: intel_cpufreq
  hardware limits: 800 MHz - 3.50 GHz
  available cpufreq governors: conservative
  ondemand userspace powersave performance
  schedutil
...
```

## **C-States:** CPU idle sleep states where the processor clock is inactive

- Partially deactivates unused CPUs
- Range: C0 (active) to Cn
- Deeper C-state -> exit latency duration becomes longer

```
# cpupower idle-info

CPUidle driver: intel_idle
CPUidle governor: menu
analyzing CPU 0:
Number of idle states: 4
Available idle states: POLL C1 C1E C6
...
```

**Uncore:** components of a processor that are not directly involved in the execution of instructions

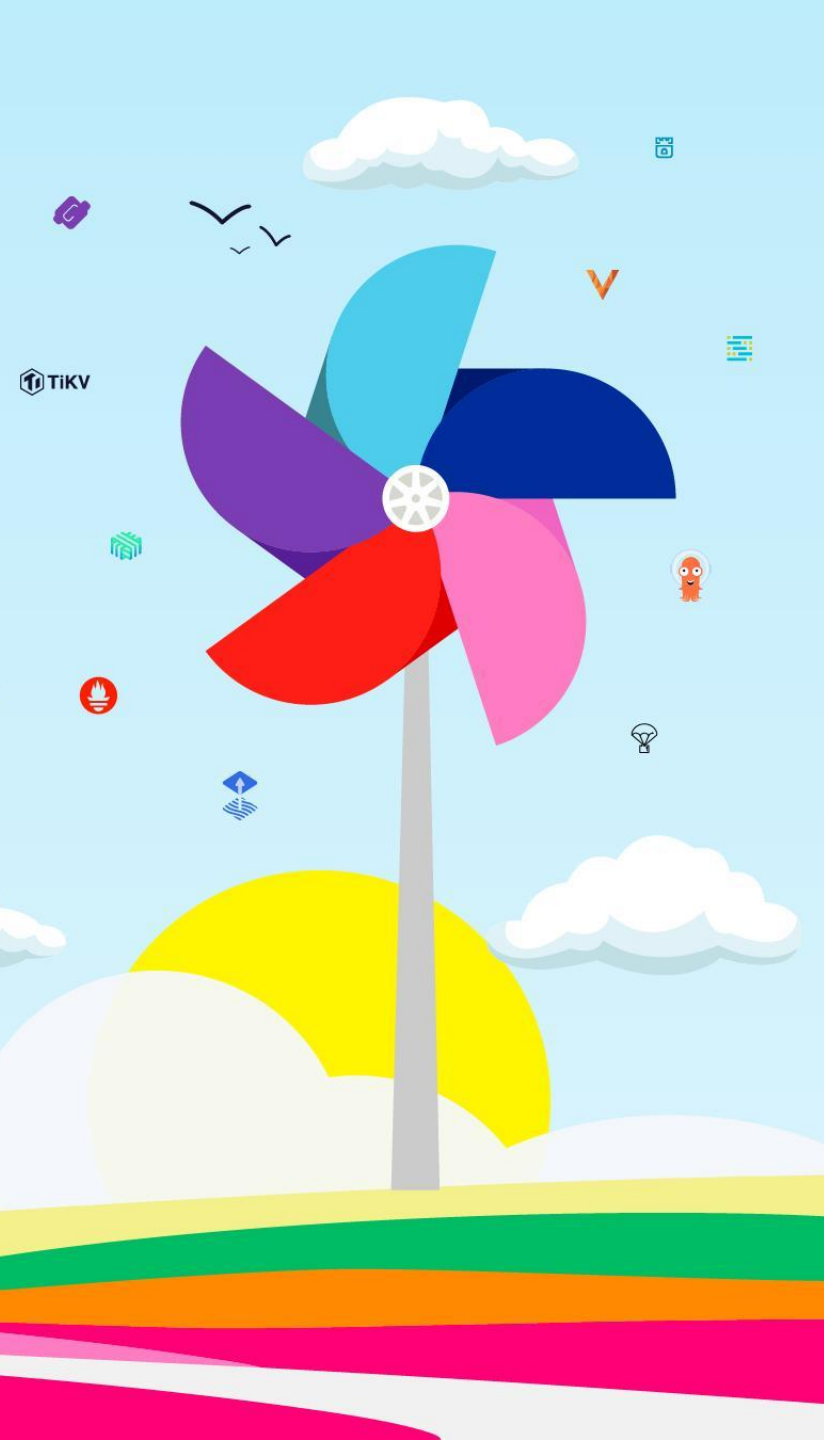
- Memory controller
- Cache coherency logic
- Interconnects between processor cores (QPI)

**Uncore Frequency Scaling (UFS):** Cores interconnect and L3 shared cache frequency scaling for energy efficiency

- BIOS Setting
- Future functionality: OS controls with TuneD or Intel's Kubernetes Power Manager<sup>1</sup>

<sup>1</sup><https://github.com/intel/kubernetes-power-manager>



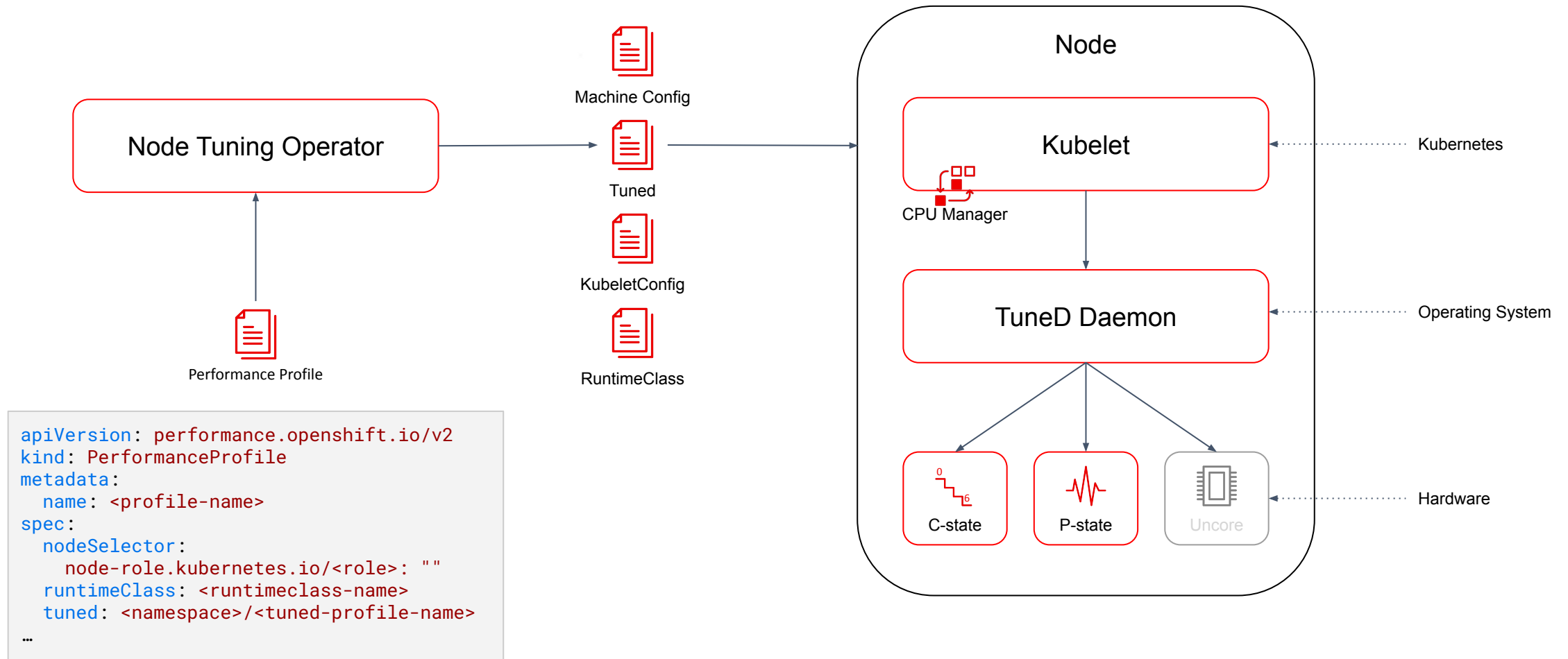


# Kubernetes Power Configuration

# Kubernetes Components

- Kubelet
  - **CPU manager** (cpu pinning and isolation)
  - Topology manager (NUMA affinity)
  - Memory manager (Hugepages NUMA affinity)
- Cluster Node Tuning Operator (NTO)
  - Maintains tuning rules for the distributed OCP environment
  - Runs **TuneD** that applies the tuning rules to every node
- Machine Config Operator (MCO)
  - Abstracts operating system and CRI-O configuration changes
- Performance Profile Controller
  - Computes tuning rules from the provided high level description
  - Provides configuration snippets to all lower level components

# Under the Hood



# Node Tuning Operator

Node Tuning Operator energy optimizations:

- **CPU core** disablement/offline
- **CPU Governor** selector per core
- **CPU Frequency** tuning for group of cores
- Granular **power optimizations** for mixed workloads

# Permanently Offline CPUs

Use case: the worker nodes of the cluster have been deployed with extra CPU capacity that will be used in the future. How to turn them off until we need them?

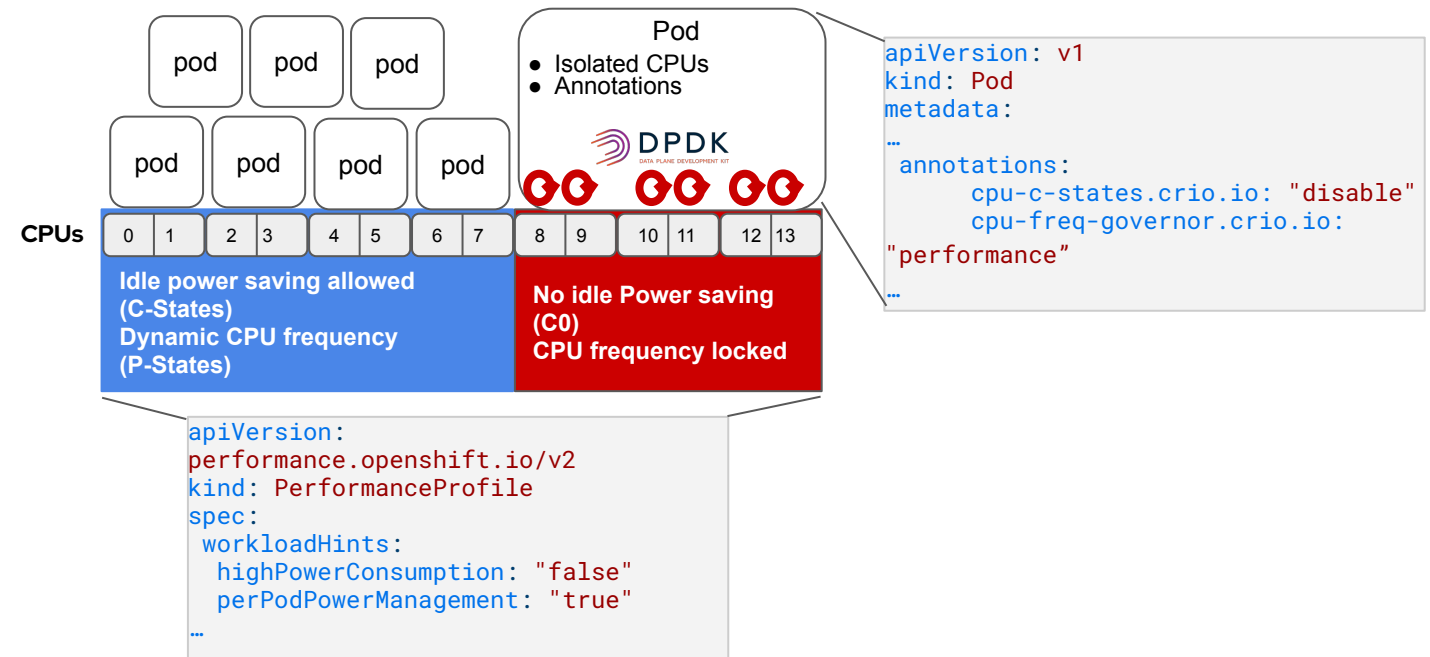
- Shut down select CPUs with PerformanceProfile
- Done at boot time, so any configuration change requires a reboot
  - Sets `maxcpus=X` in kernel boot arguments
  - Changes `/sys/devices/system/cpu/cpuX/online`

```
apiVersion:
performance.openshift.io/v2
kind: PerformanceProfile
metadata:
  name: <profile-name>
spec:
  cpu:
    isolated: "2-21,26-37"
    reserved: "0-1,24-25"
    offline: "38-42"
...
```

# Power optimizations for mixed workloads

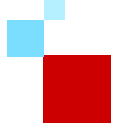
Configure CPU Power states **per pod** to support mixed workloads

- Enable CPU power savings features on all CPUs by default
- Apply performance optimizations per pod through annotations





# Kubernetes Power Manager

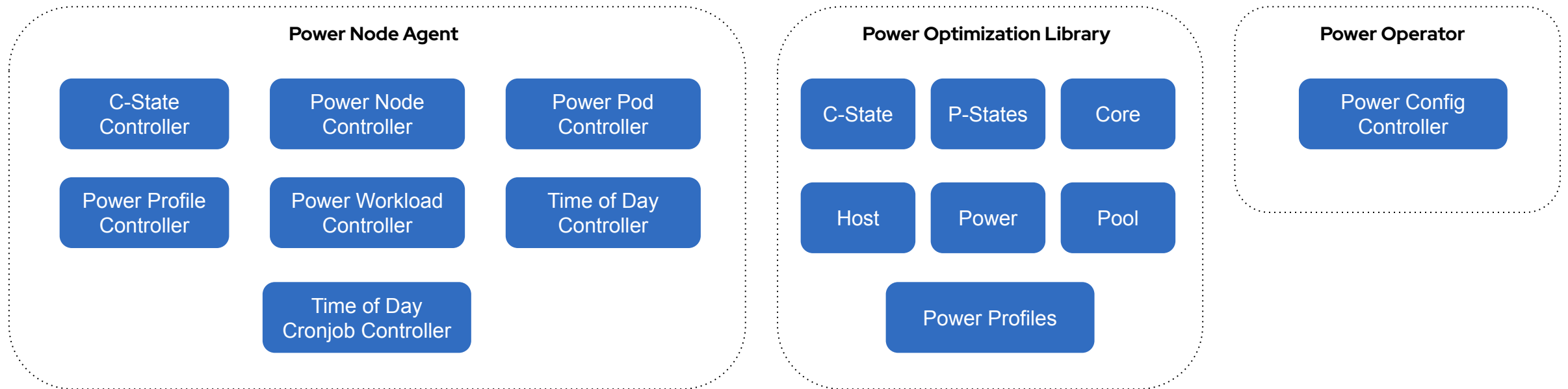


- Reduce operational overheads
- Lower power consumption by controlling the frequencies of the shared pool cores
- Turn off uncore functionality, as needed
- Choose specific governors
- Play with various sleep states

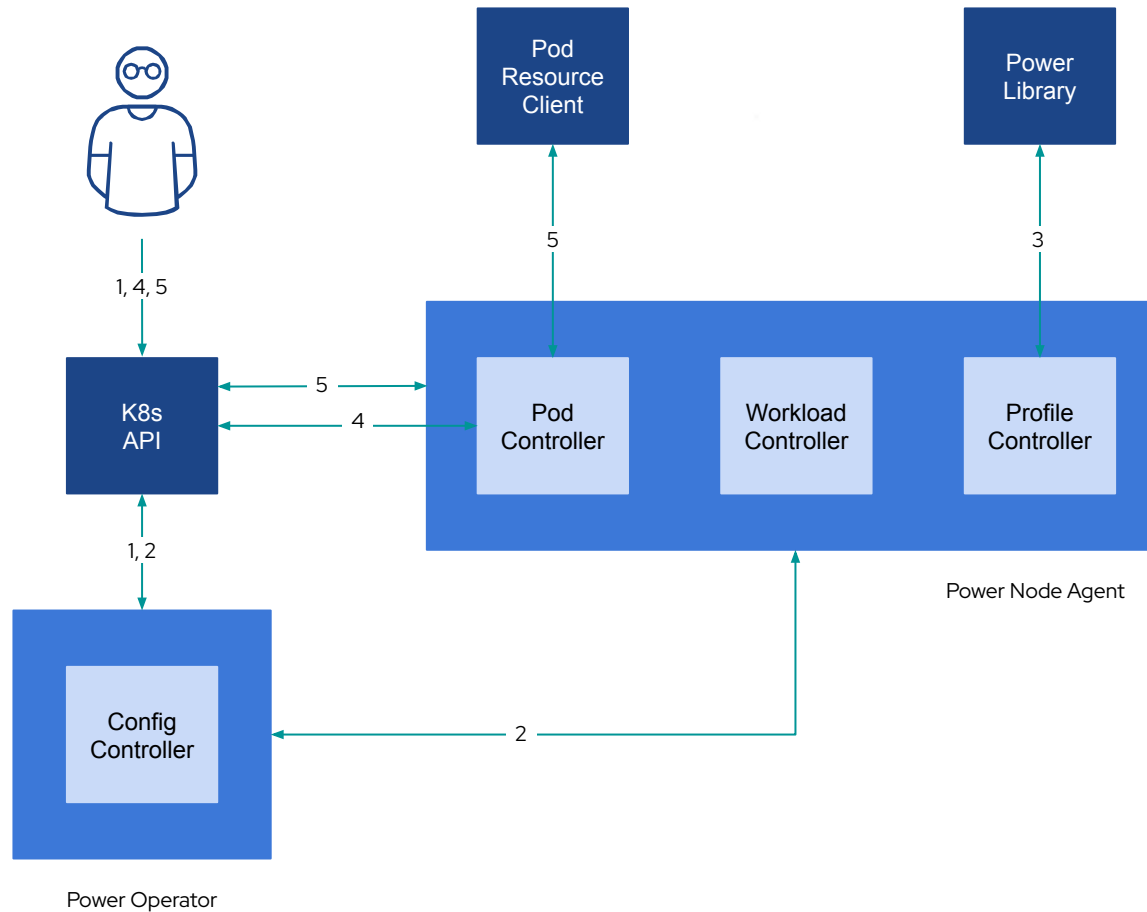
Benefits of the  
K8s Power  
Manager to our  
Ecosystem



# K8s Power Manager Component Diagram



# K8s Power Manager Flow Diagram



1. User creates Config custom resource
2. Config controller creates Power Node Agent / Profiles
3. Profiles / Pools created in Power Optimization Library
4. Shared Profile / Workload created by the user
5. User creates Pod requesting Profile. Power Node Agent configures the associated cores

# Example of Power Profile Spec

## Power Profile Spec

```
apiVersion:"power.intel.com/v1"  
Kind: Power Profile  
Metadata:  
  name: performance  
Spec:  
  name: "Performance"  
  max: 3200  
  min: 2800  
  epp: "performance"
```

## Power Profile Spec

```
apiVersion:"power.intel.com/v1"  
Kind: Power Profile  
Metadata:  
  name: balance-performance  
Spec:  
  name: "Balance-Performance"  
  max: 3000  
  min: 2500  
  epp: "balance-performance"
```

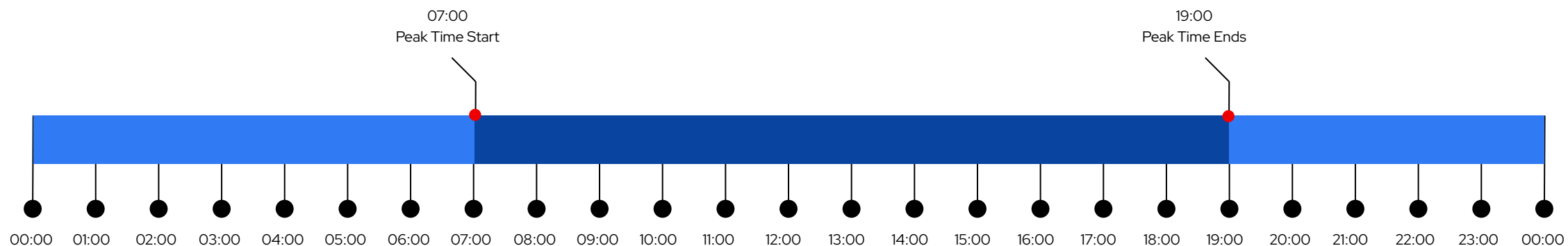
## Power Profile Spec

```
apiVersion:"power.intel.com/v1"  
Kind: Power Profile  
Metadata:  
  name: power  
Spec:  
  name: "Power"  
  max: 1500  
  min: 1000  
  epp: "power"
```

## Power Profile Spec

```
apiVersion:"power.intel.com/v1"  
Kind: Power Profile  
Metadata:  
  name: balance-power  
Spec:  
  name: "Balance-Power"  
  max: 2200  
  min: 1800  
  epp: "balance-power"
```

# Time of Day

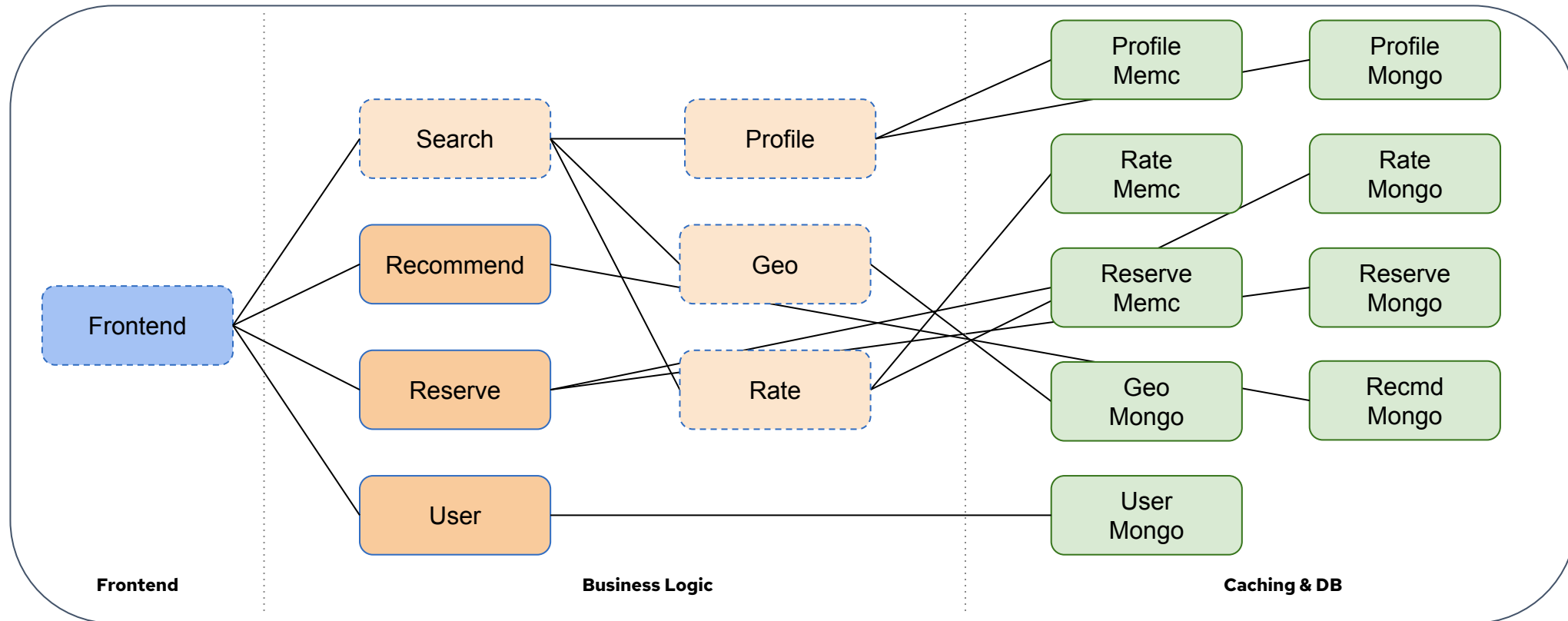


Sleep Time

Active Time

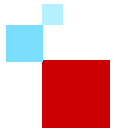
```
apiVersion: time-of-day/v1
kind: peak-time
metadata:
  name: peak-time
spec:
  weekdays: "1-5"
  sleep: "19:00"
  active: "07:00"
  timeZone: "Europe/Ireland"
```

# Use Case – DSB Hotel Reservation



- DSB(DeathStarBench) includes representative microservices workloads open sourced by Cornell University
  - Hotel Reservation, Social Network Service, Media Service
  - Workloads to simulate real world services

- Wrk is used as load generator, traffic specified via lua script
- Created larger dataset with 10M users/properties
- Uses gRPC for communication between services



# Run Power Manager in OCP

1. Configure TuneD
2. Deploy Power Manager
3. Deploy Power Profiles
4. Deploy Power Workload



# 1. TuneD Profile

## 2. Deploy Power Manager

```
apiVersion: tuned.openshift.io/v1
kind: Tuned
metadata:
  name: intel-kpm-hotfixes
  namespace: openshift-cluster-node-tuning-operator
spec:
  profile:
  - data: |
      [main]
      summary=Configuration changes profile inherited from performance created tuned

      include=performance
      # ensure intel_pstate is loaded and in active mode
      [bootloader]
      cmdline_removeKernelArgs=-intel_pstate=disable -intel_pstate=no_hwp -intel_pstate=passive

      [cpu]
      # disabled as it clashes with power manager
      enabled=false

      # ensure required modules are loaded,
      # no module options
      [modules]
      intel_cstate=
    name: openshift-intel-kpm-hotfixes
  recommend:
  - machineConfigLabels:
      machineconfiguration.openshift.io/role: intel-kpm
    priority: 15
    profile: openshift-intel-kpm-hotfixes
```

<https://github.com/intel/kubernetes-power-manager>

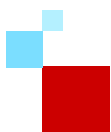


### 3. Deploy Power Profiles

```
apiVersion: "power.intel.com/v1"
kind: PowerProfile
metadata:
  name: shared-example-node1
spec:
  name: "shared-example-node1"
  max: 1500
  min: 1000
  epp: "power"
  governor: "powersave"
```

```
apiVersion: "power.intel.com/v1"
kind: PowerProfile
metadata:
  name: performance-example-node
spec:
  name: "performance-example-node"
  max: 3700
  min: 3300
  epp: "performance"
  governor: "performance"
```





## 4. Deploy Workload

```
! performance.yaml X
root > demo > ! performance.yaml
1  apiVersion: v1
2  kind: Pod
3  metadata:
4    name: performance-pod
5  spec:
6    containers:
7      - name: performance-container
8        image: ubuntu
9        command: ["/bin/sh"]
10       args: ["-c", "sleep 15000"]
11       resources:
12         requests:
13           memory: "200Mi"
14           cpu: "2"
15           power.intel.com/performance: "2"
16         limits:
17           memory: "200Mi"
18           cpu: "2"
19           power.intel.com/performance: "2"
20
```

```
! balance-power.yaml X
root > demo > ! balance-power.yaml
1  apiVersion: v1
2  kind: Pod
3  metadata:
4    name: balance-power-pod
5  spec:
6    containers:
7      - name: balance-power-container
8        image: ubuntu
9        command: ["/bin/sh"]
10       args: ["-c", "sleep 15000"]
11       resources:
12         requests:
13           memory: "200Mi"
14           cpu: "2"
15           power.intel.com/balance-power: "2"
16         limits:
17           memory: "200Mi"
18           cpu: "2"
19           power.intel.com/balance-power: "2"
20
```



1. *Journal of Management Studies*, 1997, 34(1), 1-14.  
 2. *Journal of Management Studies*, 1997, 34(1), 15-30.  
 3. *Journal of Management Studies*, 1997, 34(1), 31-46.  
 4. *Journal of Management Studies*, 1997, 34(1), 47-62.  
 5. *Journal of Management Studies*, 1997, 34(1), 63-78.  
 6. *Journal of Management Studies*, 1997, 34(1), 79-94.  
 7. *Journal of Management Studies*, 1997, 34(1), 95-110.  
 8. *Journal of Management Studies*, 1997, 34(1), 111-126.  
 9. *Journal of Management Studies*, 1997, 34(1), 127-142.  
 10. *Journal of Management Studies*, 1997, 34(1), 143-158.  
 11. *Journal of Management Studies*, 1997, 34(1), 159-174.  
 12. *Journal of Management Studies*, 1997, 34(1), 175-190.  
 13. *Journal of Management Studies*, 1997, 34(1), 191-206.  
 14. *Journal of Management Studies*, 1997, 34(1), 207-222.  
 15. *Journal of Management Studies*, 1997, 34(1), 223-238.  
 16. *Journal of Management Studies*, 1997, 34(1), 239-254.  
 17. *Journal of Management Studies*, 1997, 34(1), 255-270.  
 18. *Journal of Management Studies*, 1997, 34(1), 271-286.  
 19. *Journal of Management Studies*, 1997, 34(1), 287-302.  
 20. *Journal of Management Studies*, 1997, 34(1), 303-318.  
 21. *Journal of Management Studies*, 1997, 34(1), 319-334.  
 22. *Journal of Management Studies*, 1997, 34(1), 335-350.  
 23. *Journal of Management Studies*, 1997, 34(1), 351-366.  
 24. *Journal of Management Studies*, 1997, 34(1), 367-382.  
 25. *Journal of Management Studies*, 1997, 34(1), 383-398.  
 26. *Journal of Management Studies*, 1997, 34(1), 399-414.  
 27. *Journal of Management Studies*, 1997, 34(1), 415-430.  
 28. *Journal of Management Studies*, 1997, 34(1), 431-446.  
 29. *Journal of Management Studies*, 1997, 34(1), 447-462.  
 30. *Journal of Management Studies*, 1997, 34(1), 463-478.  
 31. *Journal of Management Studies*, 1997, 34(1), 479-494.  
 32. *Journal of Management Studies*, 1997, 34(1), 495-510.  
 33. *Journal of Management Studies*, 1997, 34(1), 511-526.  
 34. *Journal of Management Studies*, 1997, 34(1), 527-542.  
 35. *Journal of Management Studies*, 1997, 34(1), 543-558.  
 36. *Journal of Management Studies*, 1997, 34(1), 559-574.  
 37. *Journal of Management Studies*, 1997, 34(1), 575-590.  
 38. *Journal of Management Studies*, 1997, 34(1), 591-606.  
 39. *Journal of Management Studies*, 1997, 34(1), 607-622.  
 40. *Journal of Management Studies*, 1997, 34(1), 623-638.  
 41. *Journal of Management Studies*, 1997, 34(1), 639-654.  
 42. *Journal of Management Studies*, 1997, 34(1), 655-670.  
 43. *Journal of Management Studies*, 1997, 34(1), 671-686.  
 44. *Journal of Management Studies*, 1997, 34(1), 687-702.  
 45. *Journal of Management Studies*, 1997, 34(1), 703-718.  
 46. *Journal of Management Studies*, 1997, 34(1), 719-734.  
 47. *Journal of Management Studies*, 1997, 34(1), 735-750.  
 48. *Journal of Management Studies*, 1997, 34(1), 751-766.  
 49. *Journal of Management Studies*, 1997, 34(1), 767-782.  
 50. *Journal of Management Studies*, 1997, 34(1), 783-798.  
 51. *Journal of Management Studies*, 1997, 34(1), 799-814.  
 52. *Journal of Management Studies*, 1997, 34(1), 815-830.  
 53. *Journal of Management Studies*, 1997, 34(1), 831-846.  
 54. *Journal of Management Studies*, 1997, 34(1), 847-862.  
 55. *Journal of Management Studies*, 1997, 34(1), 863-878.  
 56. *Journal of Management Studies*, 1997, 34(1), 879-894.  
 57. *Journal of Management Studies*, 1997, 34(1), 895-910.  
 58. *Journal of Management Studies*, 1997, 34(1), 911-926.  
 59. *Journal of Management Studies*, 1997, 34(1), 927-942.  
 60. *Journal of Management Studies*, 1997, 34(1), 943-958.  
 61. *Journal of Management Studies*, 1997, 34(1), 959-974.  
 62. *Journal of Management Studies*, 1997, 34(1), 975-990.  
 63. *Journal of Management Studies*, 1997, 34(1), 991-1006.  
 64. *Journal of Management Studies*, 1997, 34(1), 1007-1022.  
 65. *Journal of Management Studies*, 1997, 34(1), 1023-1038.  
 66. *Journal of Management Studies*, 1997, 34(1), 1039-1054.  
 67. *Journal of Management Studies*, 1997, 34(1), 1055-1070.  
 68. *Journal of Management Studies*, 1997, 34(1), 1071-1086.  
 69. *Journal of Management Studies*, 1997, 34(1), 1087-1102.  
 70. *Journal of Management Studies*, 1997, 34(1), 1103-1118.  
 71. *Journal of Management Studies*, 1997, 34(1), 1119-1134.  
 72. *Journal of Management Studies*, 1997, 34(1), 1135-1150.  
 73. *Journal of Management Studies*, 1997, 34(1), 1151-1166.  
 74. *Journal of Management Studies*, 1997, 34(1), 1167-1182.  
 75. *Journal of Management Studies*, 1997, 34(1), 1183-1198.  
 76. *Journal of Management Studies*, 1997, 34(1), 1199-1214.  
 77. *Journal of Management Studies*, 1997, 34(1), 1215-1230.  
 78. *Journal of Management Studies*, 1997, 34(1), 1231-1246.  
 79. *Journal of Management Studies*, 1997, 34(1), 1247-1262.  
 80. *Journal of Management Studies*, 1997, 34(1), 1263-1278.  
 81. *Journal of Management Studies*, 1997, 34(1), 1279-1294.  
 82. *Journal of Management Studies*, 1997, 34(1), 1295-1310.  
 83. *Journal of Management Studies*, 1997, 34(1), 1311-1326.  
 84. *Journal of Management Studies*, 1997, 34(1), 1327-1342.  
 85. *Journal of Management Studies*, 1997, 34(1), 1343-1358.  
 86. *Journal of Management Studies*, 1997, 34(1), 1359-1374.  
 87. *Journal of Management Studies*, 1997, 34(1), 1375-1390.  
 88. *Journal of Management Studies*, 1997, 34(1), 1391-1406.  
 89. *Journal of Management Studies*, 1997, 34(1), 1407-1422.  
 90. *Journal of Management Studies*, 1997, 34(1), 1423-1438.  
 91. *Journal of Management Studies*, 1997, 34(1), 1439-1454.  
 92. *Journal of Management Studies*, 1997, 34(1), 1455-1470.  
 93. *Journal of Management Studies*, 1997, 34(1), 1471-1486.  
 94. *Journal of Management Studies*, 1997, 34(1), 1487-1502.  
 95. *Journal of Management Studies*, 1997, 34(1), 1503-1518.  
 96. *Journal of Management Studies*, 1997, 34(1), 1519-1534.  
 97. *Journal of Management Studies*, 1997, 34(1), 1535-1550.  
 98. *Journal of Management Studies*, 1997, 34(1), 1551-1566.  
 99. *Journal of Management Studies*, 1997, 34(1), 1567-1582.  
 100. *Journal of Management Studies*, 1997, 34(1), 1583-1598.  
 101. *Journal of Management Studies*, 1997, 34(1), 1599-1614.  
 102. *Journal of Management Studies*, 1997, 34(1), 1615-1630.  
 103. *Journal of Management Studies*, 1997, 34(1), 1631-1646.  
 104. *Journal of Management Studies</*



Please scan the QR Code above  
to leave feedback on this session

