Kubernetes Patterns - Chapter 6: Automated Placement (扩展版)

Reusable Elements for Designing Cloud Native Applications (Second Edition)

作者: Bilgin Ibryam & Roland Huß

本PPT详细覆盖 Automated Placement 模式，扩展为20页，深入调度过程和控制机制。

This PPT covers the Automated Placement pattern in detail, expanded to 20 pages, diving deep into the scheduling process and control mechanisms.

Chapter 6: Automated Placement (扩展版)

章节概述 (1/2)

- Chapter 6: Automated Placement (页61-75)
- 结构: Problem, Solution (Available Node Resources 等子节), Discussion, More Information

> *Automated Placement is the core function of the Kubernetes scheduler...*

关键点:

- 焦点: Pod 到节点的自动分配

Chapter 6: Automated Placement (扩展版)

章节概述 (2/2)

- 与 Foundational Patterns 相关 (Part I)
- 目标: 优化可用性、性能和容量

> *It is an area that is highly configurable, still evolving, and changing rapidly...*

关键点:

- 调度器高度可配置且演进中

Chapter 6: Automated Placement (扩展版)

Problem - 放置挑战 (1/2)

- 微服务系统有大量进程，手动放置不可扩展
- Pod 抽象好，但不解决节点分配

> *With a large and ever-growing number of microservices, assigning and placing them individually to nodes is not a manageable activity...*

关键点:

- 示例: 数十或数百微服务

Chapter 6: Automated Placement (扩展版)

Problem - 放置挑战 (2/2)

- 依赖、资源需求动态变化；集群资源变异
- 放置影响可用性、性能、容量

> *All of that makes scheduling containers to nodes a moving target that has to be shot on the move.*

关键点:

- "Moving target that has to be shot on the move."

Chapter 6: Automated Placement (扩展版)

Solution - 概述 (1/2)

- Kubernetes 调度器负责 Pod 分配
- 考虑依赖、资源、HA 策略；水平扩展和共置

> *It does this by considering runtime dependencies, resource requirements...*

关键点:

- 主要控制机制覆盖

Chapter 6: Automated Placement (扩展版)

Solution - 概述 (2/2)

- 子节: Available Node Resources, Container Demands, Scheduler Configs, Process, Affinity 等
- 确保 Pod 匹配节点容量

> *The scheduler has a default set of predicate and priority policies configured...*

关键点:

- 调度器默认策略适合多数用例

Chapter 6: Automated Placement (扩展版)

Available Node Resources - 概念

- 节点容量 (CPU, 内存, 存储)
- 节点分配 (已分配资源)
- 节点标签 (用于选择)
- 污点 (Taints) - 用于排斥

> *Node capacity, allocation, labels, and taints are all important factors in scheduling.*

关键点:

- 节点资源是调度的基础

Chapter 6: Automated Placement (扩展版)

Container Demands - 概念

- 资源请求 (requests) 和限制 (limits)
- 端口需求
- 存储卷挂载
- 环境变量和配置

> *Containers have resource demands, port requirements, volume mounts, and configuration needs.*

关键点:

- 容器需求驱动调度决策

Chapter 6: Automated Placement (扩展版)

Scheduler Configs - 概念

- 调度策略配置
- 资源评分算法
- 自定义调度器
- 多调度器支持

> *Scheduler configurations define policies, algorithms, and custom schedulers.*

关键点:

- 调度配置决定调度行为

Chapter 6: Automated Placement (扩展版)

Process - 概念

- 过滤 (Predicates) - 筛选合适节点
- 优先级 (Priorities) - 对节点评分
- 绑定 (Binding) - 分配Pod到节点
- 抢占 (Preemption) - 必要时驱逐低优先级Pod

> *The scheduling process involves filtering, prioritizing, binding, and preemption.*

关键点:

- 调度是多阶段过程

Chapter 6: Automated Placement (扩展版)

Node Selector - 概念

- 最简单的节点选择机制
- 基于节点标签匹配
- 硬性要求，不满足则Pod不会调度
- 适用于简单场景

> *Node selector is the simplest way to constrain pods to nodes with specific labels.*

关键点:

- 节点选择器是最基本的调度约束

Chapter 6: Automated Placement (扩展版)

Node Affinity - 概念

- 比Node Selector更灵活的节点选择
- 支持软约束(Preferred)和硬约束(Required)
- 基于节点标签的表达式匹配
- 支持多种操作符: In, NotIn, Exists, DoesNotExist

> *Node affinity provides more flexible node selection than node selectors.*

关键点:

- 节点亲和性提供更灵活的调度控制

Chapter 6: Automated Placement (扩展版)

Pod Affinity and Anti-Affinity - 概念

- 基于Pod标签的亲和性调度
- Affinity: 将Pod调度到相同拓扑域
- Anti-Affinity: 将Pod分散到不同拓扑域
- 支持硬约束和软约束

> *Pod affinity and anti-affinity allow co-locating or separating pods based on labels.*

关键点:

- Pod亲和性控制Pod间的共置关系

Chapter 6: Automated Placement (扩展版)

Topology Spread Constraints - 概念

- 控制Pod在拓扑域间的分布
- 支持跨区域、可用区、节点等维度
- 确保应用的高可用性
- 避免资源浪费和单点故障

> *Topology spread constraints control how pods are distributed across topology domains.*

关键点:

- 拓扑分布约束优化Pod分布

Chapter 6: Automated Placement (扩展版)

Taints and Tolerations - 概念

- Taint: 节点上的排斥标记
- Toleration: Pod上的容忍标记
- 实现节点专用化和污染隔离
- 支持NoSchedule, PreferNoSchedule, NoExecute

> *Taints and tolerations allow nodes to repel pods, enabling node specialization.*

关键点:

- 污点和容忍实现节点专用化

Chapter 6: Automated Placement (扩展版)

Node Selector - 配置示例

```
spec:
  nodeSelector:
    disktype: ssd
    accelerator: nvidia-tesla-p100
```

> *Node selector constrains pods to nodes with matching labels.*

关键点:

- 节点选择器配置简单直观

Chapter 6: Automated Placement (扩展版)

Node Affinity - 配置示例

```
spec:
  affinity:
    nodeAffinity:
      requiredDuringSchedulingIgnoredDuringExecution:
        nodeSelectorTerms:
        - matchExpressions:
          - key: kubernetes.io/hostname
            operator: In
            values:
            - node-1
            - node-2
```

> *Node affinity provides more flexible node selection than node selectors.*

关键点:

- 节点亲和性配置支持复杂的匹配规则

Chapter 6: Automated Placement (扩展版)

Pod Affinity and Anti-Affinity - 配置示例

```
spec:
  affinity:
    podAntiAffinity:
      requiredDuringSchedulingIgnoredDuringExecution:
      - labelSelector:
          matchExpressions:
          - key: app
            operator: In
            values:
            - web
        topologyKey: kubernetes.io/hostname
```

> *Pod affinity and anti-affinity allow co-locating or separating pods based on labels.*

关键点:

- Pod亲和性配置控制Pod间的共置关系

Chapter 6: Automated Placement (扩展版)

Topology Spread Constraints - 配置示例

```
spec:
  topologySpreadConstraints:
  - maxSkew: 1
    topologyKey: topology.kubernetes.io/zone
    whenUnsatisfiable: DoNotSchedule
    labelSelector:
      matchLabels:
        app: web
```

> *Topology spread constraints control how pods are distributed across topology domains.*

关键点:

- 拓扑分布约束配置优化Pod分布