CLOUDERA

APACHE YUNIKORN ENHANCED SCHEDULING IN THE CLOUD

Wilfred Spiegelenburg
Manikandan Ramaraj

APACHECON
North America

AGENDA

Why Apache YuniKorn

Architecture

Queues: hierarchies and quotas

Gang Scheduling

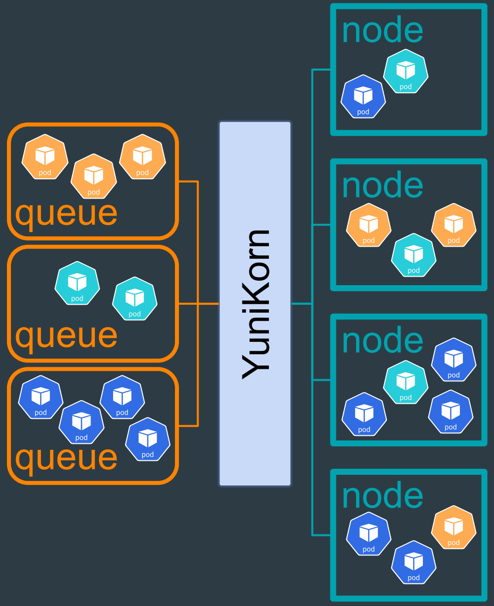
User and Group Quotas

Roadmap

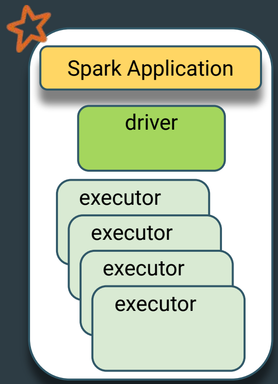
© 2022 Cloudera, Inc. All rights reserved.

WHY APACHE YUNIKORN

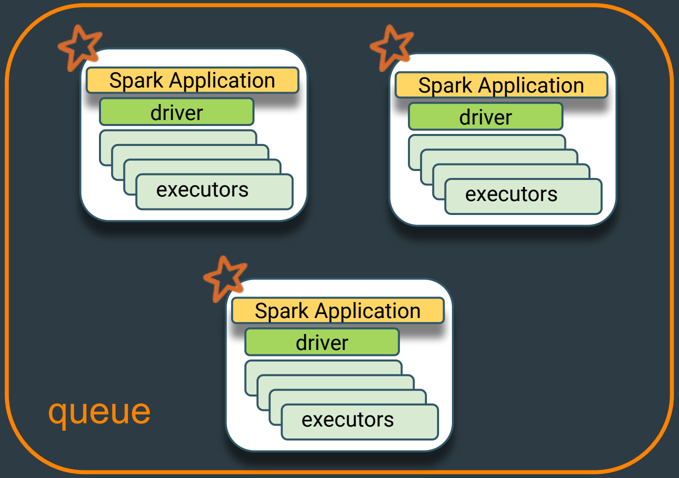
Advanced scheduling requirements



Workload Queueing



Gang Scheduling



Application Sorting

© 2022 Cloudera, Inc. All rights reserved.

AGENDA

Why Apache YuniKorn

Architecture

Queues: hierarchies and quotas

Gang Scheduling

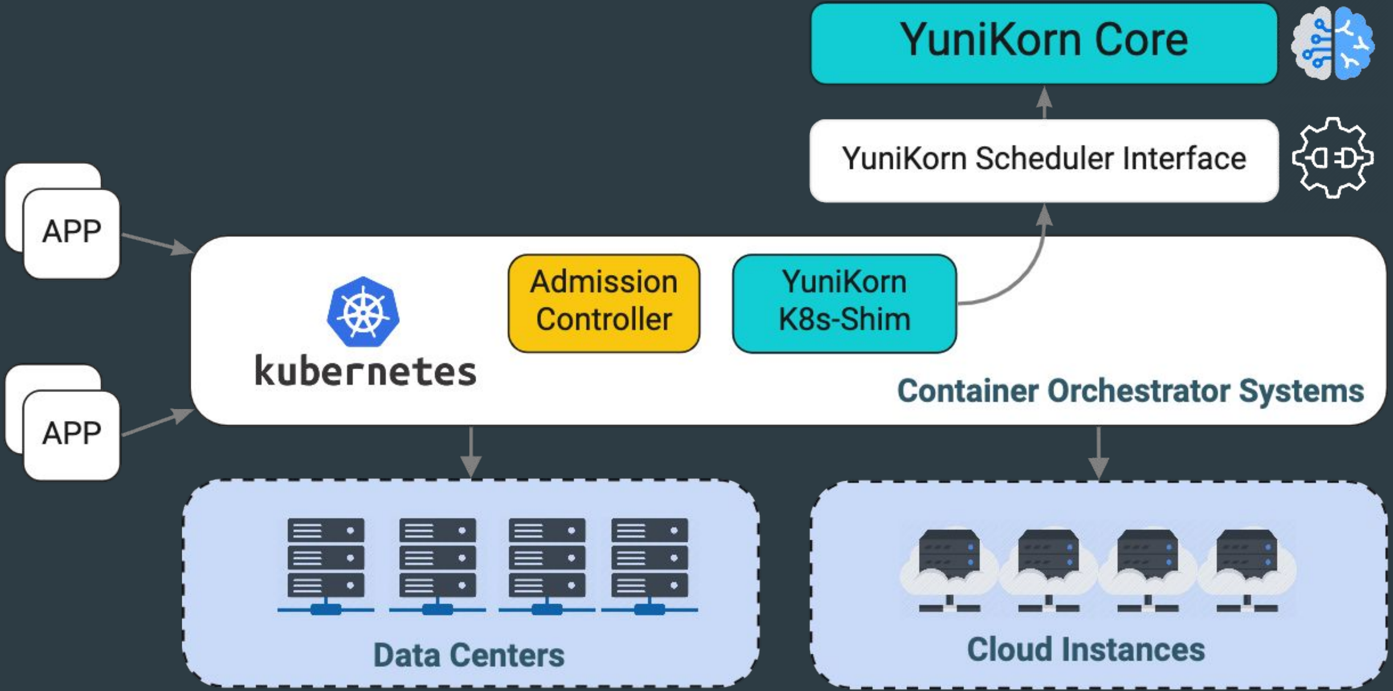
User and Group Quotas

Roadmap

© 2022 Cloudera, Inc. All rights reserved.

ARCHITECTURE

Basics

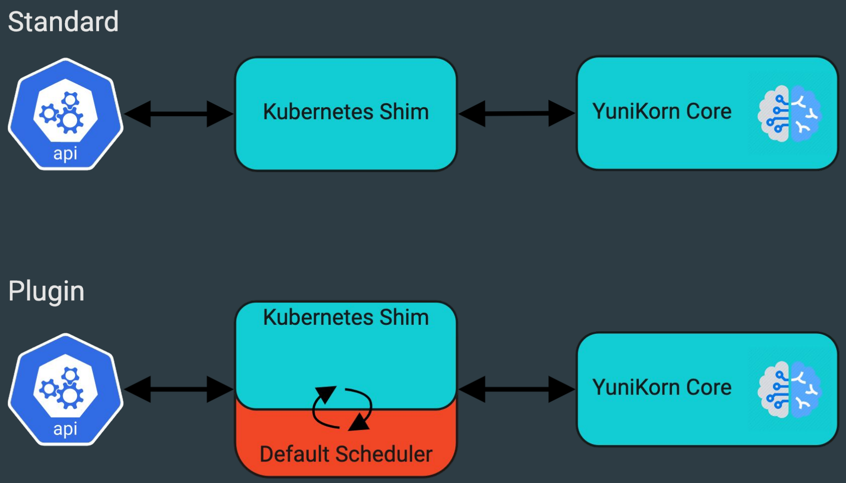


© 2022 Cloudera, Inc. All rights reserved.

ARCHITECTURE

Deployment models

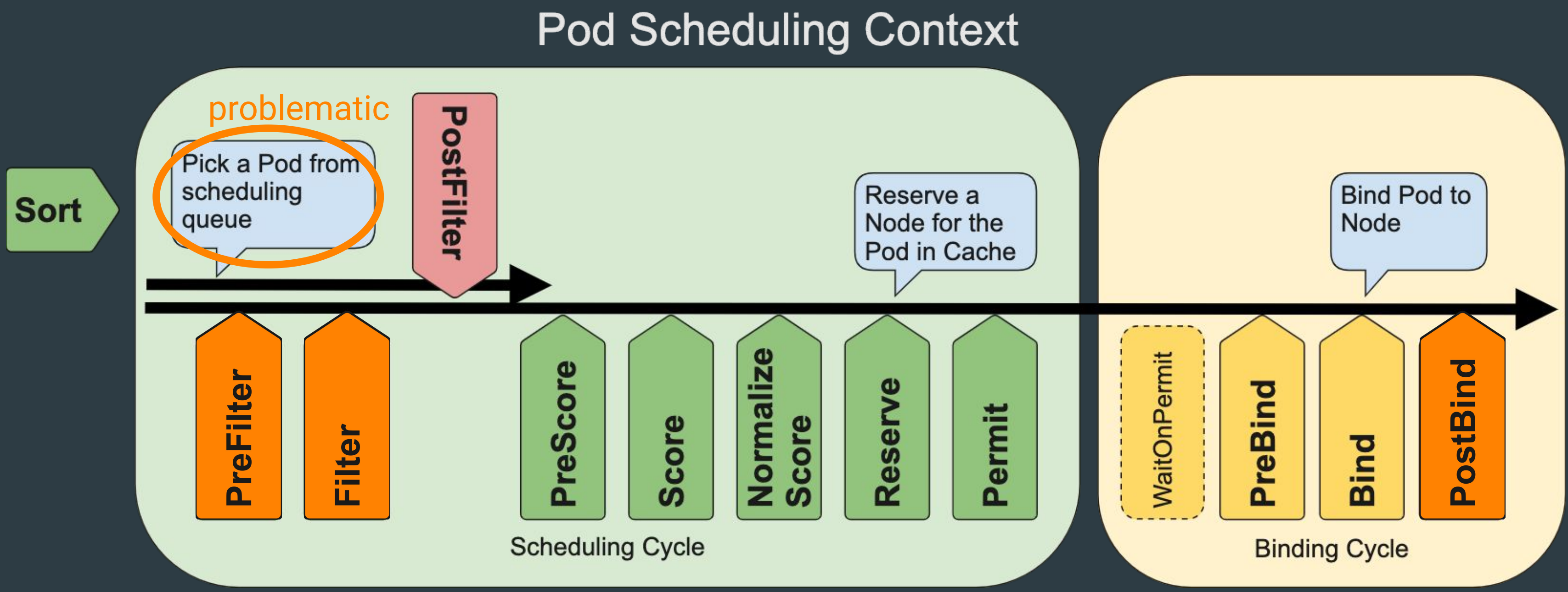
- STANDARD:
  - Custom scheduler
  - Replaces *default* scheduler
- PLUGIN:
  - Kubernetes Scheduling Framework (API)
  - Replace or augment limited functionality



© 2022 Cloudera, Inc. All rights reserved.

PLUGIN ARCHITECTURE

Extension points



© 2022 Cloudera, Inc. All rights reserved.

KUBERNETES COMMUNITY

- Multiple initiatives around batch and HPC schedulers
  - Apache YuniKorn
  - Volcano
  - Kueue
  - Armada
- Scheduler plugins for specific point solutions
- Fragmentation concerns: KubeCON NA panel discussion
- Pre-enqueue plugin
  - KEP-3521: Pod Schedule Readiness
- Pre-emption behaviour: priority only
  - K8s has: do not preempt other pods during scheduling
  - Batch needs: do not preempt the tagged pod while running

© 2022 Cloudera, Inc. All rights reserved.

AGENDA

Why Apache YuniKorn

Architecture

Queues: hierarchies and quotas

Gang Scheduling

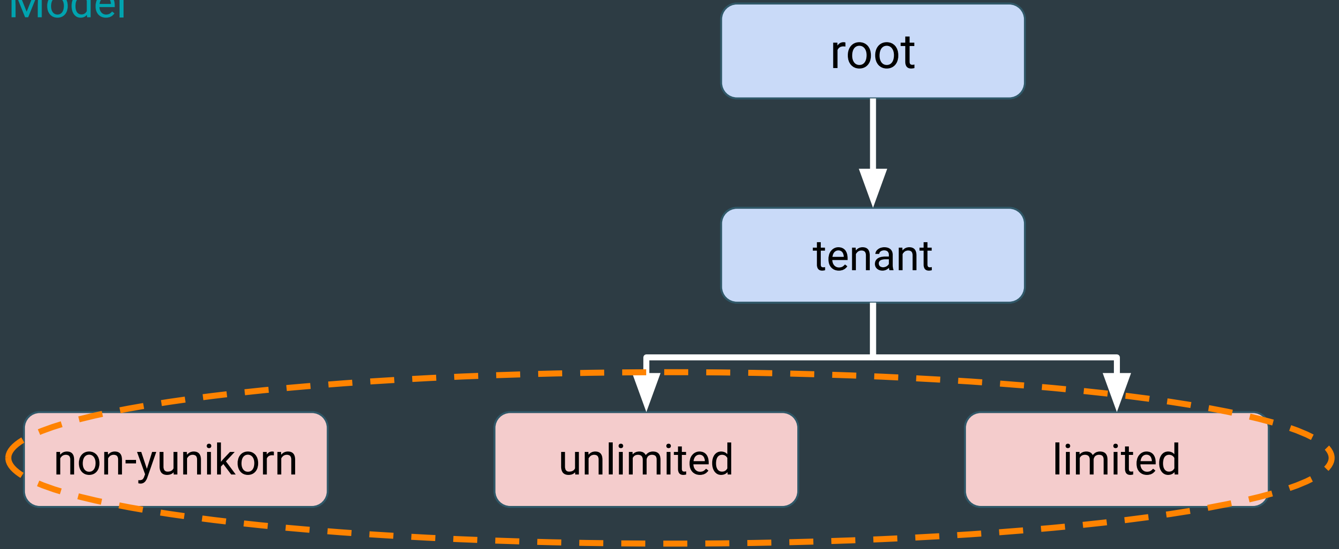
User and Group Quotas

Roadmap

© 2022 Cloudera, Inc. All rights reserved.

QUEUES, HIERARCHIES & QUOTAS

Hierarchical Model

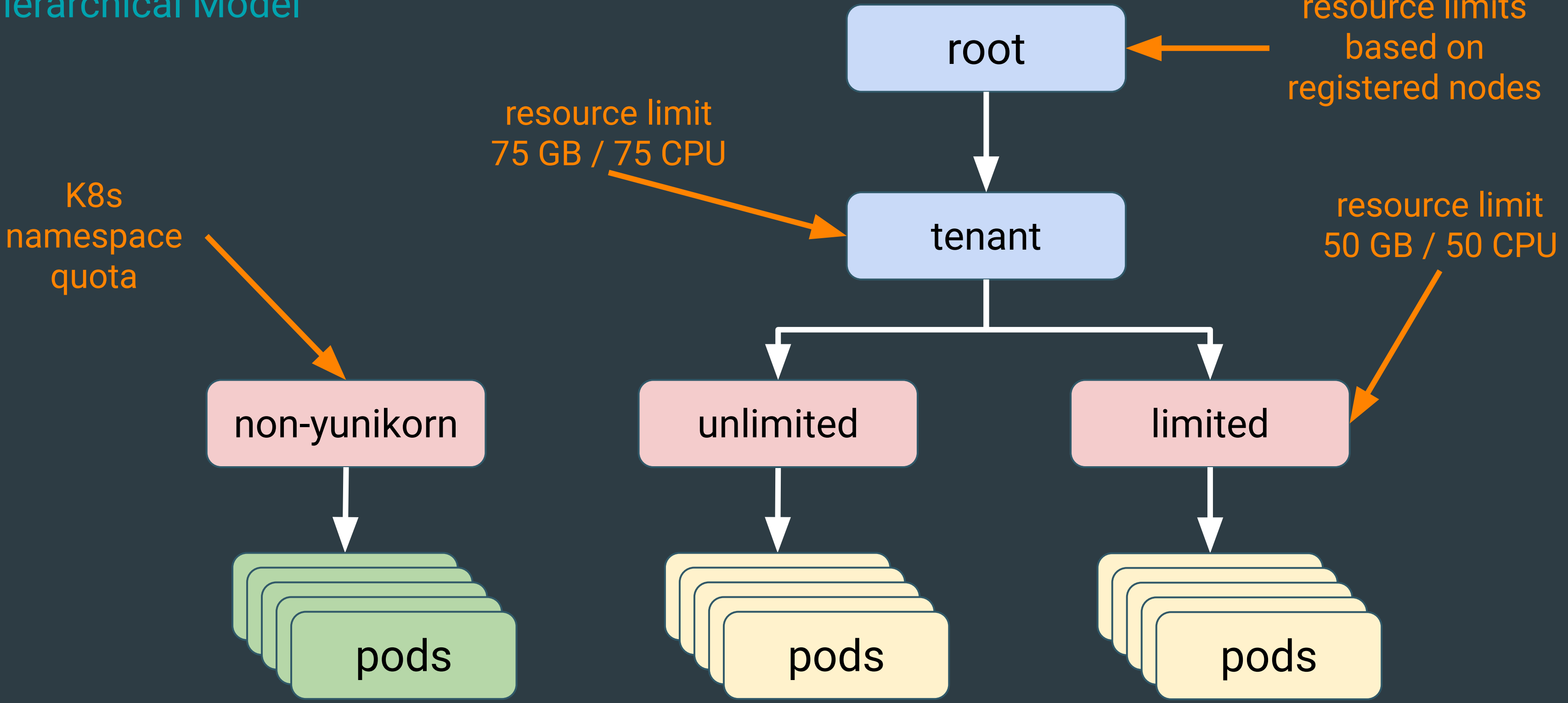


namespaces
defined in k8s

© 2022 Cloudera, Inc. All rights reserved.

QUEUES, HIERARCHIES & QUOTAS

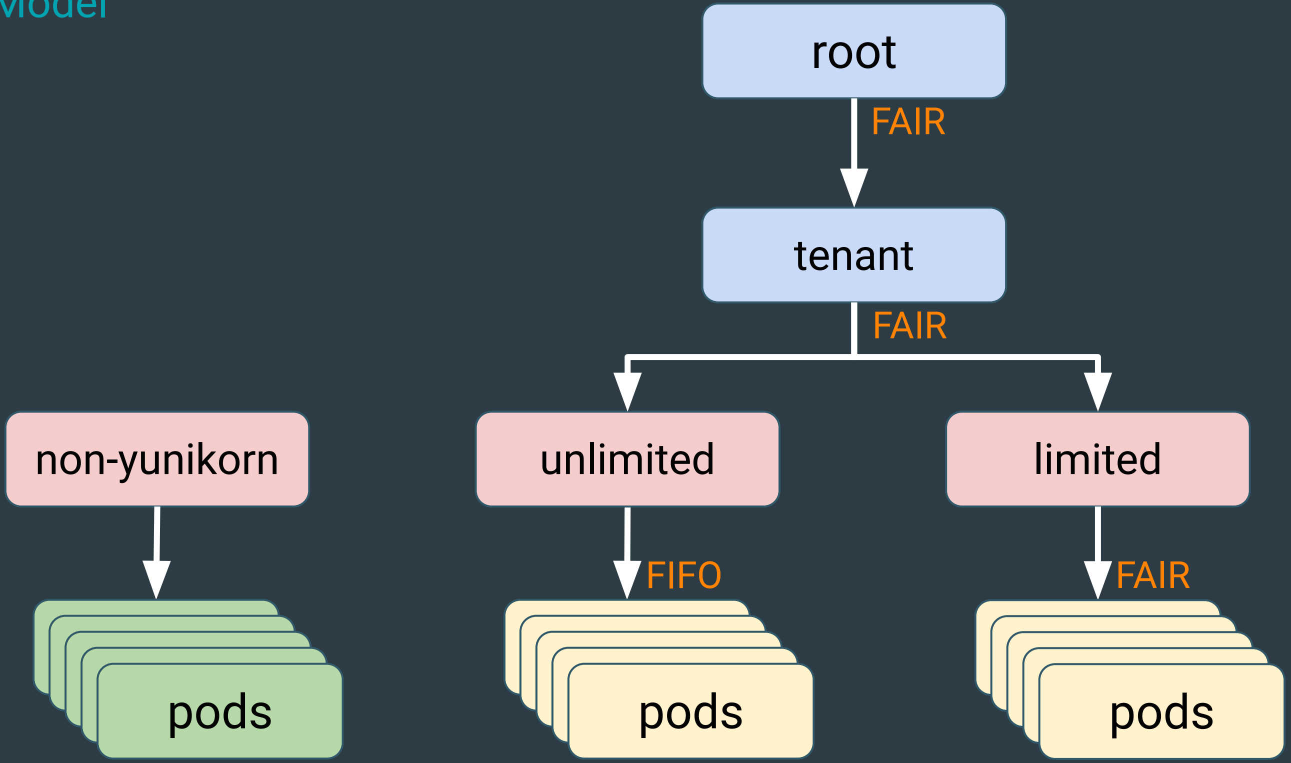
Hierarchical Model


root
resource limits based on registered nodes
resource limit 75 GB / 75 CPU
tenant
K8s namespace quota
resource limit 50 GB / 50 CPU
non-yunikorn
unlimited
limited
pods
pods
pods

© 2022 Cloudera, Inc. All rights reserved.

QUEUES, HIERARCHIES & QUOTAS

Hierarchical Model



© 2022 Cloudera, Inc. All rights reserved.

AGENDA

Why Apache YuniKorn

Architecture

Queues: hierarchies and quotas

■ **Gang Scheduling**

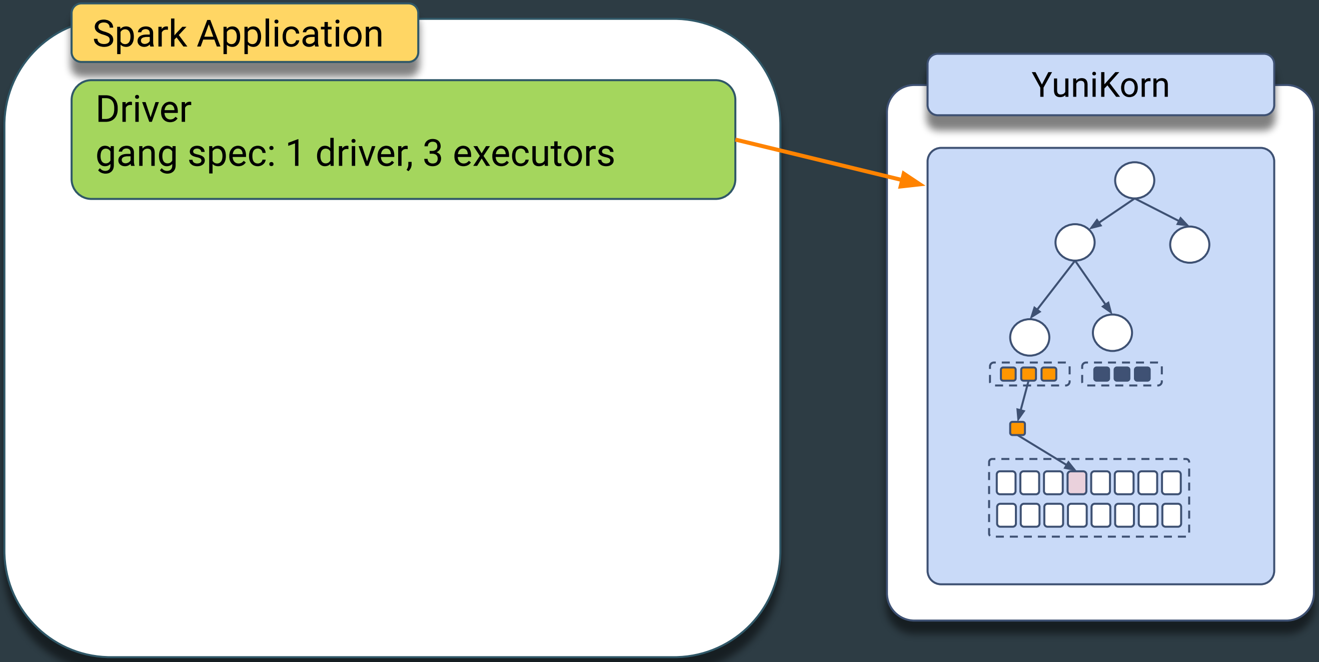
User and Group Quotas

Roadmap

© 2022 Cloudera, Inc. All rights reserved.

GANG SCHEDULING

Gang example: Spark

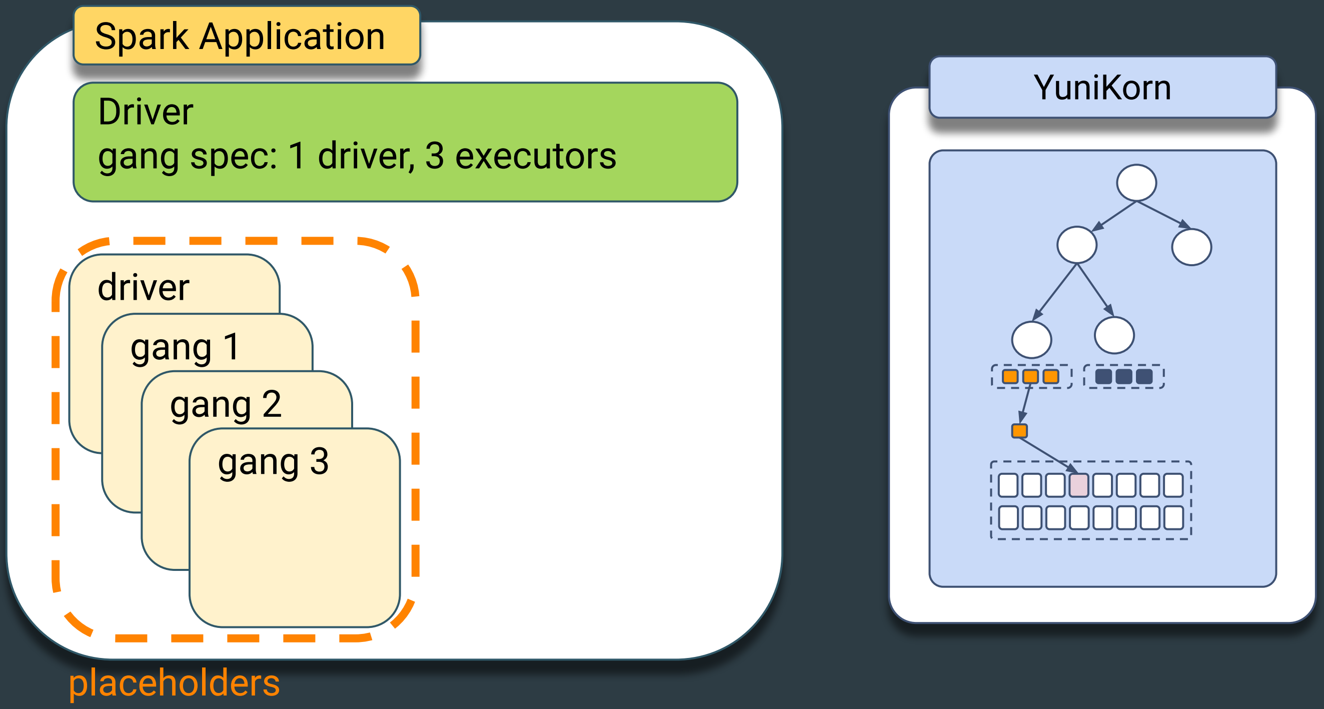


placeholders

© 2022 Cloudera, Inc. All rights reserved.

GANG SCHEDULING

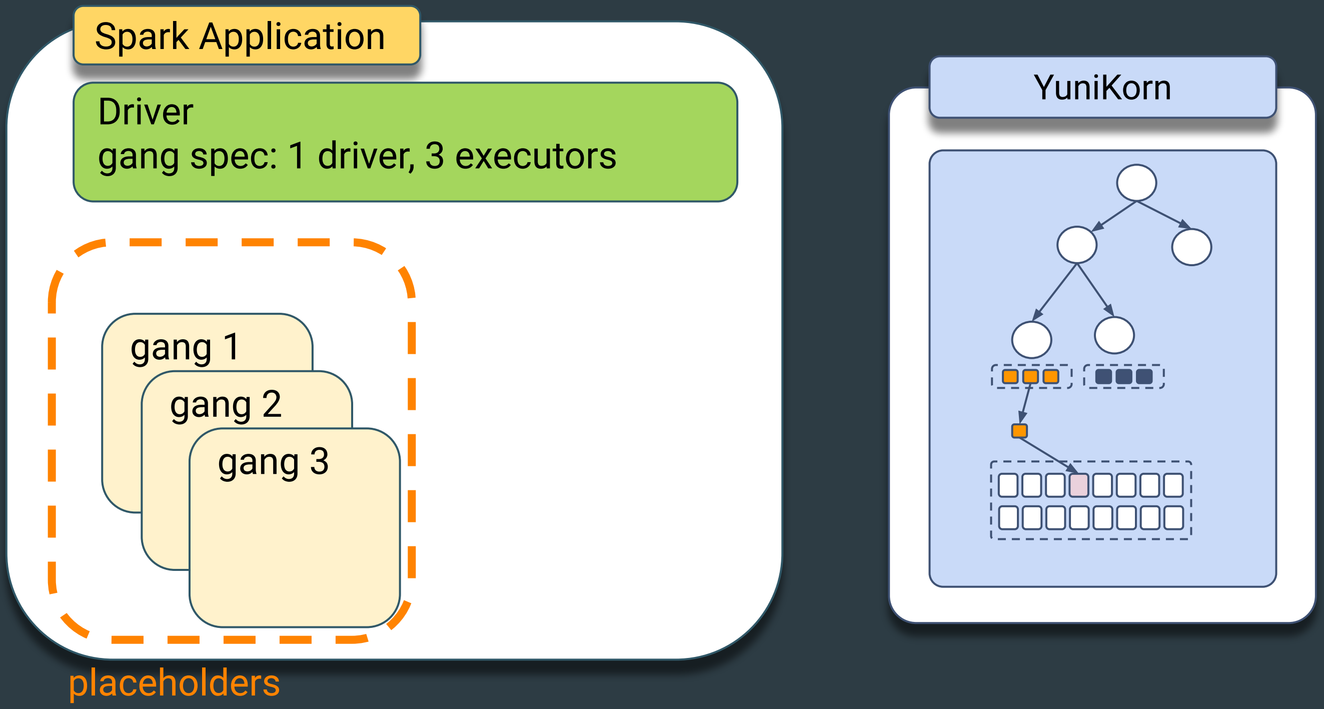
Gang example: Spark



© 2022 Cloudera, Inc. All rights reserved.

GANG SCHEDULING

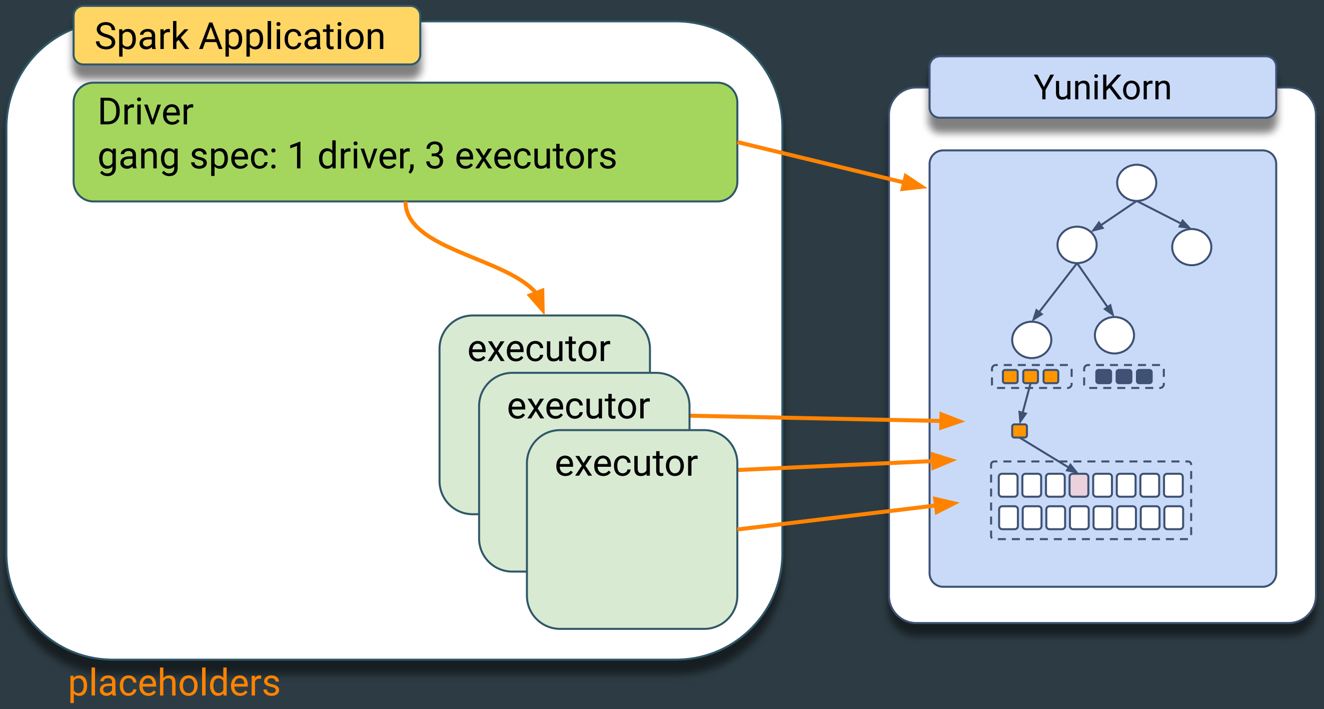
Gang example: Spark



© 2022 Cloudera, Inc. All rights reserved.

GANG SCHEDULING

Gang example: Spark



© 2022 Cloudera, Inc. All rights reserved.

AGENDA

Why Apache YuniKorn

Architecture

Queues: hierarchies and quotas

Gang Scheduling

User and Group Quotas

Roadmap

© 2022 Cloudera, Inc. All rights reserved.

USER AND GROUP INFORMATION

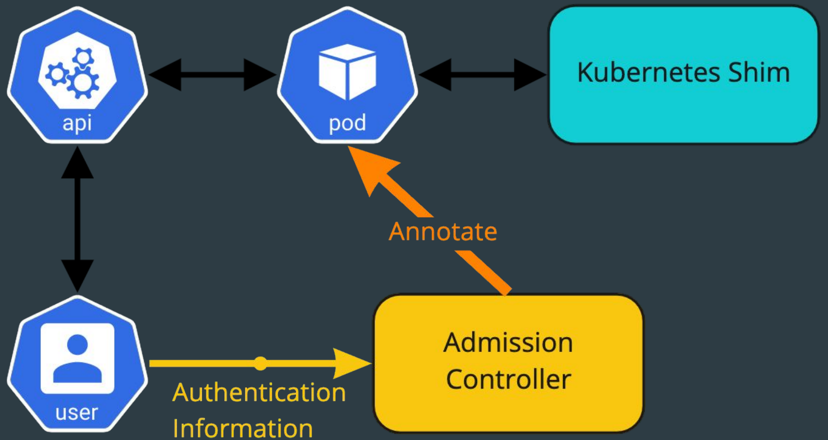
Two options

Authenticated user:

- User information based on authenticated user
- Set by admission controller on pod creation

Annotation on pod:

- User information provided on pod creation
- Check & override by admission controller



© 2022 Cloudera, Inc. All rights reserved.

USER AND GROUP INFORMATION

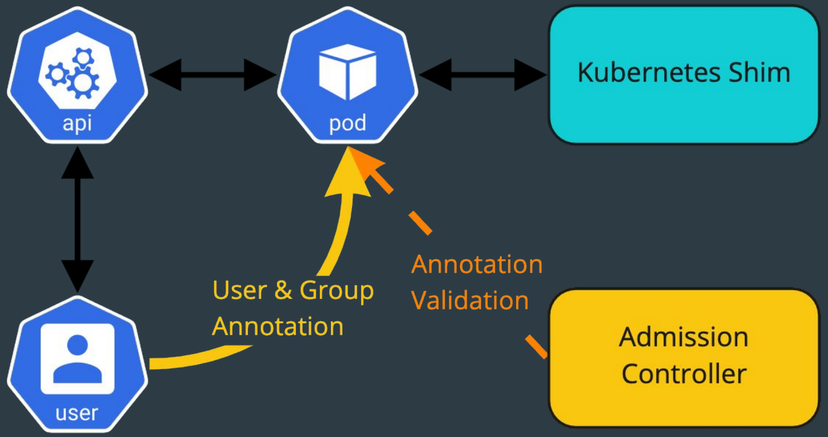
Two options

Authenticated user:

- User information based on authenticated user
- Set by admission controller on pod creation

Annotation on pod:

- User information provided on pod creation
- Check & override by admission controller

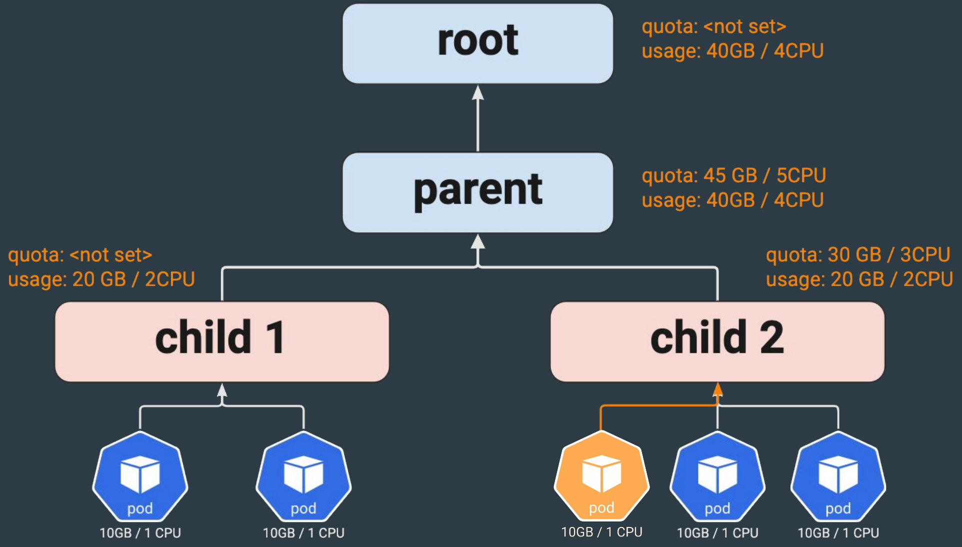


© 2022 Cloudera, Inc. All rights reserved.

USER AND GROUP QUOTAS

Queues vs User & Groups

- Queue quota: root to leaf
  - Integrated in scheduling cycle
  - No quota available queue will be skipped
- User and group: leaf to root
  - Request fits in queue *headroom*
  - Last check before confirming node placement



© 2022 Cloudera, Inc. All rights reserved.

AGENDA

Why Apache YuniKorn

Architecture

Queues: hierarchies and quotas

Gang Scheduling

User and Group Quotas

Roadmap

© 2022 Cloudera, Inc. All rights reserved.

Roadmap

With all appropriate caveats…

Apache YuniKorn 1.2

- Application limits
  - Maximum running applications per Queue
- User and Group quotas
  - Implementation started for:
    - User retrieval
    - Usage tracking
  - Enforcement design in final stage

Future release

- Priority support
  - Early design stage
- Pre-emption
  - Obsolete code removed (YuniKorn 1.1)
  - Design starting

© 2022 Cloudera, Inc. All rights reserved.

CLOUDERA

Q & A

© 2022 Cloudera, Inc. All rights reserved.

THANK YOU

CLOUDERA