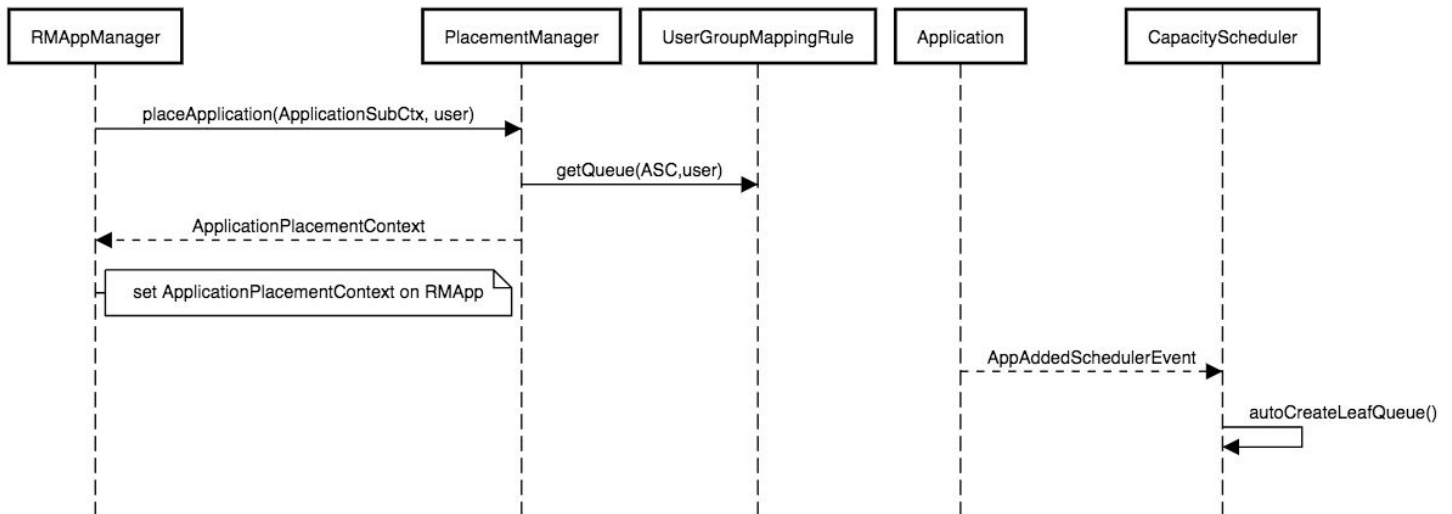


# YARN-7117 Workflow

## Queue Creation



On application submission, **RMAAppManager** calls a queue mapping service for the application to determine which queue the application should be placed. This is delegated to the **PlacementManager** which goes through a set of pre-configured placement rule/policies. The default is **UserGroupPlacementRule** which can return a specific queue based on the user/group mapping defined in [YARN-3635](#).

PlacementManager returns an **ApplicationPlacementContext** which contains the queue which has been mapped, its parent queue(if it is specified in placement rule) and the **PlacementRule** that was used to resolve the queue mapping.

The queues created automatically are all instantiated as **AutoCreateEnabledLeafQueue** which are managed by its parent queue **AutoCreateEnabledParentQueue**.

# Capacity Management

A scheduled service **QueueEntitlementDynamicEditPolicy** is invoked at a configurable interval (`yarn.resourcemanager.monitor.capacity.queue-entitlement.monitoring_interval`) by **SchedulingMonitor** in RM to compute the queue capacity entitlements for all **AutoCreatedLeafQueue(s)** under a given **AutoCreateEnabledParentQueue**.

All auto created leaf queues are initially assigned **0** capacity and the above service computes new queue entitlements and returns a **List<QueueEntitlementChange>** based on pending/schedule-able applications at that point in time. Since the queues are created dynamically, there might be more number of queues created than the parent's capacity guarantees and some of the queues may not be used after their creation. To address this, these leaf queues are allotted capacity only when there is a pending workload and deactivated when they do not have any pending applications. While determining the order in which leaf queues can be allocated guaranteed capacity, the applications are sorted by their submission time and the corresponding leaf queues are allocated capacity. For more details on capacity management for auto created leaf queues within a parent queue, refer the [design](#) document.

The queue entitlements are then submitted to the scheduler which validates and applies the queue entitlements.

