Title: [YARN-4091] Improve debug/diagnostic messages in CapacityScheduler
Authors: Sunil Govind, Rohith Sharma, Nijel S F with inputs from Wangda Tan, Varun and Jian
Last modified: August 21 2015

Preamble

To Improve debug capability in YARN, specifically in schedulers, is one of an area where we are lagging now. One of the main reason for this is the support of various configurations which tunes the schedulers to take actions such as limit assigning containers to an application, or introduce delay to allocate container etc. There are no clear information passed down from scheduler to outerworld under these various scenarios hence making debugging tougher.

This is a proposal to add more debug and diagnostic informations in various places in Capacity Scheduler so that user/admin will have more information about what is happening inside scheduler when different NMs updates its resource usage in regular heartbeat.

Problem Statement

Scheduler is tend to skip or reject few container assignments or restrict application itself become active due to the fine tuning features present in Capacity Scheduler. Few cases are

- node locality delay
- application priority
- capacity limits of queue
- user limits (includes node label partition)
- FIFO waiting based on application submission time
- Fairness waiting (based on resource usage)
- application master resource limit
- reservation

Such cases are not informed back to user in a clear way, hence it is hard to derive why a specific container is not allocated under certain circumstances.

Proposal

Main intention of this proposal is to supply more information about the state of an application or container and this information has to be taken to outside world via REST or GUI. Hence one of the main focus will be to make sure that performance and maintainability of scheduler is not tampered with.

- Storage

Few bytes storage per application or container level is what can be afford now. Hence the proposal is to store the debug or diagnostic information as an enum. This enum state can be kept per container/application level to indicate what has happened with it lately.

Meaningful enum states will be defined as per the various conditions mentioned in Problem Statement earlier. In future, more defined debug state can be introduced and with such abundant information set, user will be able to know the current state of each container/application within scheduler lively.

Two states will be helpful here to store some primary details and detailed information. For Example, some toplevel hardcoded debug states are,

```
enum ApplicationState {
SKIP_CONTAINER_ASSIGNMENT,
REJECT_CONTAINER_ASSIGNMENT,
SUCCESS_CONTAINER_ASSIGNMENT,
DISALLOW_ACTIVATE_APPLICATION,
RUNNING_APPLICATION
}

enum DiagnosticsInfo {
AM_RESOURCE_LIMIT_PER_QUEUE_EXCEEDED,
USER_LIMIT_EXCEEDED,
LOCALITY_WAIT,
PRIORITY_WAIT,
FIFO_WAIT
}
```

More enums will be added as we go along with this proposal. This is a sample one to show the basic information.

- Accessibility

Information about the last state of a container/application can be fetched via below mechanisms

  - REST api can have this additional information
  - UI to display debug information along with container/application

Others

This proposal focus on schedulers now, but few more useful informations can be updated to RMApp and RMAppAttempt as needed.