CLOUD COMPUTING APPLICATIONS

YARN: Node Manager

Roy Campbell & Reza Farivar

Node Manager Responsibilities

- Launch containers
- Localize files for containers
- Report container status to RM
- Kill containers and clean up local storage
- Monitor container resource usage and kill misbehaving containers
- Run health check and notify RM if unhealthy
- Log aggregation
- Web app for reporting node status and applications/containers
- Web services to avoid screen-scraping
- Runs auxiliary services for application frameworks as plugins
  - MapReduce ShuffleHandler serves map outputs to reducers

Node Manager Container Localization

- Download files/directories to local filesystem for the container
  - .jar file or executable for container
  - Supporting files, libraries
  - Known as the Distributed Cache in MapReduce
- Container launch context lists entries needing to be localized
- Localization requests have three privilege modes
  - Public: all users can access
  - Private: only the user can access
  - Application: only the application can access
- Downloaded entries are symlinked into the container working directory

Node Manager Log Aggregation

- NM collects logs from all containers for an application
- Collated into a file per node per application and uploaded to HDFS
- Benefits:
  - Allows long-term application log retention
  - Increases log availability
  - Easier to post-process logs via MapReduce jobs