Task Execution Environment

- Hadoop provide info to M/R task about env it's running
- map task can discover name of file it is processing
- M/R task can find out the attempt no of task.
- 7-3 properties can be accessed from
  - JobConfiguration - configure() of MR - old API
  - Context object - new API

Speculative Execution

- MR model break ~~tas~~ Jobs into task and run in parallel to make Job exe time smaller
- Realworld a Job - 100/1000 of task so few task can be slow
- Task may be slow for various reason
  - hardware degradation
  - S/w misconfiguration
- Hard to detect / Hadoop doesn't diagnose and fix slow task
- Instead detect and launches another task (speculative execution)
- when task complete successfully - duplicate killed
- S.E is optimization / not feature to make Job run more reliably.
- If bugs - cause task to hang / slow - rely of S.E to avoid problem not wise
- Turn (default) - .
- Enabled / Disable - independent for M/R task
- 7.4 fig
- Goal - reduce Job Exe time but at the cost of Cluster efficiency
- on busy cluster it can reduce overall throughput - redundant tasks

Date: Page:

- Usualy turnit off on cluster - user explicitly tun it on for individual Jobs.
- Good to turn off for reduce task - duplicate reduce task to have fetch same map o/p as original task
  Can increase n/w traffic on cluster
- Another (off) is non idempotent tasks.