Desenvolvimento de um Escalonador Sensível ao Contexto para o *Apache Hadoop*

Guilherme Weigert Cassales[1]
Orientador: Profª Drª Andrea Schwertner Charão[1]

[1]Ciência da Computação
Universidade Federal de Santa Maria

13 de novembro de 2013

informáticaufsm
Ciência da Computação

Roteiro

informáticaufsm
Ciência da Computação

Roteiro

informáticaufsm
Ciência da Computação

Introdução

- Relevância do *framework Apache Hadoop*.
- Problemas do *Hadoop*:
  - desenvolvimento muito específico;
  - pouco adaptável.
- Diversas soluções:
  - **escalonamento** de acordo com as *características* dos nós;
  - volatilidade dos nós utilizados no *cluster*;
  - auto configuração do ambiente de execução.

informáticaufsm
Ciência da Computação

Objetivo e Justificativas

- Objetivo
  - Desenvolvimento de um escalonador sensível ao contexto.
- Justificativas:
  - problemas de performance em situações de heterogeneidade;
  - problemas em manter um sistema homogêneo atualmente.

informáticaufsm
Ciência da Computação

Roteiro

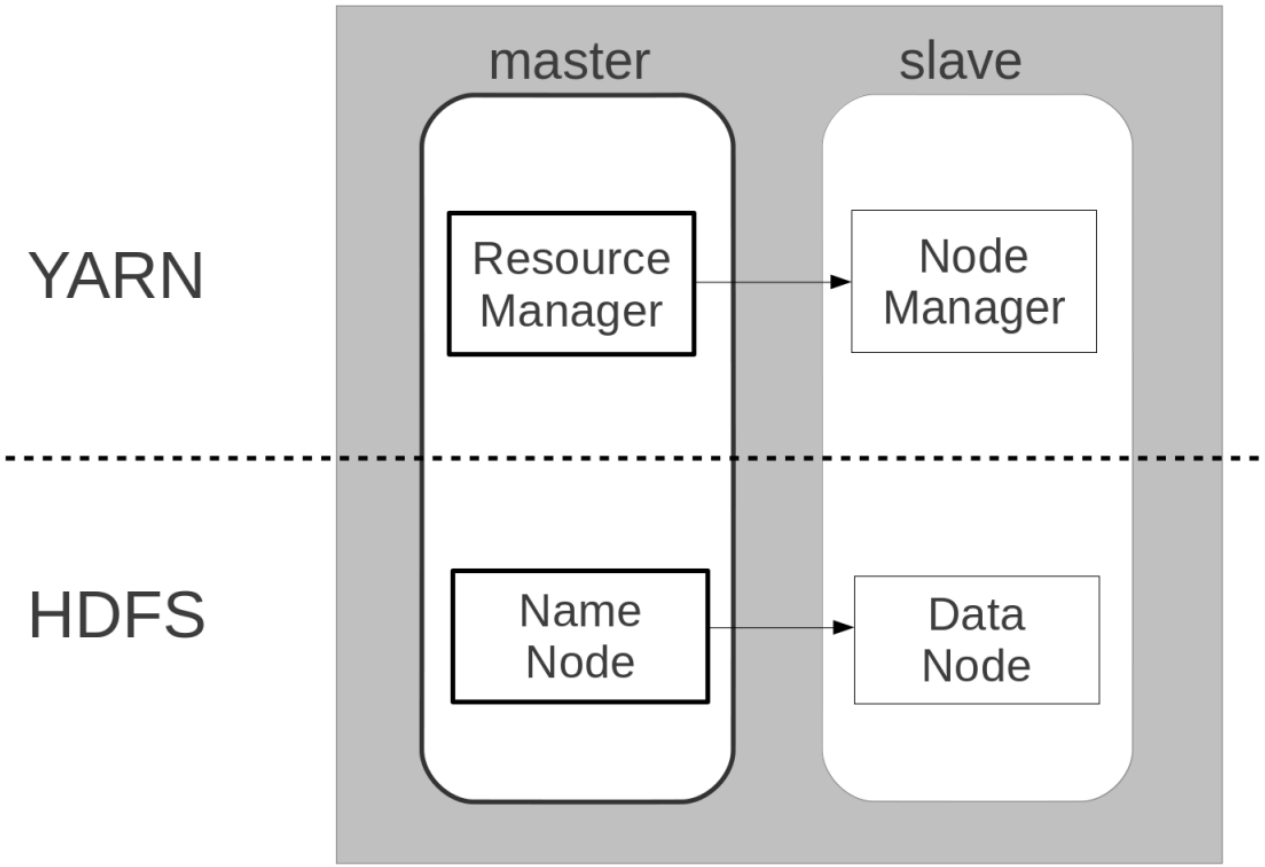
informáticaufsm
Ciência da Computação

Sensibilidade ao Contexto

- Contexto
  - "qualquer informação que pode ser utilizada para caracterizar a situação de uma entidade (pessoa, lugar ou objeto) considerada relevante para a interação entre usuário e aplicação"(DEY, 2001).
- Sensibilidade ao Contexto
  - "se refere a habilidade de uma aplicação de detectar e responder as mudanças no ambiente de execução"(Maamar; Benslimane; Narendra, 2006).
- Ganhos na utilização
- Como um *software* utiliza?

informáticaufsm
Ciência da Computação

Arquitetura Geral do *Hadoop*

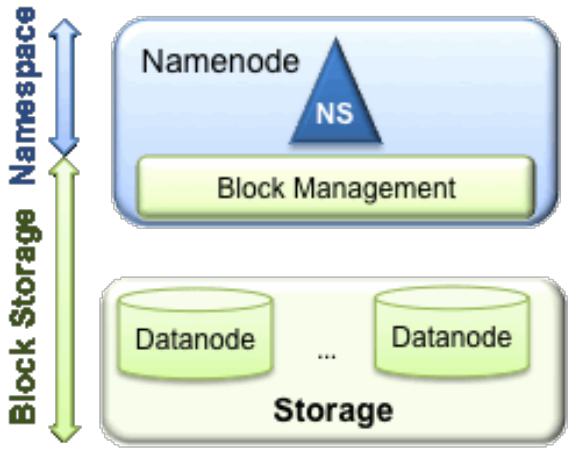


informáticaufsm
Ciência da Computação

Arquitetura Geral do *Hadoop*

HDFS

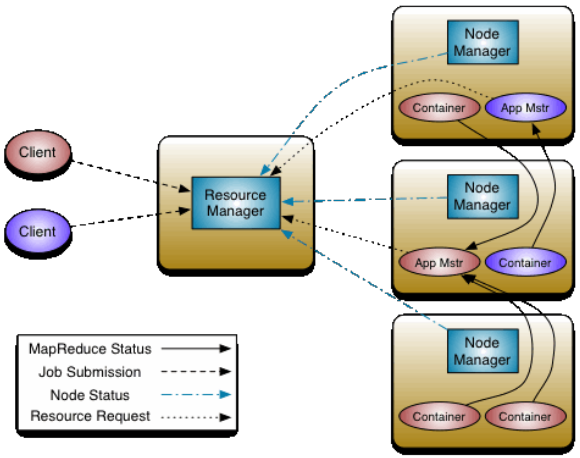
- NameNode
- DataNode



(HADOOP, 2013)

YARN

- ResourceManager
- NodeManager



(HADOOP, 2013)

informáticaufsm
Ciência da Computação

Escalonadores Padrão

- Internal
  - FIFO.
- Fair
  - Distribuição igualitária de recursos.
  - Dois níveis de fila.
- Capacity
  - Divisão de um *cluster* entre várias empresas.
  - Política de *MinShare*.

informáticaufsm
Ciência da Computação

Trabalhos relacionados

- (Kumar et al., 2012).
- (Rasooli; Down, 2012).
- (Chen et al., 2010).
- (Xie et al., 2010).
- (Tian et al., 2009).
- (Isard et al., 2009).
- (Zaharia et al., 2008).

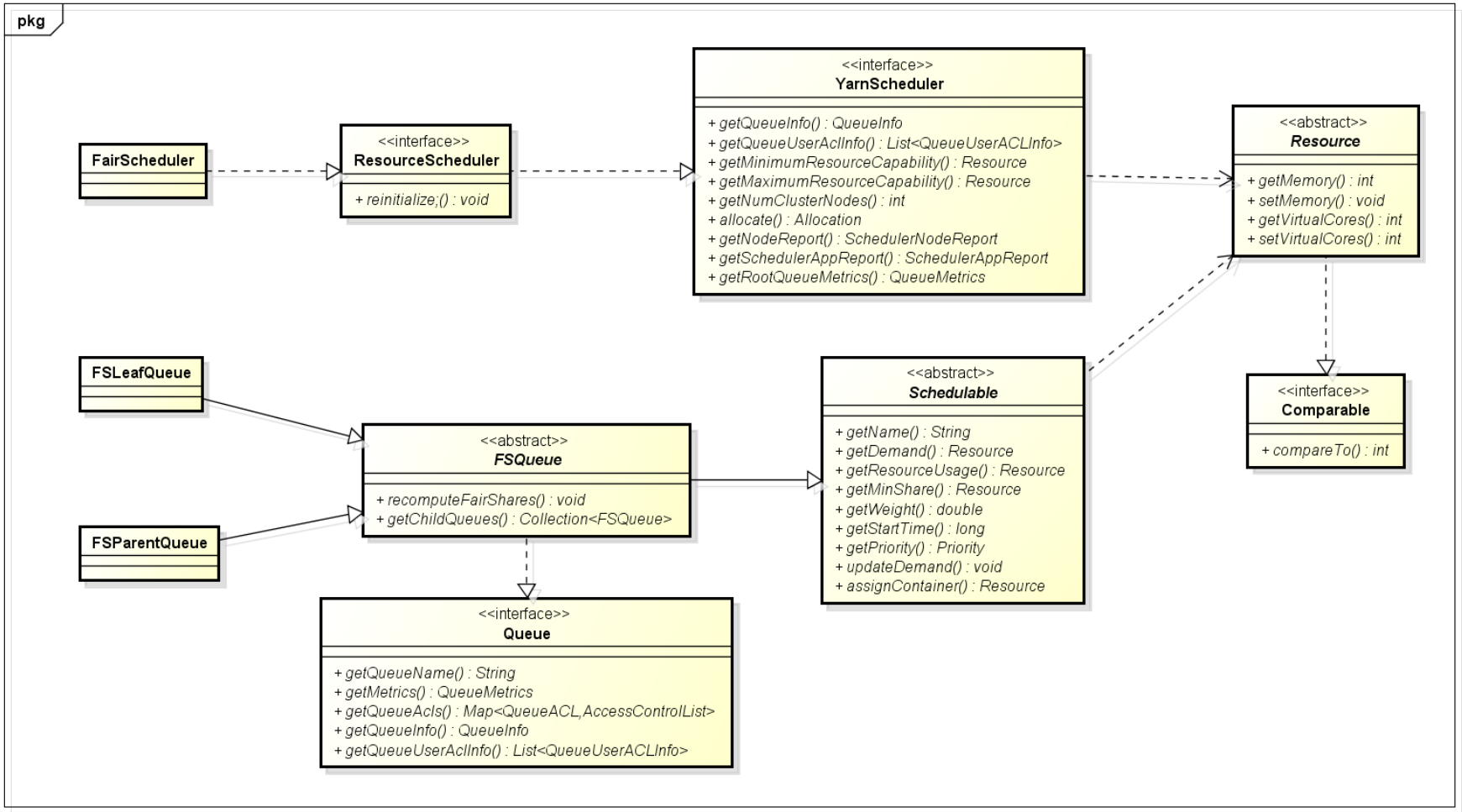
- Contextos mais comuns:
  - classificação de *jobs* e nós quanto ao potencial de E/S ou CPU;
  - avaliação do progresso da *task* na decisão de lançar ou não uma *task* especulativa.
- Objetivos mais comuns:
  - melhorar o *throughput*;
  - diminuir o tempo de resposta.

informáticaufsm
Ciência da Computação

Roteiro

informáticaufsm
Ciência da Computação

Estudo da Arquitetura



Etapas de Preparação

Realizadas com intuito de entender como o novo escalonador iria ser incluso e utilizado numa nova versão do *Hadoop*.

- Grid'5000
  - Diferenças na instalação/configuração.
  - Material/Recursos produzidos.
- Compilação e testes
  - Compilar uma nova classe.
  - Utilizar nova classe no Grid'5000.

informáticaufsm
Ciência da Computação

Escalonamento

```java
public class FifoAppComparator implements Comparator<AppSchedulable>, Serializable {
  private static final long serialVersionUID = 3428835083489547918L;

  public int compare(AppSchedulable a1, AppSchedulable a2) {
    int res = a1.getPriority().compareTo(a2.getPriority());
    if (res == 0) {
      if (a1.getStartTime() < a2.getStartTime()) {
        res = -1;
      } else {
        res = (a1.getStartTime() == a2.getStartTime() ? 0 : 1);
      }
    }
    if (res == 0) {
      // If there is a tie, break it by app ID to get a deterministic order
      res = a1.getApp().getApplicationId().compareTo(a2.getApp().getApplicationId());
    }
    return res;
  }
}
```

informáticaufsm
Ciência da Computação

Escalonamento

```
public Resource assignContainer(FSSchedulerNode node, boolean reserved) {
  LOG.debug("Node offered to queue: " + getName() + " reserved: " + reserved);
  // If this queue is over its limit, reject
  if (Resources.greaterThan(getResourceUsage(),
      queueMgr.getMaxResources(getName()))) {
    return Resources.none();
  }

  // If this node already has reserved resources for an app, first try to
  // finish allocating resources for that app.
  if (reserved) {
    for (AppSchedulable sched : appScheds) {
      if (sched.getApp().getApplicationAttemptId() ==
          node.getReservedContainer().getApplicationAttemptId()) {
        return sched.assignContainer(node, reserved);
      }
    }
    return Resources.none(); // We should never get here
  }

  // Otherwise, chose app to schedule based on given policy (fair vs fifo).
  else {
    Comparator<Schedulable> comparator;
    if (schedulingMode == SchedulingMode.FIFO) {
      comparator = new SchedulingAlgorithms.FifoComparator();
    } else if (schedulingMode == SchedulingMode.FAIR) {
      comparator = new SchedulingAlgorithms.FairShareComparator();
    } else {
      throw new RuntimeException("Unsupported queue scheduling mode " +
          schedulingMode);
    }

    Collections.sort(appScheds, comparator);
    for (AppSchedulable sched: appScheds) {
      if (sched.getRunnable()) {
        Resource assignedResource = sched.assignContainer(node, reserved);
        if (!assignedResource.equals(Resources.none())) {
          return assignedResource;
        }
      }
    }

    return Resources.none();
  }
}
```

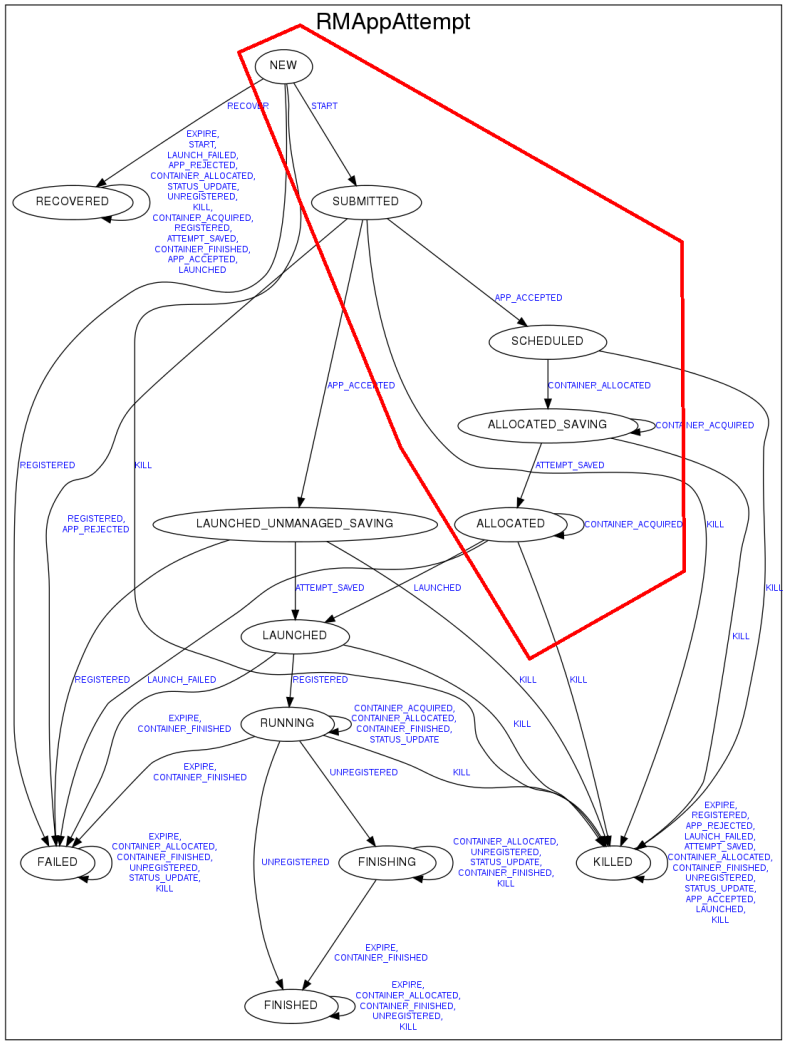
informáticaufsm
Ciência da Computação

Desenvolvimento

- Grafos de máquinas de estado contendo ciclos de execução
  - Como foram gerados?
  - Por que são importantes?

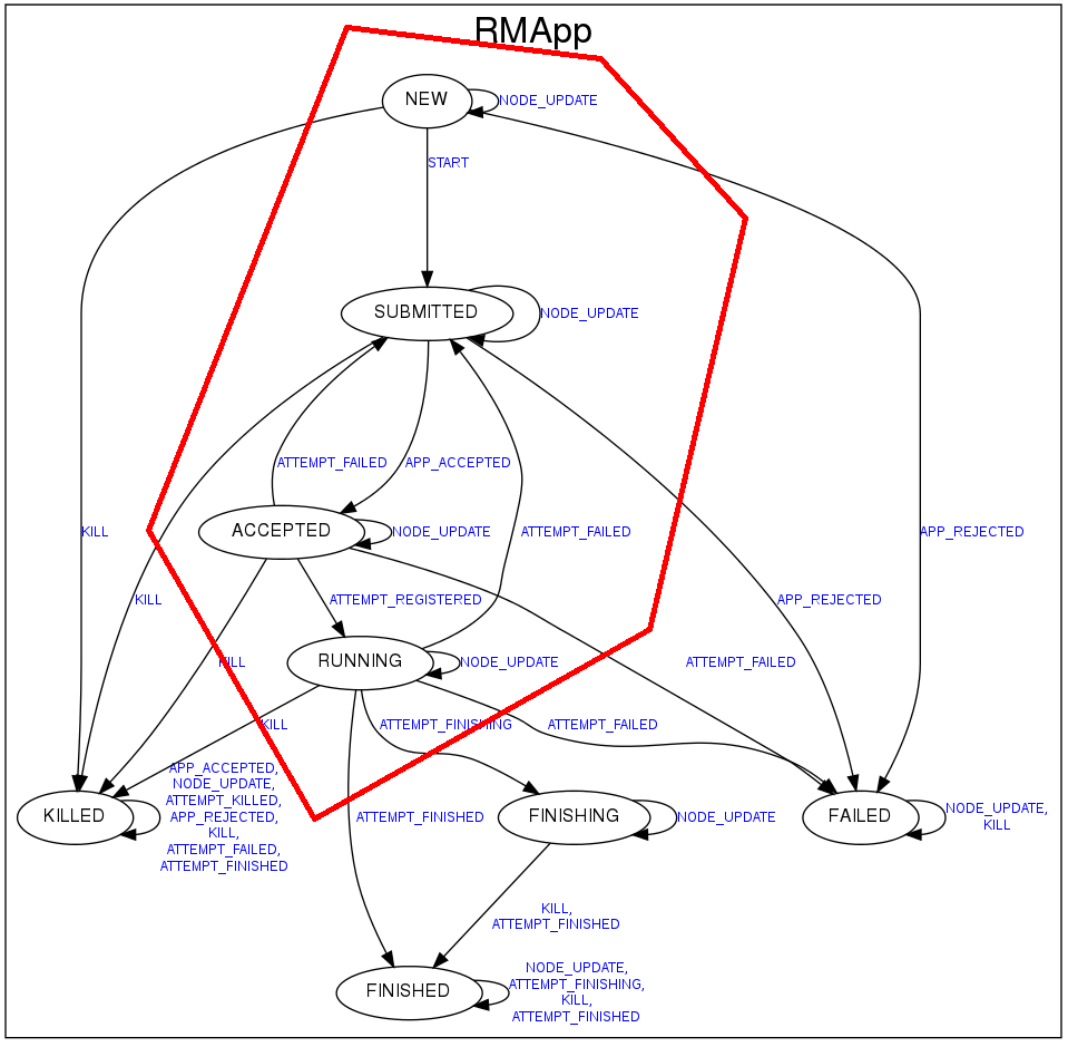
informáticaufsm
Ciência da Computação

RMAppAttempt



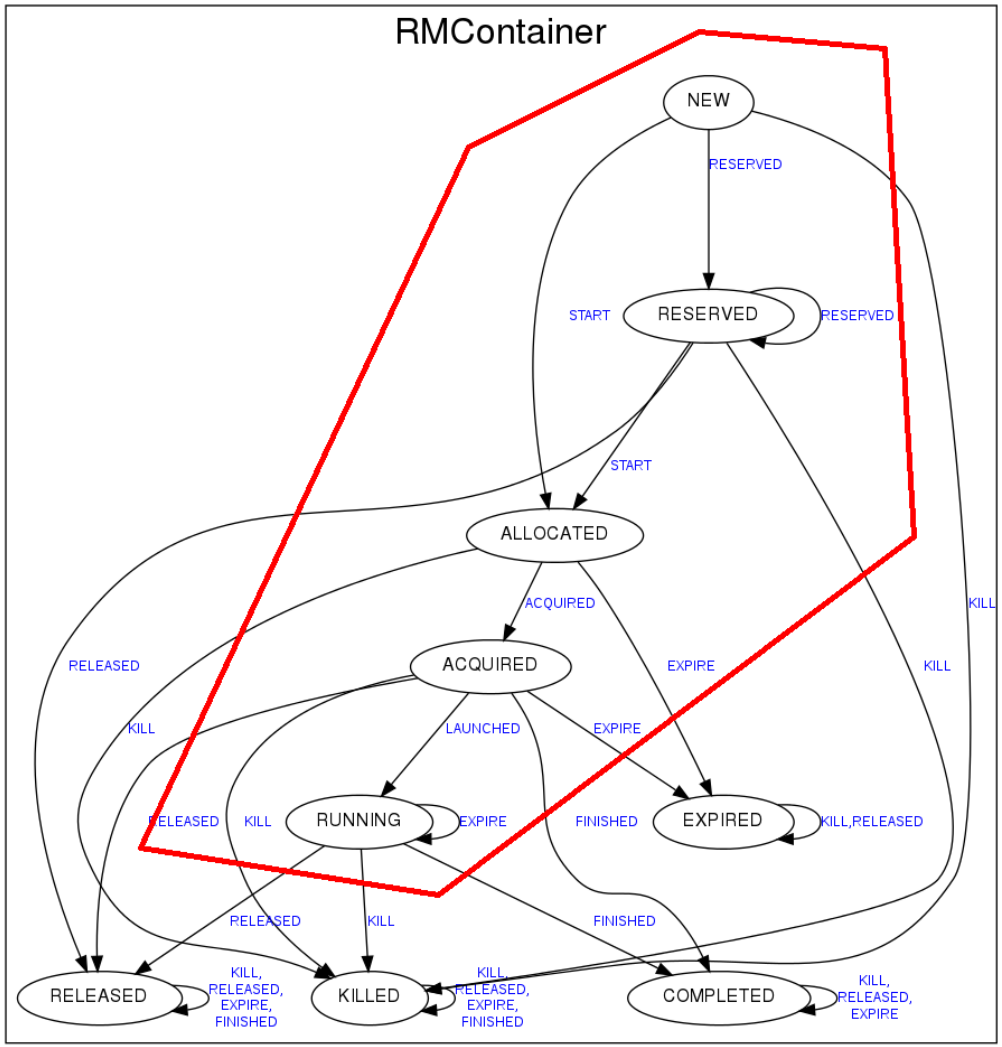
informáticaufsm
Ciência da Computação

RMApp



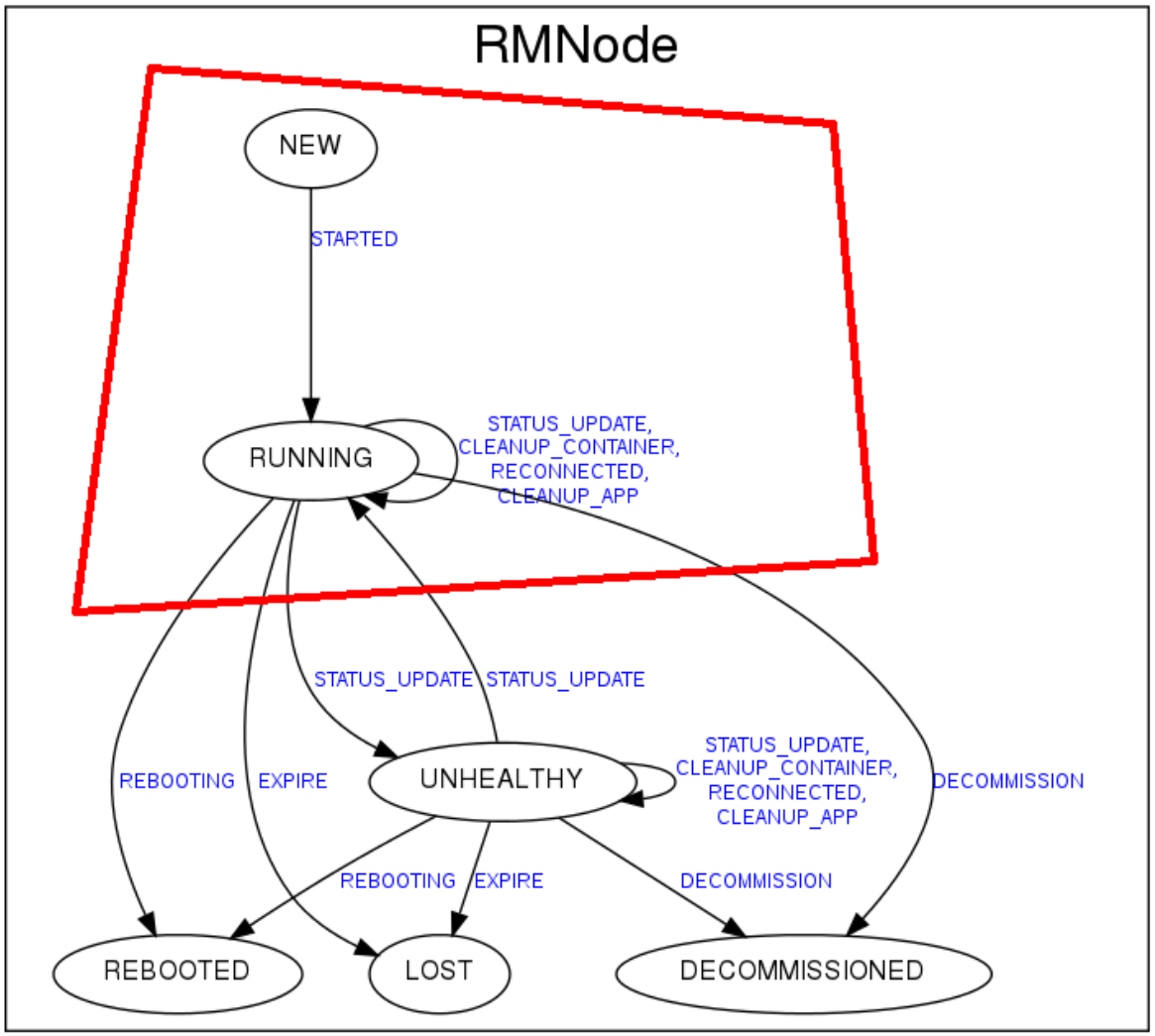
informáticaufsm
Ciência da Computação

RMContainer



informáticaufsm
Ciência da Computação

RMNode



informáticaufsm
Ciência da Computação

Roteiro

informáticaufsm
Ciência da Computação

Desenvolvimento Futuro

- Contexto possível de ser utilizado.
- Métricas adequadas para utilização.
- Implementação.

Roteiro

informáticaufsm
Ciência da Computação

Referências

- DEY, A. K. Understanding and Using Context. Personal Ubiquitous Comput., London, UK, UK, v.5, n.1, p.4-7, Jan. 2001.
- MAAMAR, Z.; BENSLIMANE, D.; NARENDRA, N. C. What can context do for web services? Commun. ACM, New York, NY, USA, v.49, n.12, p.98-103, Dec. 2006.
- KUMAR, K. A. et al. CASH: context aware scheduler for hadoop. In: INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTING, COMMUNICATIONS AND INFORMATICS, New York, NY, USA. Proceedings. . . ACM, 2012. p.52-61. (ICACCI '12).
- ZAHARIA, M. et al. Improving MapReduce performance in heterogeneous environments. In: USENIX CONFERENCE ON OPERATING SYSTEMS DESIGN AND IMPLEMENTATION, 8., Berkeley, CA, USA. Proceedings. . . USENIX Association, 2008. p.29-42. (OSDI'08).
- TIAN, C. et al. A Dynamic MapReduce Scheduler for Heterogeneous Workloads. In: EIGHTH INTERNATIONAL CONFERENCE ON GRID AND COOPERATIVE COMPUTING, 2009., Washington, DC, USA. Proceedings. . . IEEE Computer Society, 2009. p.218-224. (GCC '09).
- CHEN, Q. et al. SAMR: a self-adaptive mapreduce scheduling algorithm in heterogeneous environment. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION TECHNOLOGY, 2010., Washington, DC, USA. Proceedings. . . IEEE Computer Society, 2010. p.2736-2743. (CIT '10).

Referências (Continuação)

- RASOOLI, A.; DOWN, D. G. Coshh: a classification and optimization based scheduler for heterogeneous hadoop systems. In: SC COMPANION: HIGH PERFORMANCE COMPUTING, NETWORKING STORAGE AND ANALYSIS, 2012., Washington, DC, USA. Proceedings. . . IEEE Computer Society, 2012. p.1284-1291. (SCC '12).
- ISARD, M. et al. Quincy: fair scheduling for distributed computing clusters. In: ACM SIGOPS 22ND SYMPOSIUM ON OPERATING SYSTEMS PRINCIPLES, New York, NY, USA. Proceedings. . . ACM, 2009. p.261-276. (SOSP '09).
- XIE, J. et al. Improving MapReduce performance through data placement in heterogeneous Hadoop clusters. In: PARALLEL AND DISTRIBUTED PROCESSING, WORKSHOPS AND PHD FORUM (IPDPSW). Anais. . . IEEE International Symposium, 2010.
- HADOOP, A. Arquitetura do HDFS. http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/Federation.html, Acesso em novembro de 2013.
- HADOOP, A. Arquitetura do YARN. http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html, Acesso em novembro de 2013.

Desenvolvimento de um Escalonador Sensível ao Contexto para o *Apache Hadoop*

Guilherme Weigert Cassales[1]
Orientador: Profª Drª Andrea Schwertner Charão[1]

[1]Ciência da Computação
Universidade Federal de Santa Maria

13 de novembro de 2013