

CollA: Uma Plataforma Colaborativa para Programas Java e Java Cá & Lá

Bruno Cerqueira Hott

Prof. Joubert de Castro Lima



○ problema

- Como permitir que diferentes especialistas possam colaborar cedendo computadores, programas diversos e serviços especializados em análises de dados num ambiente integrado e escalável ?

Justificativas

- Muitos especialistas possuem problemas de grande porte, porém não detêm infra computacional para resolvê-los
- Muitos especialistas possuem bases de dados reais e massivas, porém não possuem métodos escaláveis de análise
- Muitas tarefas são executadas repetidas vezes por grupos diversos sem qualquer possibilidade de interação entre grupos e usuários.

CollA – Uma plataforma colaborativa para análise de dados

Cedendo computadores

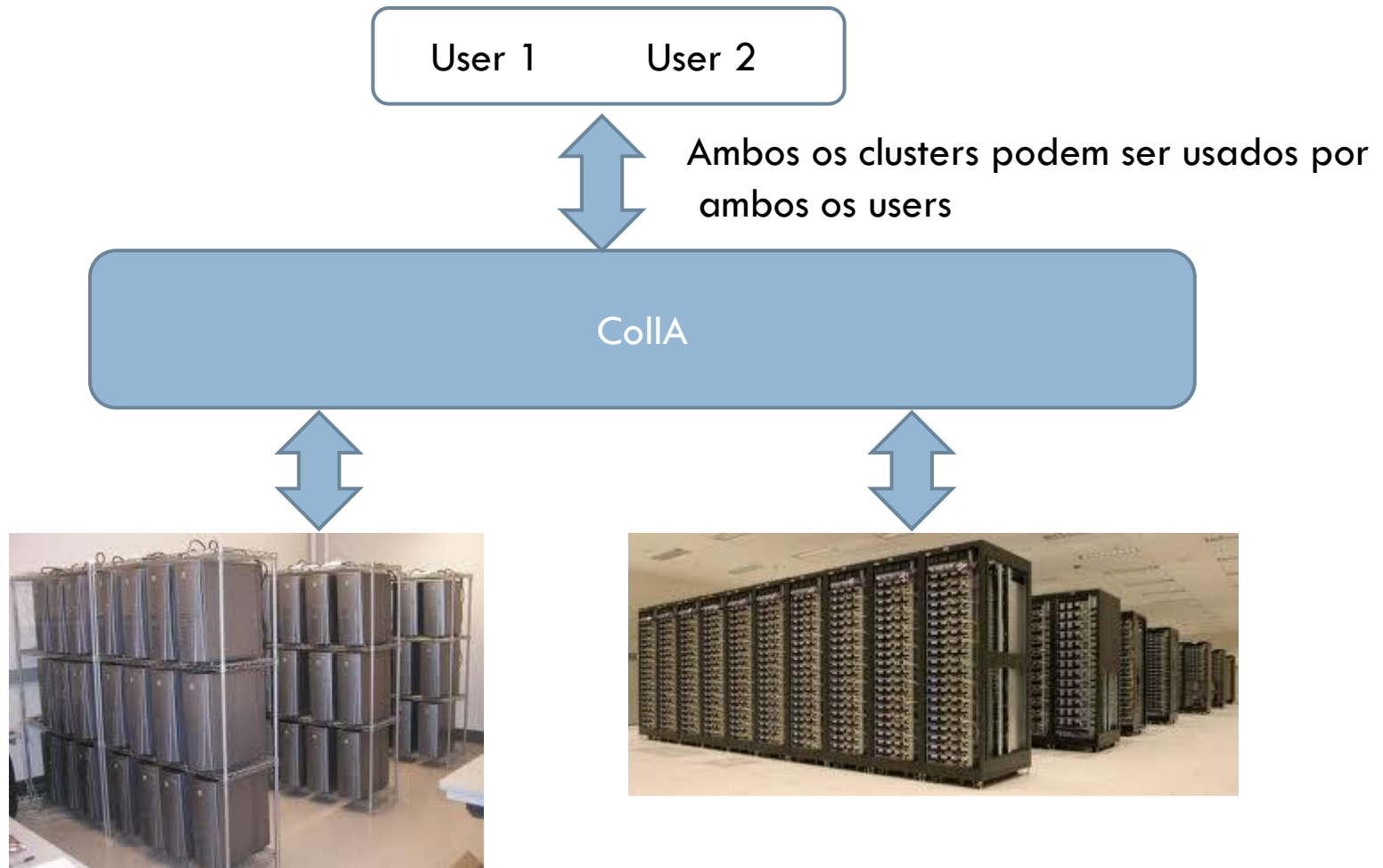
UFOP - User 1 – cluster de médio porte, porém insuficiente para o problema da dengue de Ouro Preto e Região



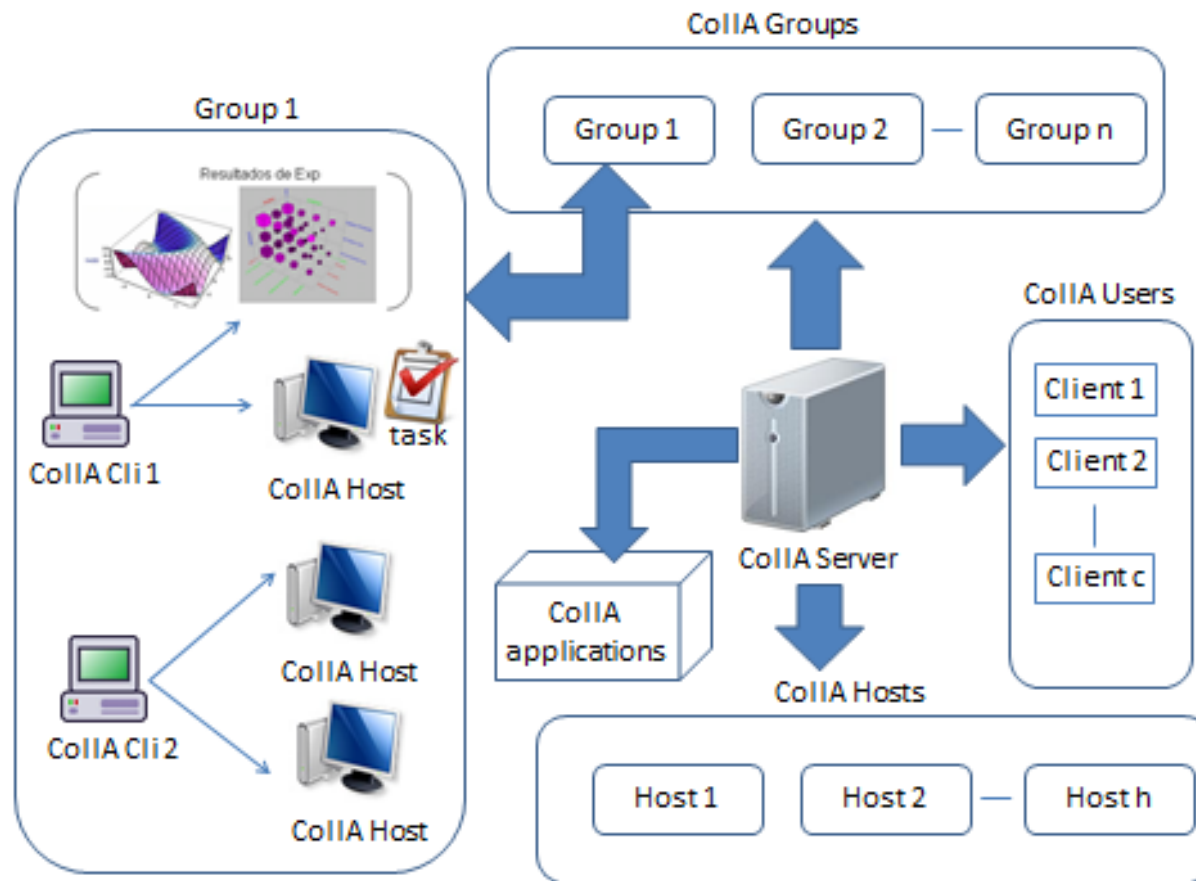
LNCC - User 2 – cluster de grande porte, porém sendo usado apenas em alguns turnos



CollA – Uma plataforma colaborativa para análise de dados



Quais as demais formas de colaboração do ColIA ?



- ✓ **Jobs ou Tarefas**
- ✓ **Resultados**
- ✓ **Fontes de dados**

O que existe na literatura?



☐ Globus

- ☐ <http://globus.org/>
- ☐ Plataforma para construir GRID
- ☐ Monitor de recursos e tarefas
- ☐ GRID individual (sem colaboração)
- ☐ Custoso para usuários esporádicos



Worldwide Virtual Computer
e pluribus unum: one out of many

☐ Legion

- ☐ <http://www.cs.virginia.edu/~legion/>
- ☐ Semelhante ao Globus
- ☐ Principal característica é sua arquitetura orientada a objetos

O que existe na literatura?



☐ SETI@home

- ☐ <http://setiathome.berkeley.edu/>
- ☐ Busca de vida extraterrestre
- ☐ Utiliza PC's espalhados pelo mundo
- ☐ Computação voluntária
- ☐ Não é possível que se execute outros problemas além do SETI

APLICAÇÃO

☐ BOINC

- ☐ <http://boinc.berkeley.edu/>
- ☐ Criado para suprir deficiências do SETI
- ☐ Provê ferramentas para criar e gerenciar projetos
- ☐ Projetado para ser utilizado por cientistas, não por programadores
- ☐ É similar ao CollA, porém não prevê grupos de usuários num projeto e com isto não prevê colaboração de jobs e resultados.

O que existe na literatura?



☐ Condor

- ☐ <http://research.cs.wisc.edu/condor/>
- ☐ Utiliza máquinas ociosas
- ☐ Prioriza usuários que menos conseguem máquinas
- ☐ Não se tem garantia de quando a aplicação será executada

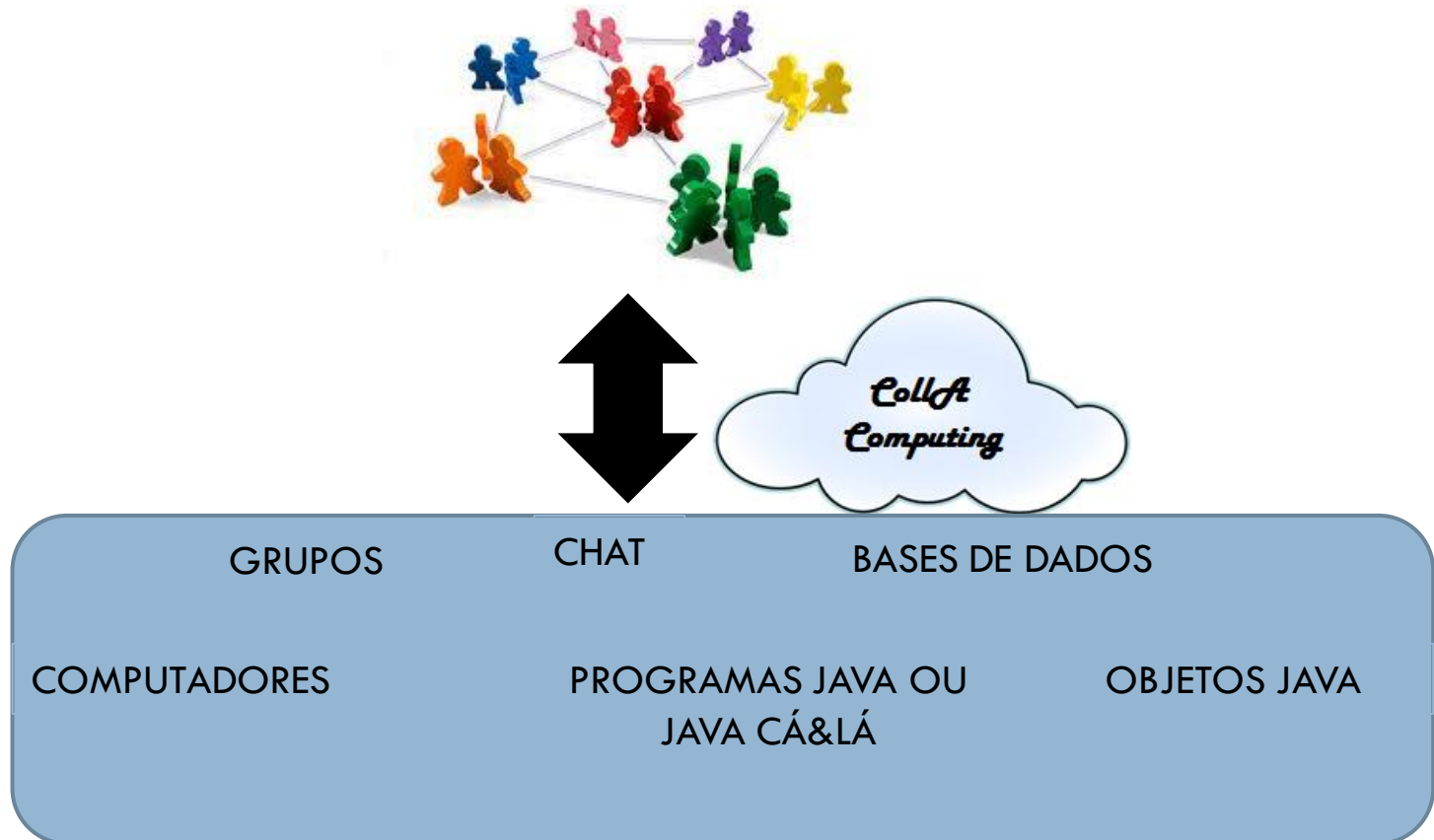


☐ BonjourGrid

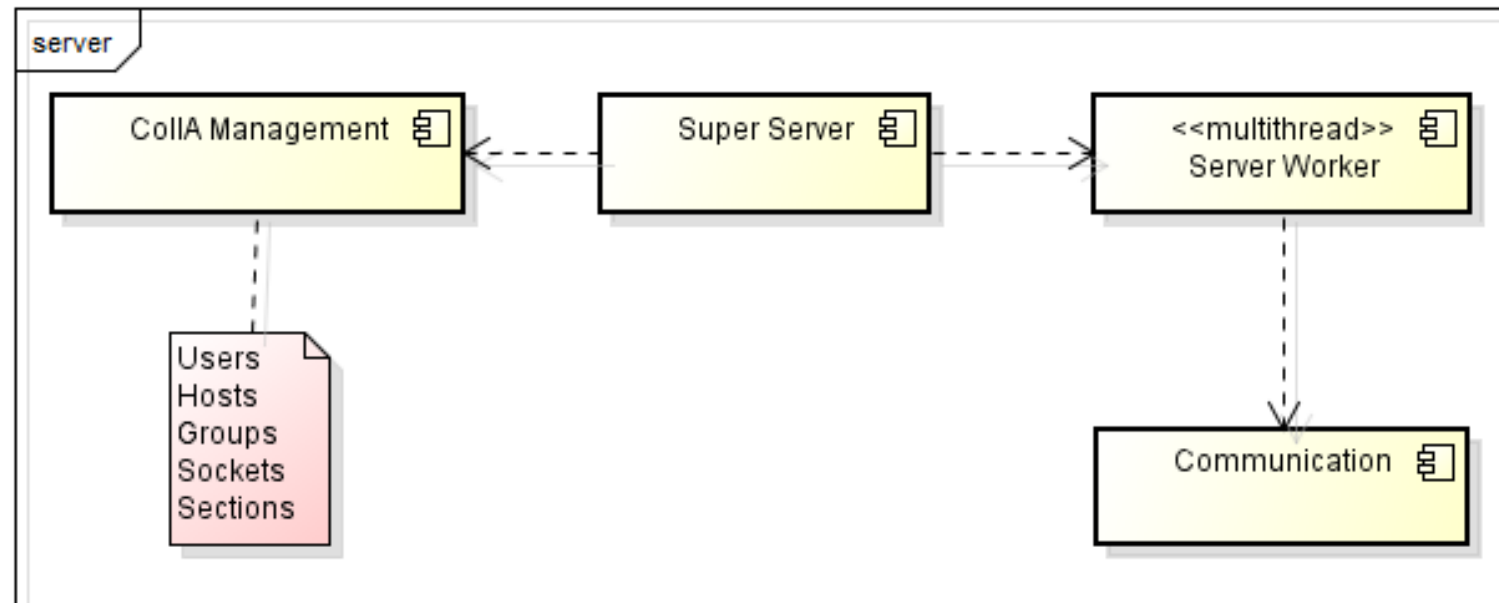
- ☐ <http://sourceforge.net/projects/bonjour-grid/>
- ☐ Orquestra múltiplas instâncias de middlewares para grades computacionais
- ☐ Distribuição de máquinas é desbalanceada
- ☐ Não permite interação entre os usuários

O que existe na literatura?

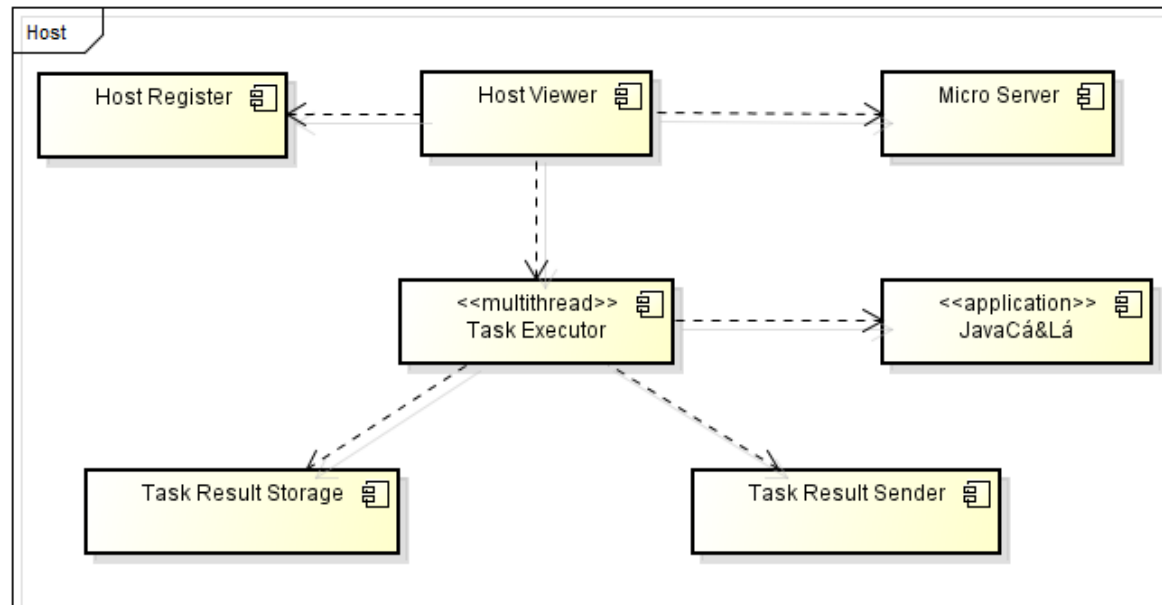
- Nenhum deles é colaborativo de forma que os usuários possam interagir e trabalhar juntos.



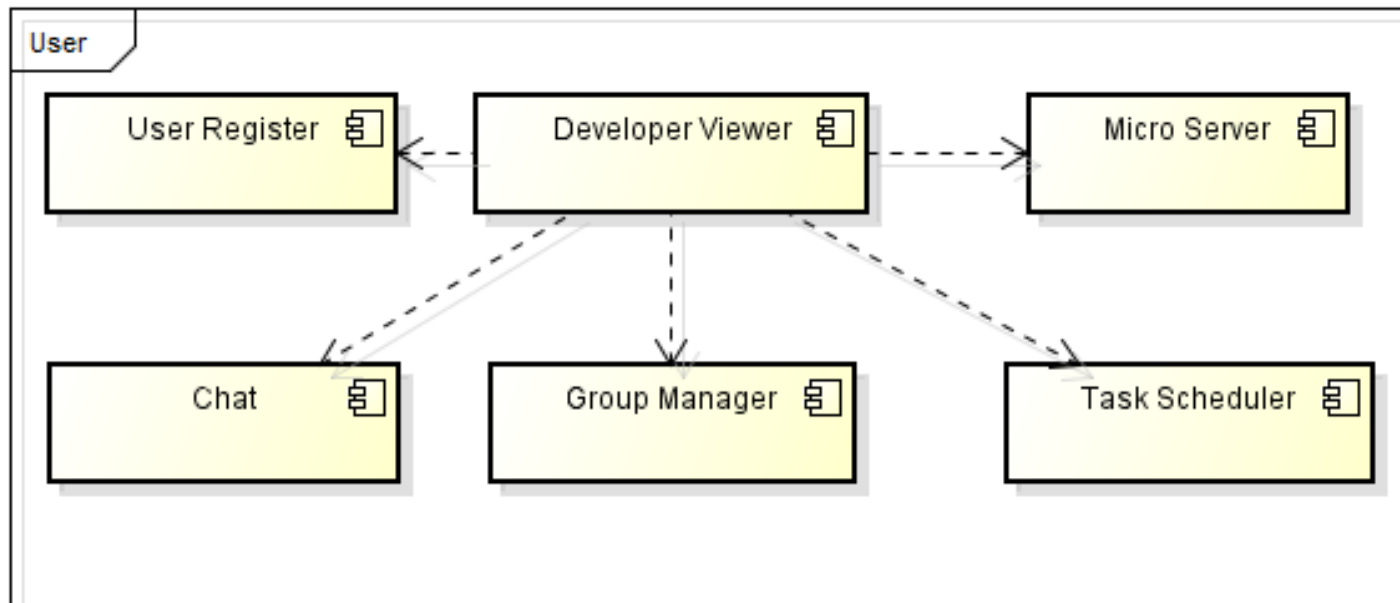
Arquitetura CollA – Servidor



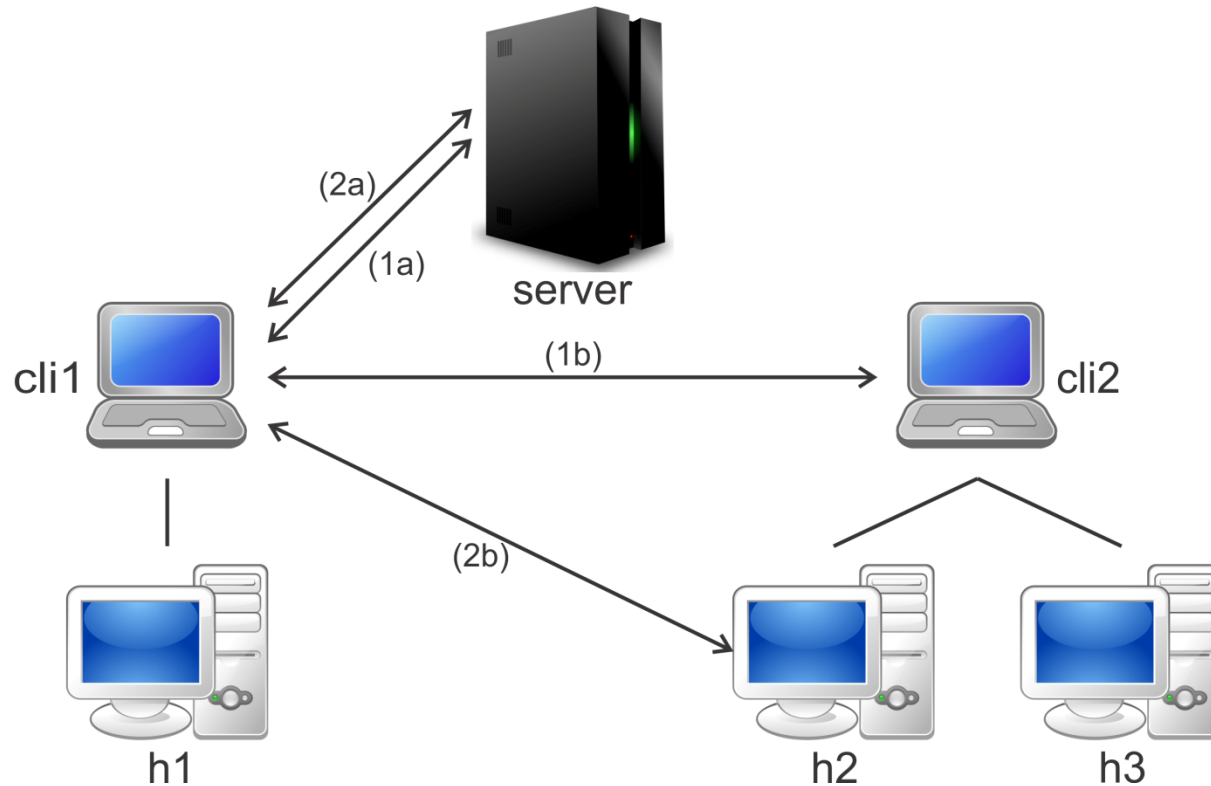
Arquitetura Colla – Host



Arquitetura CollA – Cliente



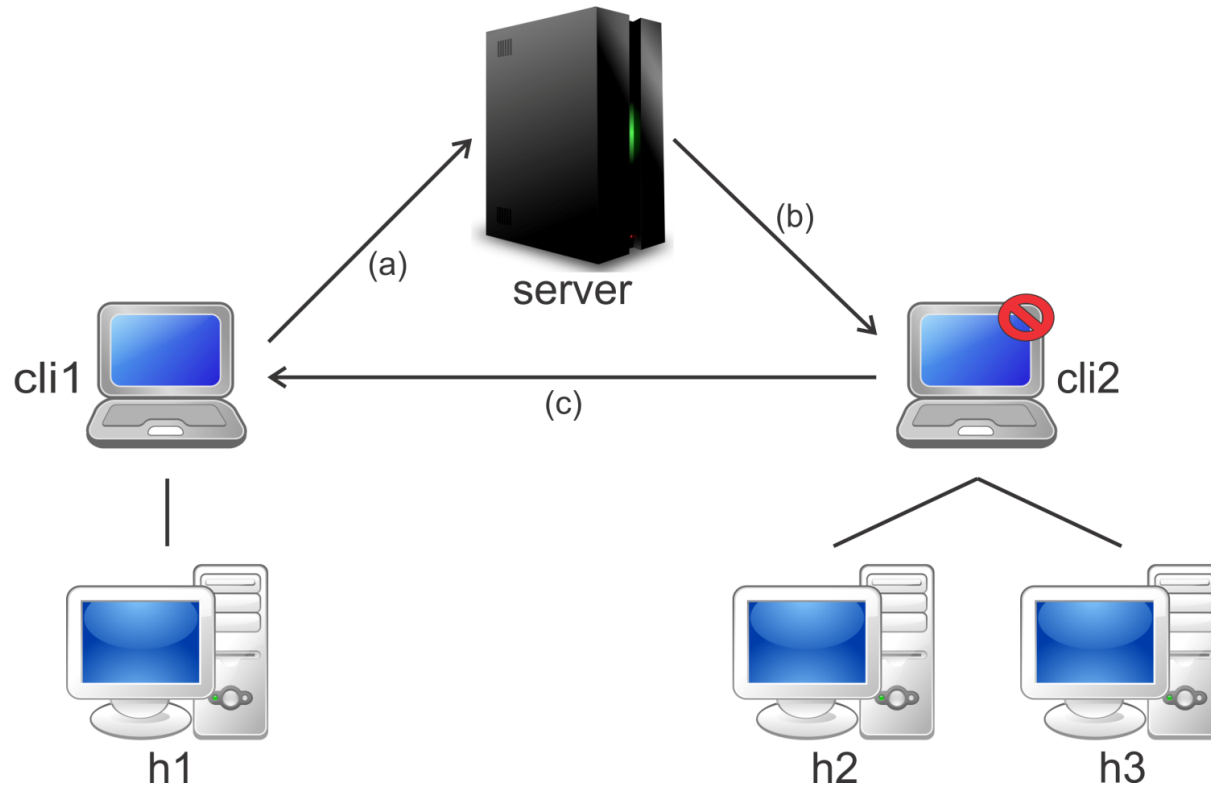
Arquitetura CollA – IP válido



- ▶ (1a) Lista de contatos
- ▶ (1b) Chat

- (2a) Host
- (2b) Task/Result

Arquitectura CollA – IP inválido

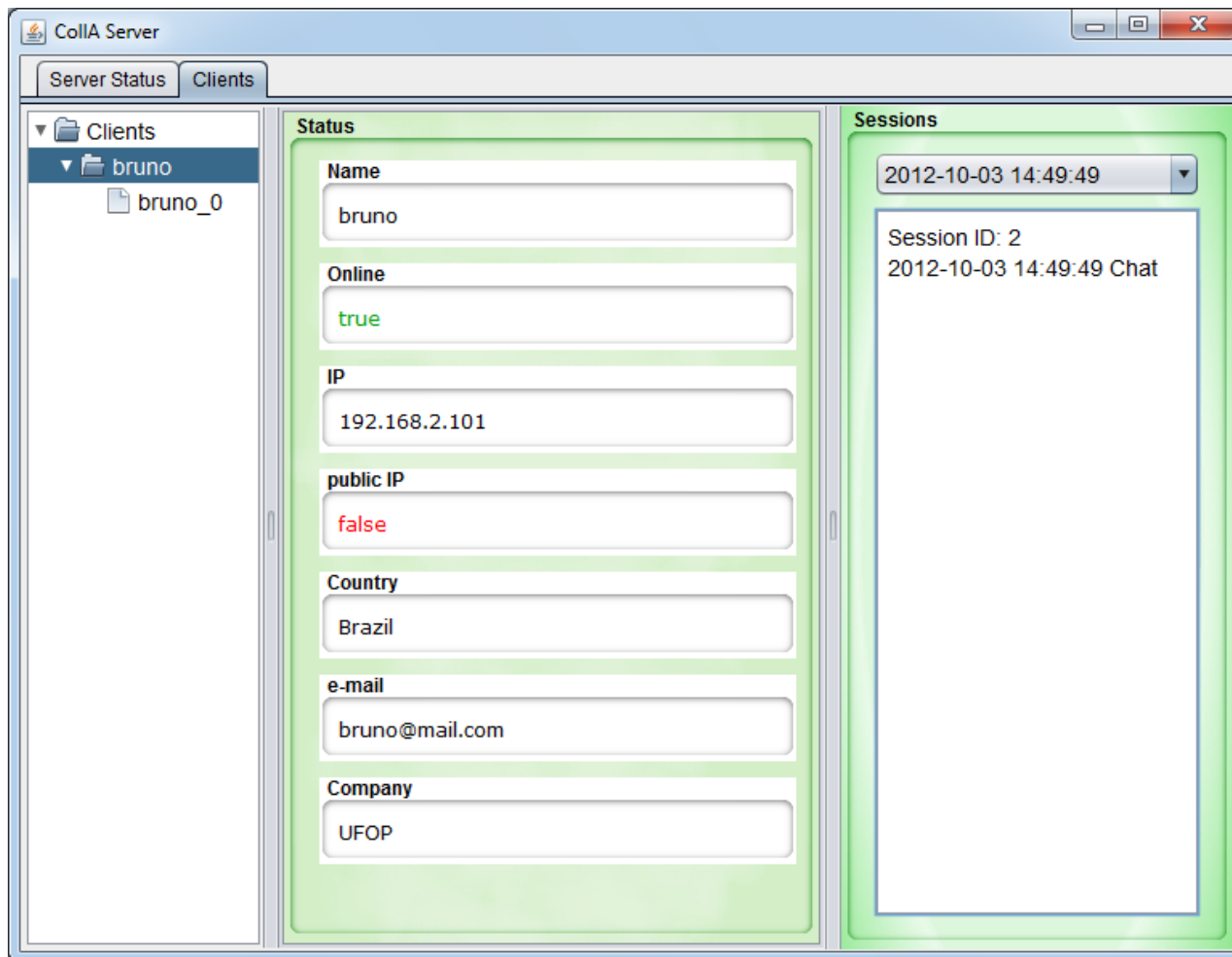


- ▶ (a) Cliente/Servidor
- ▶ (b) Servidor/Cliente

□ (c) Cliente/Cliente

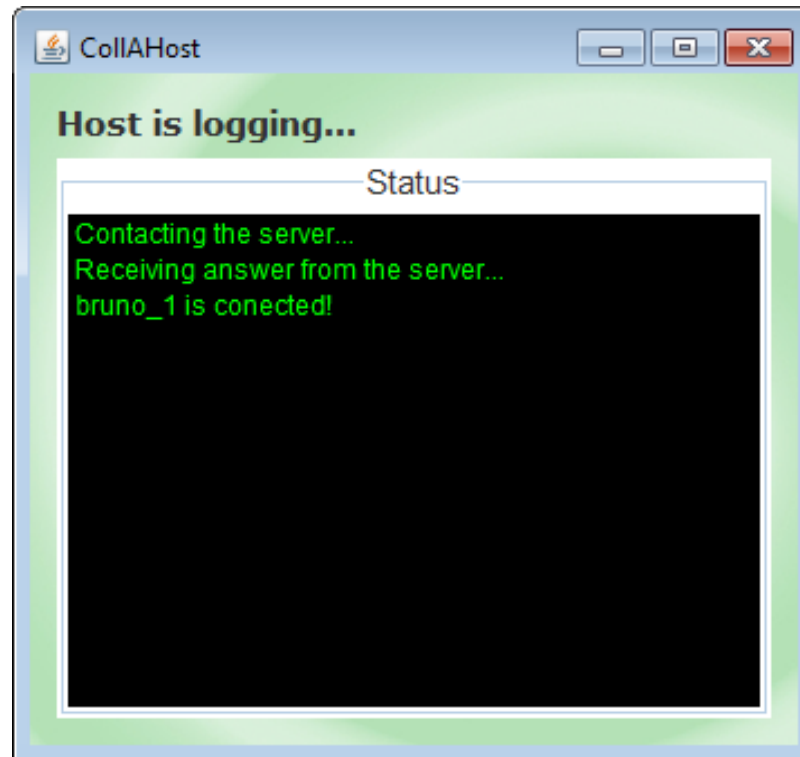
O CollA em detalhes

❑ Servidor



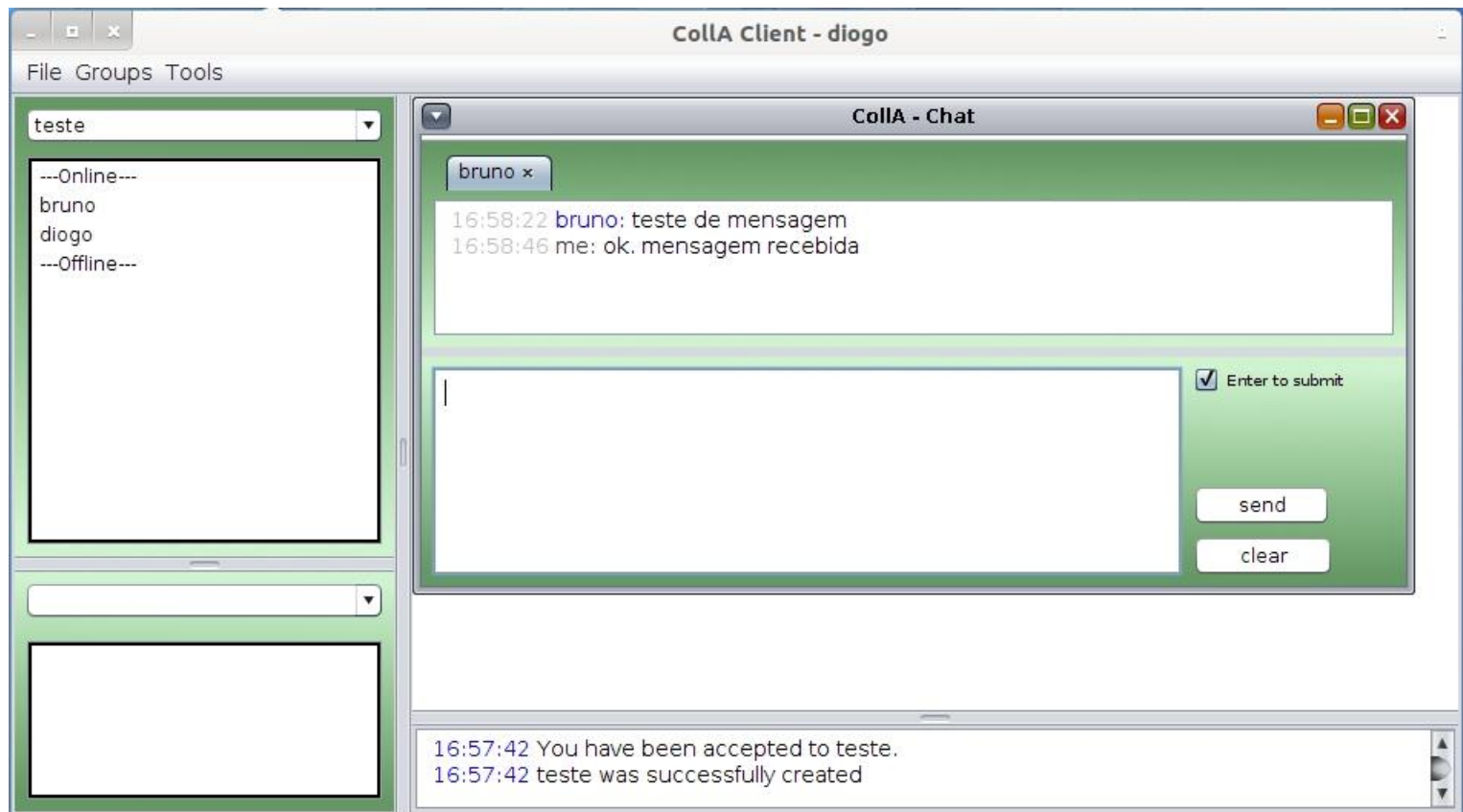
○ CollA em detalhes

□ Host



○ CollA em detalhes

□ Chat



O CollA em detalhes

□ Grupos

The screenshot shows the 'Groups management' window with the 'Requests' tab selected. The 'Group' dropdown menu is set to 'teste'. The 'Requests' list contains two entries: 'marcio' and 'pablo'. To the right of this list are two buttons with arrows: a right-pointing arrow and a left-pointing arrow. Below these are two empty boxes labeled 'Accept' and 'Refuse'. At the bottom are 'Cancel', 'Apply', and 'OK' buttons.

Groups management

Group: teste

Requests Members

Requests

- marcio
- pablo

Accept

Refuse

Cancel Apply OK

The screenshot shows the 'Groups management' window with the 'Members' tab selected. The 'Group' dropdown menu is set to 'teste'. The 'Members' list contains three entries: 'diogo', 'bruno', and 'marcio', with 'marcio' highlighted. To the right of this list is an 'Informations' box containing the following text: 'username: marcio', 'e-mail: marcio@ufop.br', 'Country: Brazil', and 'Offering 0 hosts'. At the bottom are 'Cancel', 'Apply', and 'OK' buttons.

Groups management

Group: teste

Requests Members

Members

- diogo
- bruno
- marcio

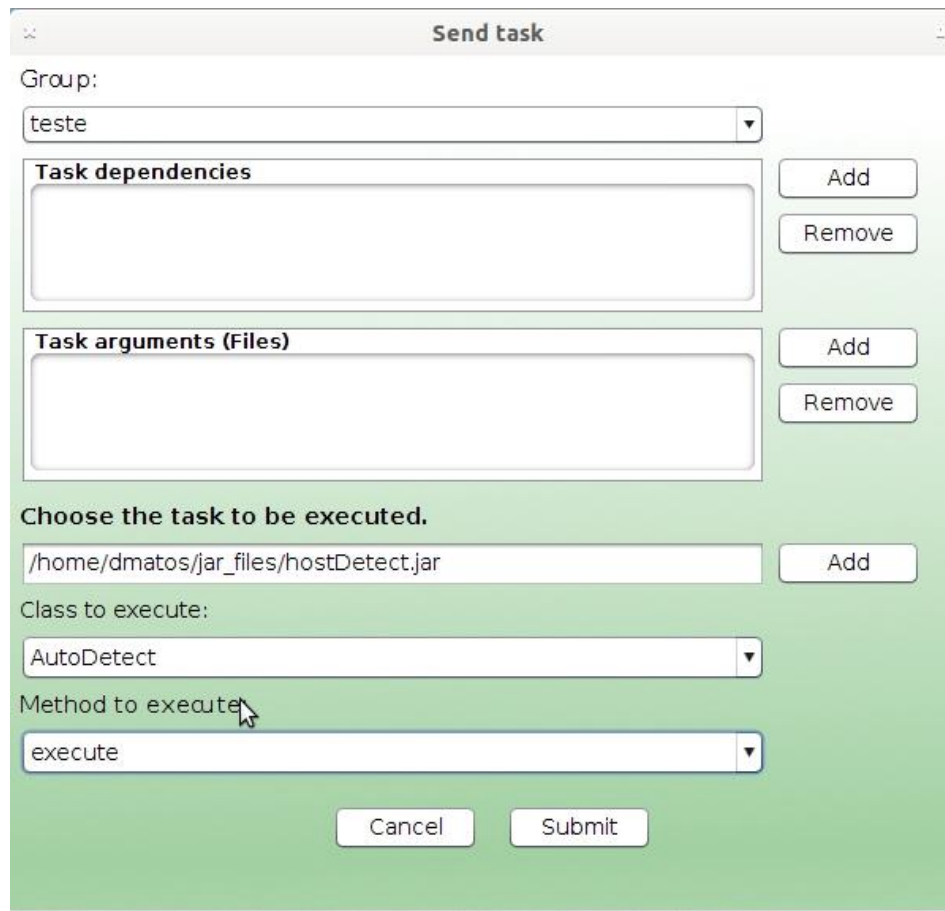
Informations

username: marcio
e-mail: marcio@ufop.br
Country: Brazil
Offering 0 hosts

Cancel Apply OK

○ CollA em detalhes

□ Submissão de Tasks

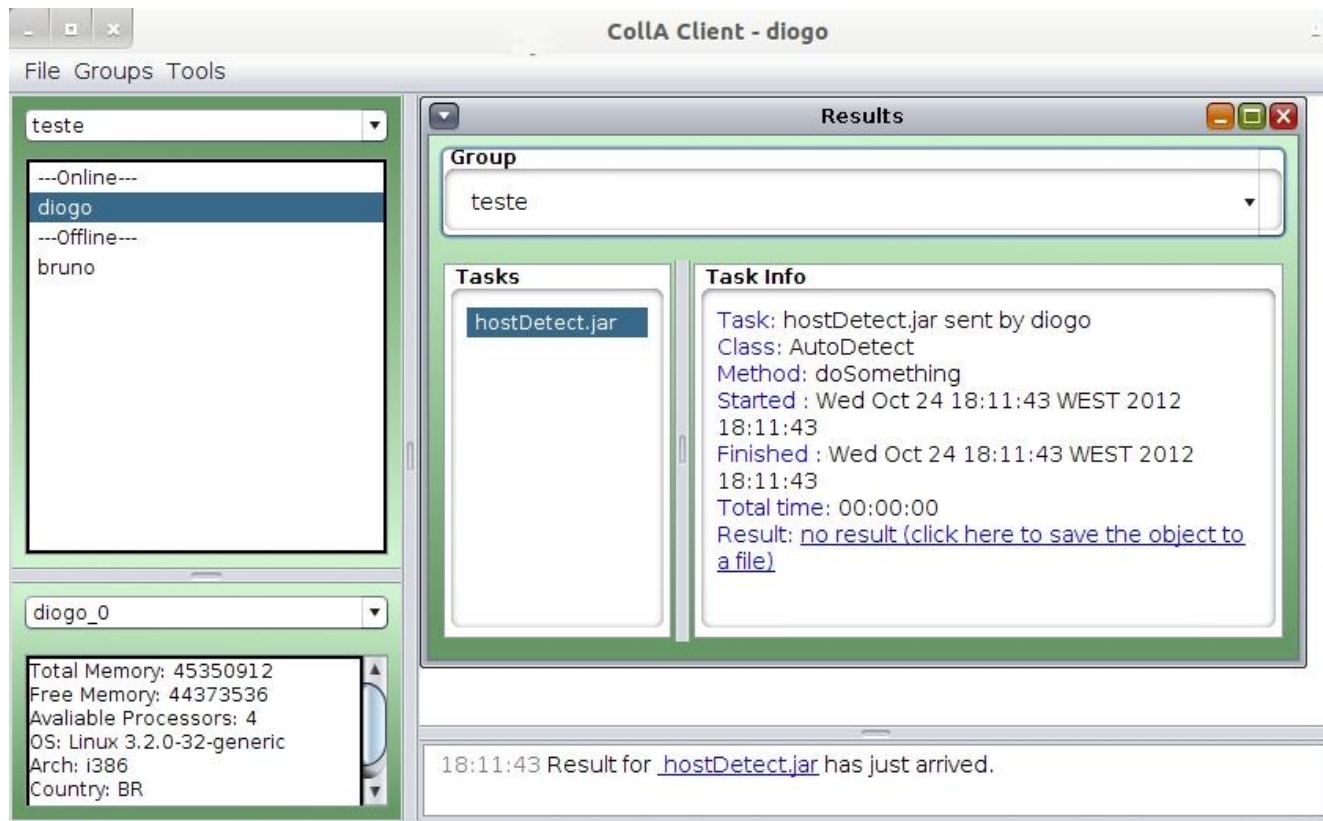


The screenshot shows a 'Send task' dialog box with the following fields and controls:

- Group:** A dropdown menu with 'teste' selected.
- Task dependencies:** A text area for dependencies, with 'Add' and 'Remove' buttons to its right.
- Task arguments (Files):** A text area for file arguments, with 'Add' and 'Remove' buttons to its right.
- Choose the task to be executed:** A text field containing '/home/dmatos/jar_files/hostDetect.jar' and an 'Add' button to its right.
- Class to execute:** A dropdown menu with 'AutoDetect' selected.
- Method to execute:** A dropdown menu with 'execute' selected, with a mouse cursor pointing at it.
- Buttons:** 'Cancel' and 'Submit' buttons at the bottom.

○ CollA em detalhes

□ Resultados de Tasks



Testes - Ambiente

- Ambos com Servidor Dedicado
- Primeiro Ambiente:
 - ▣ 35 máquinas com IP válido
- Segundo Ambiente:
 - ▣ + 3 notebook em Ouro Preto + 1 notebook em Porto, Portugal

Tarefas enviadas:

Nome	Dificuldade	Complexidade
Jar0	Tarefa fácil	Instantâneo
Jar1	Tarefa média	17~70s
jar2	Tarefa difícil	17~70s c/ 1Mb de tráfego

Testes – Resultado 1 (IPs válidos)

Resultados dos experimentos no ambiente 1:

Tarefas à enviar	Tarefas enviadas	Respostas obtidas
19.200	18.633	18.632
100,00%	97,05%	97,04%

Tempo médio de cada tarefa:

Tarefa	Tempo médio de execução
Jar0	29,8s
Jar1	87,9s
Jar2	87,4s

Testes – Resultado 2 (IPs inválidos)

Resultados dos experimentos no ambiente 2:

Tester	Tarefas à enviar	Tarefas enviadas	Respostas obtidas
Tester1	5.760	5.619	5.353
	100,00%	97,55%	92,93%
Tester2	5.760	5.600	5.364
	100,00%	97,22%	93,13%

Tempo médio de cada tarefa:

Tarefa	Tempo médio de execução
Jar0	7,6s
Jar1	64,2s
Jar2	69,1s

Conclusões

- CollA é compartilhamento e colaboração.
- Trabalhos correlatos não suportam colaboração.
- Versão atual:
 - ▣ Compartilhamento de computadores.
 - ▣ Execução de tarefas.
 - ▣ Chat.
 - ▣ Manutenção de usuários e grupos.
 - ▣ Escalonador de tarefas.
 - ▣ Suporte a máquinas com IP inválido.
- Testes:
 - ▣ foram executados testes com duração de 8 horas
 - ▣ ~20 mil tarefas foram submetidas.
 - ▣ O CollA se comportou satisfatoriamente.

Trabalhos futuros

- Tolerância a falha
- Duplicação de tarefas
- Super Hosts
- Integração ao JCL distribuído
- Novos serviços nativos CollA:
 - ▣ Data Splitter
 - ▣ Olap
 - ▣ Mining
 - ▣ Simulação de sensores
 - ▣ ...

Agradecimentos

