

# Gaudi

## Grupo de Computação Distribuída

---

*Fabíola Gonçalves Pereira Greve*

DCC - UFBA

<http://gaudi.dcc.ufba.br>

---

Como construir aplicações distribuídas confiáveis, que apesar da ocorrência de falhas, continuam funcionando de maneira correta, além de possuírem bom desempenho ?

# Linhas de Pesquisa

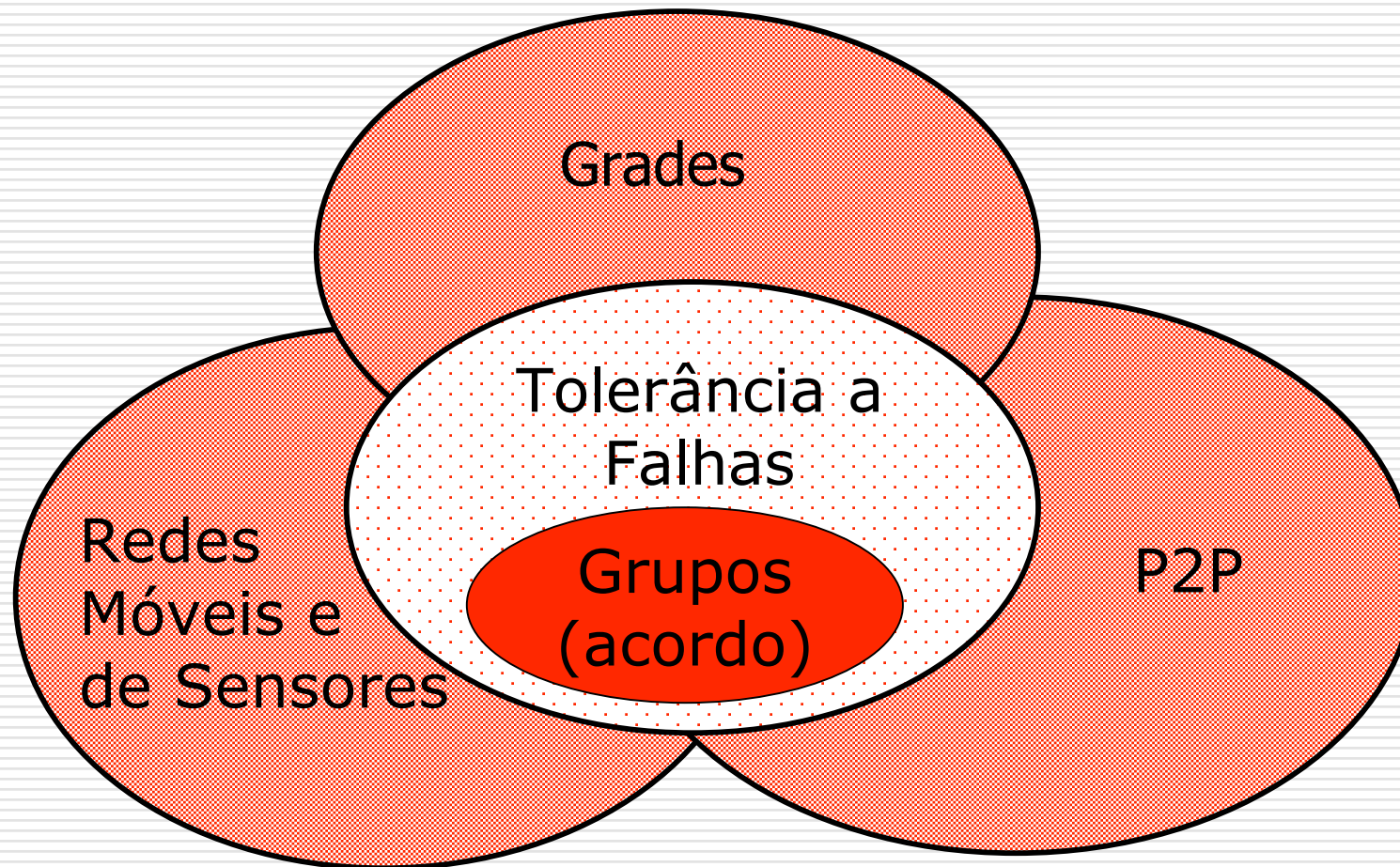
---

## Tolerância a Falhas

- Algoritmos Distribuídos
- Sistemas Distribuídos
  - Grades Computacionais
  - Redes Móveis Ad-Hoc
  - Sistemas P2P
- Middleware
  - Gestão de Grupos
  - Componentes de Acordo

# Projetos

---



# Projetos

---

- TOPGRID, VCG, SimeGrid
  - Grades Computacionais
    - Cooperação Internacional CAPES-COFECUB
    - Cooperação Nacional coordenada pelo LNCC
- AMON, Gaudi-MOTO
  - Redes Móveis Auto-Organizáveis (Manets) e P2P
    - Paris-Sud, Lip6
    - CNPQ Edital Universal, Produtividade Pesquisa CNPQ
- EDEN
  - Serviço de Gestão de Grupos
    - IRISA
    - Université de Rennes, CNPQ

# Tolerância a Falhas em Grids

---

- Parte I
  - Projetos Existentes
- Parte II
  - Alguns Resultados

# Parte I

---

## **Projetos Existentes**

# TOPGRID

## Confiabilidade em Grids

---

- ❑ Criar ambiente de execução de aplicações em grade que contemple características de alta disponibilidade e confiabilidade
- ❑ Rede de Cooperação Internacional CAPES - COFECUB, 2004 a 2008
- ❑ UFBA
- ❑ UFCG - **OurGrid**
- ❑ IRISA-Rennes - **GenoGrid, Paradis**
- ❑ INRIA-Roquencourt e LIP6 - **Darx**
- ❑ Université Paris Sud, **XtremWeb**



# VCG (Virtual Community Grid)

## Ambiente Execução em Grids

---

- ❑ Oferecer ambiente para execução distribuída de tarefas computacionais que aproveite recursos ociosos de máquinas ligadas à rede da RNP
- ❑ GT da RNP 2008, **Globus**
- ❑ Coordenação LNCC, Bruno Schulze
- ❑ UFBA, CBPF, UFF, PUC-Rio, UNICAMP, Unisantos, UFBA, UFC, UFPR, UFRGS, UFU, USP, IME, UFScar, INPE

# SimeGrid

## Simulações Médicas em Grids

---

- ❑ Desenvolver e implantar um serviço de processamento de alto desempenho que dê suporte à execução de aplicações de simulação médica.
- ❑ CNPQ CT-INFO 2007 - 2009
- ❑ Coordenação Bruno Schulze - LNCC
- ❑ UFF, UNICAMP, UFRGS, UFBA, UFC, UFSCar, e EPFL, Suíça.

# Grupo Gaudi na UFBA

## Objetivos

---

- ❑ Prover tolerância a falhas em grades para suportar classes de aplicações que, além de recursos de processamento, demandam requisitos de confiança no funcionamento

# Parte II

---

## **Alguns Resultados**

# PARADIS

## Confiabilidade em Grades

---

- Mecanismos de tolerância a falhas
  - Gestão dos recursos da grade
    - entrada, retirada, falhas de recursos
  - Realocação de tarefas
    - variações de carga e desconexões
- Provê
  - Disponibilidade, melhoria de desempenho, adaptação da grade às variações

# PARADIS

## Mecanismos de Tolerância a Falhas

---

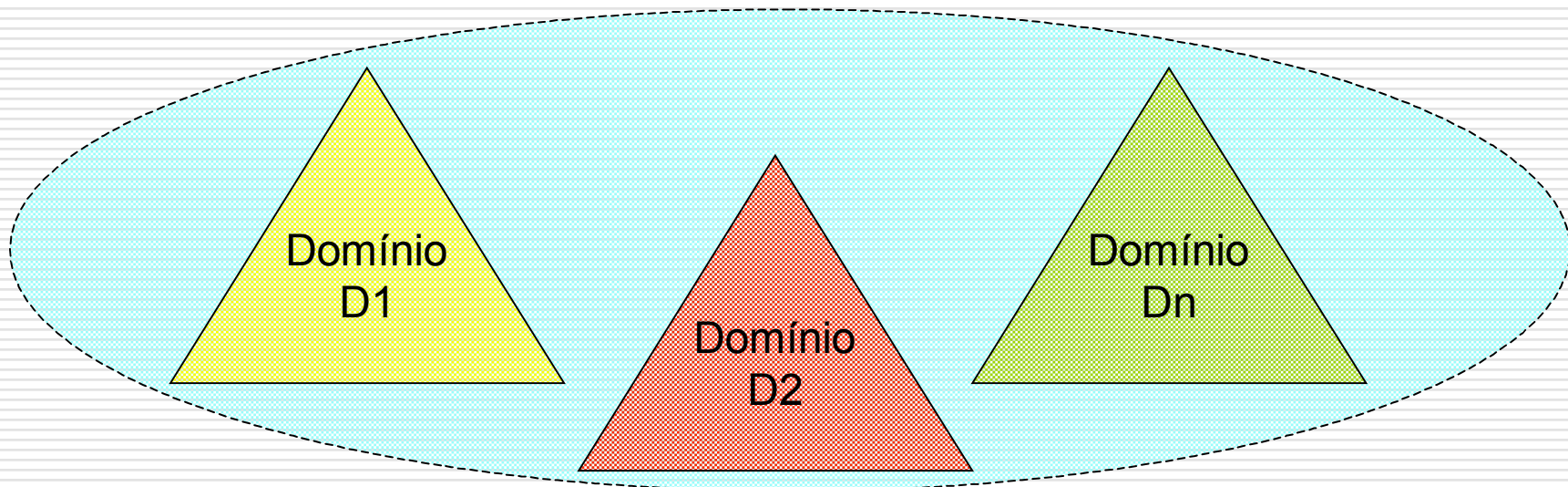
- ❑ PARADIS apoia-se em EDEN
- ❑ EDEN é um serviço de gestão de grupos
  - Desenvolvido por IRISA na França e Grupo Gaudi no DCC-UFBA
- ❑ Utilização de Protocolos de Acordo
  - Group Membership, Consensus, Detecção de Falhas
  - Construir observação consistente
    - ❑ Carga da Grade
    - ❑ Evolução dos seus Recursos

# PARADIS

## Arquitetura da Grade

---

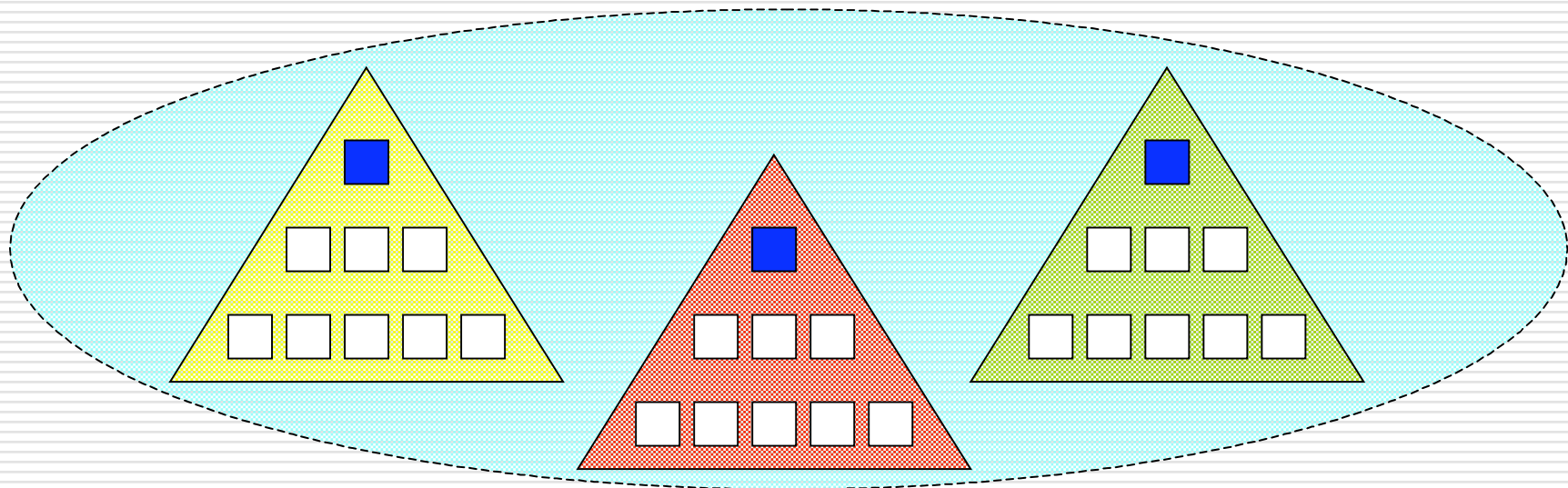
- Subredes síncronas numa rede globalmente assíncrona
  - « *ilhas síncronas num oceano assíncrono* »



# PARADIS: Dois Níveis

---

- ❑ Domínio (Intranet) - síncrono
- ❑ Grade (Internet) - assíncrono

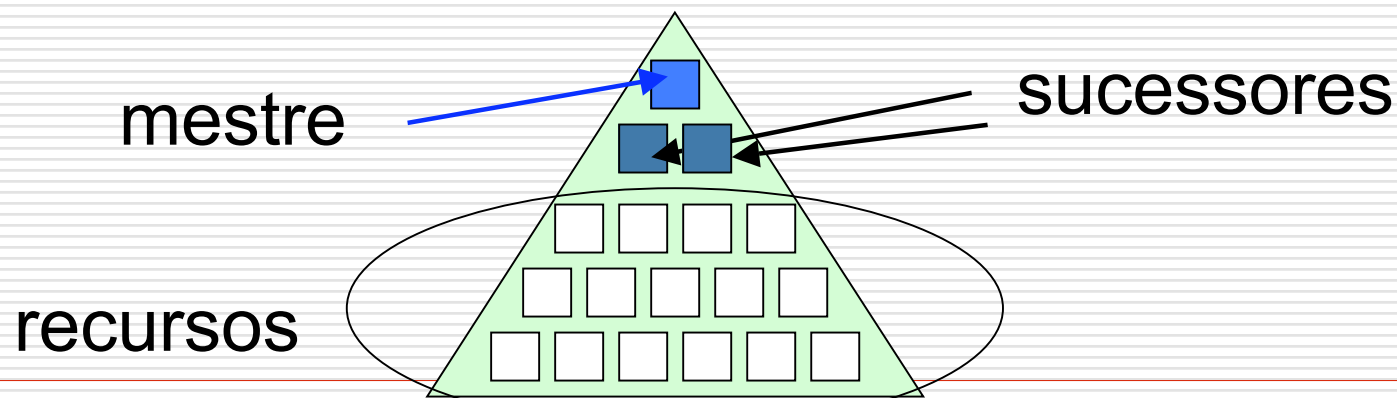




# PARADIS: Nível de Domínio

---

- ❑ Entidade administrativa
- ❑ *1 mestre*: escalonar tarefas
- ❑  $n \geq 0$  sucessores: assumem em caso falha
- ❑ diversos recursos por domínio



# Tolerância a Falhas no Domínio

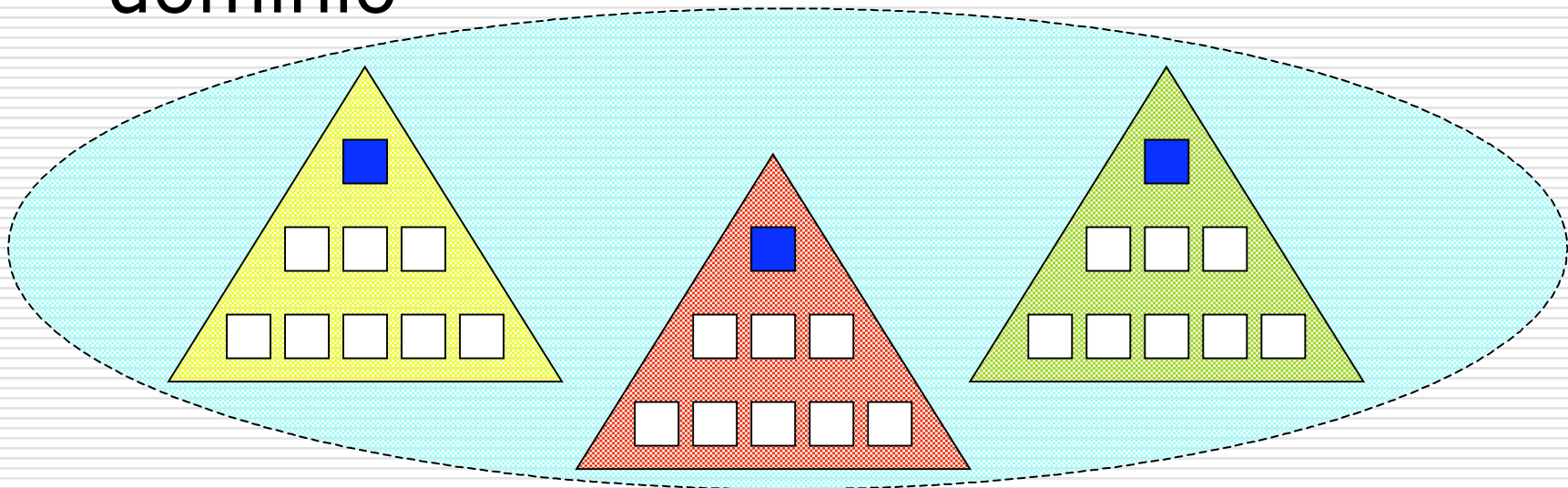
---

- Uso de detectores de falhas perfeitos
- Protocolos de Eleição
  - Escolha de novo mestre
  - Uso de serviço de Gestão da Composição
- ou Replicação Passiva
  - Troca de mestre
  - Controle consistente da atividade de escalonamento

# PARADIS: Nível de Grade

---

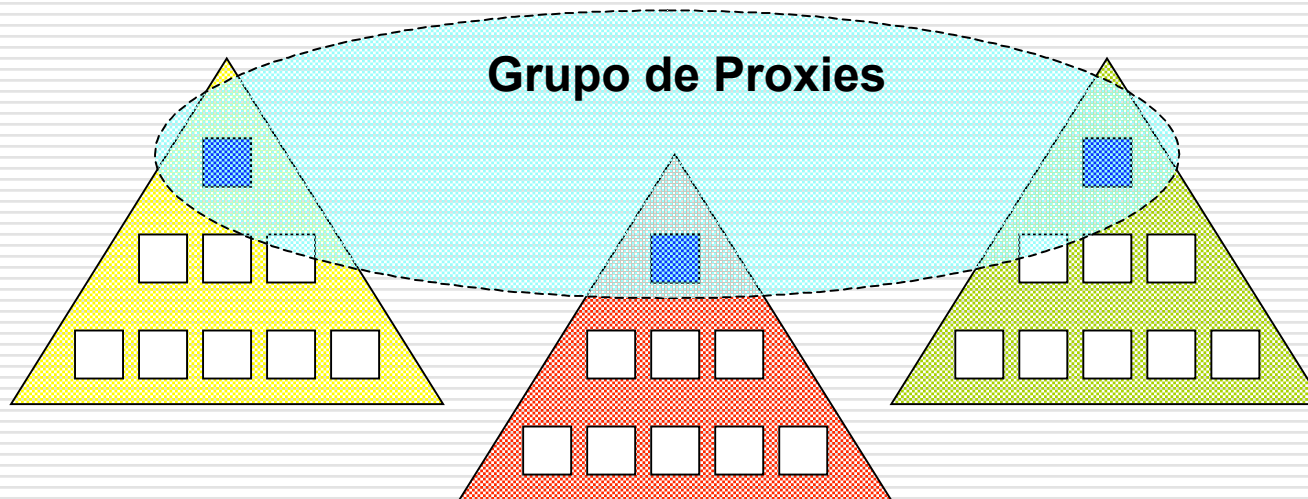
- 1 representante por domínio
  - Mestre (proxy)
- Cada representante administra seu domínio



# PARADIS : Nível de Grade

---

- Apenas os representantes (proxies) comunicam-se entre si e formam um grupo



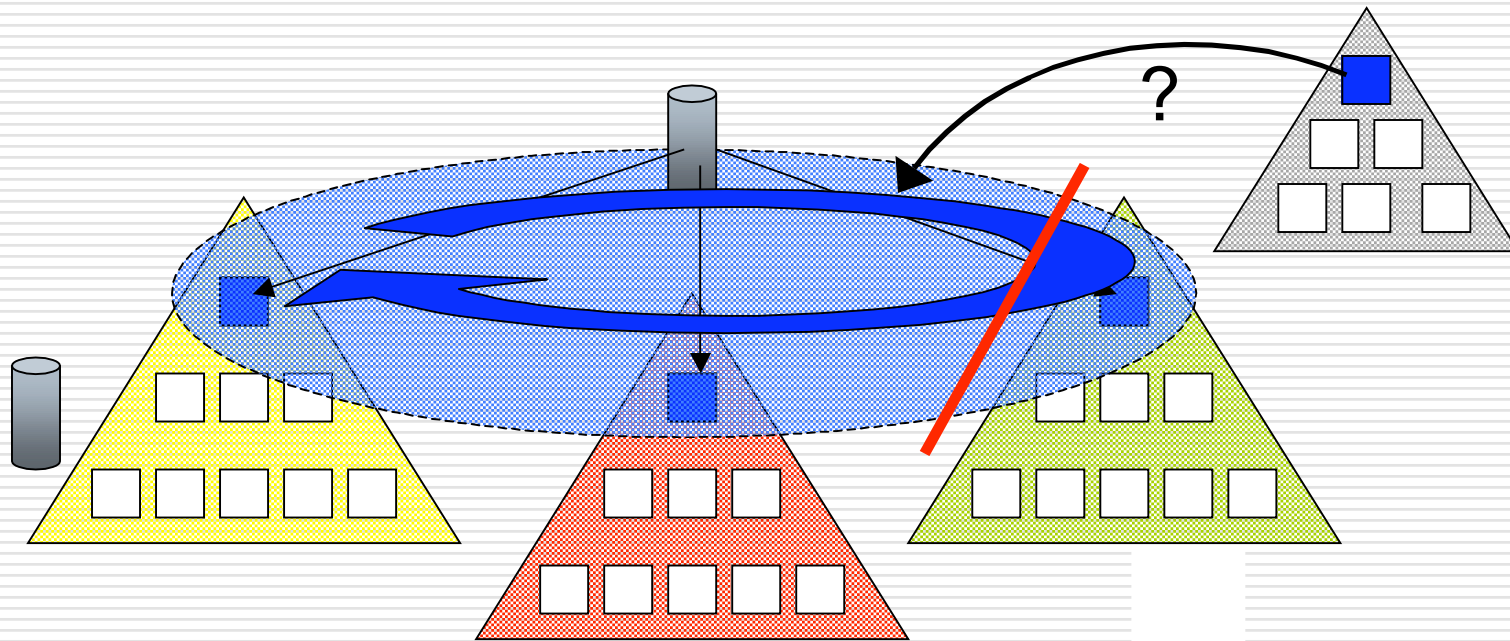
# Tolerância a Falhas no Nível de Grades

---

- Uso de detectores de falhas não confiáveis
  - Manutenção de lista de domínios inacessíveis
- Gestão da Composição de EDEN
  - Todos os domínios terão visão única da grade
  - Gestão da entrada/saída de domínios e falhas de proxies
  - PARADIS deve realocar tarefas para domínios que entram e saem da grade

# Uso de Acordo para Realocação de Tarefas na Grade

## ☐ Serviço de Consenso de EDEN

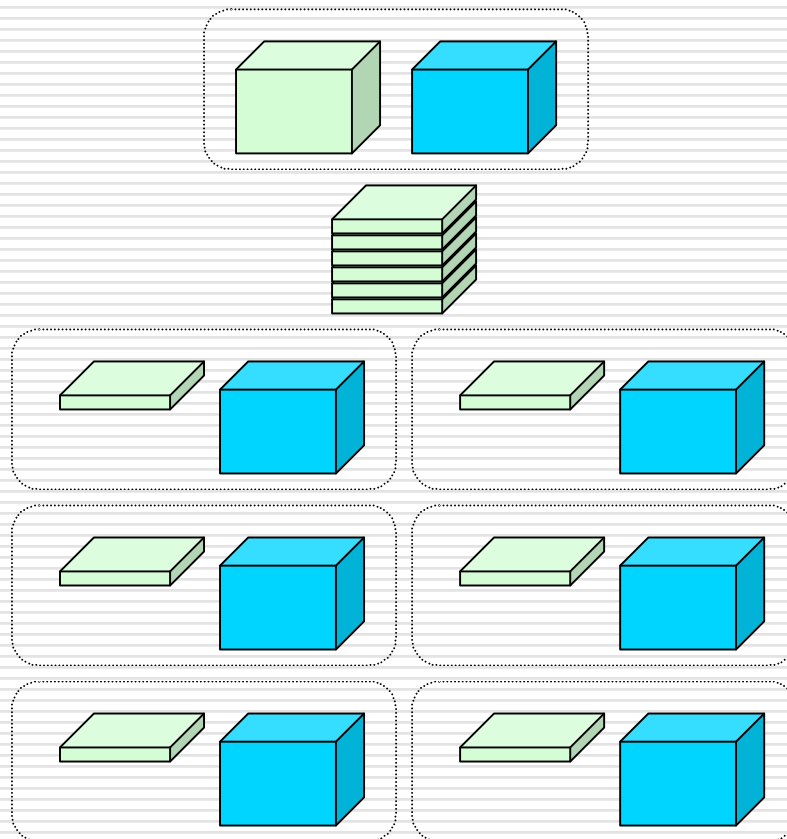


# Aplicações BoT (Tarefas Independentes)

---

- Aplicações com tarefas independentes
  - Não há necessidade de comunicação durante execução
- Aplicação representada por tarefa principal T
  - Formada por tarefas secundárias  $T_1, T_2, \dots, T_k$
  - Tarefas principais executam nos proxies
  - Tarefas secundárias utilizam os recursos da grade
- Cada aplicação deve registrar-se no domínio antes da execução
  - Fornecendo código e base de dados necessários para cada um dos domínios  $D_j$ .

# Exemplo: Comparação de Banco de Proteínas



```
T = comparison (banc1, banc2)
begin
```

```
  creation b1..b6
```

```
  ...
  T1 = SUBMIT comp(b1, banc
```

```
  2)
  T2 = SUBMIT comp(b2, banc
```

```
  2)
  ...
  T5 = SUBMIT comp(b5, banc
```

```
  2)
  T6 = SUBMIT comp(b6, banc
```

```
  2)
  WAIT(T1,...,T6)
```

```
  ...
end
```



# Mecanismo de Alocação de Tarefas

## Uso do Consenso Distribuído

---

- Mecanismo de Leilão
- Cada domínio avalia se pode executar tarefa e propõe estimativa de tempo para executá-la (« *lance* »)
  - Escrutina carga de cada recurso  $R_j$  e compatibilidade entre recurso  $R_j$  e tarefa  $T_i$  (memória, SO, etc.)
  - menor lance, maior possibilidade de terminar  $T_i$
- Protocolo de consenso distribuído entre proxies, através de EDEN
  - Obtenção de decisão unânime sobre quem executa  $T_i$

# Algoritmo de Alocação de Tarefas

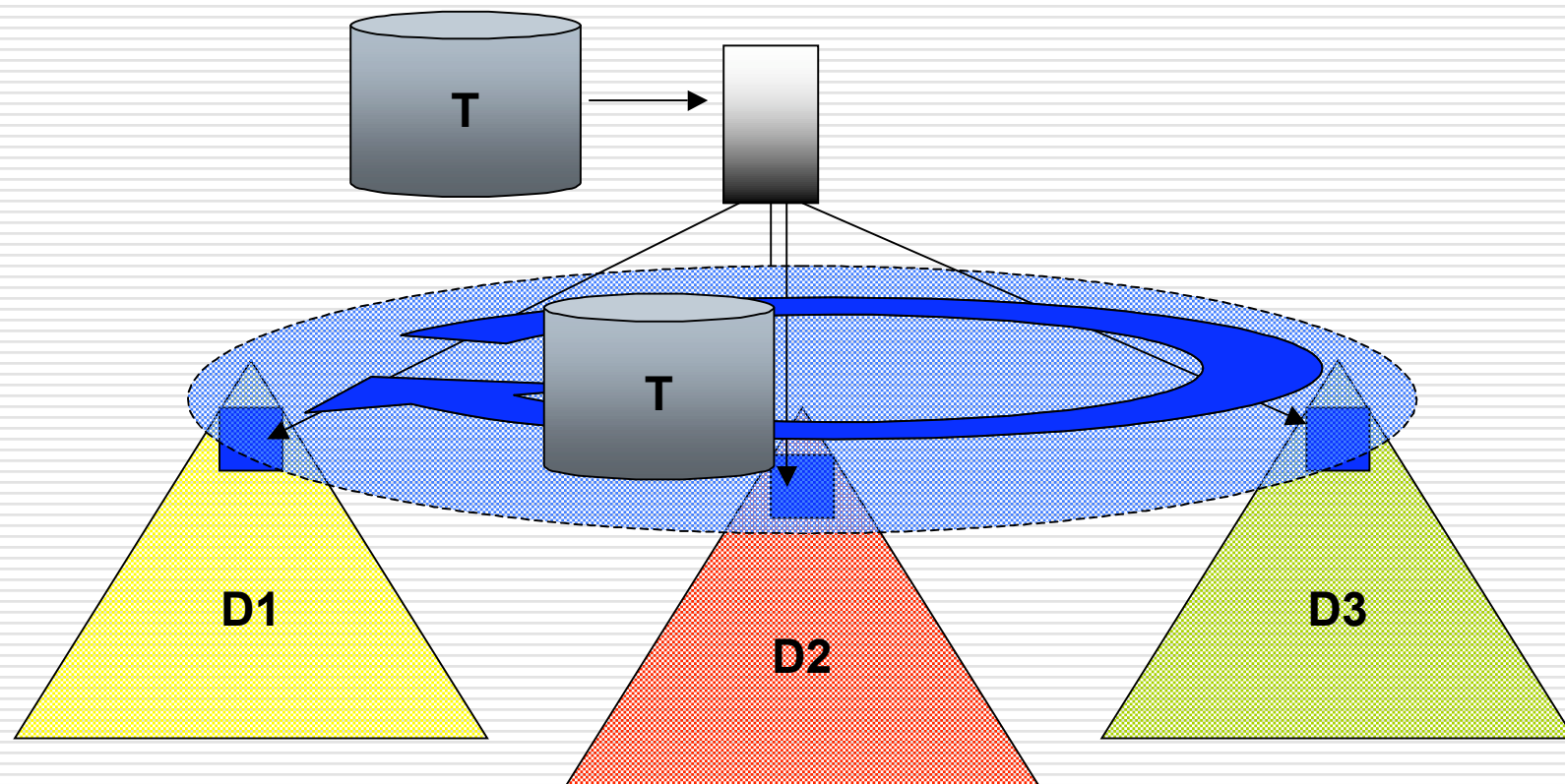
## Alocação da Tarefa Principal

---

- Uma tarefa principal  $T$  é submetida por usuário através de portal Web
- Tarefa  $T$  é difundida para o grupo de proxies
- Protocolo de consenso distribuído entre proxies
  - CO-Propose <estimativa  $e_i$  para tarefa  $T$ > do domínio  $D_i$
  - CO-Decide <[ $e_1, e_2, \dots, e_n$ ]>
- Decisão unânime sobre vetor de estimativas
- Cada proxy localmente escolhe, de maneira unívoca, o melhor domínio  $D$  para executar a tarefa  $T$

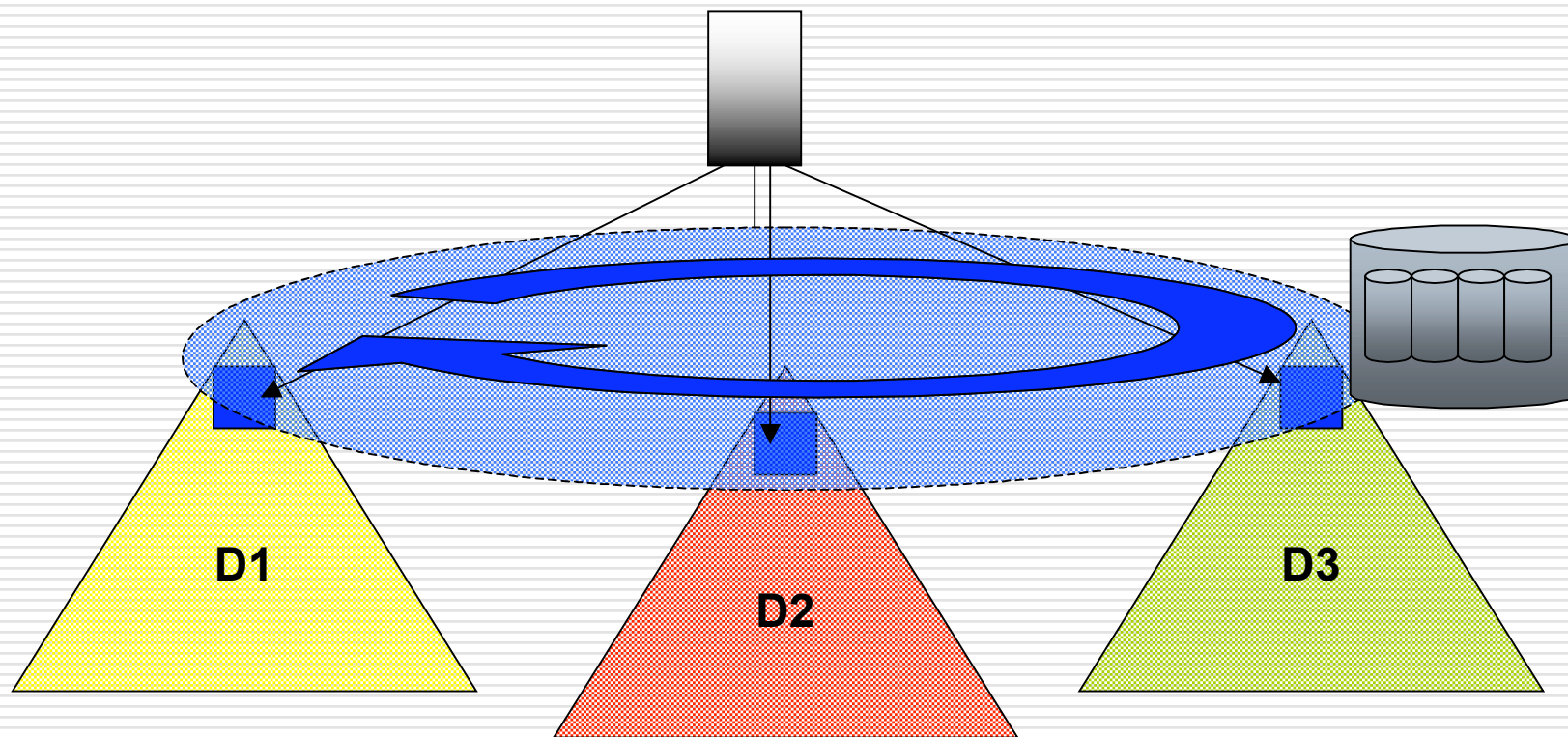
# PARADIS : Execução de Tarefa Principal

---



# PARADIS : Execução de Tarefa Principal

---



# Algoritmo de Alocação de Tarefas

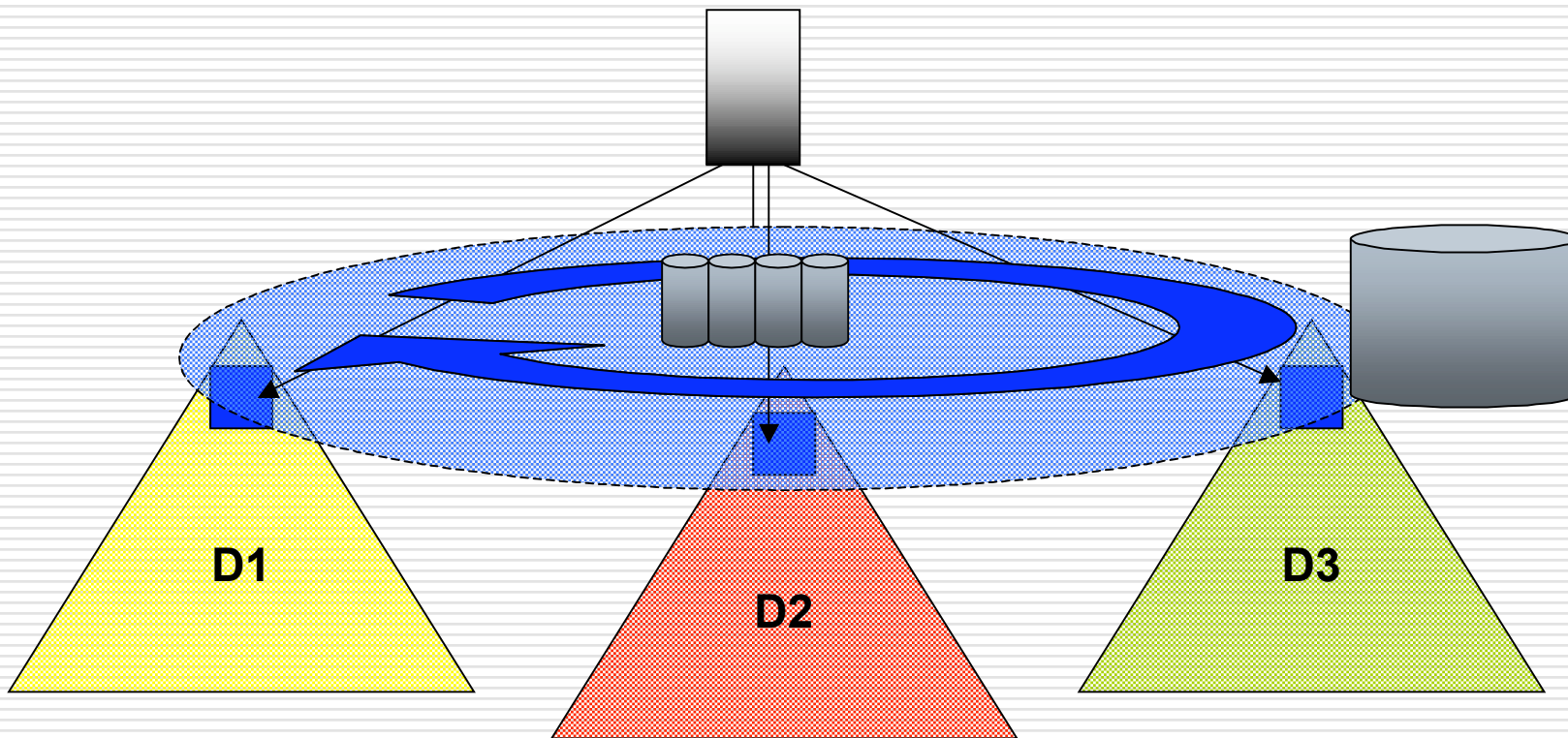
## Alocação das Tarefas Secundárias

---

- O domínio D passa a gerenciar a execução de T
- Tarefa T gera várias tarefas secundárias  $T_k$
- Para cada tarefa  $T_k$ , proxies executam consenso
  - CO-Propose <estimativa  $e_i$  para  $T_k$ > do domínio  $D_i$
  - CO-Decide <[ $e_1, e_2, \dots, e_n$ ]>
  - Decide-se domínio  $D_j$  a executar  $T_k$
  - $T_k$  movida para domínio  $D_j$
  - Ao término de  $T_k$ 
    - Envia-se resultados para domínio D que executa tarefa T
- Ao término de T
  - Usuário avisado pelo proxy onde coletar resultados

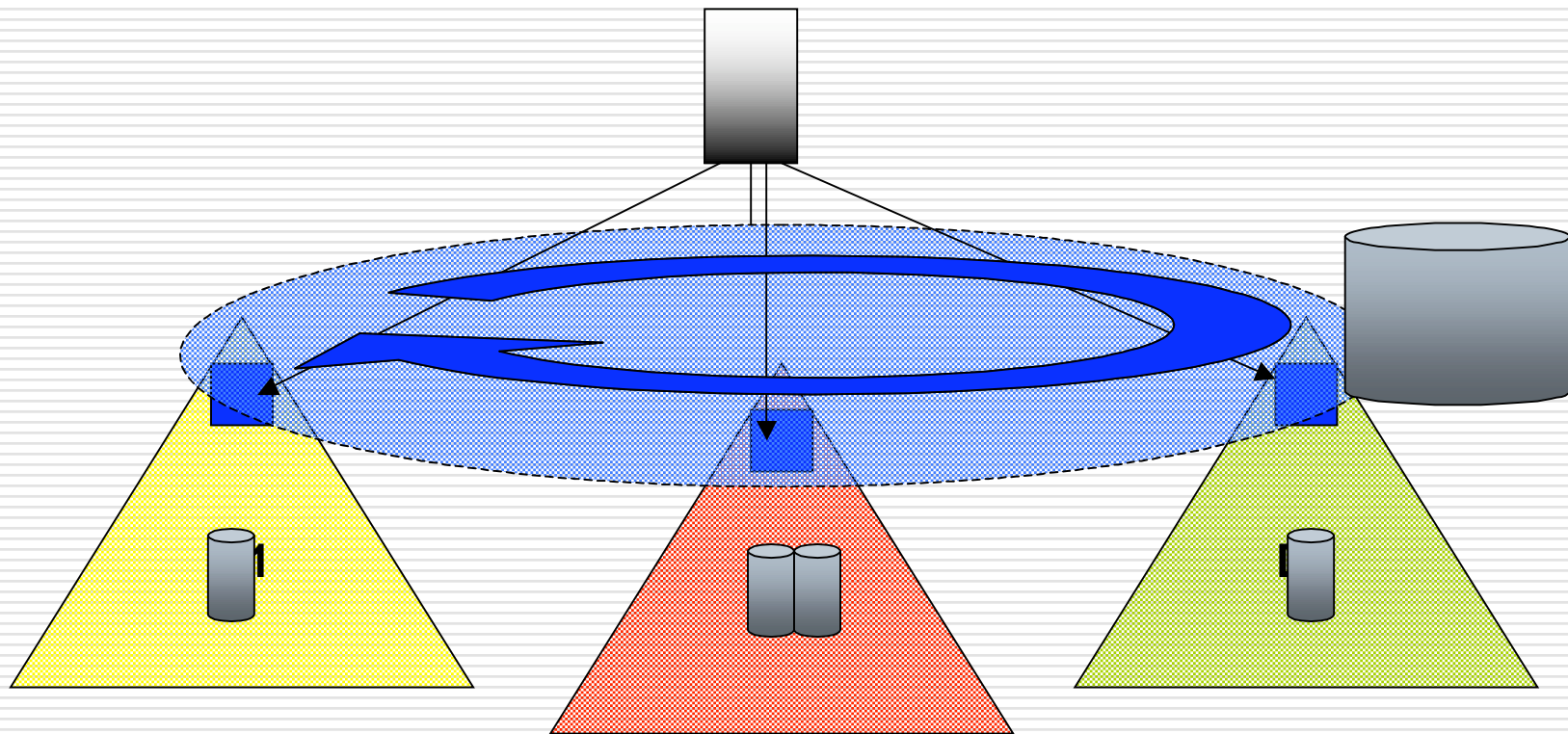
# PARADIS : Execução de Tarefas Secundárias

---



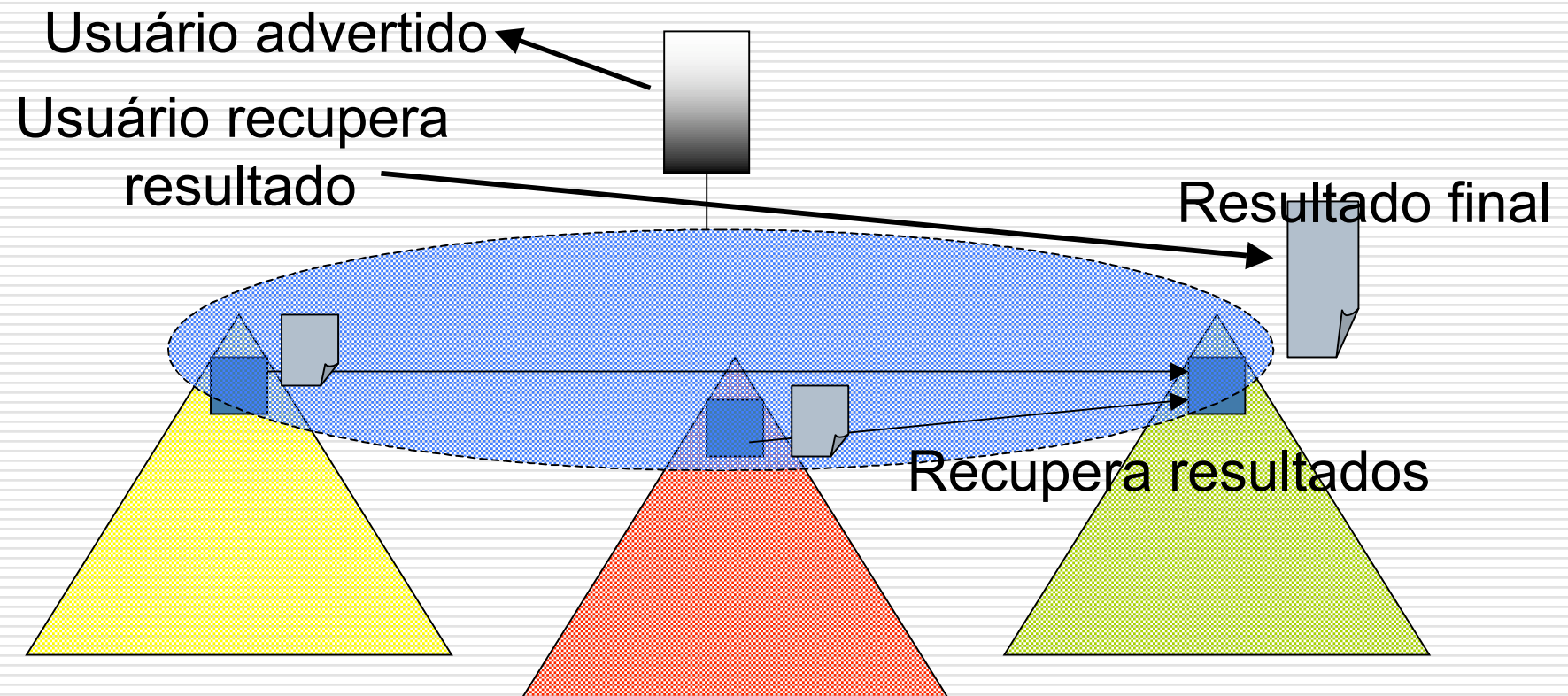
# PARADIS : Execução de Tarefas Secundárias

---



# PARADIS : Finalização e Entrega dos Resultados

---





# GAUDI

## Algumas Aplicações

---

- ❑ Reconhecimento de dados biométricos (impressão digital) em grades
- ❑ Simulação de redes de neurônios em grades

# Muito Obrigada!

---

Fabíola Gonçalves Pereira Greve

Gaudi - DCC - UFBA

[www.dcc.ufba.br/~fabiola](http://www.dcc.ufba.br/~fabiola)

<http://gaudi.dcc.ufba.br>