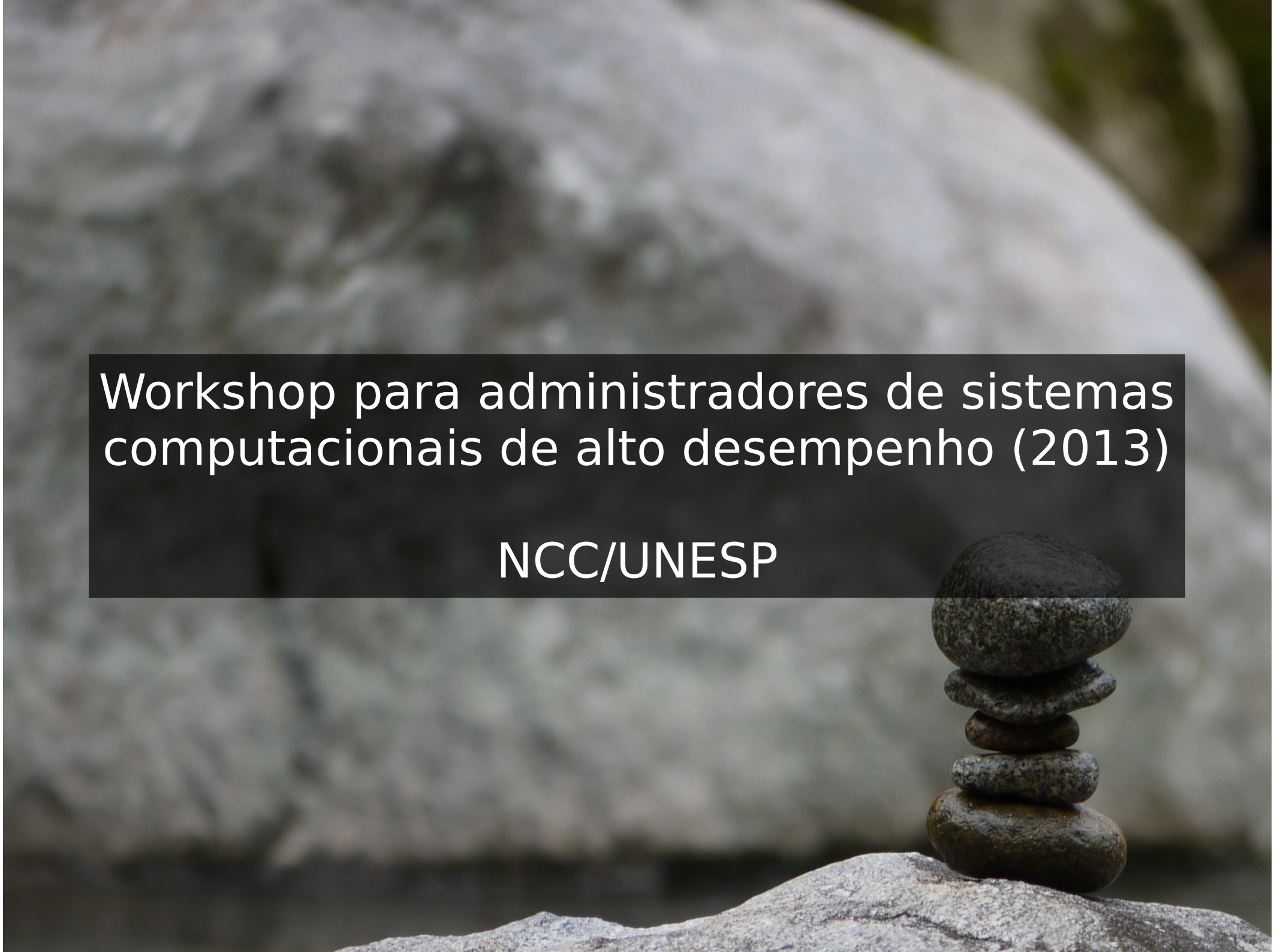
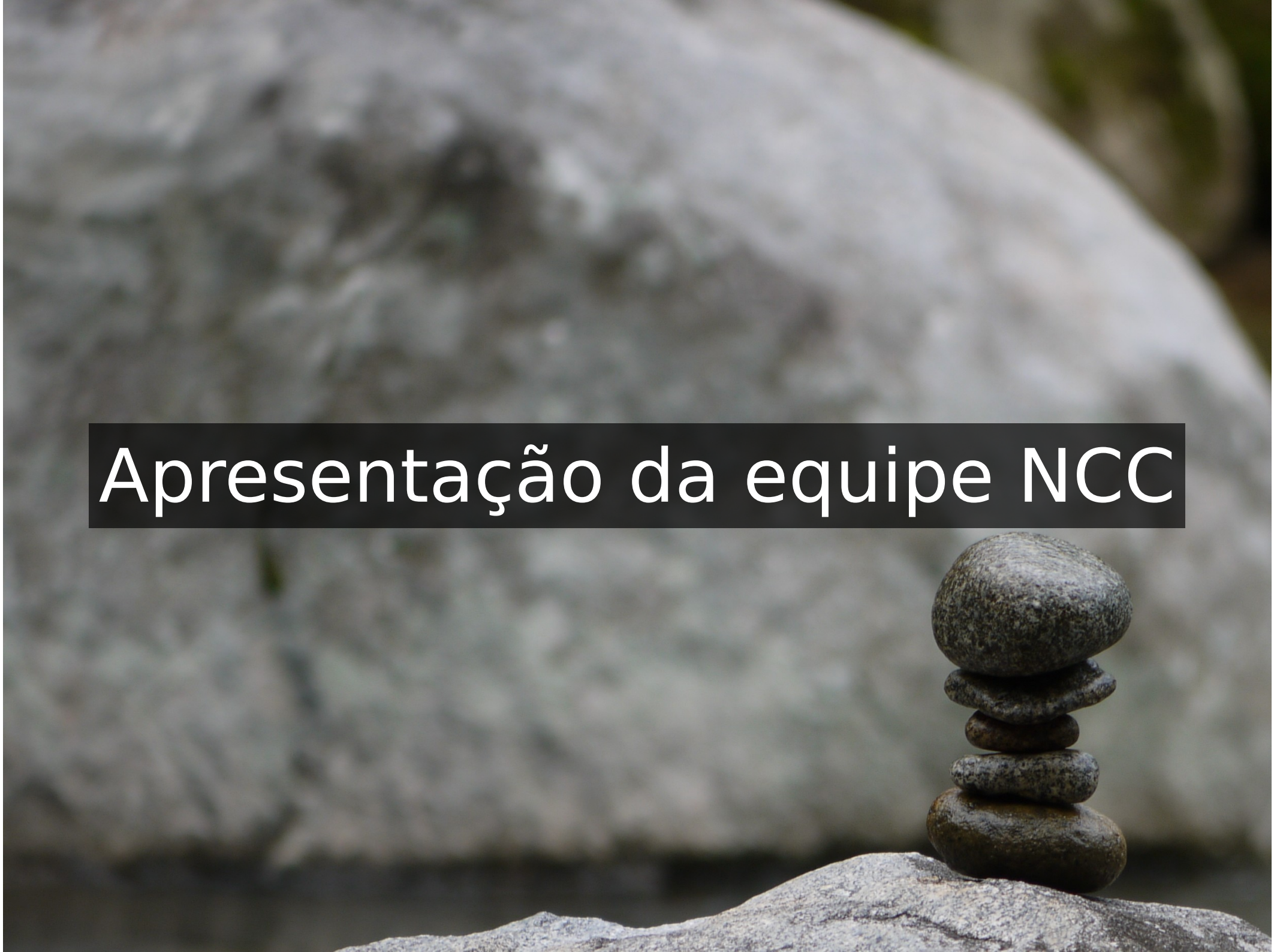
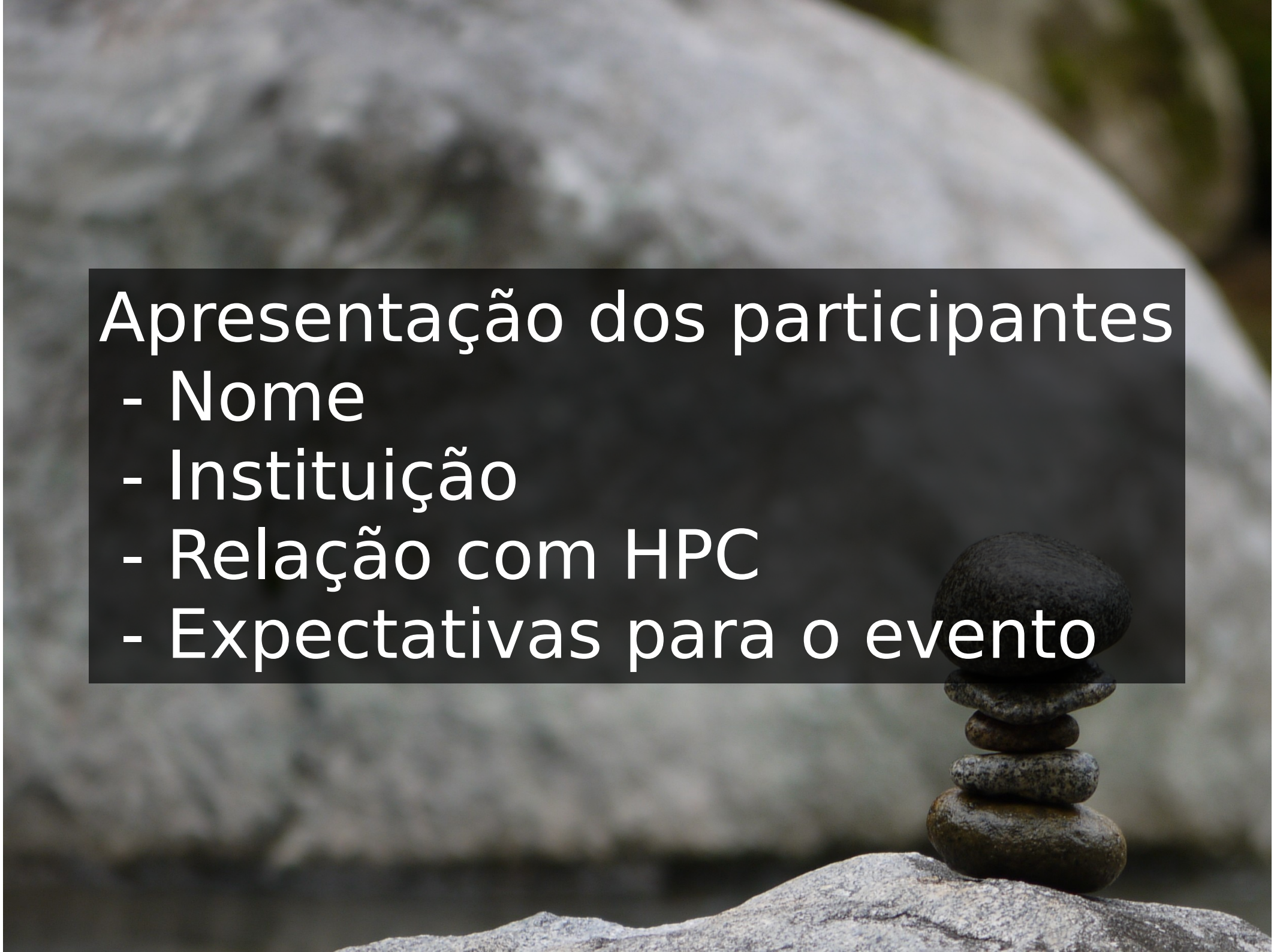
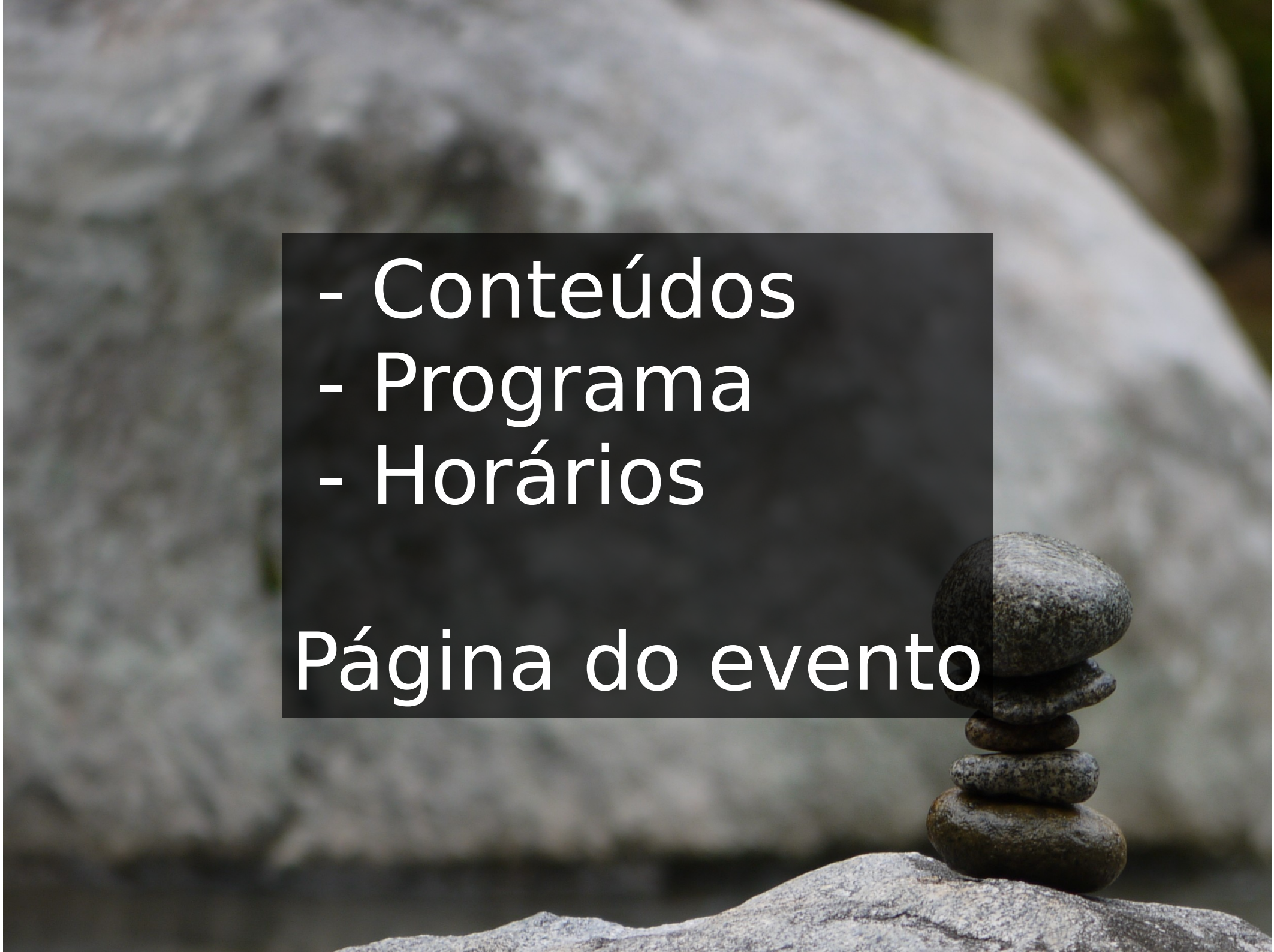
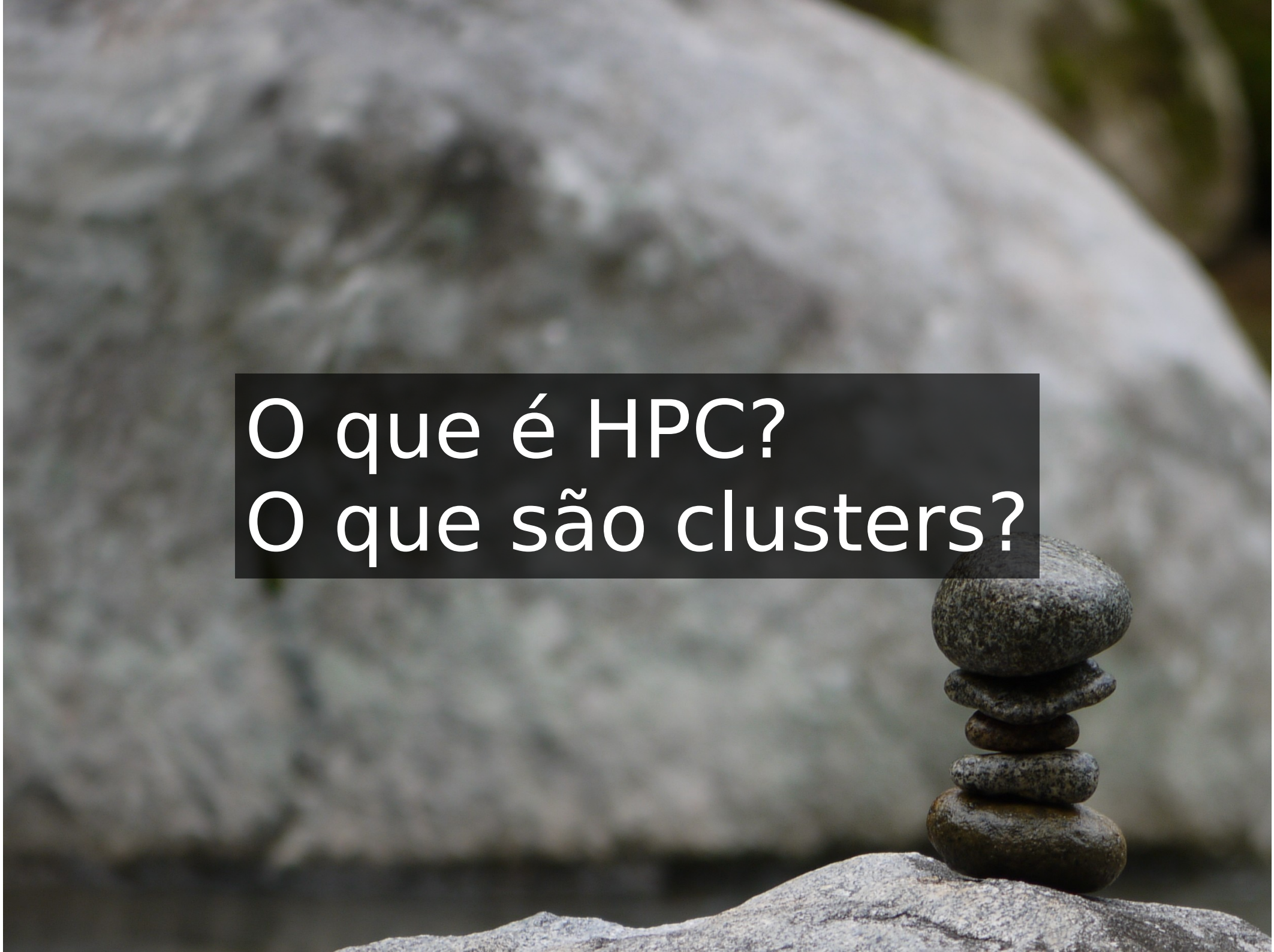

Workshop para administradores de sistemas computacionais de alto desempenho (2013)
NCC/UNESP


Apresentação da equipe NCC


Apresentação dos participantes
- Nome
- Instituição
- Relação com HPC
- Expectativas para o evento

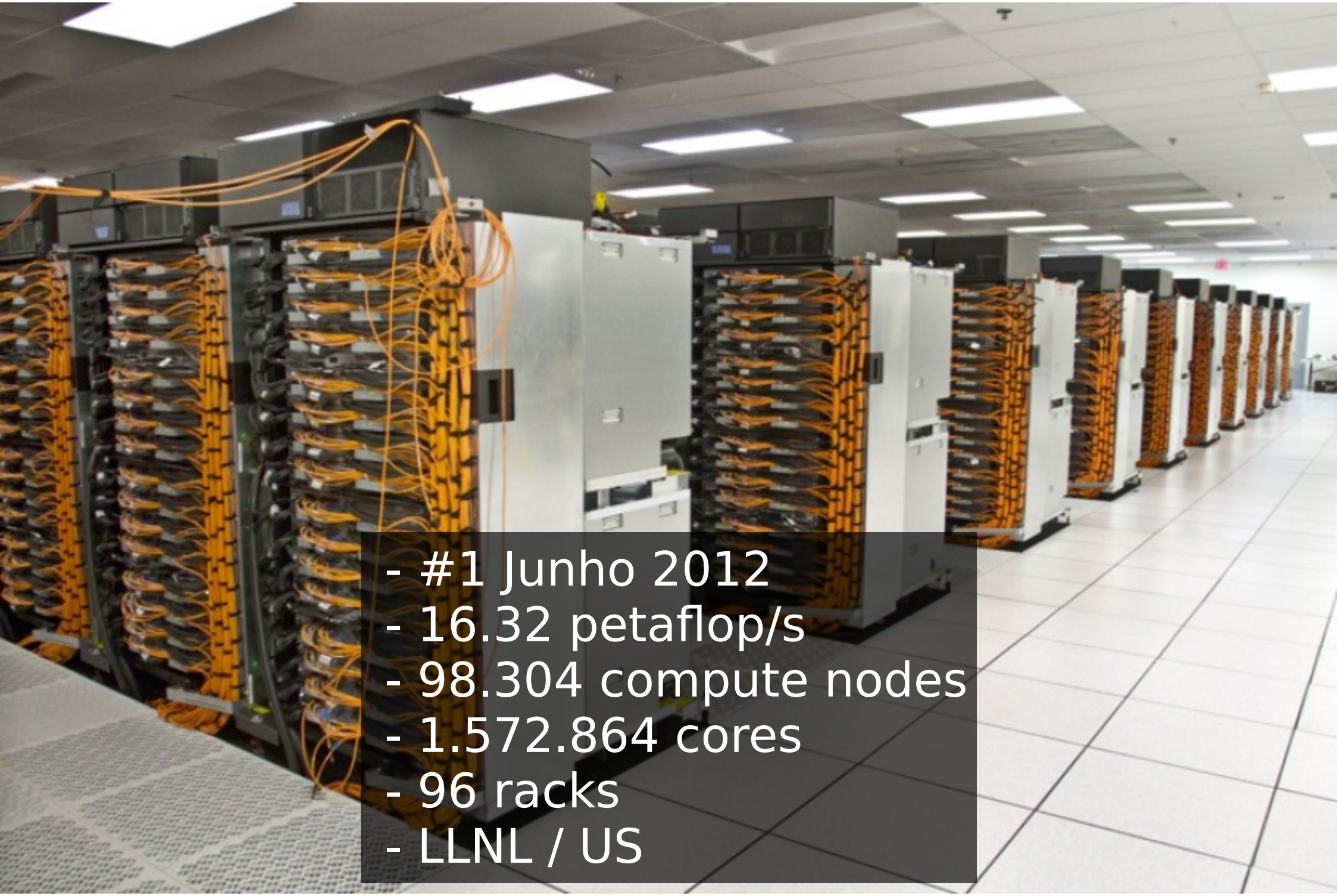

- Conteúdos
- Programa
- Horários
Página do evento


O que é HPC?
O que são clusters?


Arquitetura

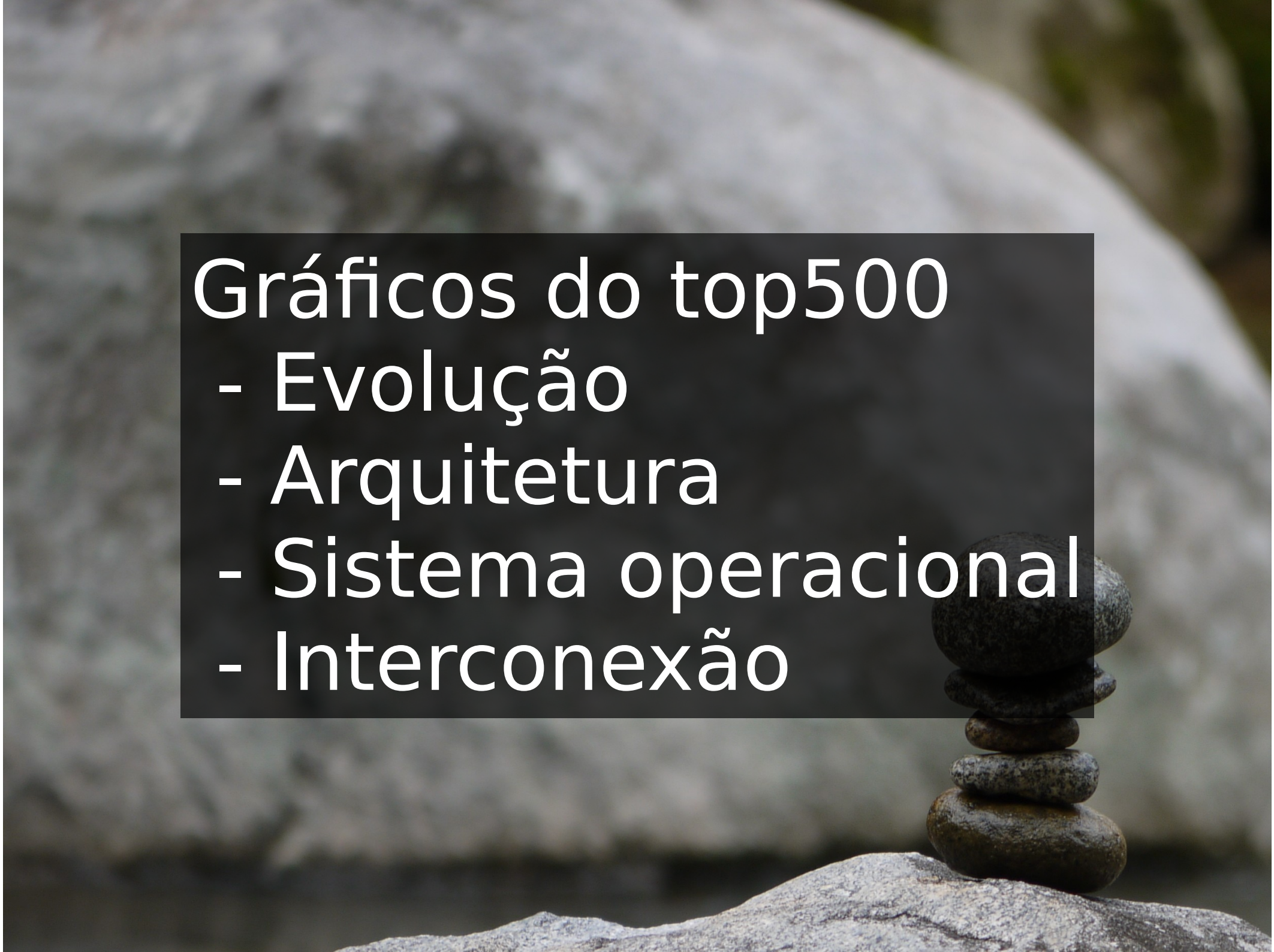

Genérico

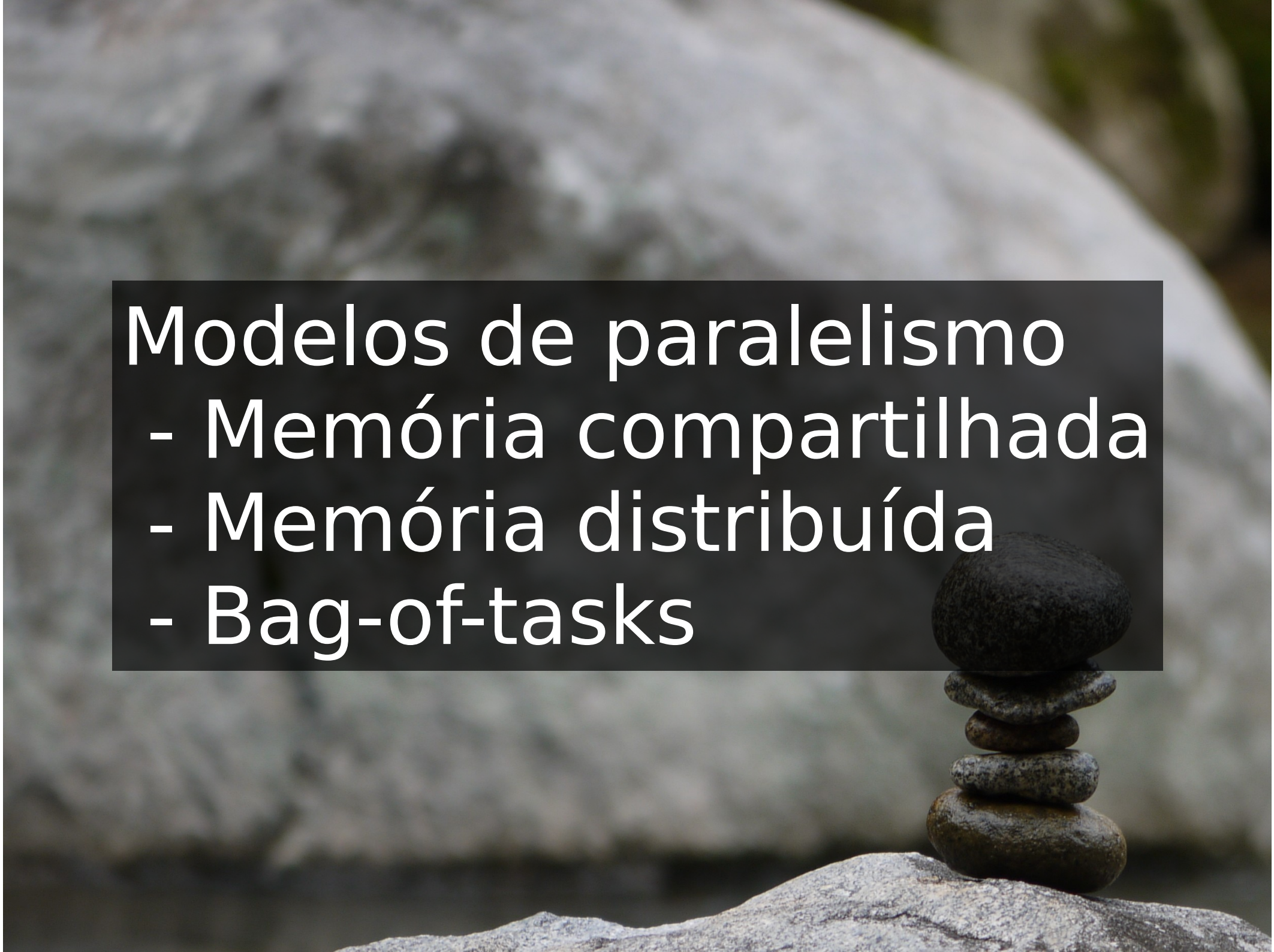

Mare Nostrum, Espanha


- #1 Junho 2012
- 16.32 petaflop/s
- 98.304 compute nodes
- 1.572.864 cores
- 96 racks
- LLNL / US



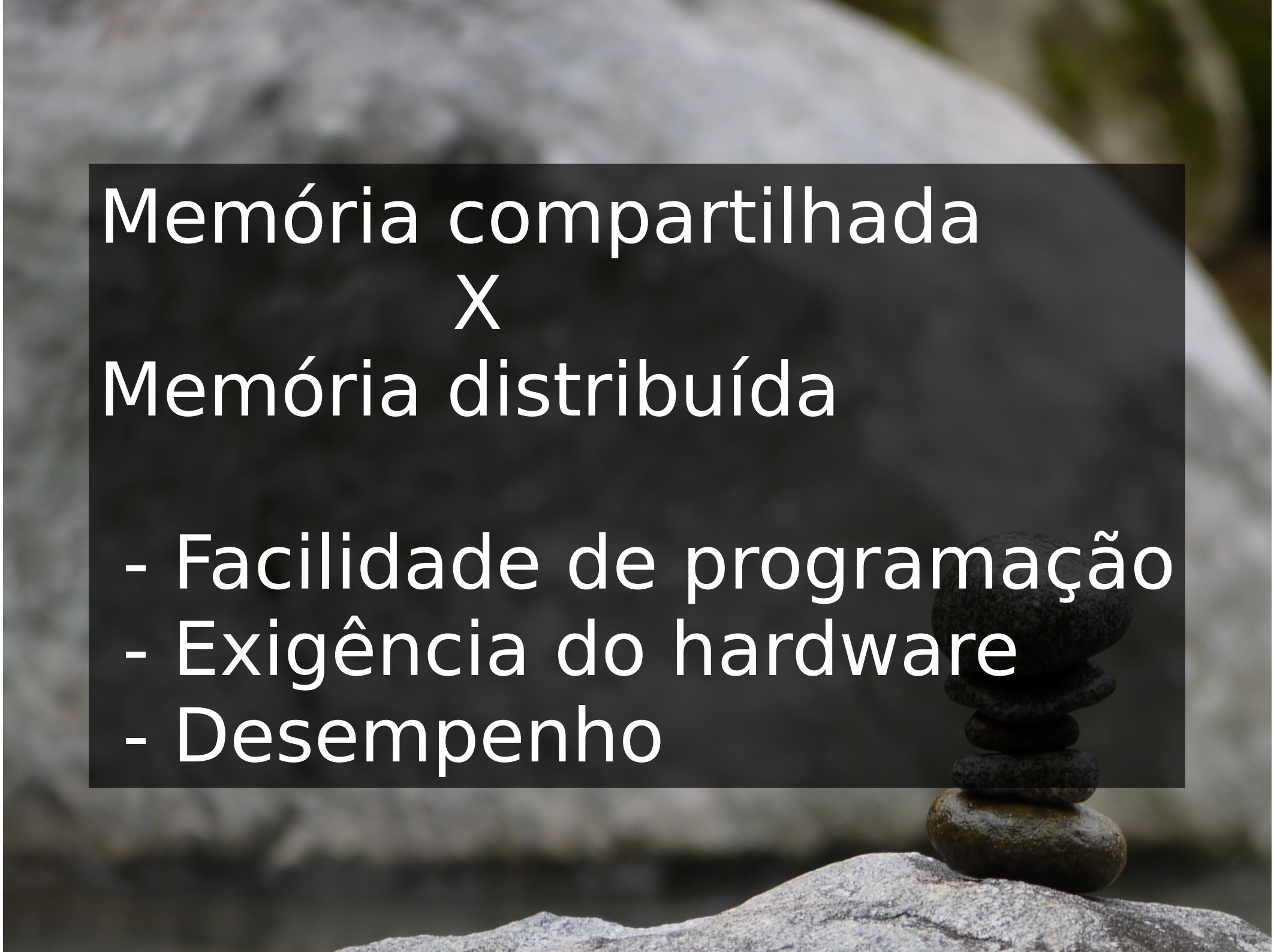

- #1 Novembro 2012
- 17.6 petaflop/s
- 552.960 processor system
- 18.688 nodes com GPGPU
- Oak Ridge National Laboratory / US

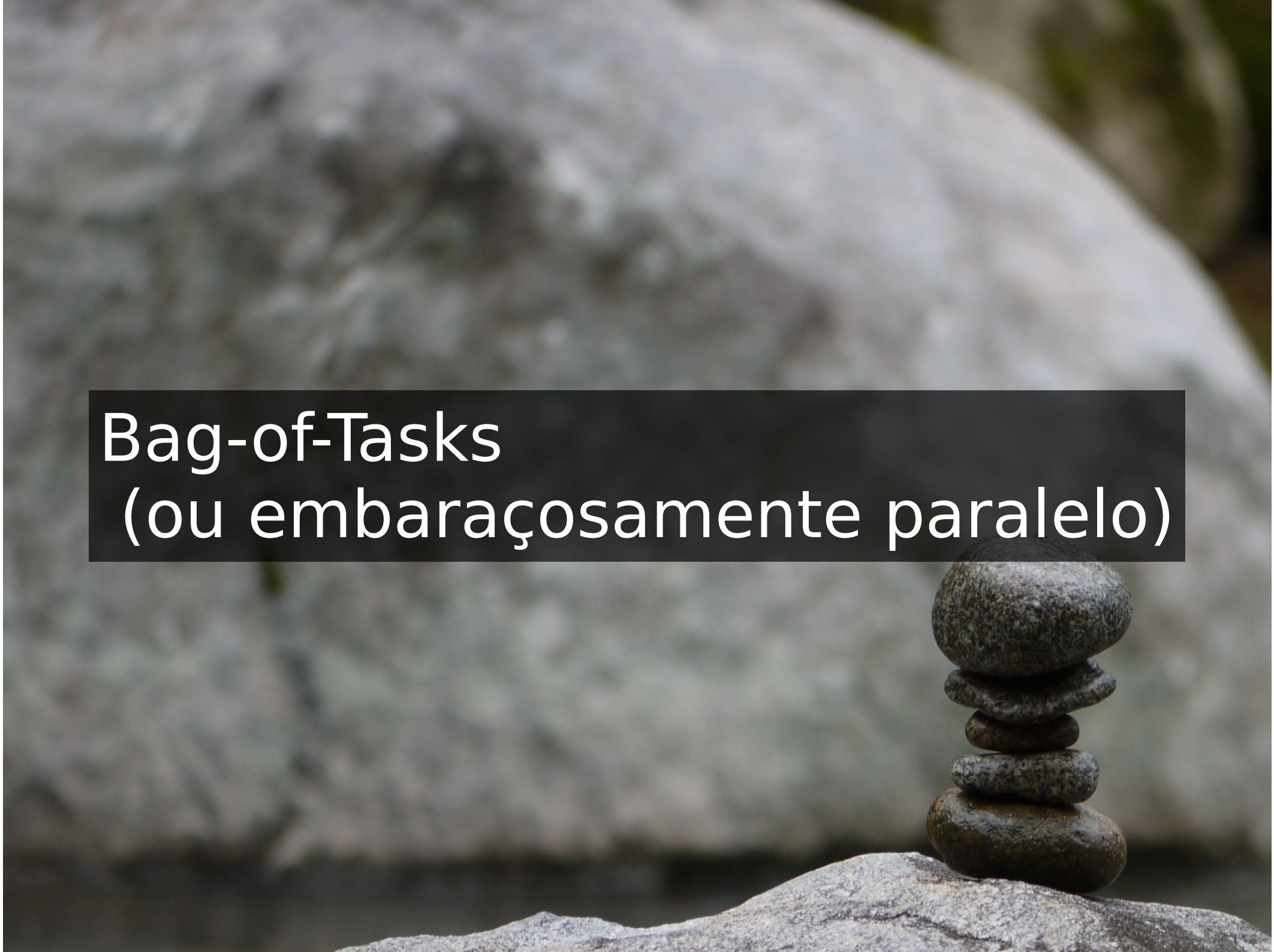

Gráficos do top500
- Evolução
- Arquitetura
- Sistema operacional
- Interconexão

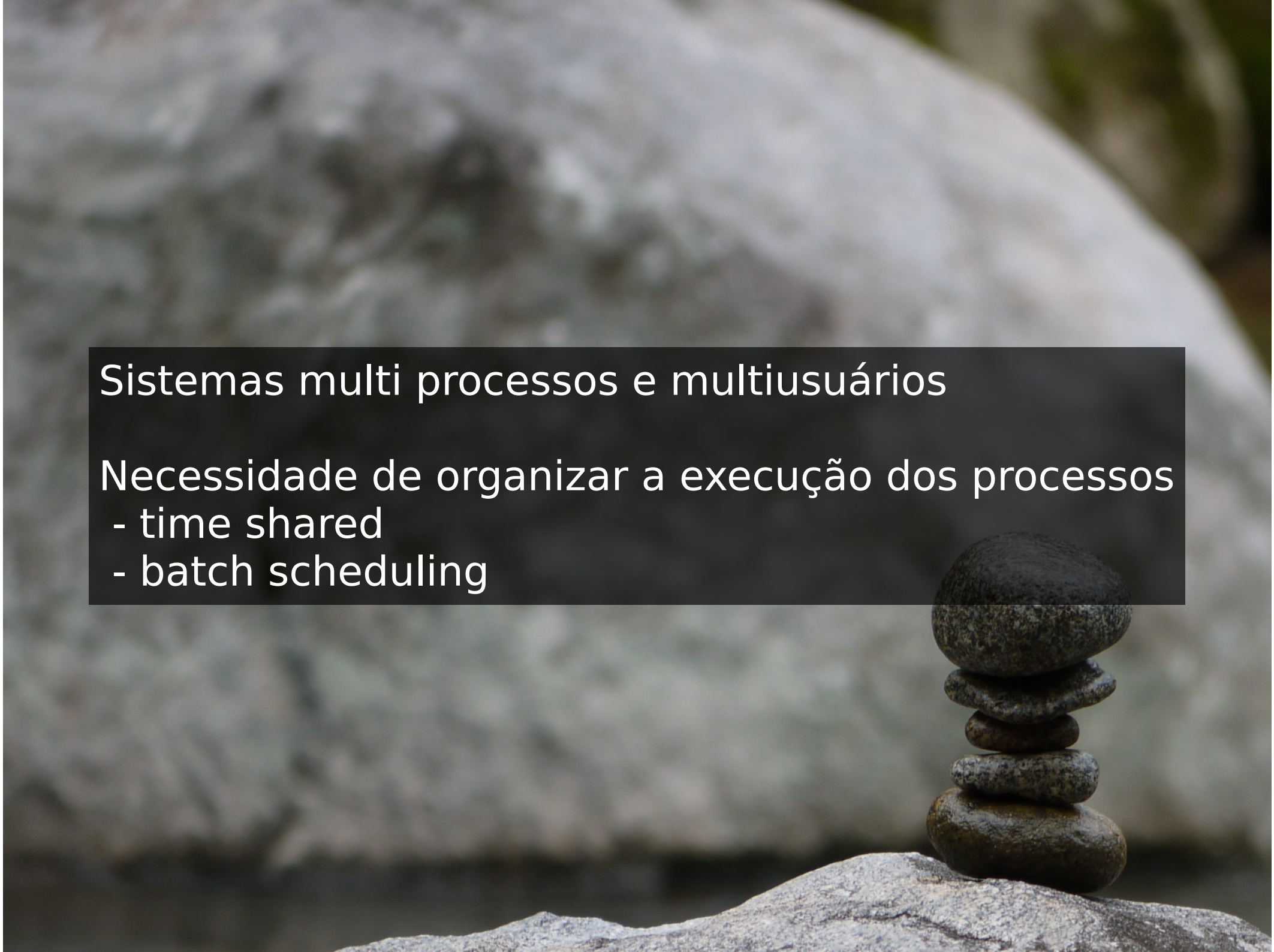

Modelos de paralelismo
- Memória compartilhada
- Memória distribuída
- Bag-of-tasks


Memória compartilhada
- SMP
- Threads
openMP


Memória distribuída
- Passagem de mensagens
MPI


Memória compartilhada
X
Memória distribuída
- Facilidade de programação
- Exigência do hardware
- Desempenho

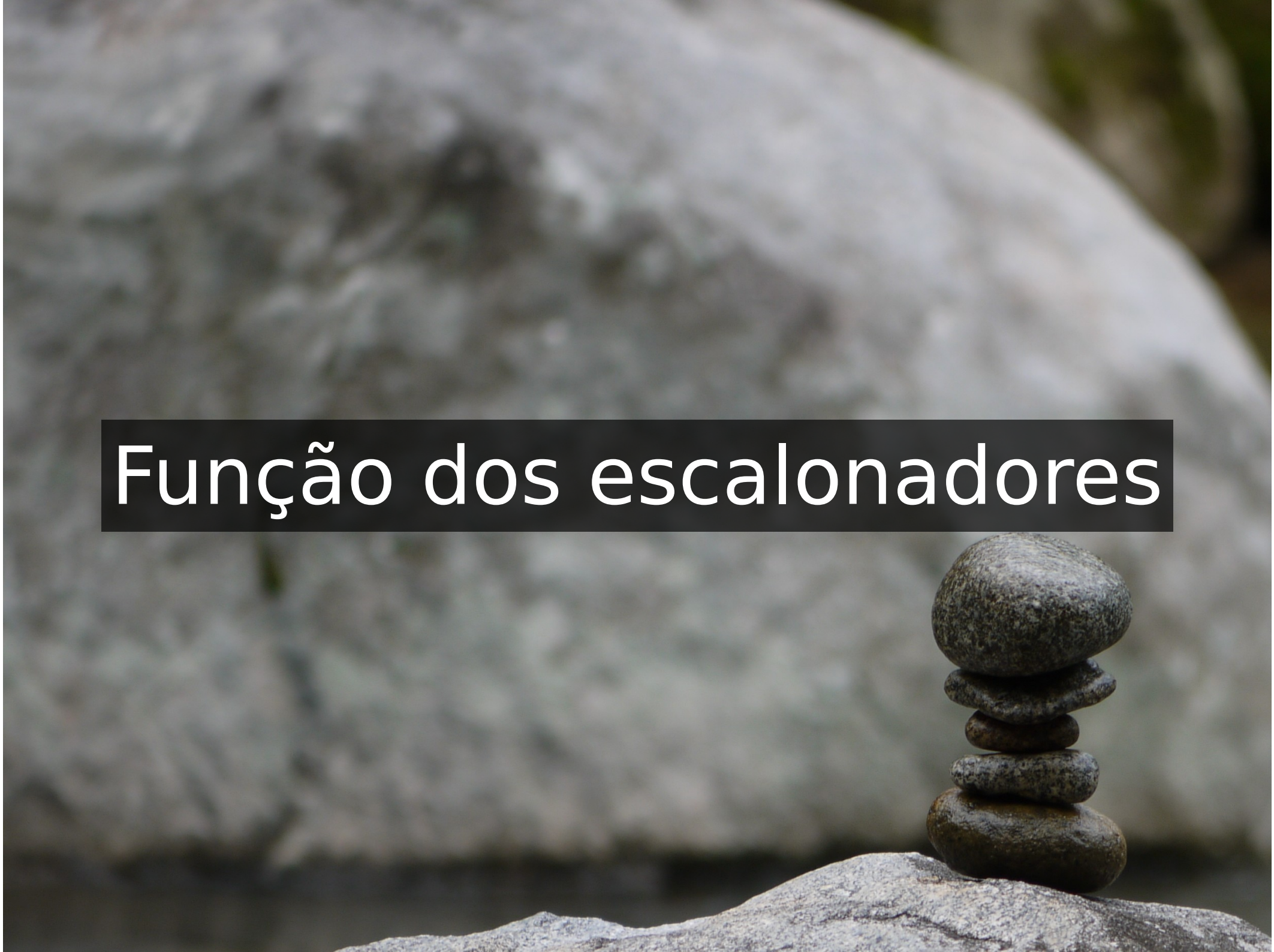

Bag-of-Tasks
(ou embaraçosamente paralelo)

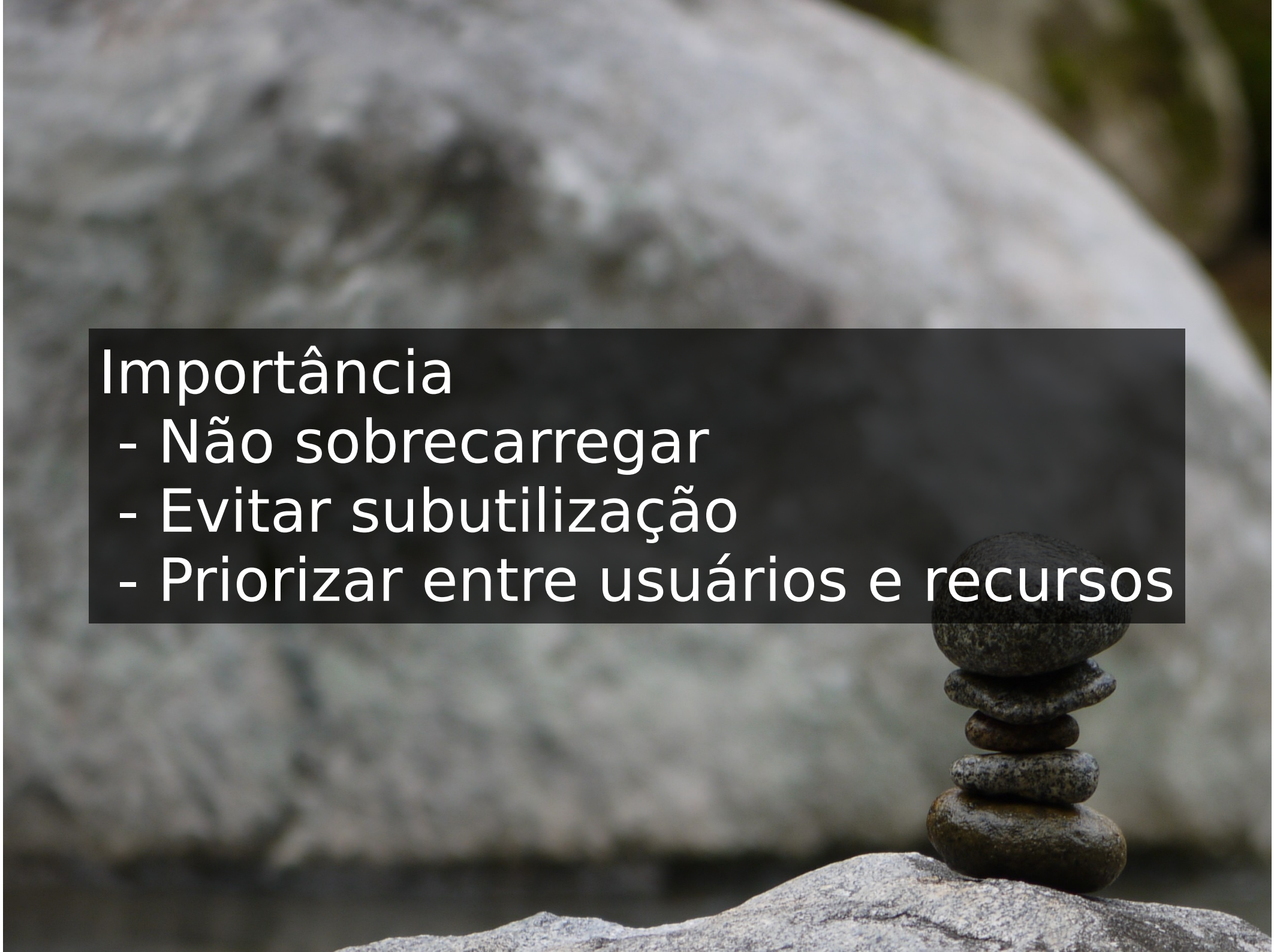


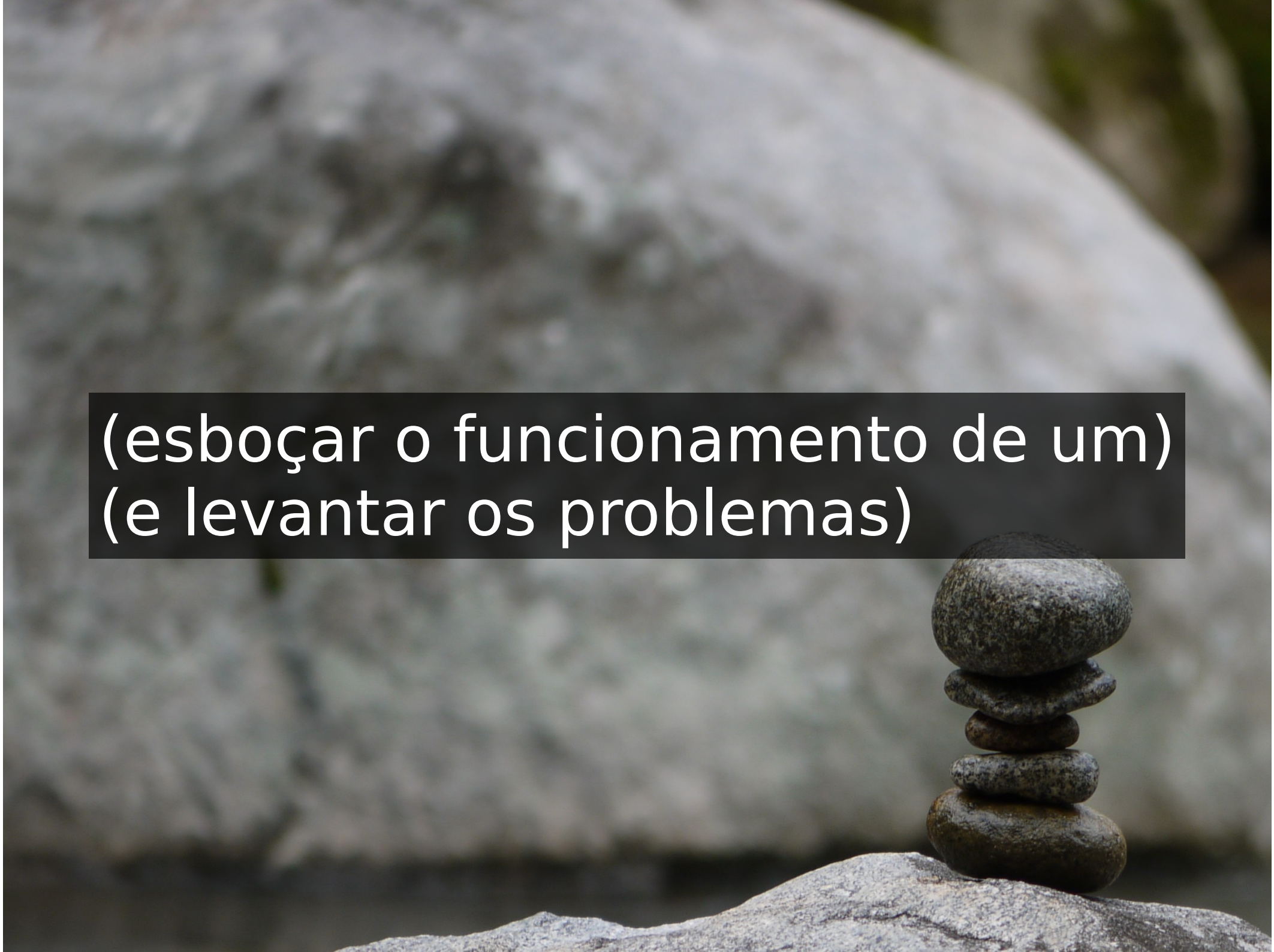
Sistemas multi processos e multiusuários

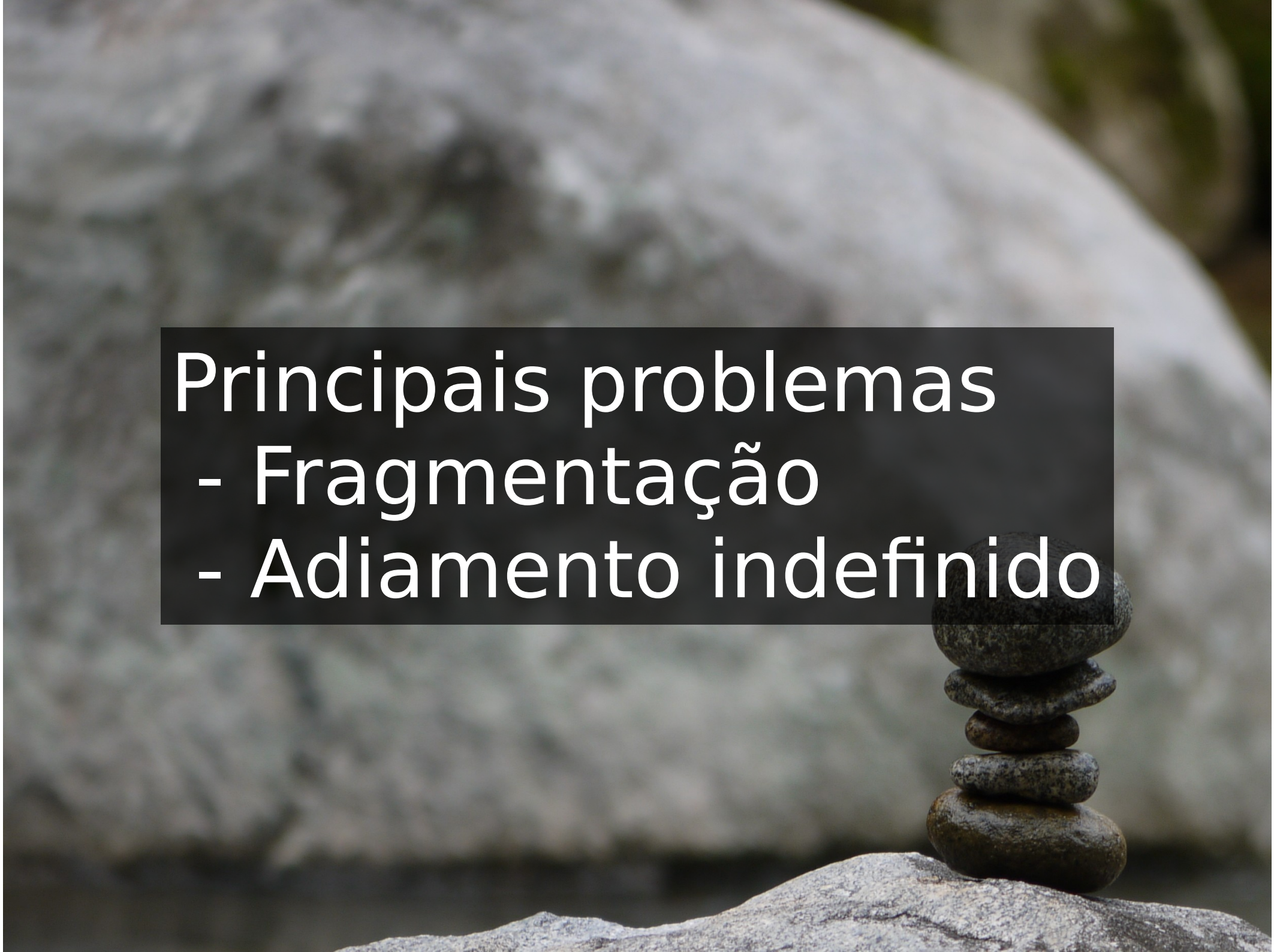
Necessidade de organizar a execução dos processos
 - time shared
 - batch scheduling

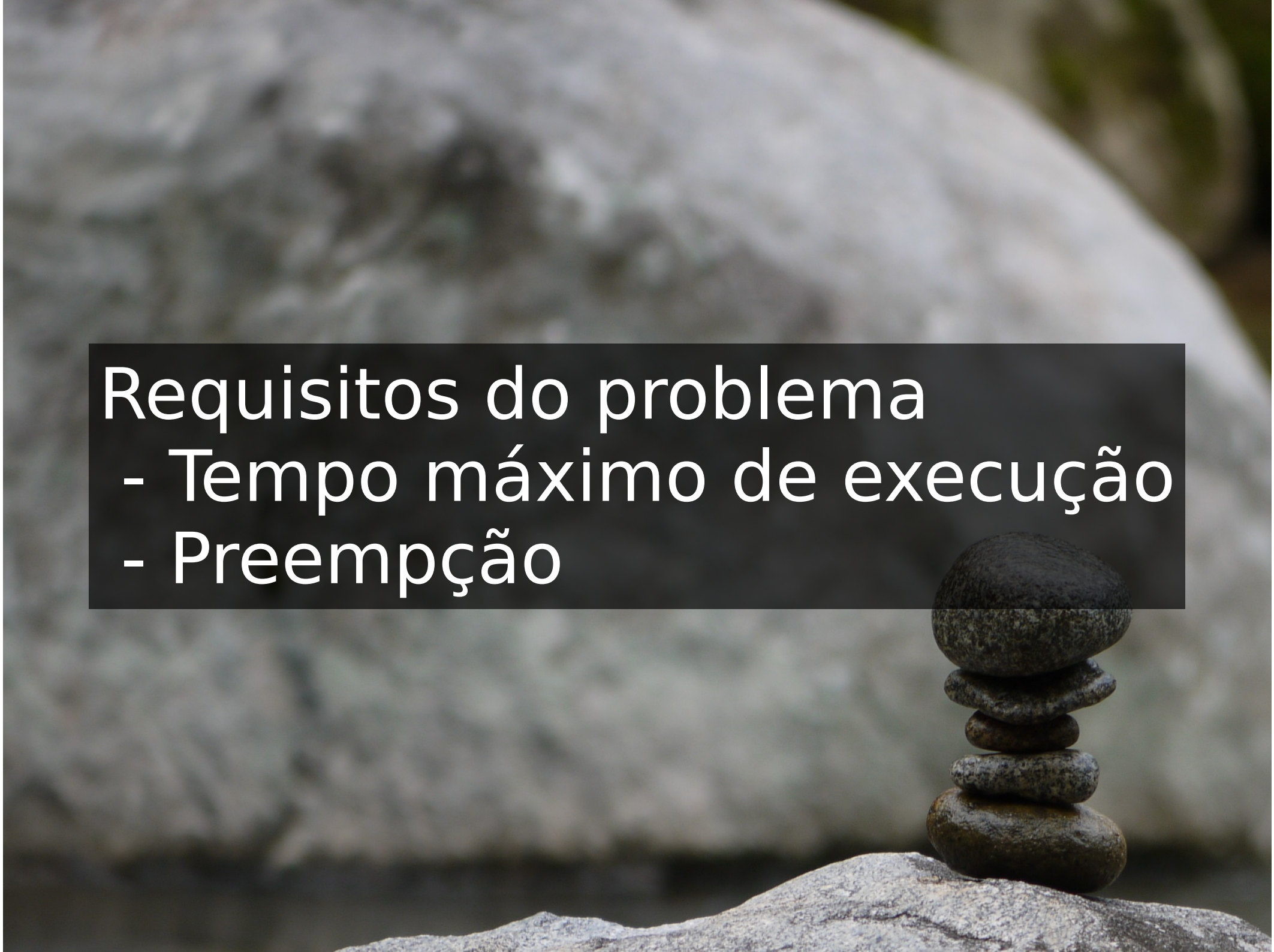

Escalonadores

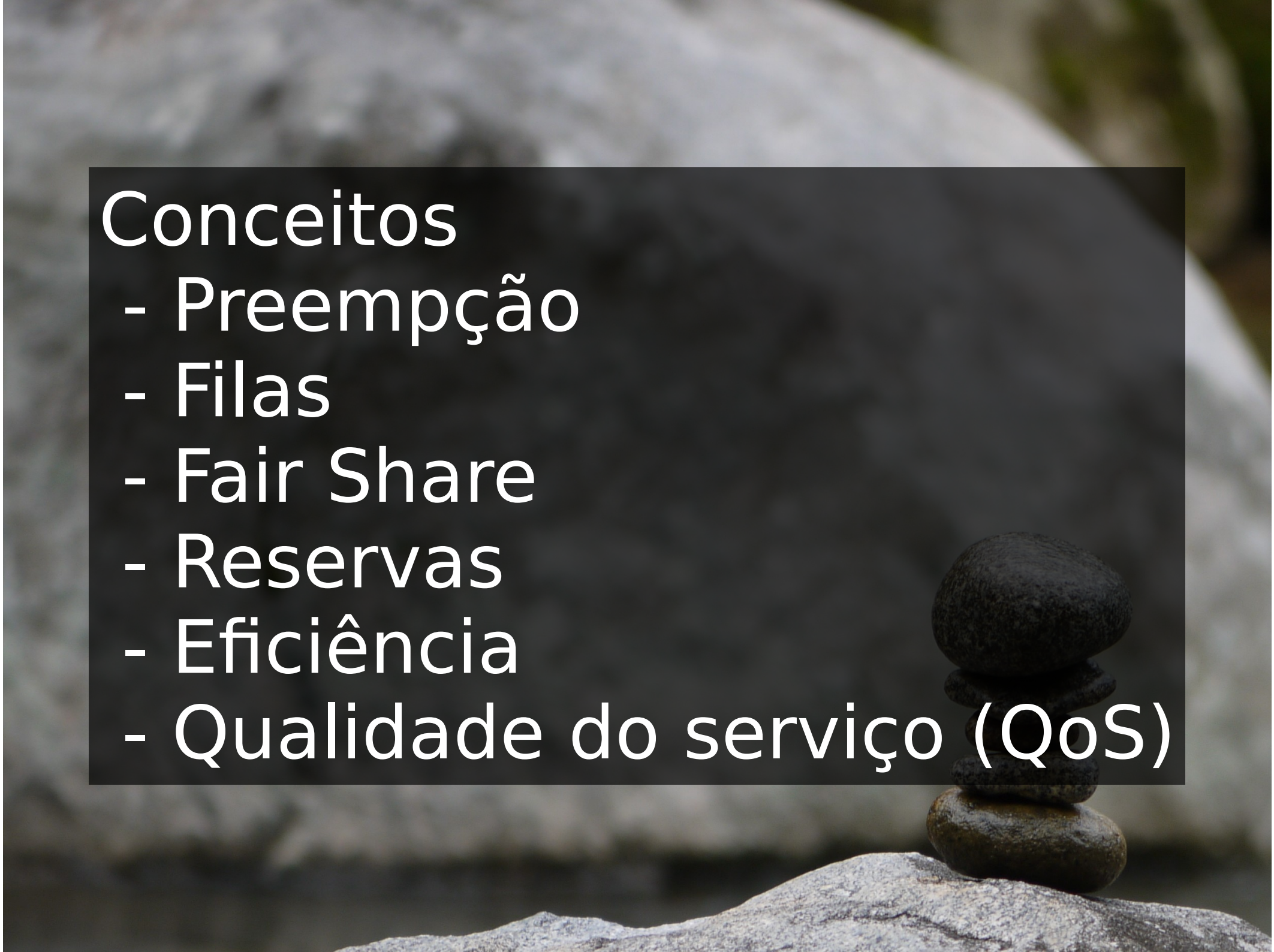

Função dos escalonadores

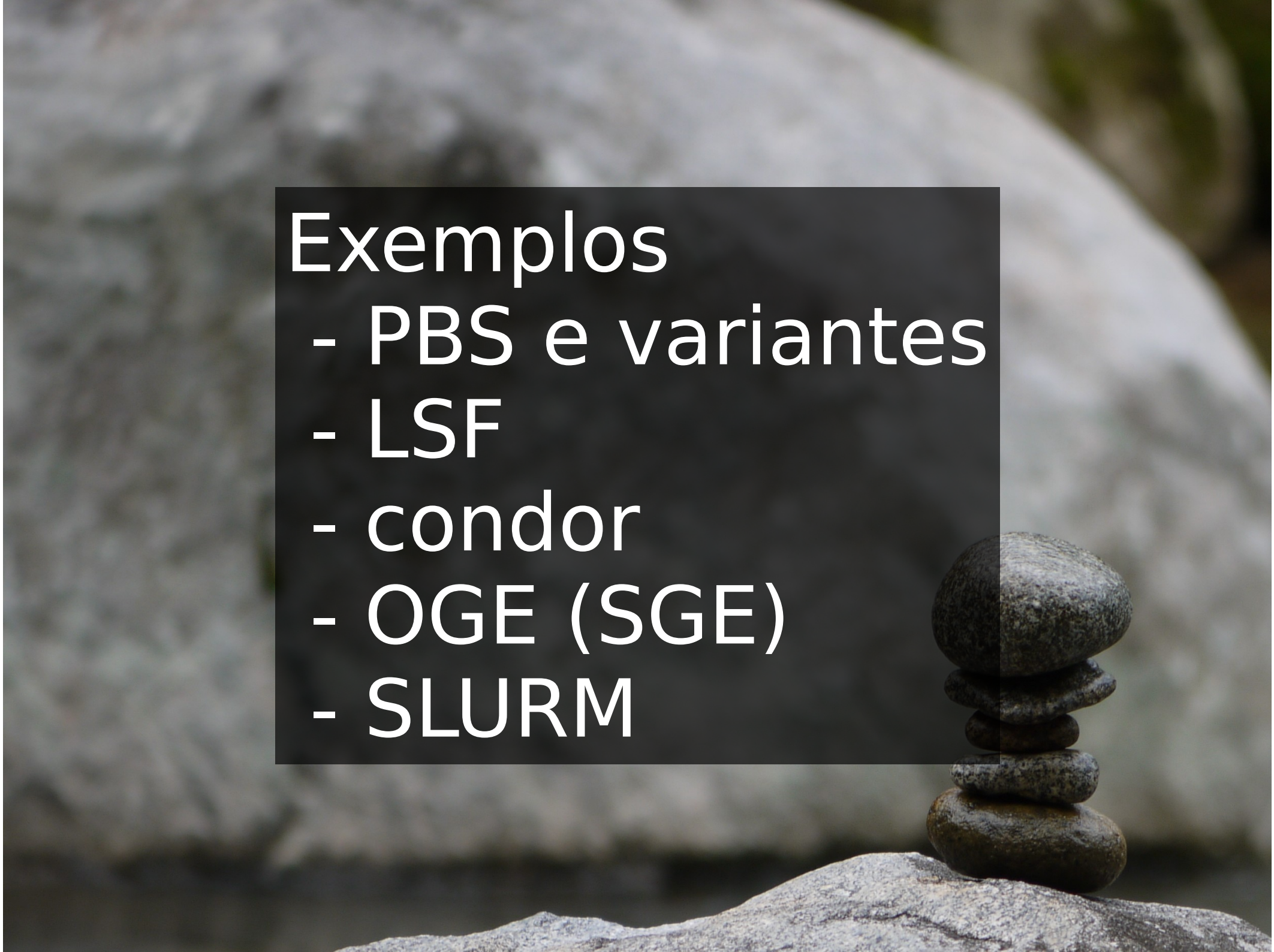

Importância
- Não sobrecarregar
- Evitar subutilização
- Priorizar entre usuários e recursos

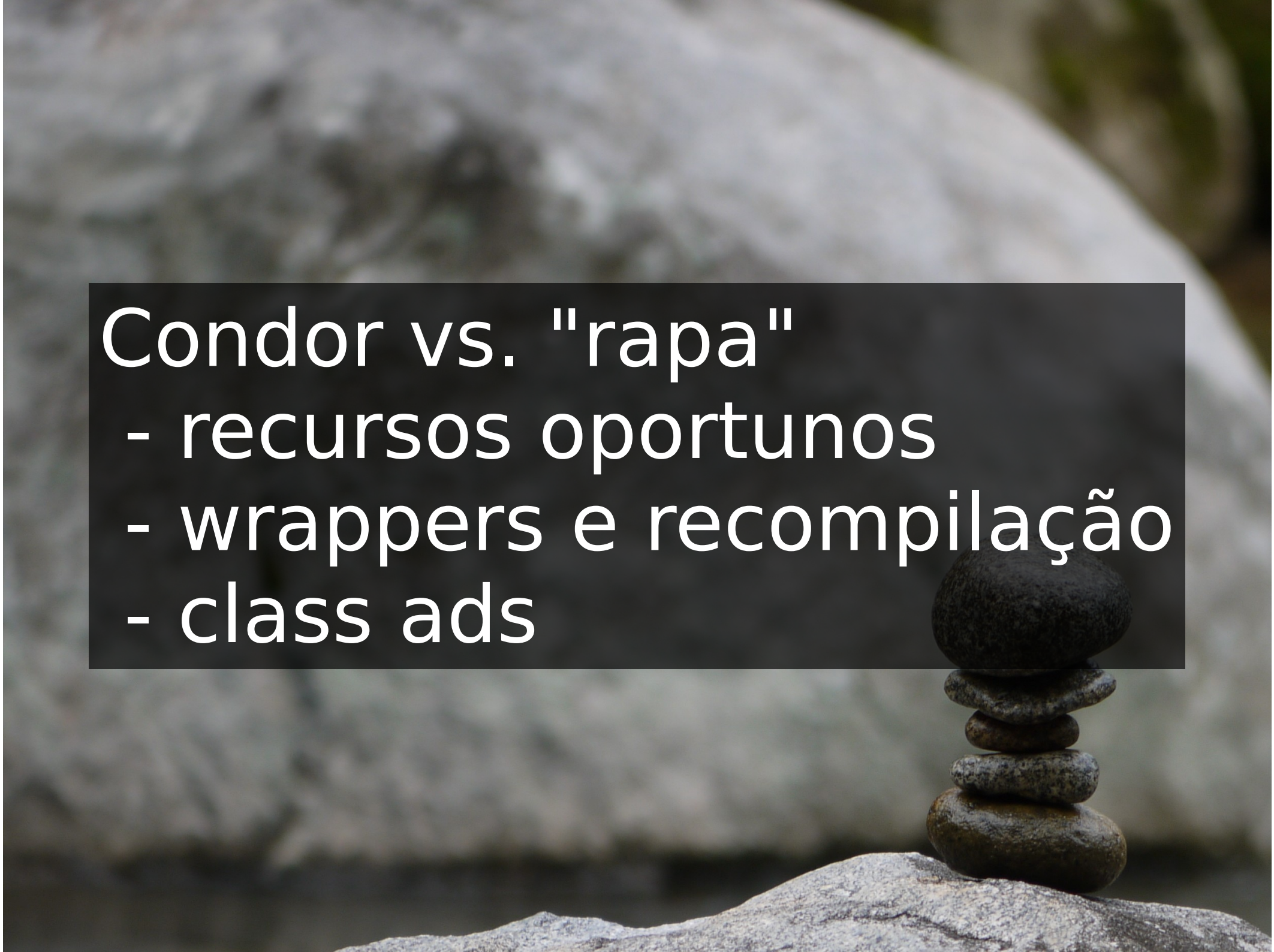

(esboçar o funcionamento de um)
(e levantar os problemas)


Principais problemas
- Fragmentação
- Adiamento indefinido


Requisitos do problema
- Tempo máximo de execução
- Preempção


Conceitos
- Preempção
- Filas
- Fair Share
- Reservas
- Eficiência
- Qualidade do serviço (QoS)

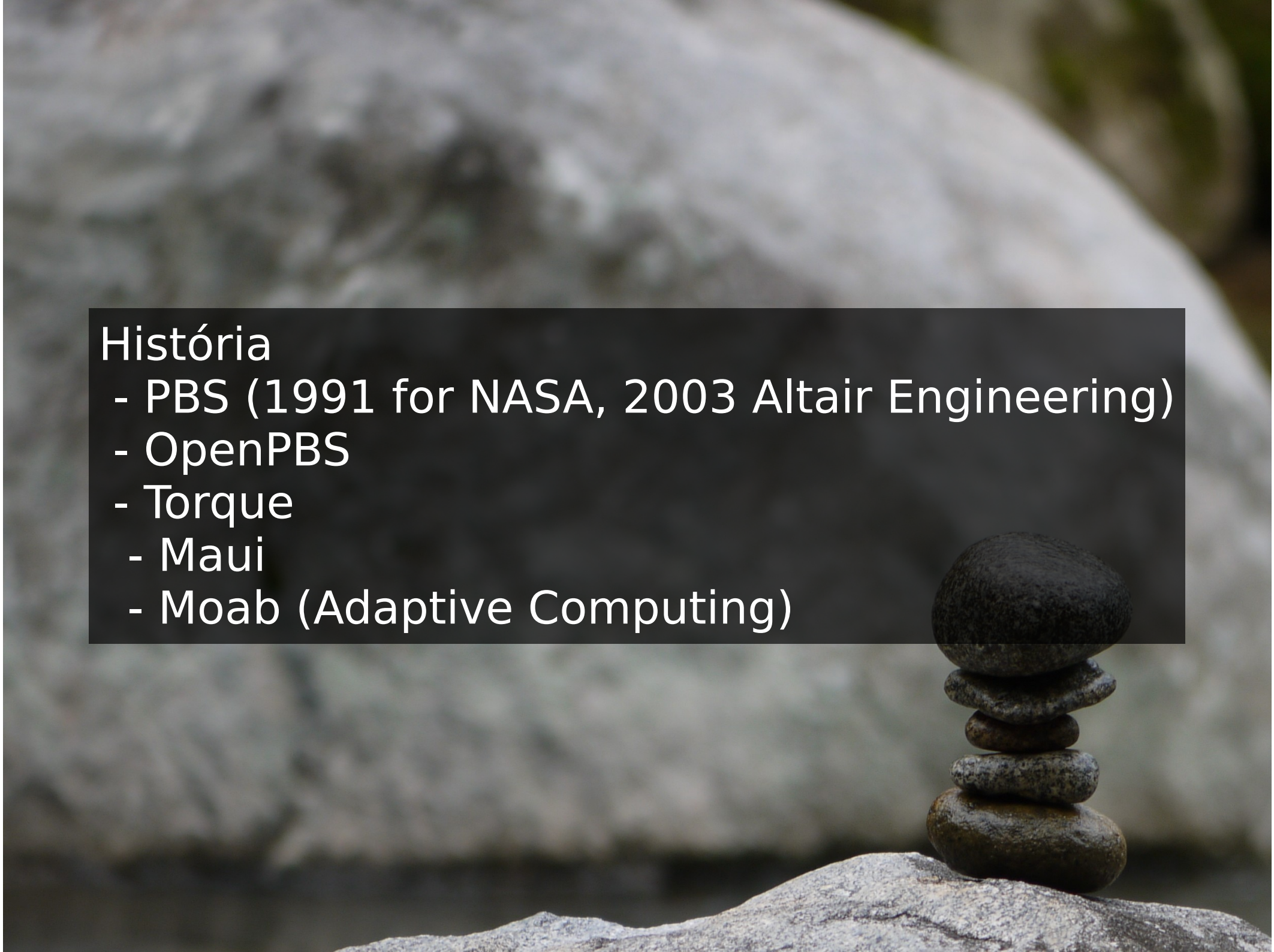

Exemplos
- PBS e variantes
- LSF
- condor
- OGE (SGE)
- SLURM

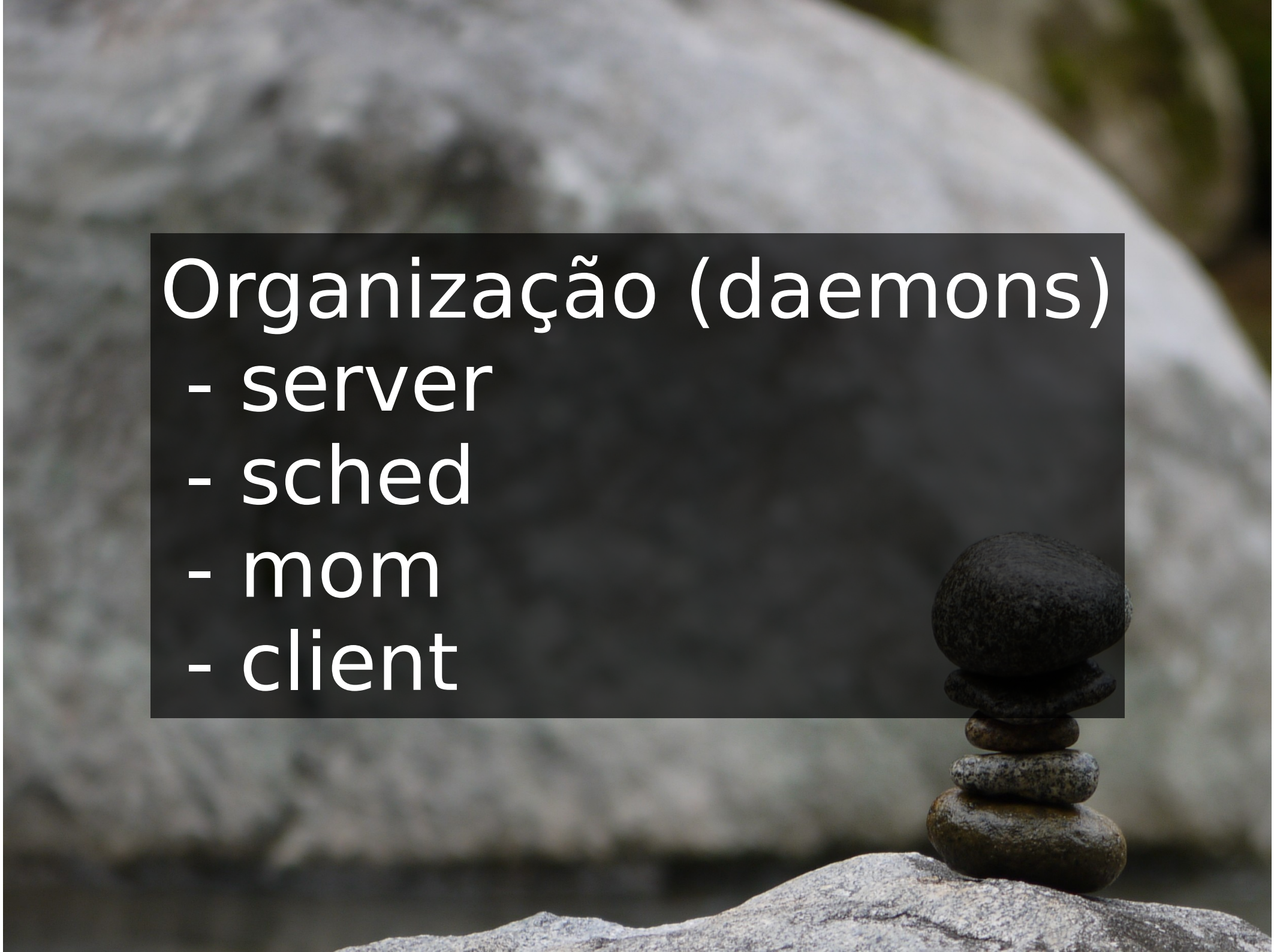

Condor vs. "rapa"
- recursos oportunos
- wrappers e recompilação
- class ads

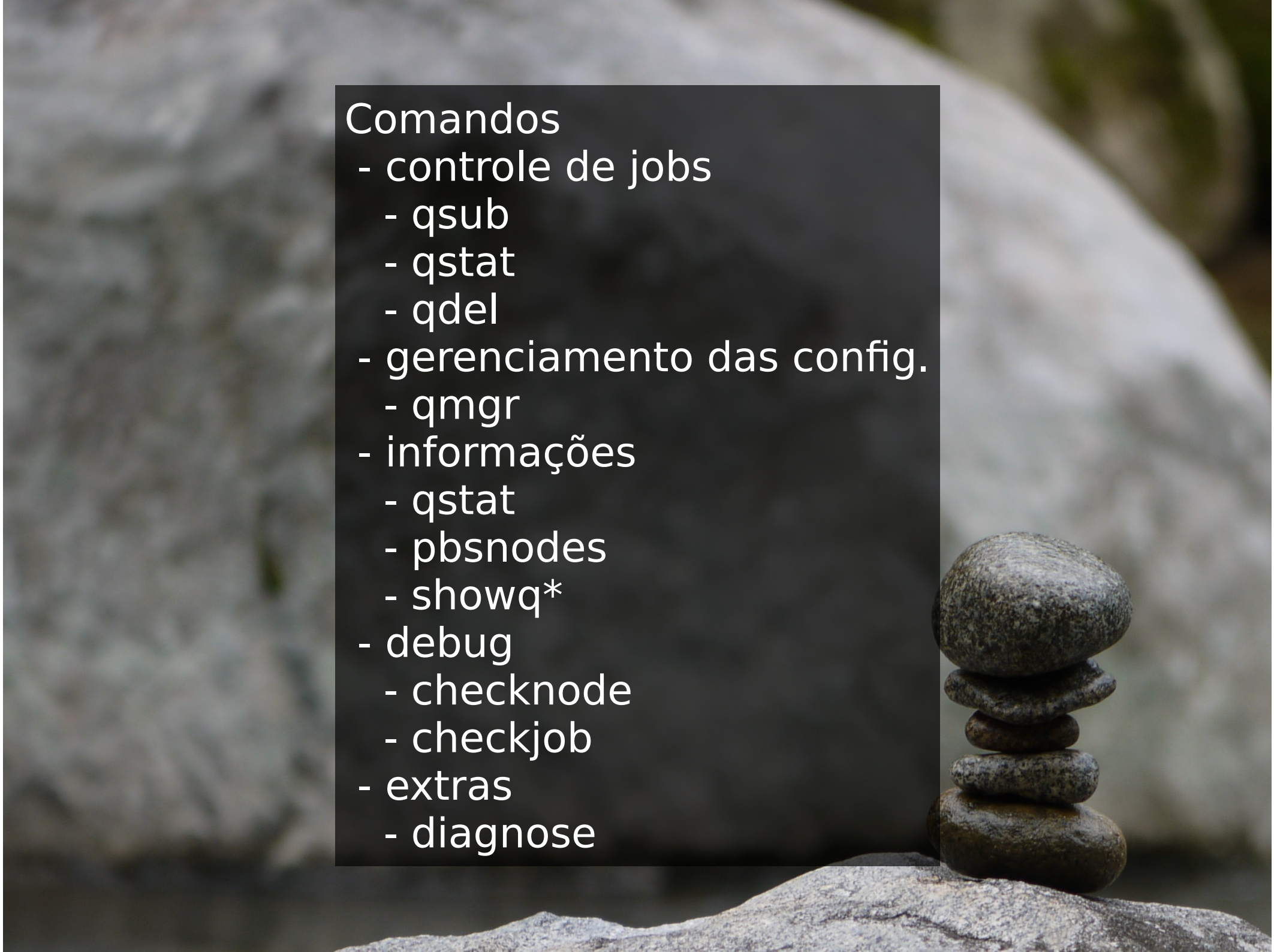

Possíveis casos


Torque (e Maui)


Terascale Open-Source Resource
and QUEue Manager

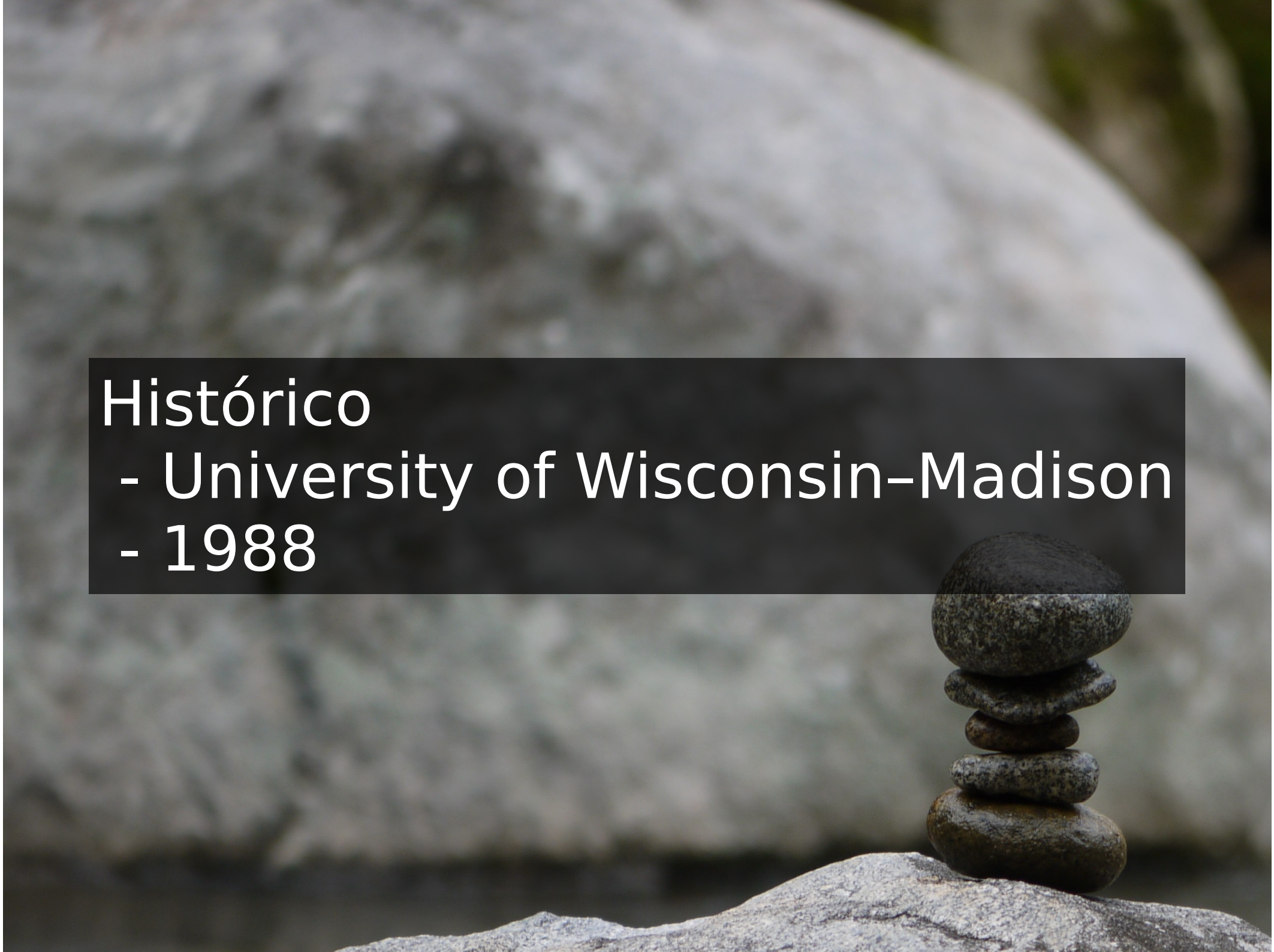

História
- PBS (1991 for NASA, 2003 Altair Engineering)
- OpenPBS
- Torque
- Maui
- Moab (Adaptive Computing)

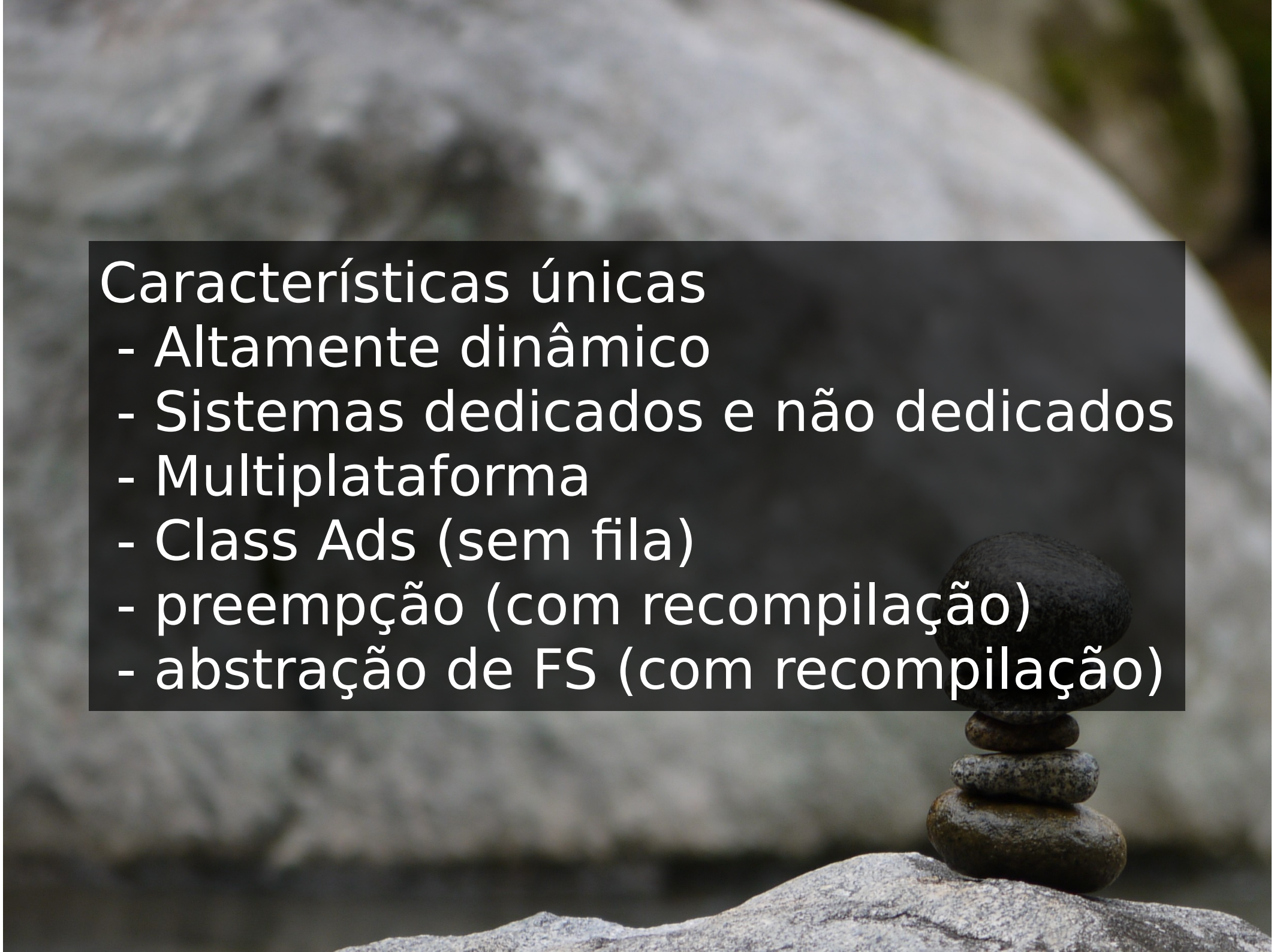

Organização (daemons)
- server
- sched
- mom
- client


Comandos
- controle de jobs
- qsub
- qstat
- qdel
- gerenciamento das config.
- qmgr
- informações
- qstat
- pbsnodes
- showq*
- debug
- checknode
- checkjob
- extras
- diagnose

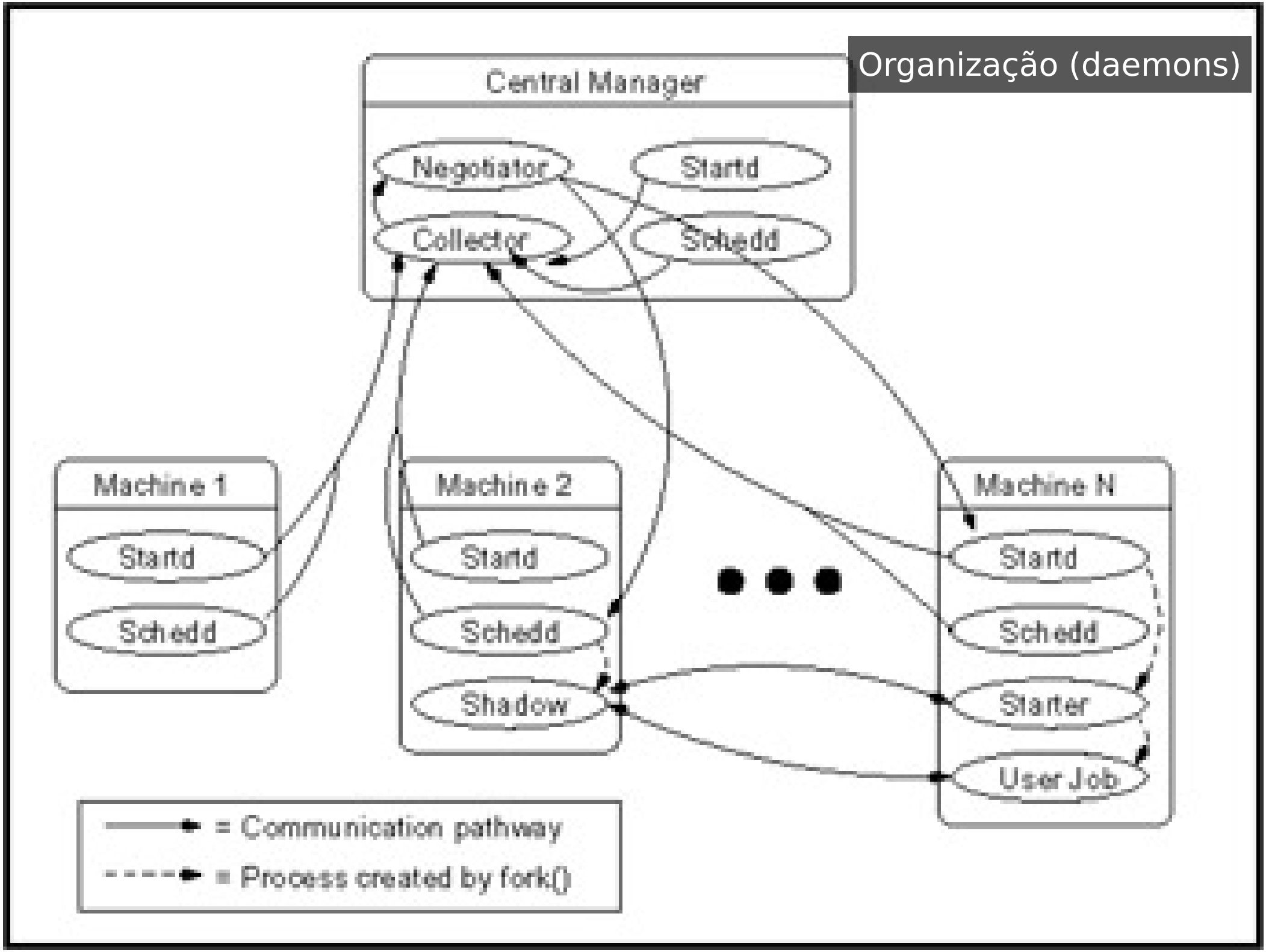

Pontos fracos
- semi estático
- "lento"
- API ruim

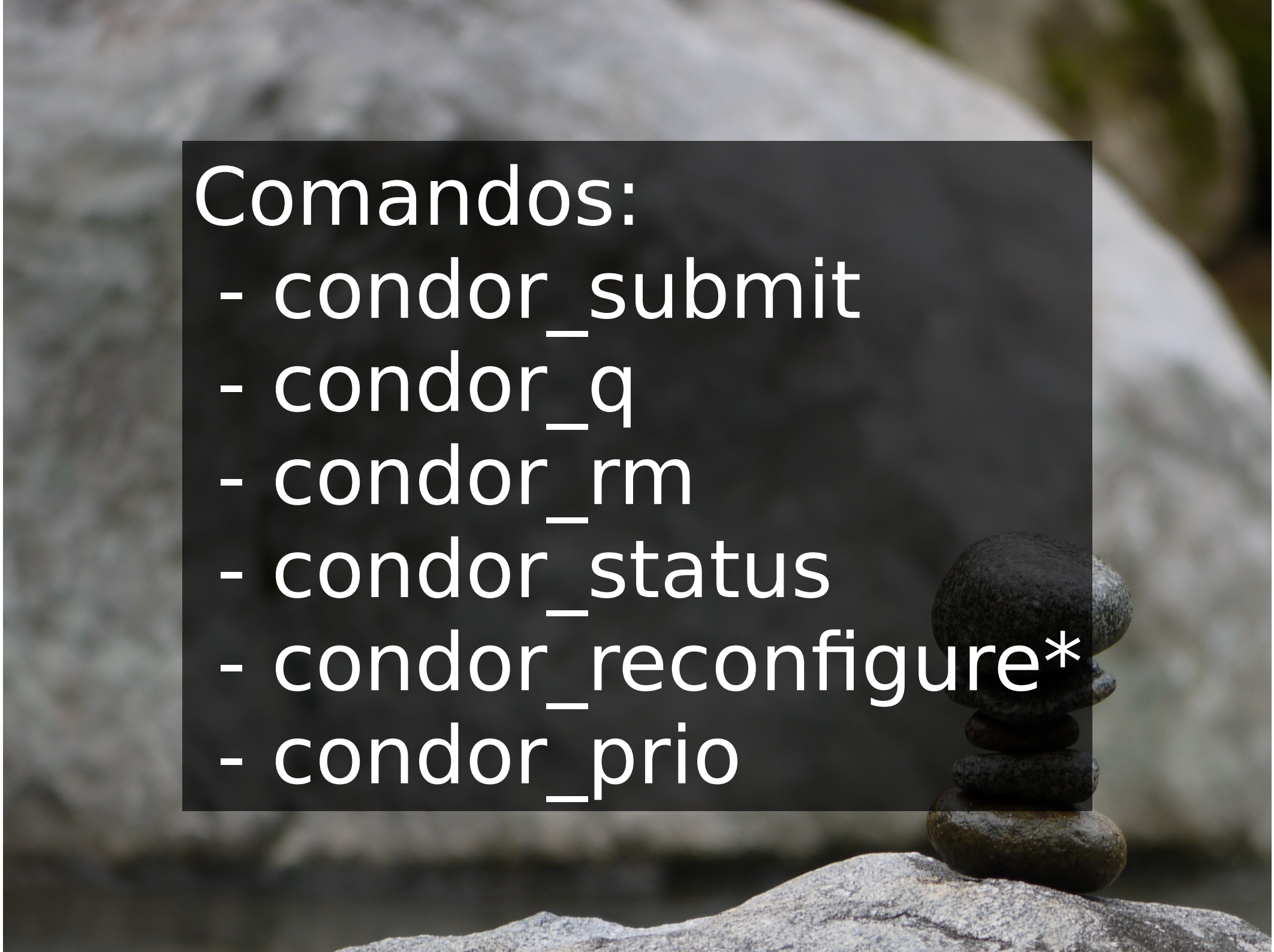

Condor (HTCondor)

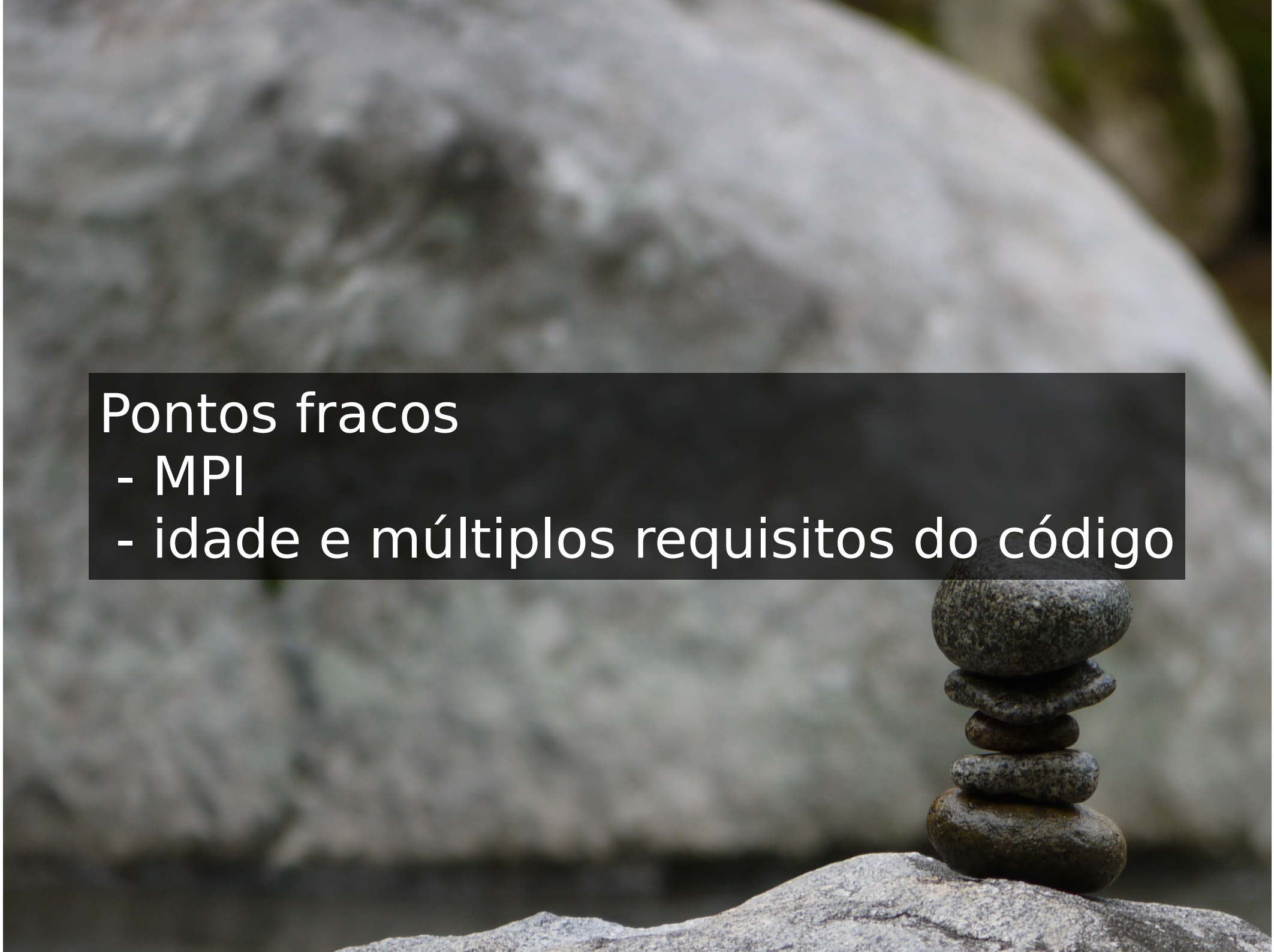

Histórico
- University of Wisconsin–Madison
- 1988

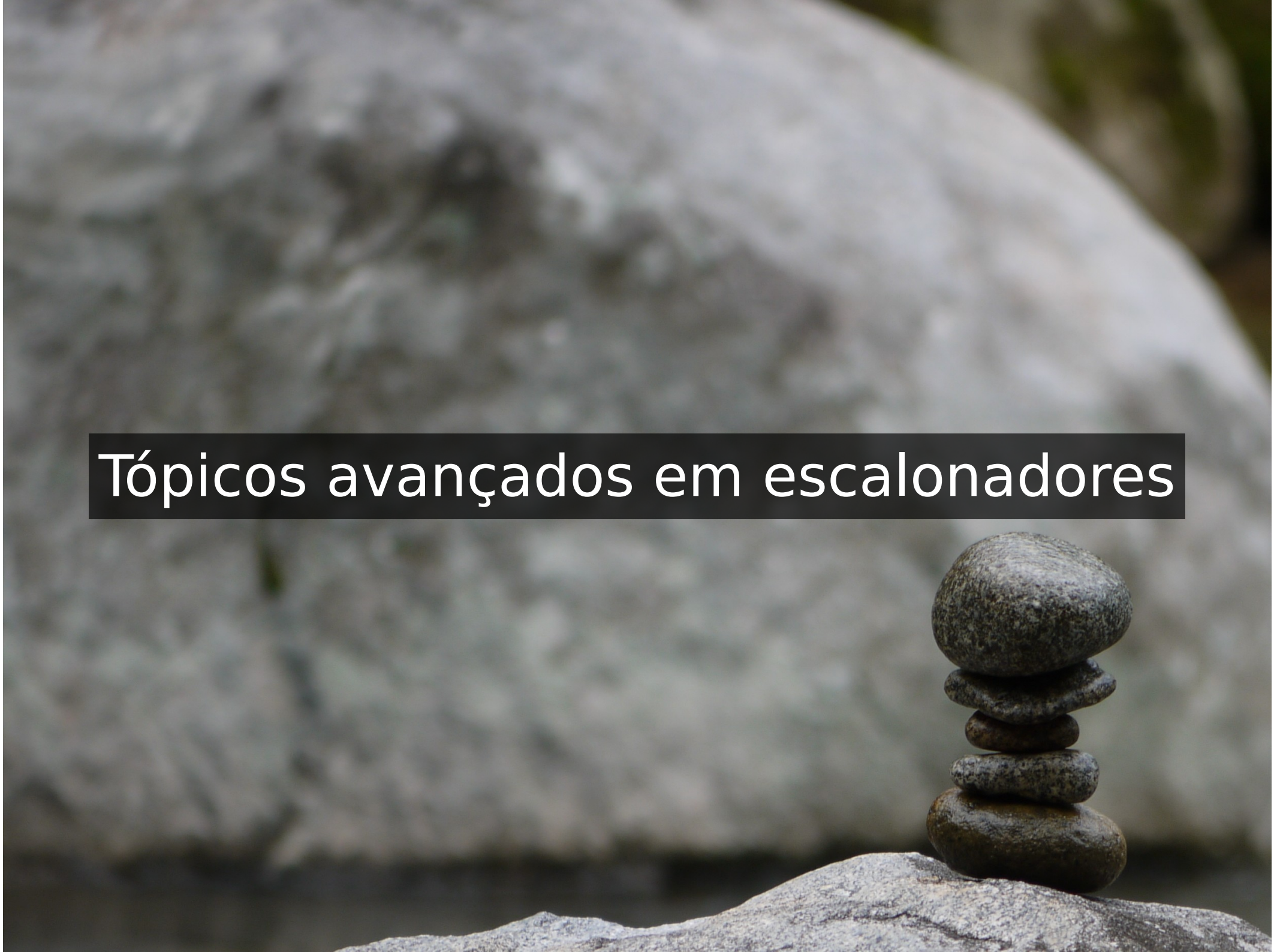

Características únicas
- Altamente dinâmico
- Sistemas dedicados e não dedicados
- Multiplataforma
- Class Ads (sem fila)
- preempção (com recompilação)
- abstração de FS (com recompilação)

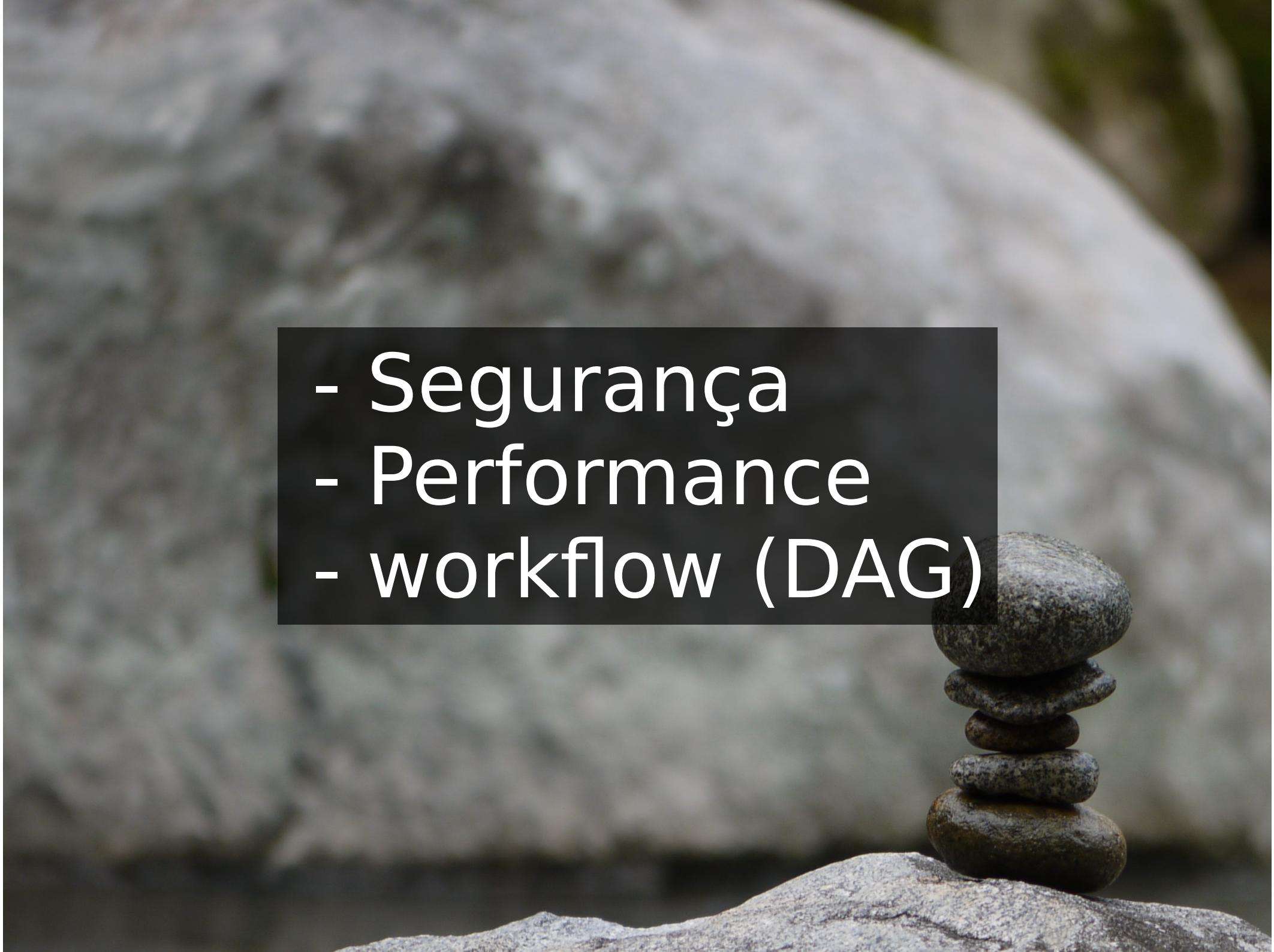
# Organização (daemons)




Comandos:
- condor_submit
- condor_q
- condor_rm
- condor_status
- condor_reconfigure*
- condor_prio


Pontos fracos
- MPI
- idade e múltiplos requisitos do código


Tópicos avançados em escalonadores


- Segurança
- Performance
- workflow (DAG)