



# Ambientes de programação distribuída

Eng. Software / PSPD

Prof. Fernando W Cruz



# Sumário

- Introdução/contextualização
- MapReduce/Hadoop



# Histórico



# Supercomputadores – anos 60



[http://upload.wikimedia.org/wikipedia/commons/c/c4/CDC\\_6600.jc.jpg](http://upload.wikimedia.org/wikipedia/commons/c/c4/CDC_6600.jc.jpg)



# Clusters – década de 1970



<http://upload.wikimedia.org/wikipedia/commons/3/3d/Us-nasa-columbia.jpg>

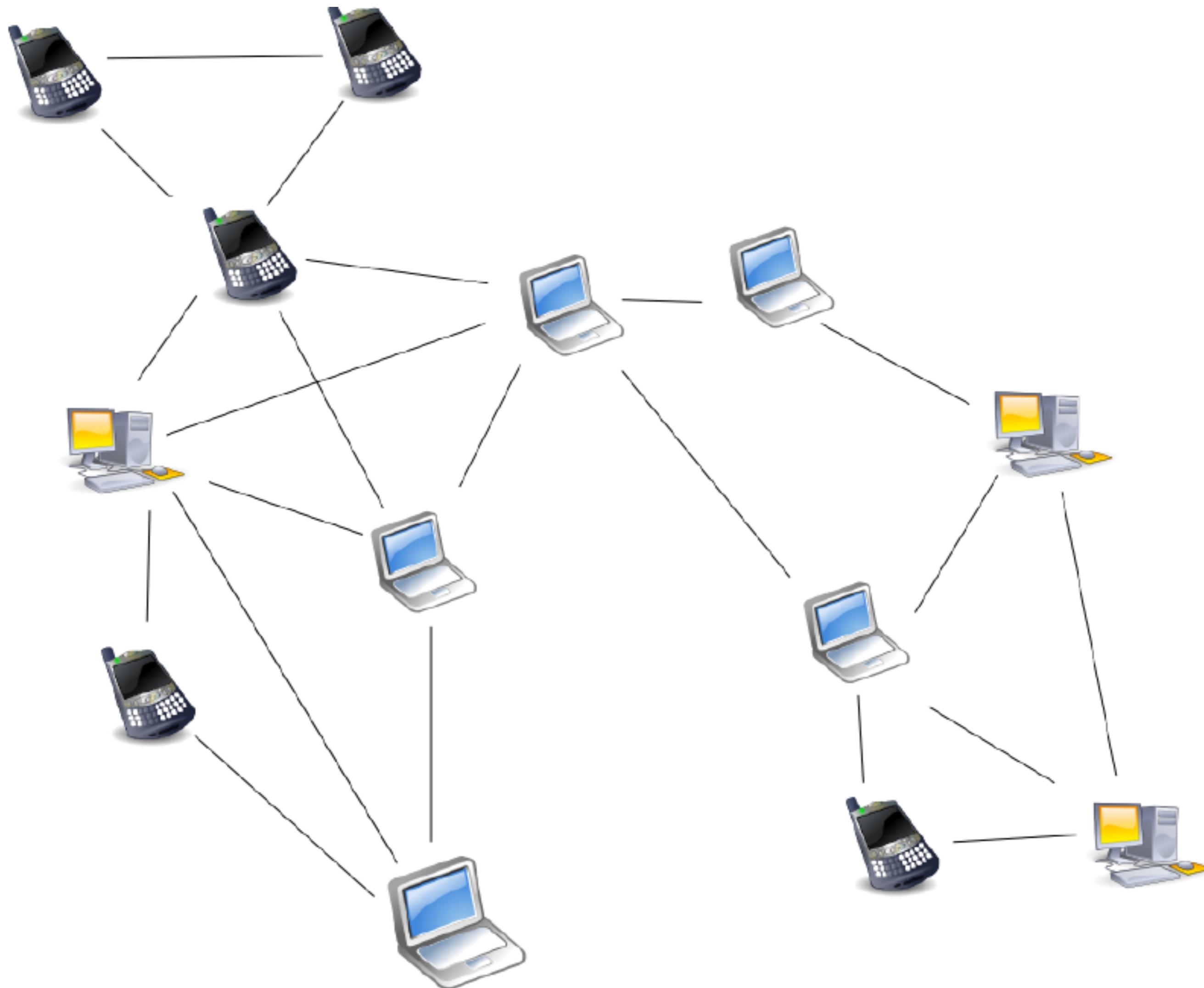


# Grids – desde a década de 1990





# Redes P2P – final dos anos 1990



[http://upload.wikimedia.org/wikipedia/en/f/fa/Unstructured\\_peer-to-peer\\_network\\_diagram.png](http://upload.wikimedia.org/wikipedia/en/f/fa/Unstructured_peer-to-peer_network_diagram.png)



# Cloud – anos 2000

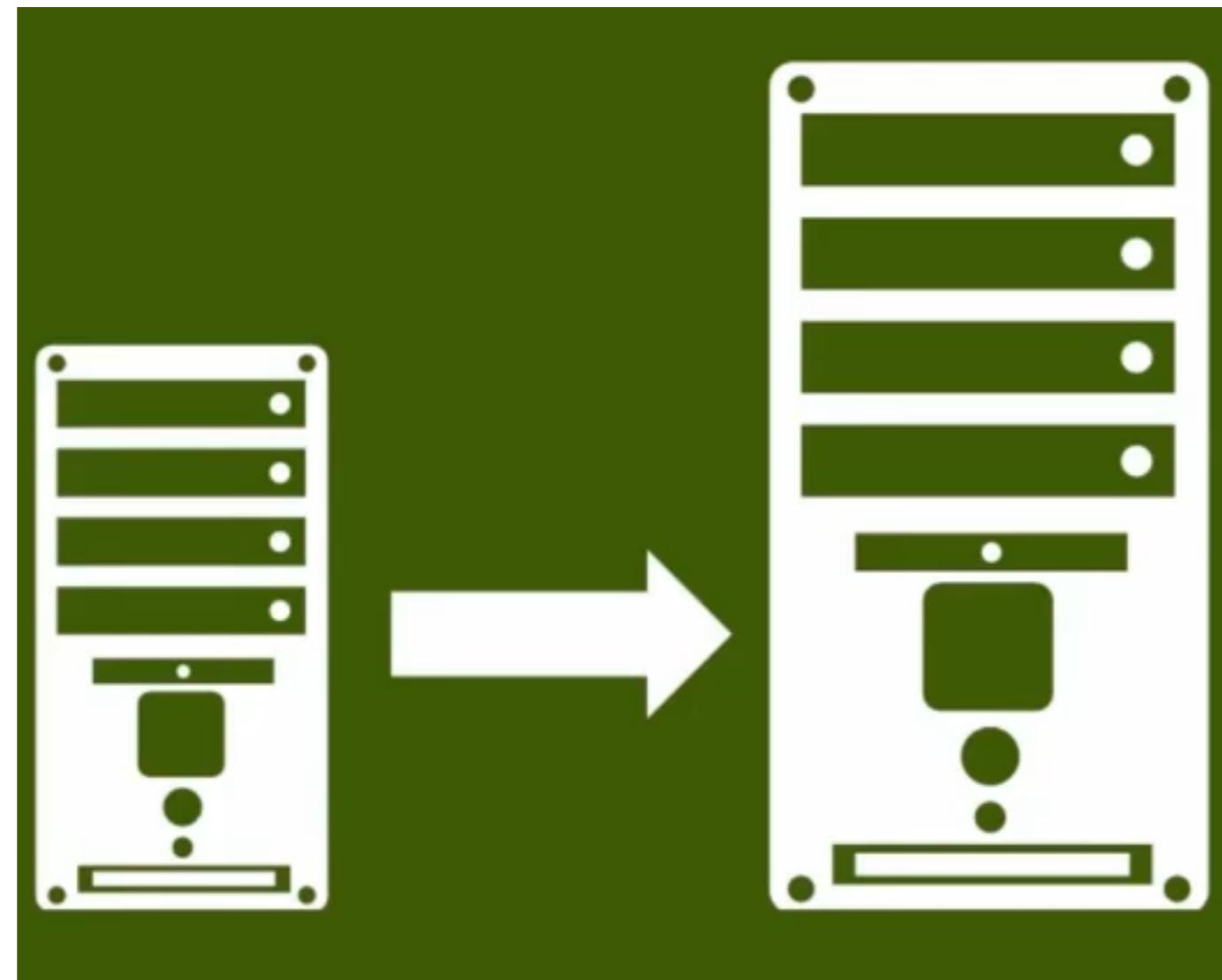
- Várias definições na literatura, mas há alguns acordos
  - Tratamento massivo de dados
  - Exploração do poder de fogo das CPUs
  - **Explora escalabilidade**





# Estratégias de escalabilidade

Escalabilidade vertical (scale up)

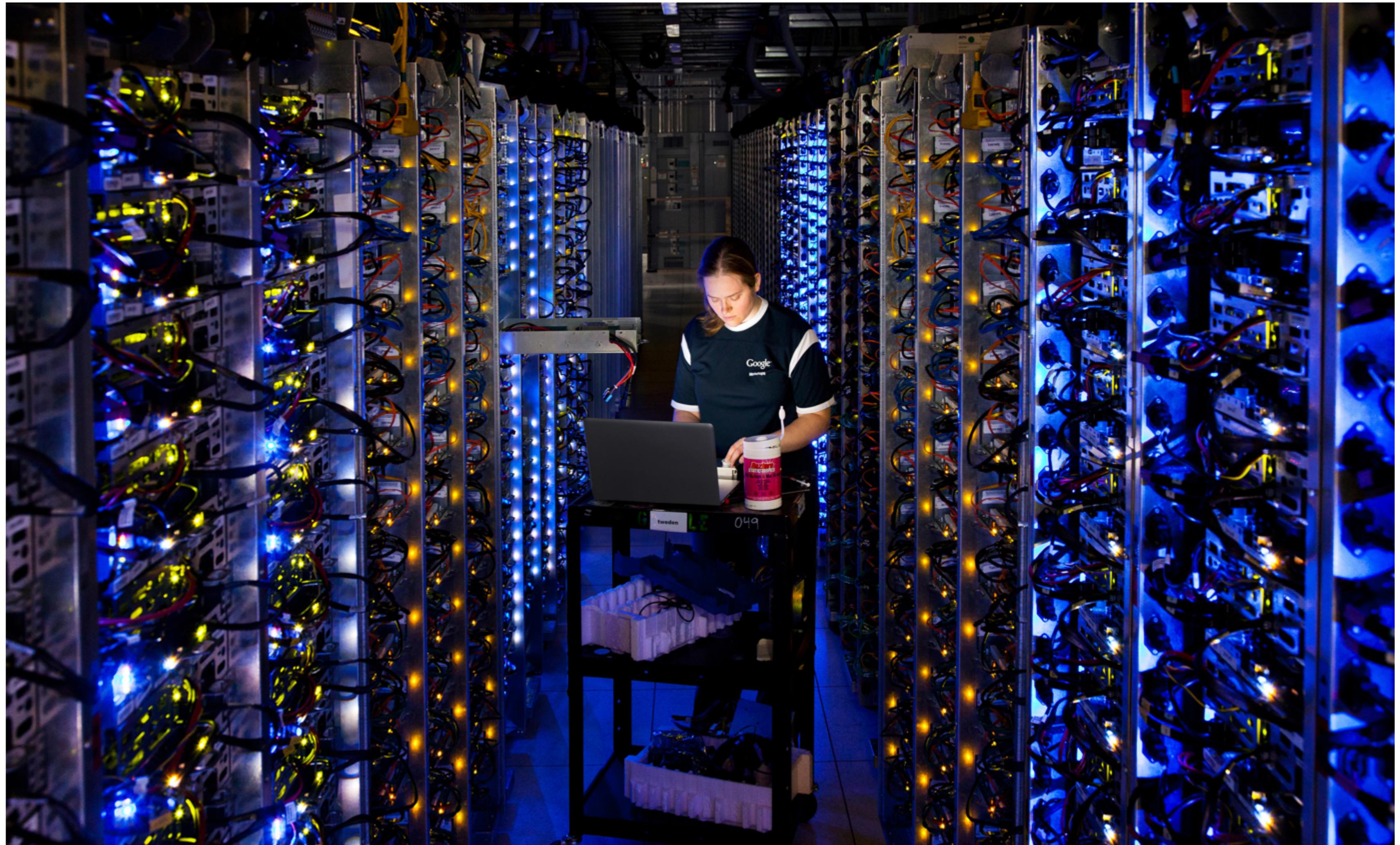


Escalabilidade horizontal (scale out)



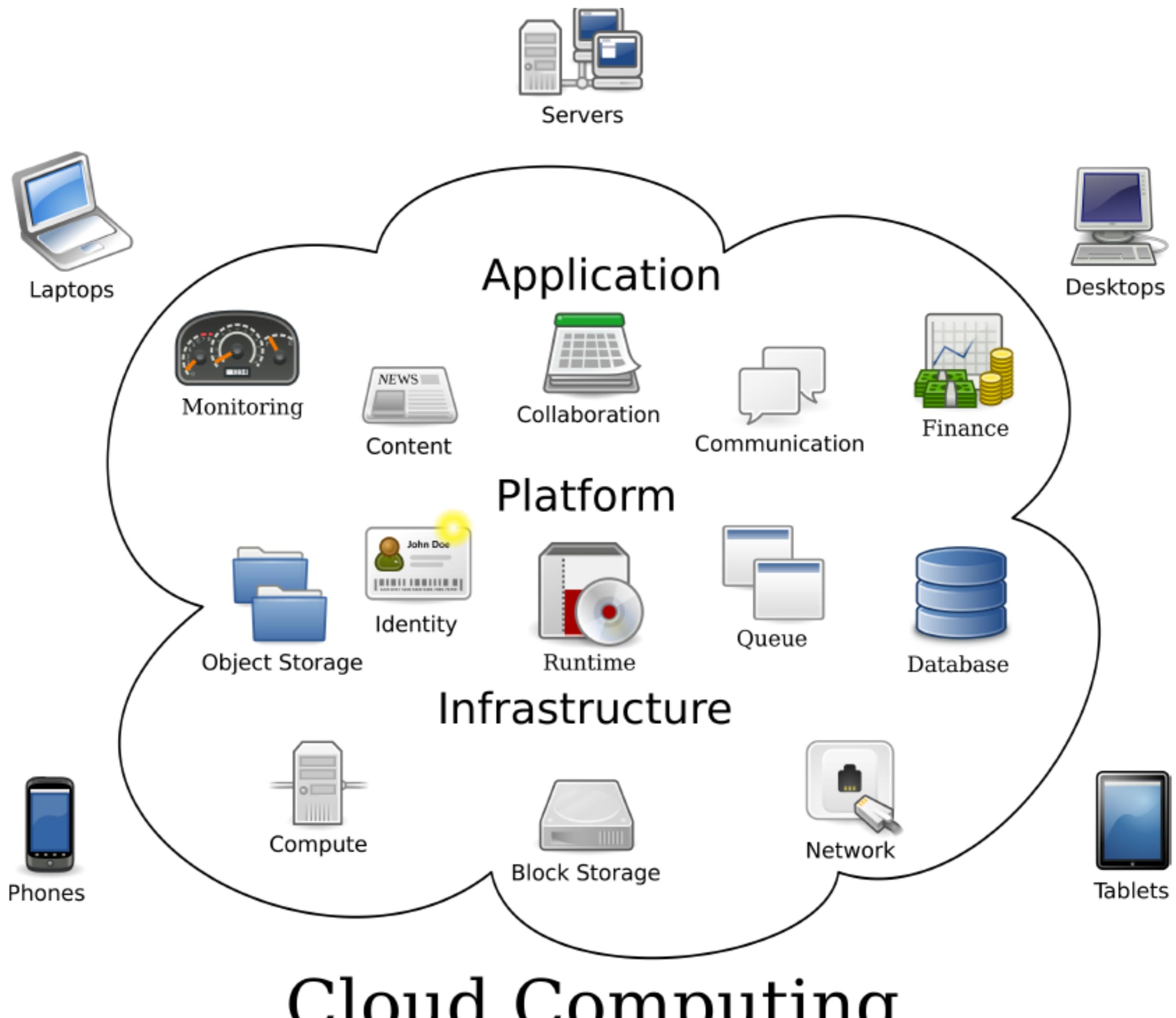


# Sala de racks/servidores (cluster)



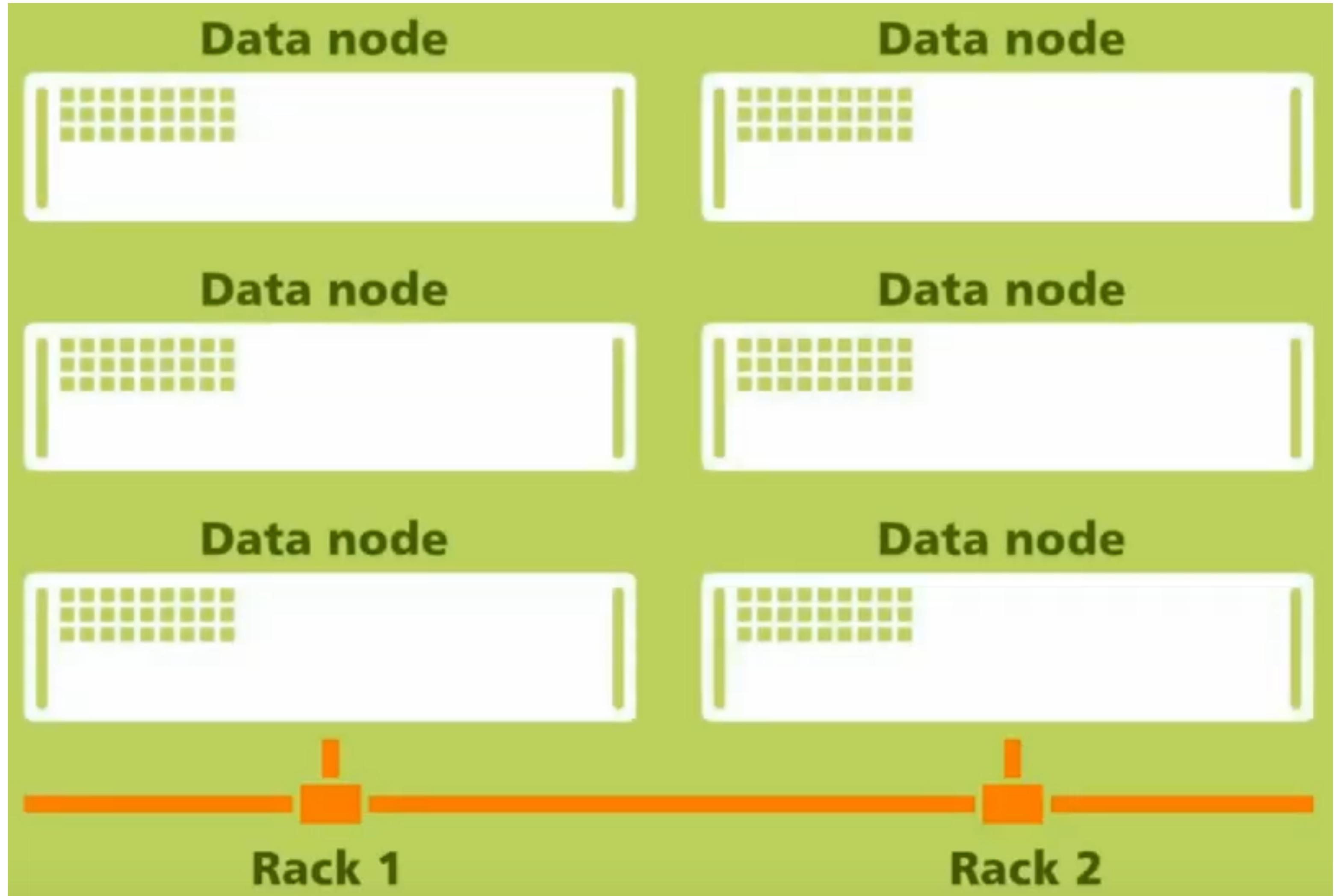


# Tipos de nuvens (propósitos)





# Nós de um Cluster





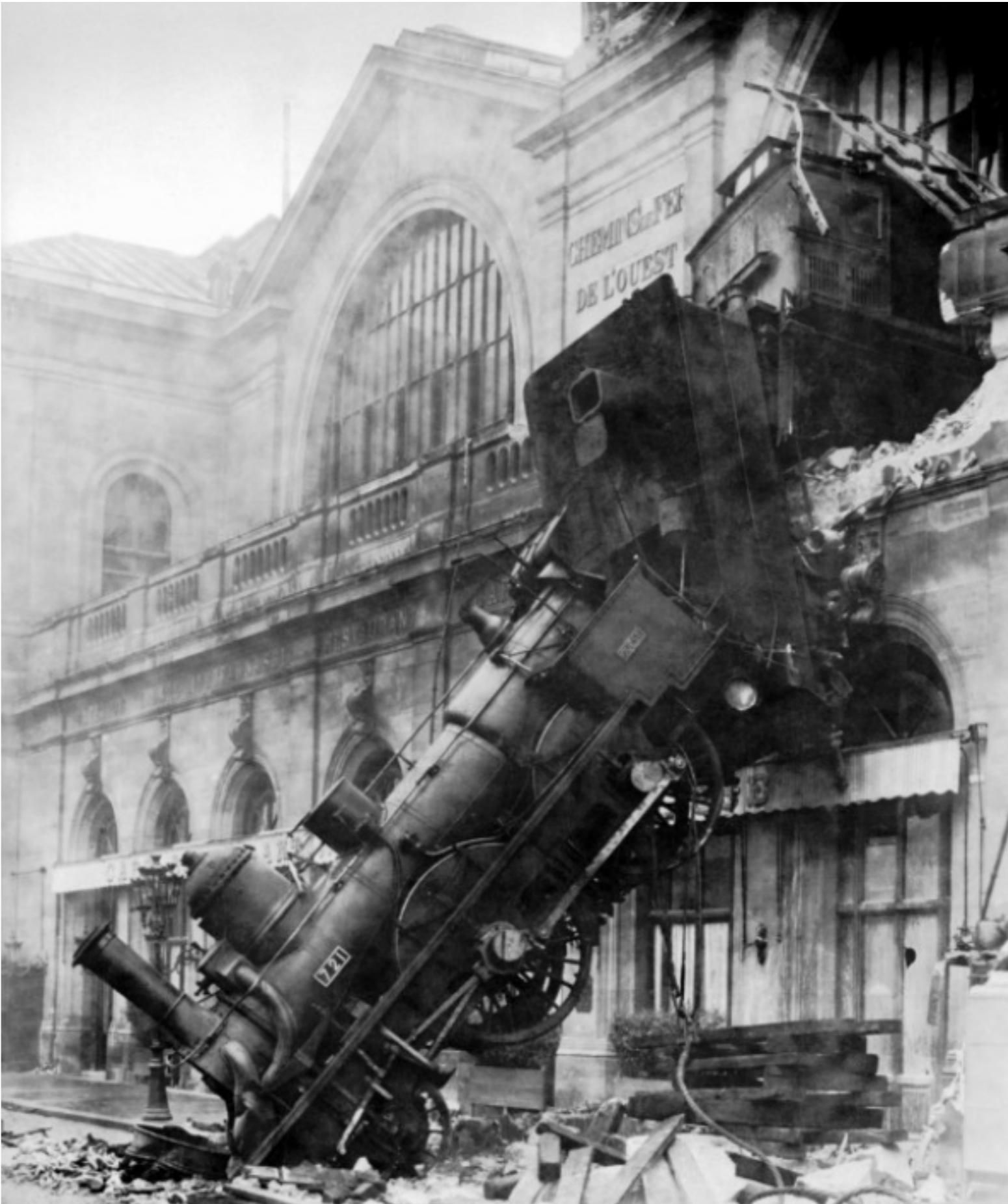
# Alguns dados sobre servidores

- MTBF (Tempo médio entre falhas)
  - Servidores falham uma vez a cada 6 anos, em média
  - Discos de servidores falham uma vez a cada 4 anos, em média
- Supondo um cluster de 10K servidores, cada um com 2 discos...
  - A cada dia, 5 servidores quebram!
  - A cada dia, 14 discos falham!



# Em clusters de larga escala...

- Erros acontecem o tempo todo e em qualquer lugar
- Por causa da escala massiva
  - Muitas formas de comunicação entre os nós
  - Redes não-confiáveis
  - Muito volume de dados
  - Etc.
- Esses problemas são geralmente endereçados por programação distribuída/paralela





PSPD

Prof. Fernando W Cruz