



Engenharia de Computação



Especialização Lato Sensu em Ciência de Dados e Analytics

Soluções em Processamento para Big Data

{ Clusters e mais... }

Prof. Jairson Rodrigues
jairson.rodrigues@univasf.edu.br

{ contexto }

- Áreas de mercado e pesquisa complexas
 - Aeronáutica, previsão do tempo, pesquisa farmacêutica e biológica, modelagem populacional, sistemas complexos adaptativos
- Aplicações cada vez mais exigentes
- Necessidade crescente de processamento paralelo massivo
- Aumento da capacidade dos computadores comuns
- Surgimento dos Clusters
- Surgimento dos Grids

{ cluster }

AGENDA

- HPC
- Cluster
- GRID
- Nuvem



{ cluster }

- Nós (computadores mono ou multiprocessados) dentro de uma LAN de alta velocidade, que funcionam como um único computador paralelo
- Computação em Cluster permite que vários computadores trabalhem juntos para resolver problemas complexos de grande porte

{ cluster }

- Componentes relativamente baratos
- Reduz custo de construção comparado com único computador paralelo de mesma capacidade
- Alto Desempenho e/ou Disponibilidade e/ou Processamento
- Maior escalabilidade que o multiprocessamento
 - Supercomputador HP 9000 Superdome: até 64 processadores
 - Cluster HP XC: até 512 processadores
 - Cluster Beowulf: até 700 processadores

{ cluster beowulf }

- Criado por Donald Becker, da NASA baseado em PC's
- Idealizado para suprir demanda por elevada capacidade de processamento (HPC) a custos economicamente viáveis
- Motivação
 - Aumento da performance computacional de PC's
 - Preço decrescente versus poder computacional crescente
 - Utiliza software gratuito de elevado desempenho: Linux/FreeBSD

{ cluster beowulf }

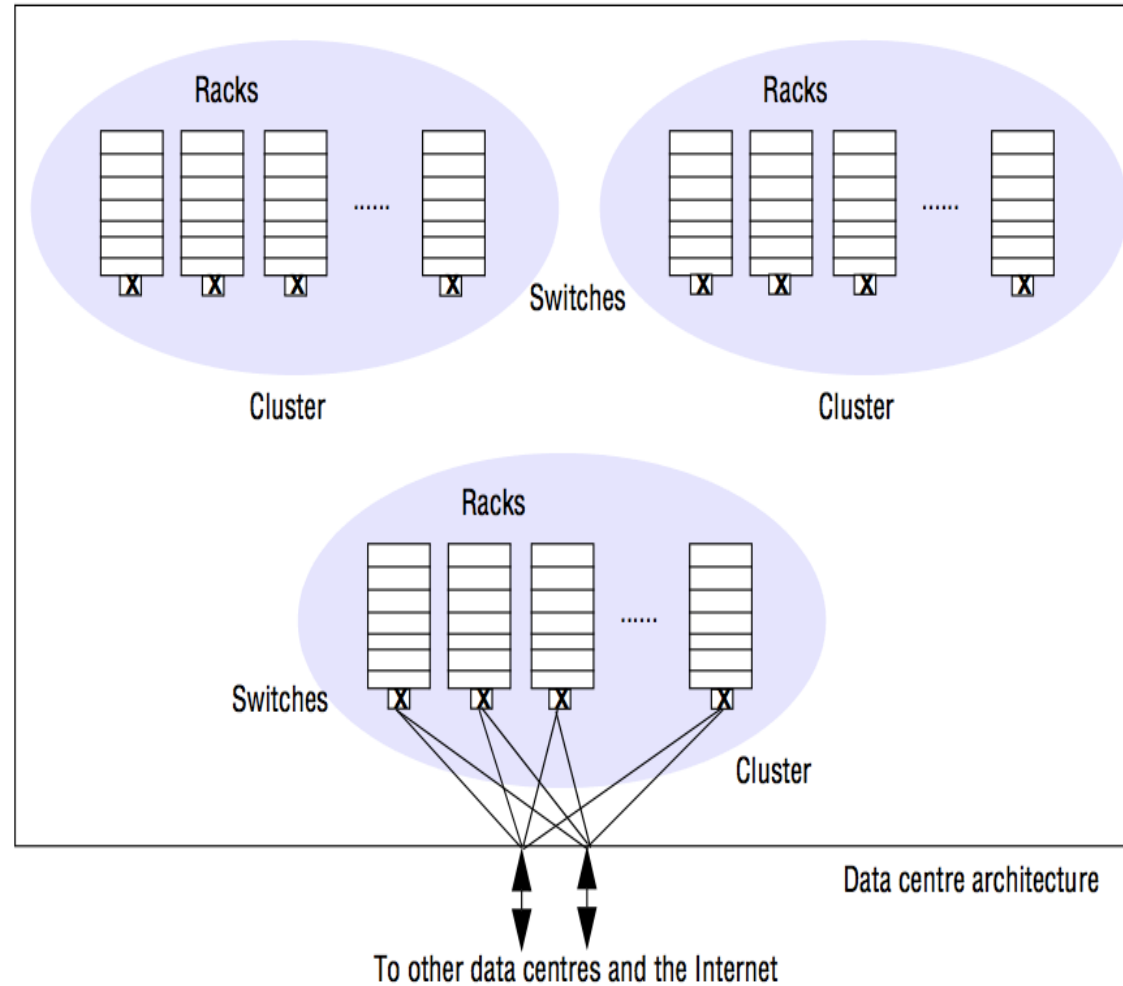


{ cluster da google (2008) ** }

- Estimando... (George Coulouris et al, Distributed Systems, Concepts and Design – Fifth Edition, dados de 2008)
 - 1 PC = 2 Terabytes
 - 1 rack de 80 PC's = 160 Terabytes
 - 1 cluster de 30 racks = 4.8 Petabytes
- Estima-se que a Google possuía 200 clusters
 - Capacidade de armazenamento próxima de 1 Exabyte = 1.099.511.627.776 Megabytes (1 trilhão de Megabytes)

** A Google não divulga detalhes de sua arquitetura, há muitas fontes na WEB que detalham a capacidade do cluster da Google, entretanto são apenas estimativas. O dado apresentado acima foi retirado de material acadêmico/universitário, entretanto encontra-se defasado. A capacidade atual (2017) é superior ao apresentado.

{ cluster da google }



{ grids }



{ grids }

- A Internet
 - Sistemas totalmente heterogêneos
 - Fracamente acoplados
 - Geograficamente distribuídos
 - Com alto índice de máquinas ligadas e ociosas
- Grid
 - Compartilhamento de ciclos de CPU
 - Quando um computador não precisasse de todo seu poder computacional outros pudesse fazer uso de da CPU ociosa

{ grids }

- Primeiros projetos da década de 1990
 - FAFNER (Factoring via Network-Enabled Recursion)
 - I-WAY (Information Wide Area Year)
- A computação em Grid teve seu marco com o workshop "*Building a Computational Grid*", no Argonne National Laboratory, EUA, 1997

{ SETI@Home }

- Iniciado em 1996
- Acrônimo para Search for Extraterrestrial Intelligence
- Utiliza radio telescópios para “escutar” sinais vindos de fora do planeta
- Utiliza computação em Grid para analisar os dados
- Atualmente foi expandido para o projeto BOINC
- 4,109,949 hosts em Jan/2017
<http://setiathome.berkeley.edu/stats.php>



{ o grande colisor de hádrons }

- Há atividades computacionais que precisam de poder computacional inimaginável
- O LHC produz cerca de 41 TB de dados por dia; até 15 PB de dados por ano
- Seria necessário um cluster com ~100.000 CPUs

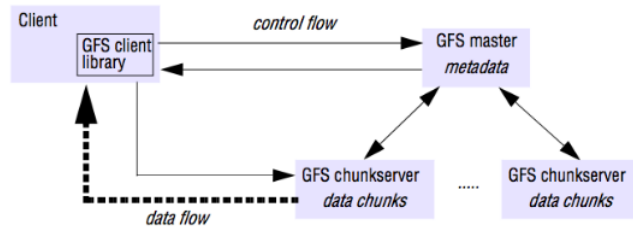


{ hpc / cluster / grid / cloud computing }

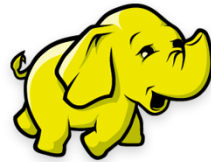
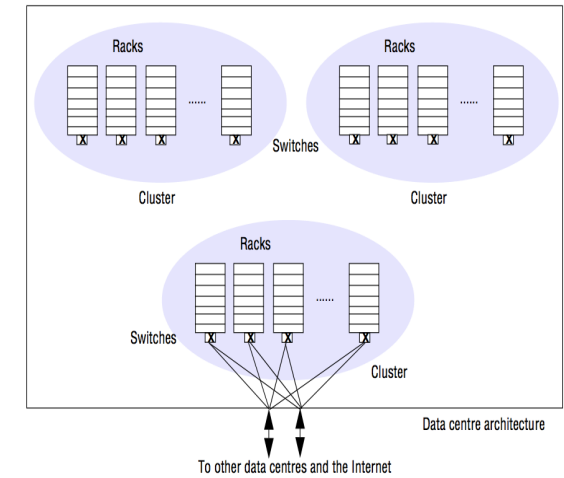


{ ... nesse contexto, anos 2000 }

Google File System



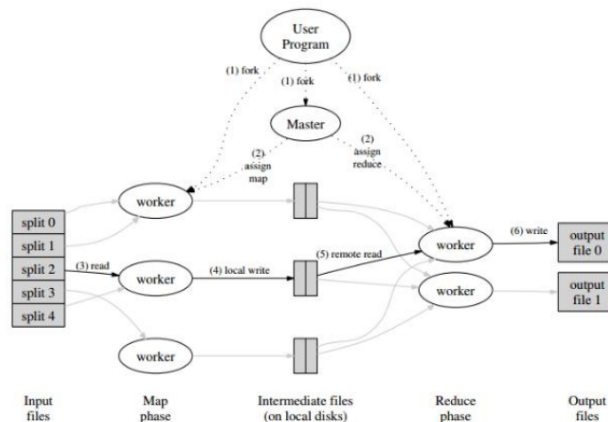
Cluster Google



Google
Brasil

YAHOO!

etc etc etc



MapReduce

HDFS



