

# 数据库系统概论新技术篇

## 数据库一体机技术分析

周烜

中国人民大学信息学院

2017年4月

# 内容概要

- ❖ 如何让数据库系统充分利用硬件？
  - 如何利用缓存？
- ❖ 什么样的硬件适合数据库系统？
  - 如何拓宽数据通道？



# 数据库是数据密集型软件

## ❖ 计算密集型软件：

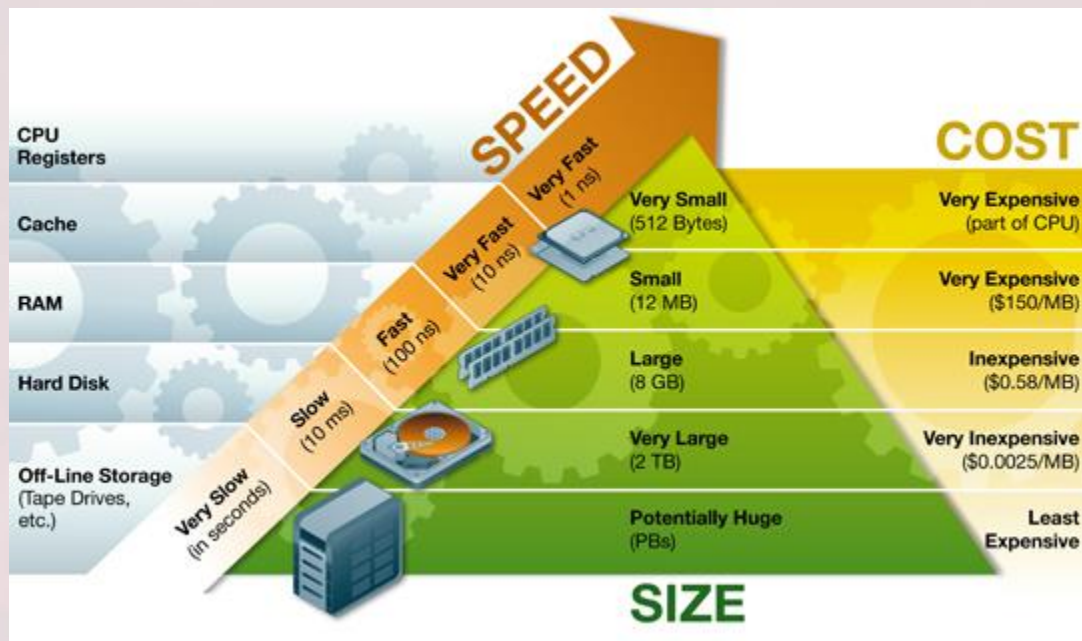
- 处理器是性能瓶颈。
- 提升性能需使用更快更多的处理器。

## ❖ 数据密集型软件：

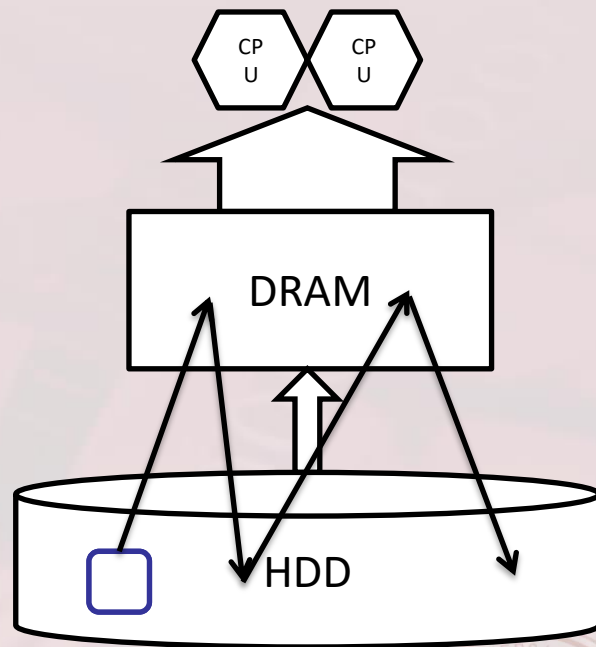
- 数据访问通道是性能瓶颈。
- 提升数据访问的速度和吞吐率。



# 如何提升数据访问效率 – 缓存

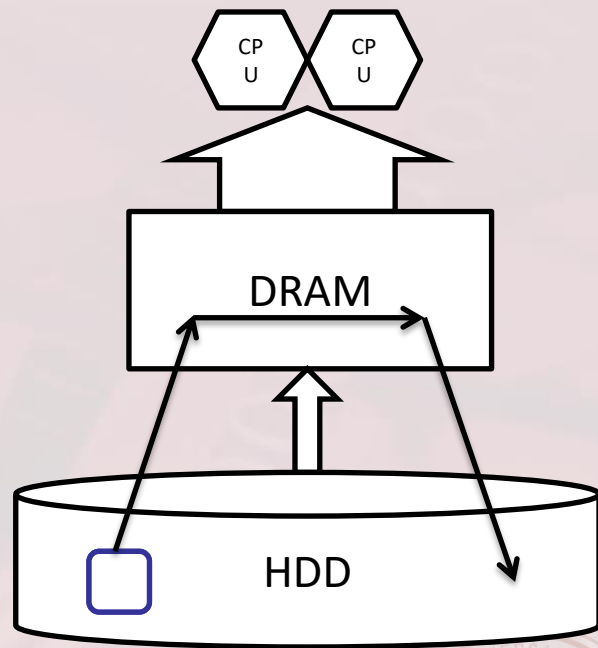
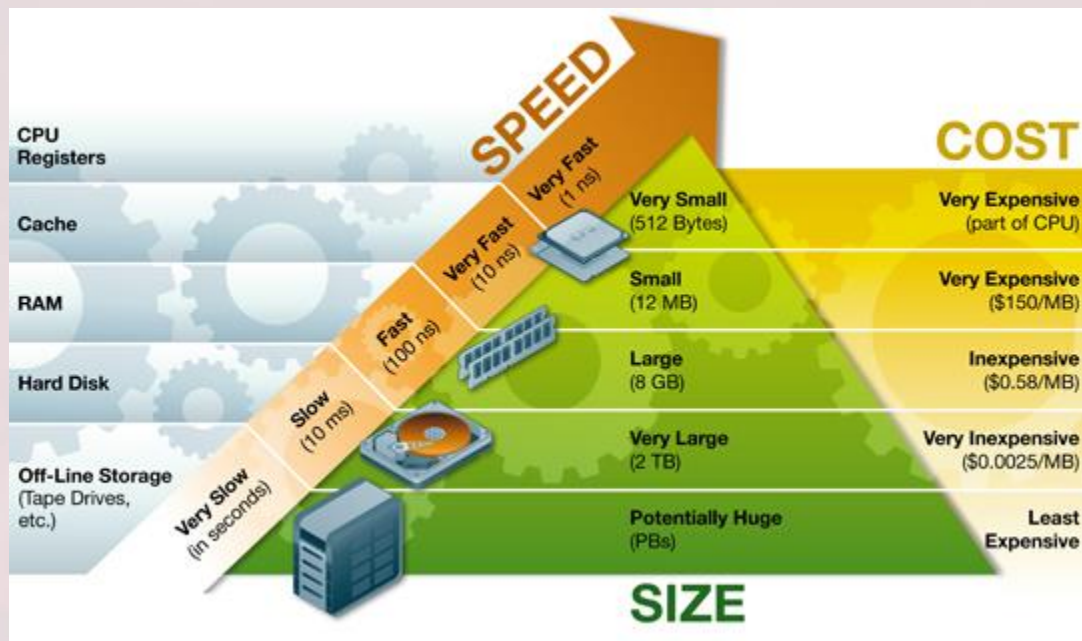


❖ 存储设备永远呈金字塔结构



数据每次访问被重新调入内存

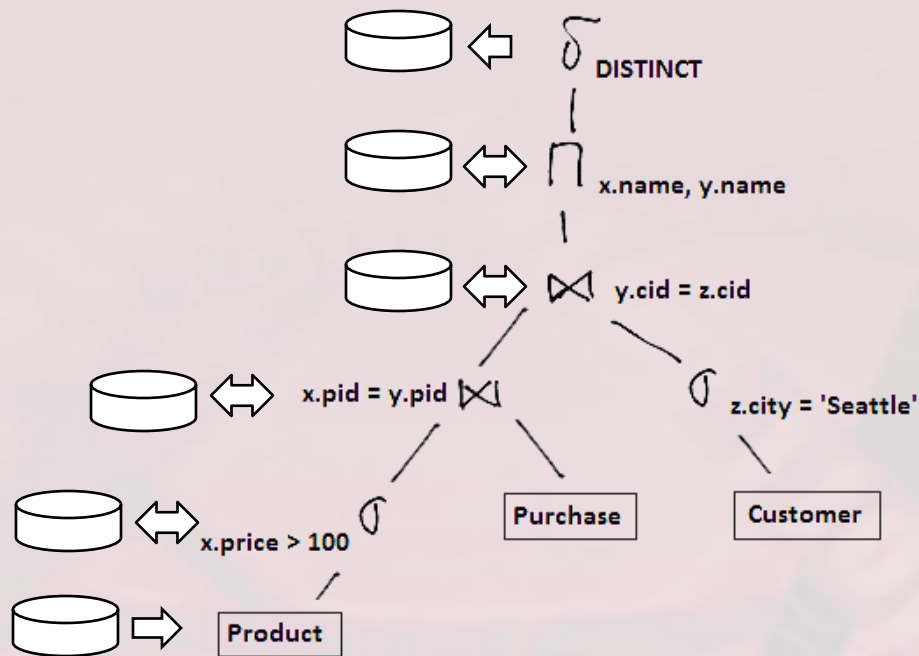
# 如何提升数据访问效率 – 缓存



❖ 存储设备永远呈金字塔结构

数据缓存在内存后  
被多次访问

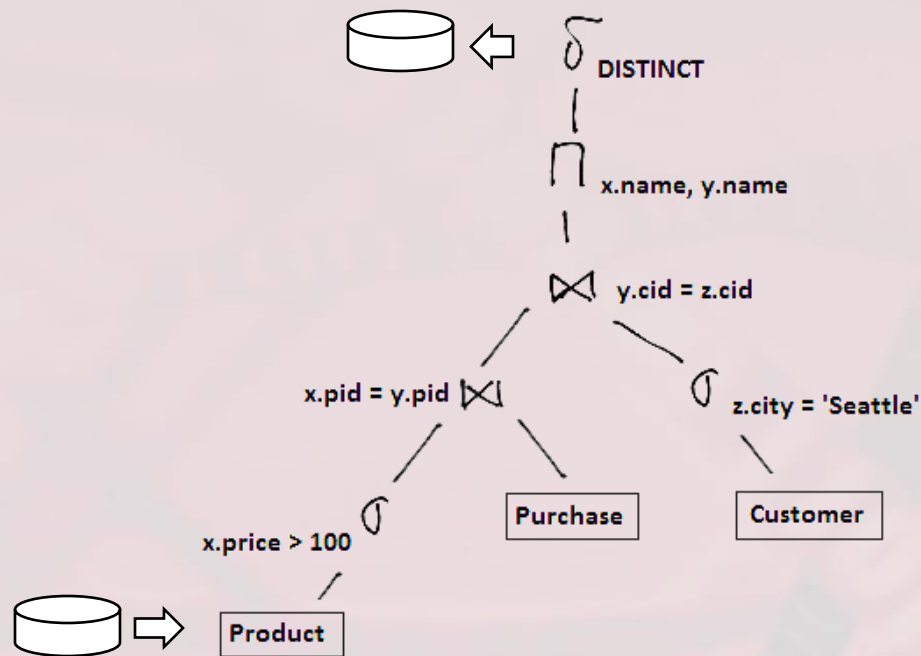
# 提升缓存的利用率 – 流水线



使用流水线之前：  
每次处理一个算子；  
每次都涉及I/O吞吐。



# 提升缓存的利用率 – 流水线

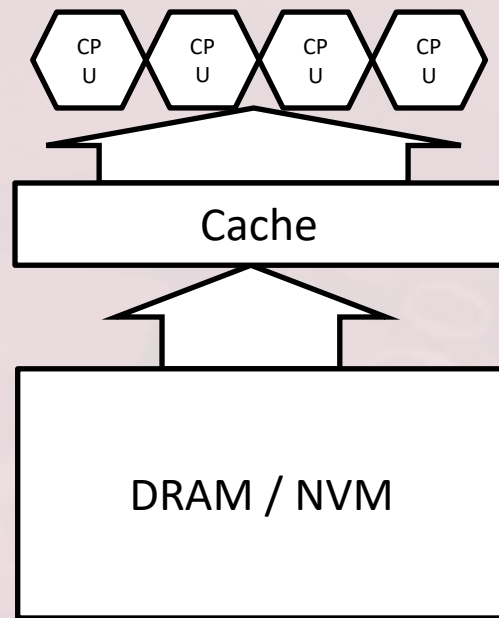
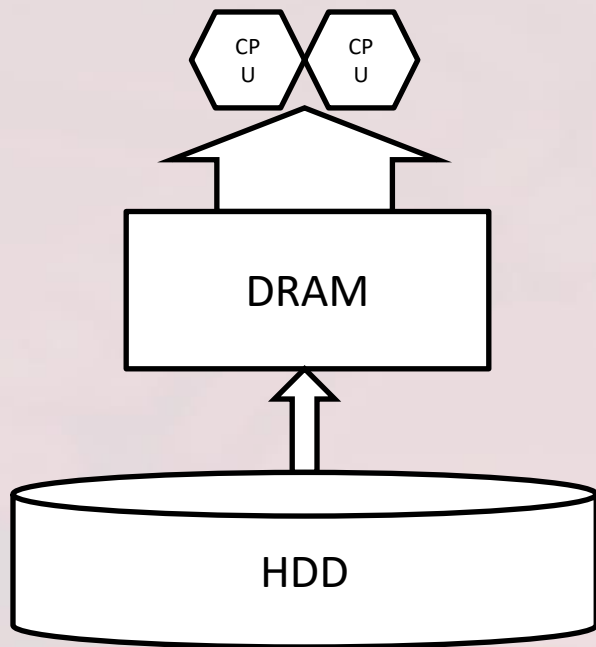


使用流水线之后：  
多个算子同时处理；  
只涉及一次I/O吞吐。



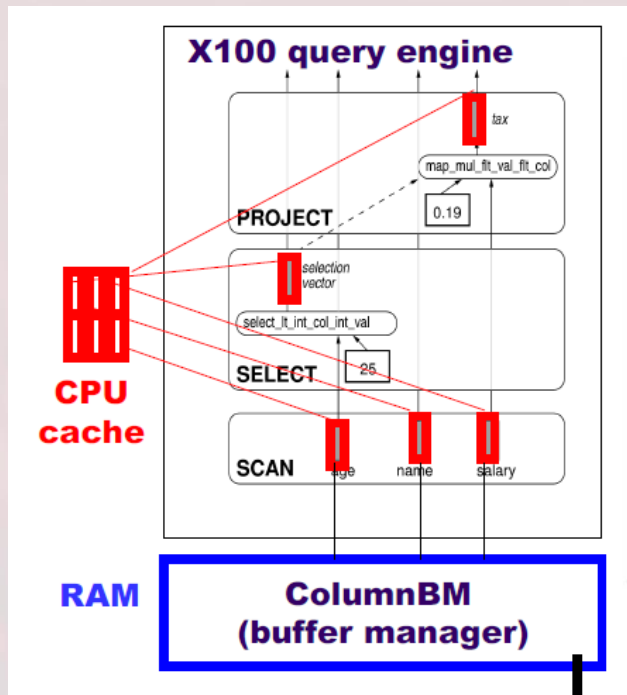


# 从磁盘到内存

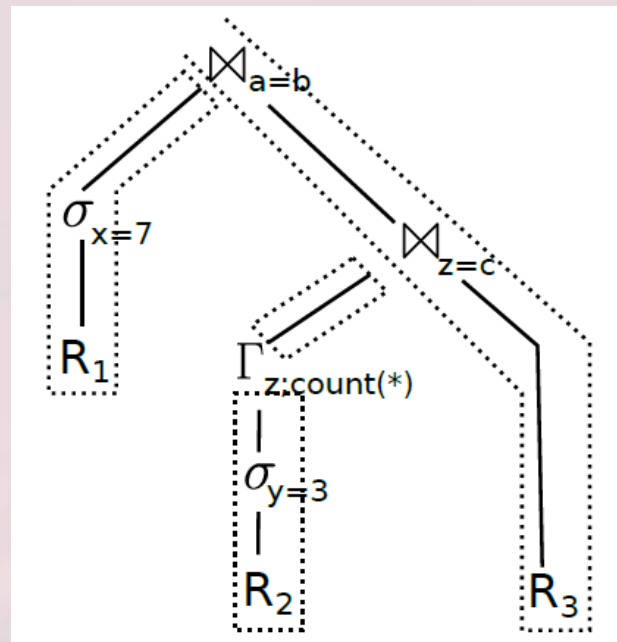




# 内存数据库的流水线优化



❖ **Vectorwise** – 缓存内的流水线



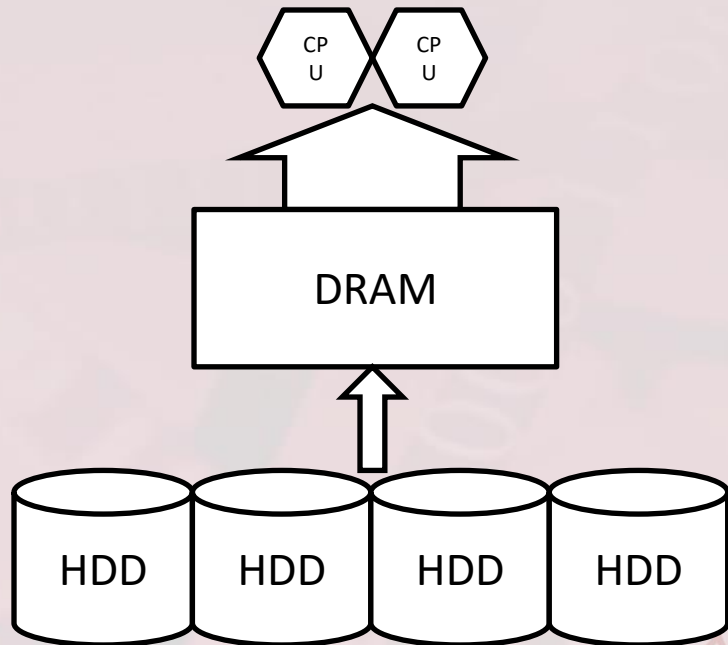
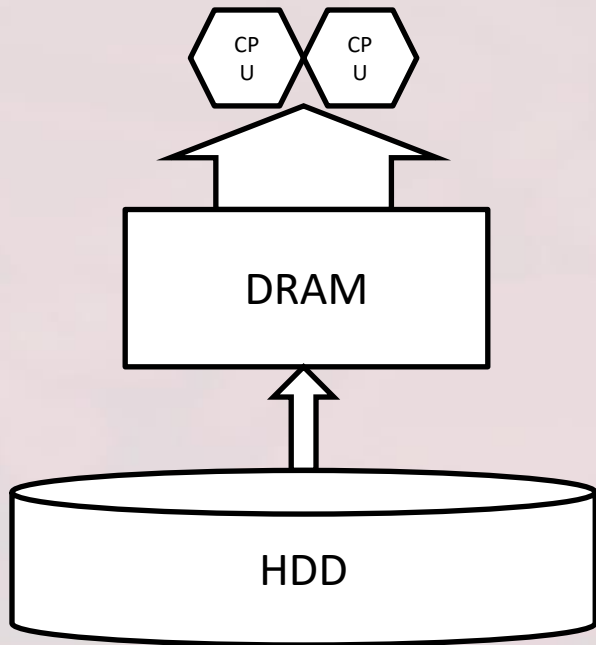
❖ **Hyper** – 寄存器内的流水线

# 内容概要

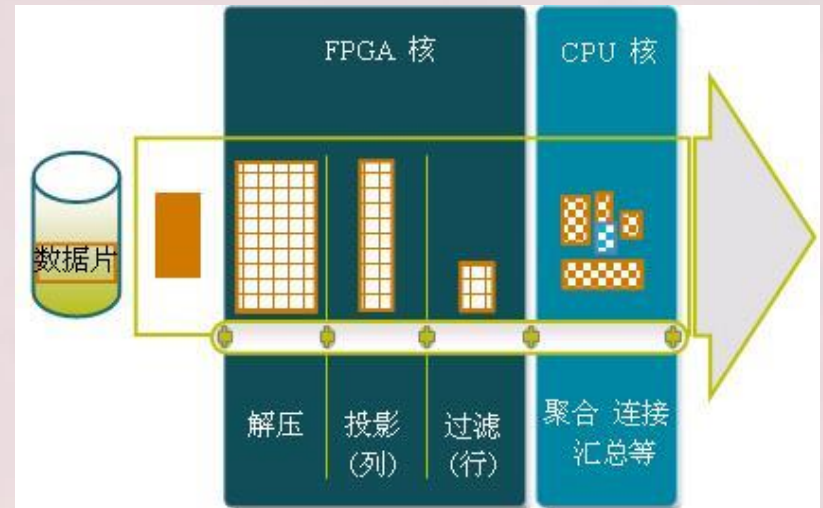
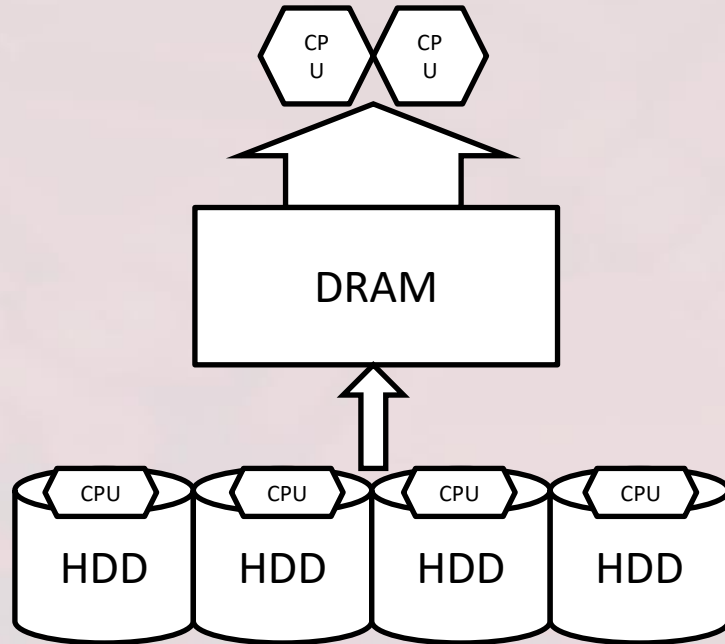
- ❖ 如何让数据库系统充分利用硬件？
  - 如何利用缓存？
- ❖ 什么样的硬件适合数据库系统？
  - 如何拓宽数据通道？



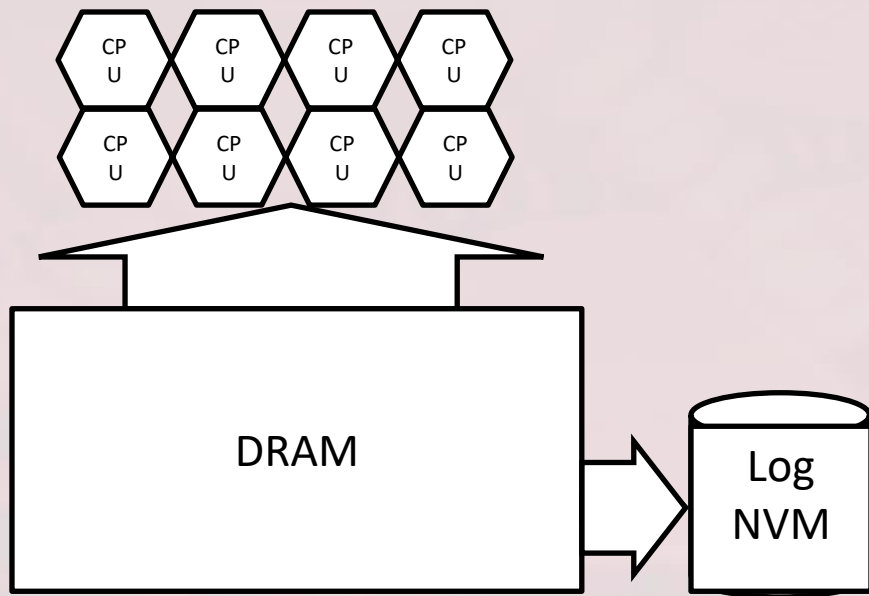
# 硬件的使用：拓宽瓶颈



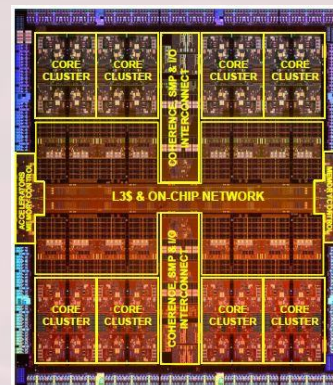
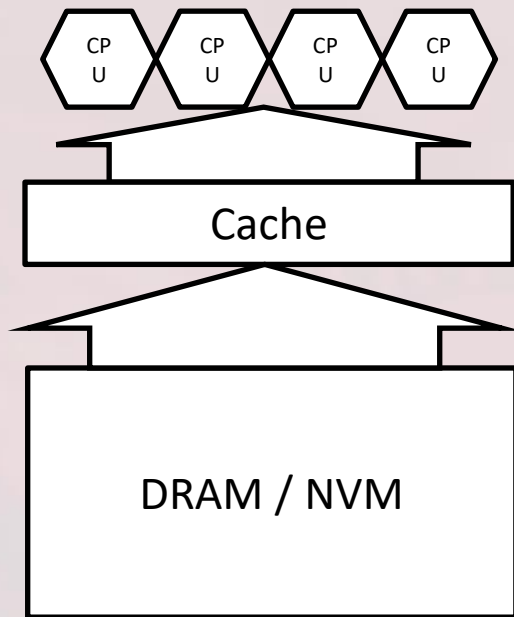
# IBM Netezza



# 硬件的使用：拓宽瓶颈



# 硬件的使用：拓宽瓶颈



# 总结：数据库一体机的优化策略

- ❖ 加宽数据访问通道
- ❖ 提高数据访问通道的效率
  - 缓存的利用
  - 计算下压—移动代码**vs**移动数据

