

COMP3027J 课程

软件架构

表现及其策略

邓永健

北京工业大学计算机学院

数据挖掘与安全实验室（DMS 实验室）



大纲

1. 绩效的意义

2. 提升绩效的策略



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

大纲

1. 绩效的意义

2. 提升绩效的策略



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

绩效的意义

担忧

- 系统对事件的响应速度

与事件的数量和到达模式相关

事件来源

- 用户请求，系统内部，系统外部



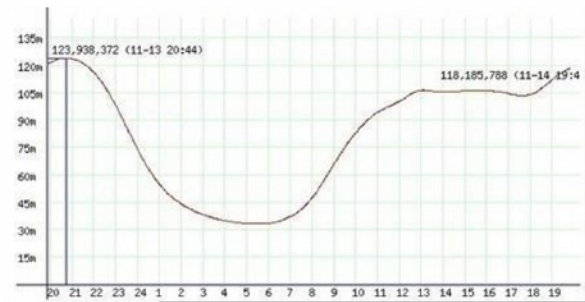
北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

绩效的意义

事件到达模式

- 随机的

在特定的时间尺度（日、月、学期、年）上定期进行



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

绩效的意义

事件到达模式



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

《表演的意义——场景》

刺激源

可能来自系统内部或外部

刺激

- 事件到达（需回复）



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

《表演的意义——场景》

文物

- 系统提供的服务

环境

系统可能处于不同的模式（正常/紧急/过载）



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

《表演的意义——场景》

反应

系统处理传入的事件，这可能会导致状态发生变化。



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

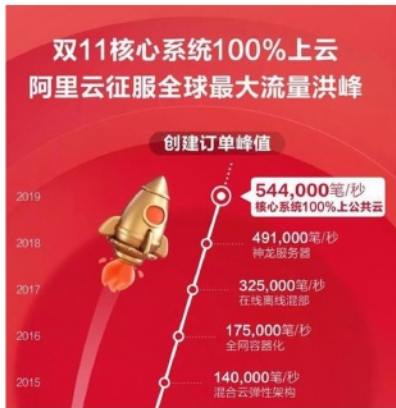
《表演的意义——场景》

应对措施

处理事件所花费的时间

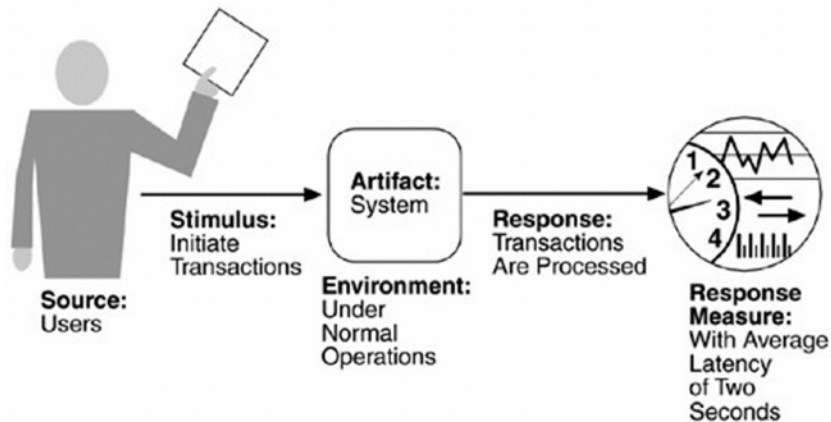
单位时间内处理的事件数量

- 处理错误率/损失率



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

性能场景示例



大纲

1. 可修改性

2. 提升绩效的策略



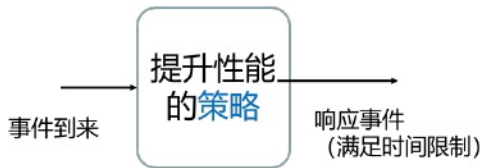
北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

提升绩效的策略概述

目标

在有限的时间内对事件作出回应

获取资源 + 使用资源



方向 1：资源需求

方向 2：资源管理

方向 3：资源仲裁



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

提高绩效的策略

-Resource Requirements

在不改变待处理数据量的情况下提高计算效率

使用更高效的算法

处理事件时减少资源占用

| 排序方法 | 平均时间 | 最好情况 | 最坏情况 | 辅助存储 | 稳定性 |
|------|---------------|---------------|---------------|---------------|-----|
| 选择排序 | $O(n^2)$ | $O(n^2)$ | $O(n^2)$ | $O(1)$ | 不稳定 |
| 插入排序 | $O(n^2)$ | $O(n)$ | $O(n^2)$ | $O(1)$ | 稳定 |
| 冒泡排序 | $O(n^2)$ | $O(n^2)$ | $O(n^2)$ | $O(1)$ | 稳定 |
| 希尔排序 | $O(n^{1.25})$ | -- | -- | $O(1)$ | 不稳定 |
| 快速排序 | $O(n \log n)$ | $O(n \log n)$ | $O(n^2)$ | $O(n \log n)$ | 不稳定 |
| 堆排序 | $O(n \log n)$ | $O(n \log n)$ | $O(n \log n)$ | $O(1)$ | 稳定 |
| 归并排序 | $O(n \log n)$ | $O(n \log n)$ | $O(n \log n)$ | $O(n)$ | 稳定 |
| 基数排序 | $O(d(n+rd))$ | $O(d(n+rd))$ | $O(d(n+rd))$ | $O(rd)$ | 稳定 |



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

减少要处理的数据总量

控制事件到达的速率

仅处理请求中的一部分



提高绩效的策略

限制执行时间

在规定时间内获得近似解

限制要处理的事件队列的长度

直接放弃处理某些事件

- 资源需求



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

提高绩效的策略

- 资源管理

利用并发机制

多线程、多进程、多核、多机……

业界首款旗舰5G SoC芯片
融合5G和AI的革命性飞跃



| | | |
|-------------------------|-----------------|-----------------|
| CPU 8-Core | NPU 2+1 Core | GPU 16-Core |
| 2G / 3G / 4G / 5G Modem | | |
| ISP 5.0 | | LPDDR 4X |
| UPS 3.0 / 2.1 | | HIFI Audio |
| 4K HDR Video | | Security Engine |

103亿晶体管

- 业界首款 7nm + EUV 5G SoC
- 业界首款 旗舰5G NSA & SA SoC
- 业界首款 16核Mali-G76 GPU
- 业界首款 大-微核架构NPU



北京工业大学
BEIHANG UNIVERSITY OF TECHNOLOGY

提高绩效的策略

- 资源管理

增加可用资源

计算资源、存储资源、带宽资源……

| 弹性计算 | 存储服务 | 数据库 | 云通信 |
|---------------------|----------------|-----------------------------|-----------------|
| 云服务器 | 云存储 | 关系型数据库 | 短信服务 HOT |
| 云服务器 ECS HOT | 对象存储 OSS | 云数据库 POLARDB | 语音服务 |
| 弹性裸金属服务器（神龙） | 块存储 | 云数据库 RDS MySQL 版 HOT | 流量服务 |
| 轻量应用服务器 | 文件存储 NAS | 云数据库 RDS MariaDB TX 版 | 物联网无线连接服务 |
| FPGA 云服务器 | 文件存储 CPFS | 云数据库 RDS SQL Server 版 | 号码隐私保护 |
| GPU 云服务器 | 文件存储 HDFS（公测中） | 云数据库 RDS PostgreSQL 版 | 号码认证服务（公测中） |
| 专有宿主机 | 归档存储 | 云数据库 RDS PPAS 版 | 云通信网络加速（公测中） |



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

提高绩效的策略

-资源仲裁

先到先得

固定优先级调度

机场和车站的军人专用通道



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

提高绩效的策略

-资源仲裁

动态优先级

- 无饥饿现象
- 最早截止日期优先



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

性能 - 概要

性能担忧

- 响应处理速度

提高绩效的策略

- 资源需求
- 资源管理

资源仲裁



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

谢谢你!



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY