

COMP3027J

Software Architecture

Performance and its Tactics

DENG, YONGJIAN

Faculty of Computer Science, BJUT

Data Mining & Security Lab (DMS Lab)



Outline

1. The Meaning of Performance

2. Tactics to Improve Performance



北京工业大学
BEIHANG UNIVERSITY OF TECHNOLOGY

Outline

1. The Meaning of Performance

2. Tactics to Improve Performance



北京工业大学
BEIHANG UNIVERSITY OF TECHNOLOGY

The Meaning of Performance

Concerns

- Speed of system response to events
- Related to the number and arrival pattern of events

Sources of Events

- User requests, inside the system, outside the system



The Meaning of Performance

Arrival Patterns of Events

- Random
- Regular at specific time scales (day, month, semester, year)



The Meaning of Performance

Arrival Patterns of Events



全国大学英语四、六级
考试成绩查询

快捷查询 免准考证查询

准考证号 准考证号

姓名 姓名

验证码 

查询



北京工业大学
BEIHANG UNIVERSITY OF TECHNOLOGY

The Meaning of Performance - Scenarios

Source of Stimulus

- May come from inside or outside the system

Stimulus

- Event arrival (requires response)



The Meaning of Performance - Scenarios

Artifacts

- Services provided by the system

Environment

- System may be in different modes
(normal / emergency / overload)



The Meaning of Performance - Scenarios

Response

- System processes incoming events, which may lead to state changes



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

The Meaning of Performance - Scenarios

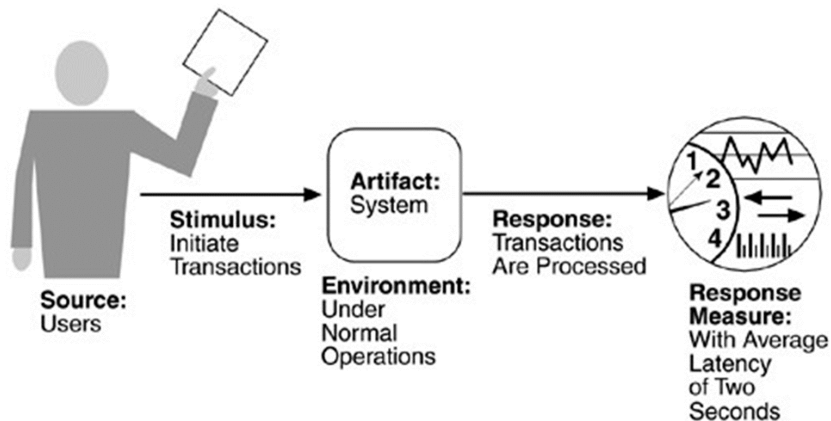
Response Measure

- Time taken to process events
- Number of events processed per unit of time
- Error rate/loss rate of processing



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

Performance Scenario Example



Outline

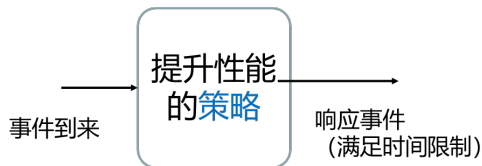
1. The Meaning of modifiability

2. Tactics to Improve Performance



北京工业大学
BEIHANG UNIVERSITY OF TECHNOLOGY

Tactics to Improve Performance-Overview



Objective

- Respond to events within a limited time
- Acquire resources + Use resources

Direction 1: Resource requirements

Direction 2: Resource management

Direction 3: Resource arbitration



Tactics to Improve Performance

-Resource Requirements

Improve computational efficiency without changing the amount of data to be processed

- Use more efficient algorithms
- Reduce resource occupation when processing events

| 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 |



Tactics to Improve Performance

-Resource Requirements

Reduce the total amount of data to be processed

- Control the rate of event arrival
- Process only a subset of requests



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

Tactics to Improve Performance

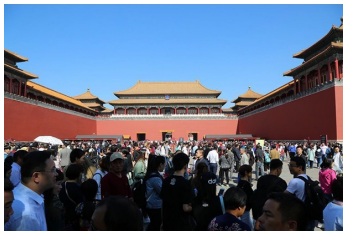
-Resource Requirements

Limit execution time

- Obtain approximate solutions within a specified time

Limit the length of the event queue to be processed

- Directly abandon processing some events



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY


Tactics to Improve Performance

-Resource Management

Utilize concurrency mechanisms

- Multithreading, multiprocessing, multicore, multi-machine...

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 |
| UFS 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

Tactics to Improve Performance

-Resource Management

Increase available resources

- Computational resources, storage resources, bandwidth resources...

| 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） |



Tactics to Improve Performance

-Resource Arbitration

First-come, first-served

Fixed priority scheduling

- Military express lanes at airports and stations



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

Tactics to Improve Performance

-Resource Arbitration

Dynamic priority

- No starvation
- Earliest deadline first



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY

Performance - Summary

Concerns of Performance

- Speed of response processing

Tactics to Improve Performance

- Resource requirements
- Resource management
- Resource arbitration



Thank you!