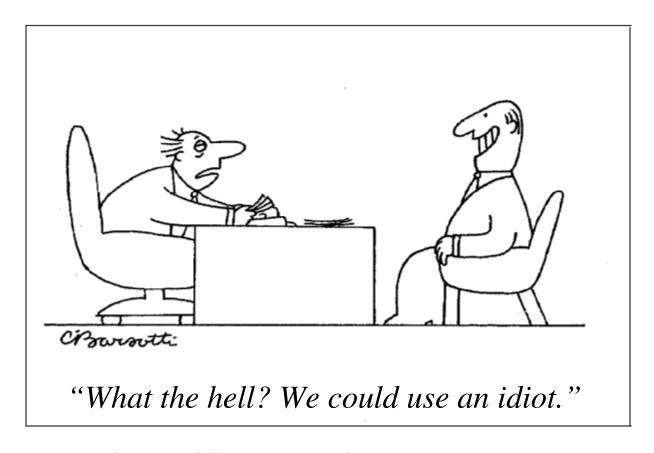
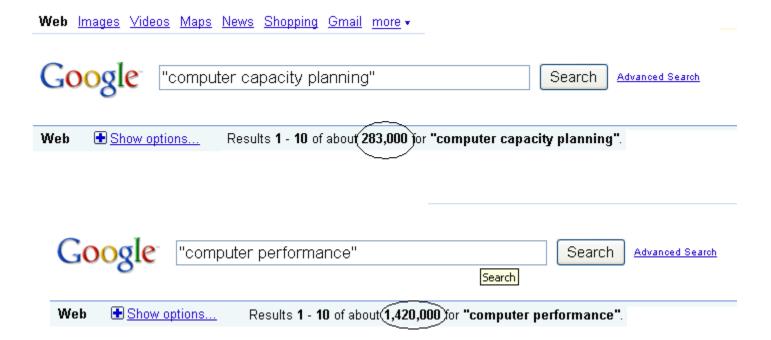
Capacity Planning & Performance for Web-based Services



by Bill Greenbaum, CSM for NYCSPIN, May 11, 2010

2010 Greenbaum Survey of Capacity Planning



4 Questions of Capacity Planning

- 1. What it is?
- 2. Why is it important to me?
- 3. Who does it?
- 4. How to do it?

What is Capacity?

- Capacity: how much traffic can your system handle before crashing?
 - *Throughput:* count of requests per period (TPM)
 - Concurrency: how many simultaneous requests

What is Capacity?

- Capacity: how much traffic can your system handle before crashing?
 - *Throughput:* count of requests per period (TPM)
 - Concurrency: how many simultaneous requests

Contrasts with:

- **Performance** = average response time (sec/req)
- **Redundancy** = capacity reserved for business continuity

What is Capacity Planning?

- Capacity Planning is good management.
 - Proactive
 - A form of risk management
 - Ensures competitiveness & profitablility

...constrasts with

• **Firefighting** = fixing issues that already have customer impact

Capacity contrast with Performance

- **Capacity** = throughput (TPM)
- **Performance** = average response time (sec/req)

	Searches in Billions		
Market	Dec 2008	Dec 2009	% Change
Worldwide	89.7	131.4	46%
Google sites	55.6	87.8	58%
Yahoo! sites	8.4	9.4	13%
Baidu.com Inc	8.0	8.5	7%
Microsoft sites	2.4	4.1	70%
eBay	1.3	2.1	58%
NHN Corp	1.9	2.1	9%
Yandex	1.0	1.9	91%
Facebook	1.1	1.6	54%
Ask Network	1.0	1.5	43%
Alibaba.com Corp	1.1	1.1	-1%

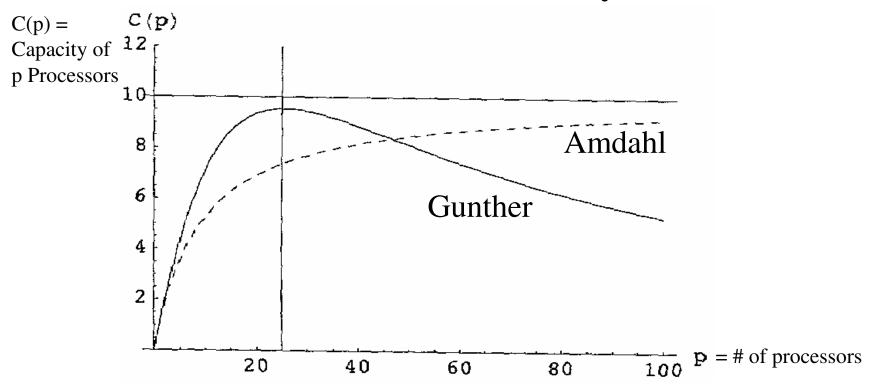
Source: comScore

Who Does Capacity Planning?

- Capacity Planner guides others thru the process, does toughest analysis, crafts proposal.
- Technical Project Managers are the glue
- Business Owners provide critical projections
- Application Developers
- System Developers
- Architecture/Hardware Resource team
- Database Admins
- Disk Storage Admins
- Network Admins
- QA validates & does load tests

How (a): Capacity Planning Theory

- Amdahl Scalability Model, 1967
- Gunther Universal Scalability Model, 2000



How (b): Practice

- Obtain Business projections (VoC)
- Model business projections into peak throughput and concurrency
- Load test / Stress test.
- Follow the data to find bottlenecks
- Monitor traffic and look for trends
- Work with lead-time for procurement

How (c): "Follow the data"

• Weak: "+11% YOY, -2% MOM"

Moderate: Regression lines

• Strong: Control charts include variation in data

Capacity Planning: Example

Further reading

- The Art of Capacity Planning, by John Allspaw, 2008, (O'Reilly Media)
- *Performance by Design*, 2006, Daniel Menasce & Virgilio Almeida, 2004, (Prentice Hall)
- Building Scalable Web Sites, by Cal Henderson, 2006 (O'Reilly Media)
- Guerrilla Capacity Planning, by Neil J. Gunther, 2007 (Springer)

Back to you, Olly!