Some Google Facts

(As of January 2003)

- Started in 1995 as research project at Stanford University by Ph.D. students **Larry Page** and **Sergey Brin**

- Private firm 1998.

- 500 employees, of which 50 hold a Ph.D.-degree.

- Claims to be responsible for around half of all internet searches.

- Three data centers: two in Silicon Valley, one on the US east coast.

$$\text{google} \approx \text{googol} = 10^{100}$$

Google Index

- +3.000.000.000 web pages.

- +35.000.000 non-HTML pages.

- +700.000.000 USENET messages (20 years).

- +2 Terabyte index, updated once a month.

- +2.000.000 terms in index.

- +150.000.000 searches a day (around 2000 per second)

- +200 file types: HTML, Microsoft Office, PDF, PostScript, WordPerfect, Lotus ...

- 28 languages

Google Hardware

- Cluster of +10.000 Intel servers
  - Single-processor
  - 256 MB–1 GB RAM
  - 2 IDE disks of size 20-40 Gb

- Error-tolerance by redundancy.

- Speed by load-balancing.

- Network
  - Gigabit ethernet backbone
  - Round-robin DNS for external load-balancing between data centers (DNS answers with limited time-to-live changes each 5 minutes).
  - Own software for internal load-balancing inside data centers.

Google Software

Red Hat Linux

- Price: Free ("bought 50 copies as a goodwill gesture").

- Widely used linux.

- Access to source.

- In-house support, large market to recruit people from.

- All machines have identical configuration (can all be webservers, index database, …).

- Stateless services.

Removed

- X windows.

- Apache Web Server.

- Telnet, …

Kept

-