

# Skalering, med open-source

# Baggrund

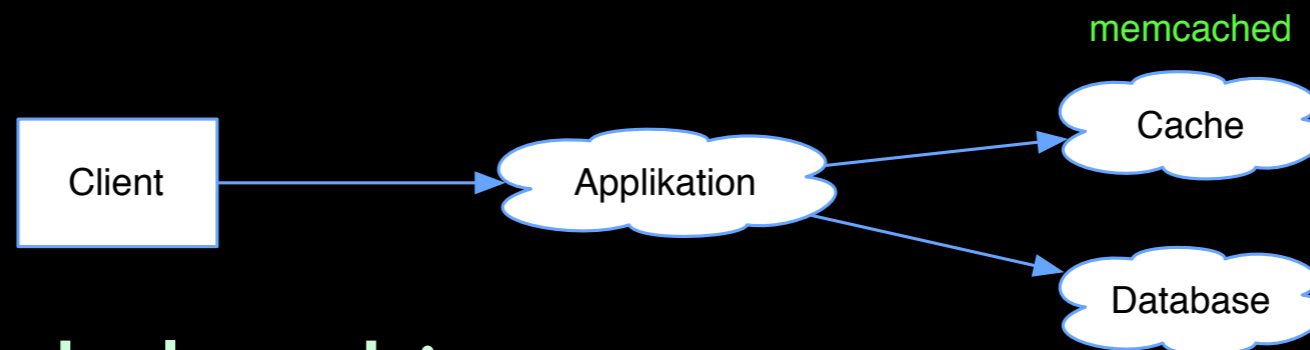
- Amir Salihefendic
- Hoved udvikler og co-founder af bl.a. Plurk.com
- Plurk.com: Top 1000 side på nettet  
ifølge Alexa.com. Flere hits end f.eks. tv2.dk eller dr.dk
- Millioner af brugere og tæt på en milliard unikke plurks
- Ren open-source stack:  
Python, MySQL, memcached, nginx, lightcloud

# Strategier

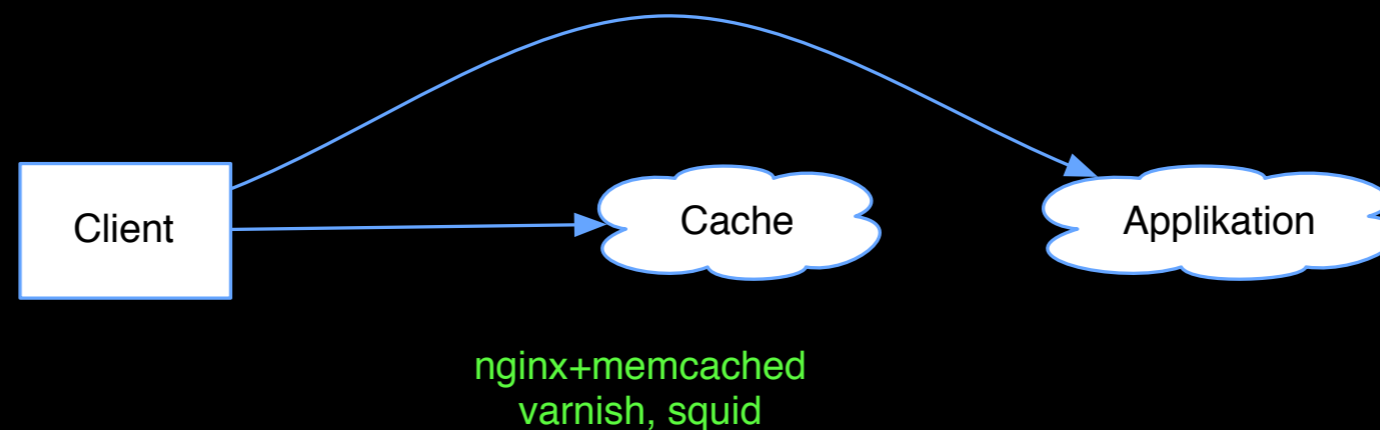
- **Caching**  
nginx+memcached, memcached, varnish
- **Skalering af applikations servere (PHP, Python m.m)**  
stateless, nginx, haproxy
- **Skalering af den relationelle database**  
partitionering, sharding
- **Skalering med NoSQL**  
alternative databaser: LightCloud, MongoDB, CouchDB

# Caching

- **memcached: Open-source caching server**  
Facebook har 800 memcached servers (28TB hukommelse)
- **Bruger specific caching:**  
nginx+memcached eller PHP/Python+memcached

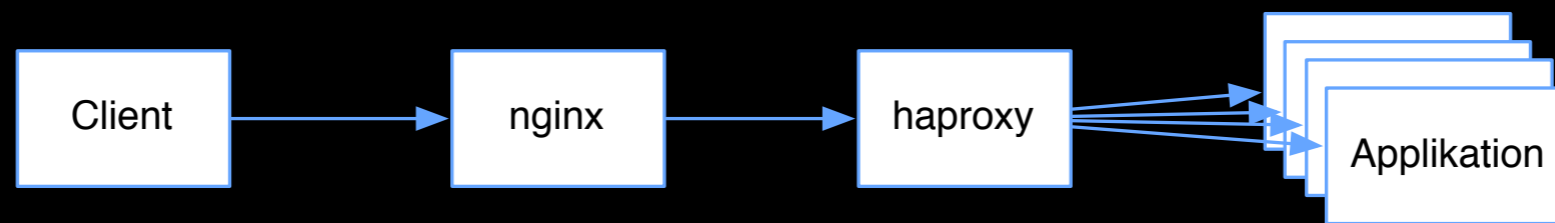


- **Mere global caching:**  
varnish, squid

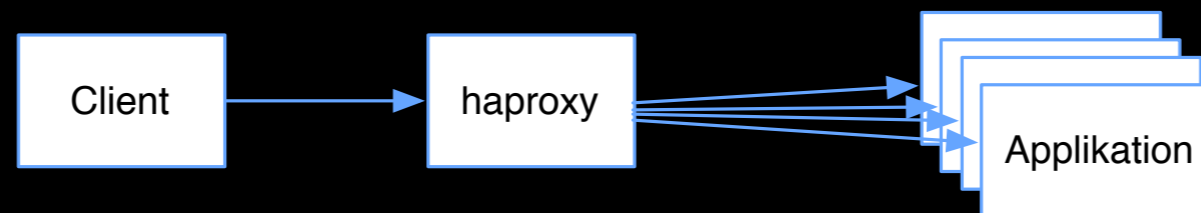


# Skalering af applikation

- Stateless applikation - godt, men ikke nødvendigt  
Secure session cookies. Dog har haproxy understøttelse for statefull routing...
- `nginx > haproxy > (Python / Perl / Ruby applikation)`



- `haproxy > (PHP+Apache)`



# Skalering af den rationelle database

- Det er Rigtigt Svært at skalere store databaser
- Kend din database rigtigt godt  
f.eks. for MySQL: MyISAM vs InnoDB. Vigtigt at vælge gode primary keys.  
Vigtigt at vælge de rigtige datatyper osv. osv.
- MySQL (og andre databaser) elsker hukommelse og meget hurtige diske
- Partitionering: Splittelse af data i flere tabeller  
understøttes direkte af MySQL
- Sharding: Splittelse af data i flere databaser  
understøttes ikke direkte og skal implementeres i applikationen. Svært og gør mange ting

# Skalering med NoSQL

- Åbenbaring: Relationelle databaser er ikke gode til at modellere alt slags data!
- Et nyt hav med nye opensource databaser som kan bidrage relationelle databaser
- Plurk LightCloud: Open source key-value database som basere sig på Tokyo Tyrant eller Redis
- Andre: CouchDB, MongoDB, Apache Cassandra, Hadoop m.fl.

# Links

- <http://opensource.plurk.com/>
- <http://en.wikipedia.org/wiki/Nosql>
- <http://amix.dk/>