

YAHOO!

Performability at Yahoo Search

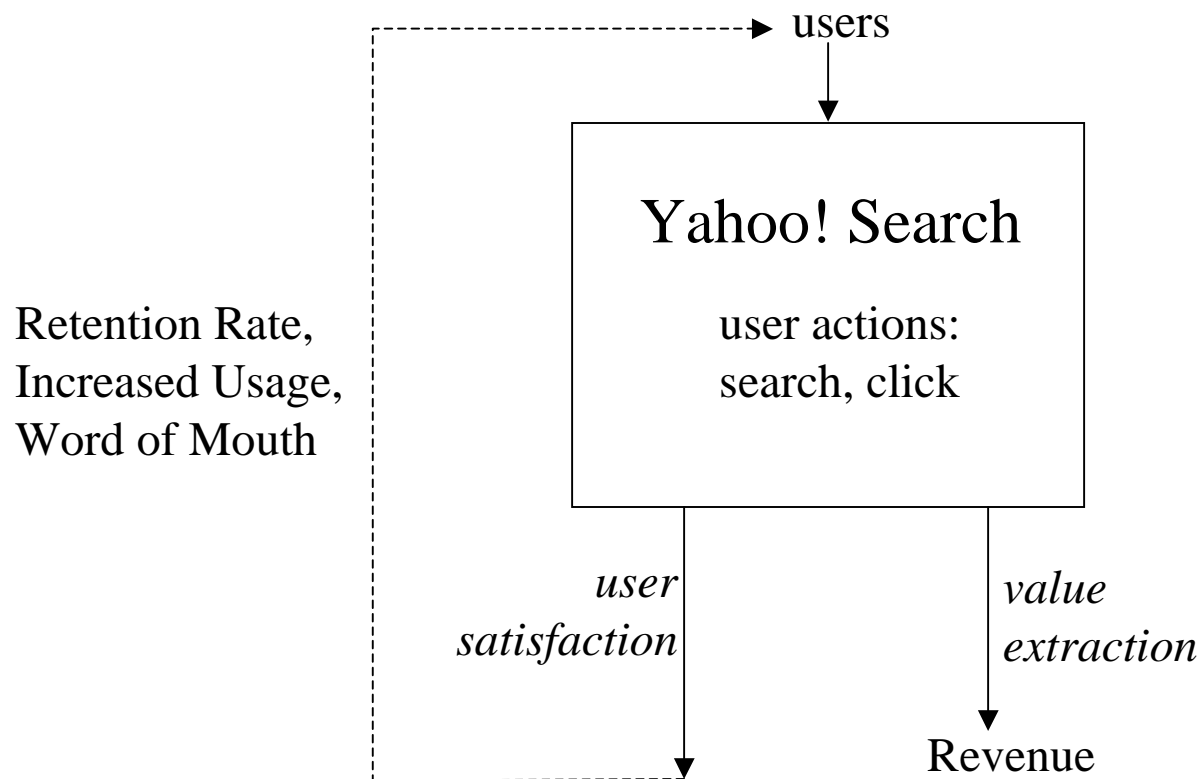
Amr Awadallah and a bunch of other yahoos

amr@yahoo-inc.com

Now, A word from our sponsor 😊

- **What is Yahoo Search ?**
 - **Web Results (Served by Google)**
 - **Direct Display (Yahoo Content)**
 - **Inside Yahoo (Yahoo Self Promotion)**
 - **Sponsored Listings (Overture)**
 - **Media Ads (e.g. North Banner)**

What to Measure?

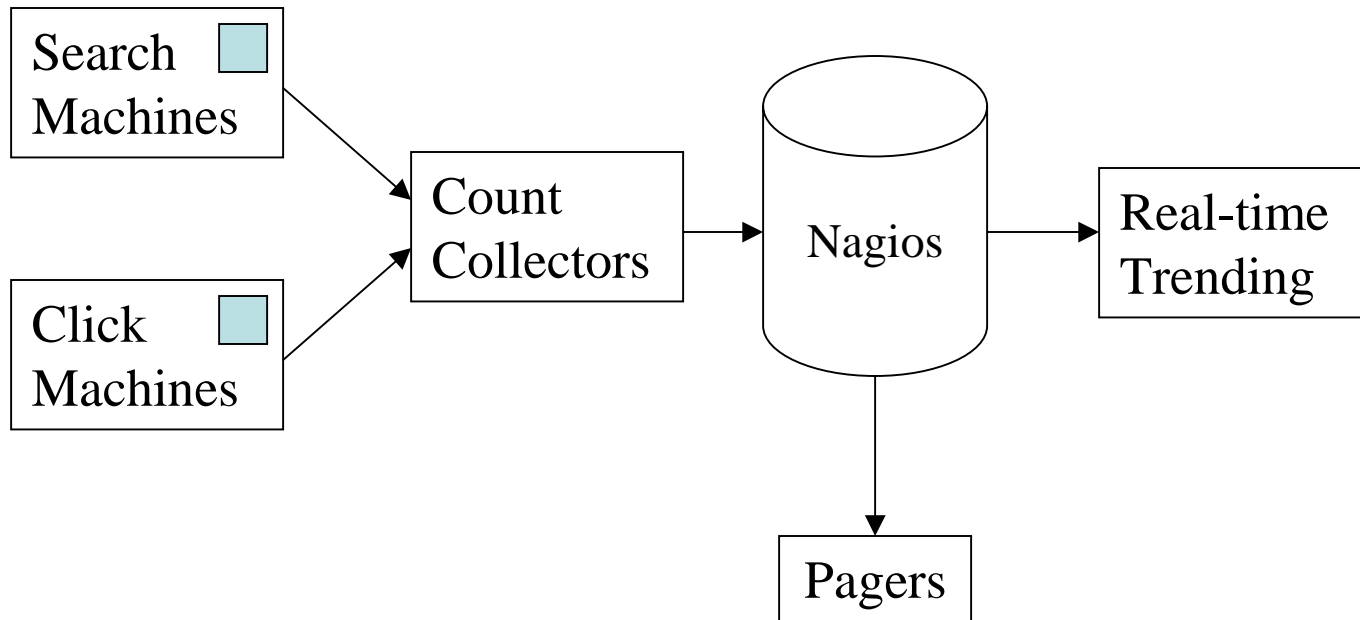


CPU, Harddisk space, Memory Map, Core Dumps, Net IO, QPS, Latency, PVs, Clicks, ...

The Holy Grail: Real-time CTR

- **CTR = Click Through Rate = Clicks/Pages**
- **Advantages:**
 - **Does not change significantly from week to week
(filters out seasonal effects)**
 - **Very sensitive to any small problem taking place**
 - **Quickly deviates from norm in case of faults**
 - **Can be done at many levels of granularity
(e.g. total CTR, Web CTR, Sponsored CTR,
Per-Server CTR, ...)**

Capture → Aggregate → Process → Report



- Quick grep agent parses apache logs on the fly and sends messages to the count collectors every 5 minutes.

Examples:

- **CSI: Crash Scene Investigation ☺**
- **Forensic evidence tend to disappear over time (CYOA principle)**

One Caveat:

- **Instrumentation for real-time metrics adds another point of failure, specially click tracking**

What next?

- **Accumulate human knowledge into rule-based systems that can follow the same diagnosis steps that a human goes through to locate the reason for the fault.**
- **Can we expand RT-CTR to other Internet Apps?**

Do YOU
YAHOO!TM
?