Fathoming the (Deep) Real-Time Web

Mike Franklin
Founder and CTO, Truviso

Defrag Conference
November 12 2009
Lots of Buzz about the Real-time Web
But, where is the information of value?
Here’s the Story for the Static Web

Exploring a ‘Deep Web’ That Google Can’t Grasp

Published: February 22, 2009

One day last summer, Google’s search engine trundled quietly past a milestone. It added the one trillionth address to the list of Web pages it knows about. But as impossibly big as that number may seem, it represents only a fraction of the entire Web.

Beyond those trillion pages lies an even vaster Web of hidden data: financial information, shopping catalogs, flight schedules, medical research and all kinds of other material stored in databases that remain largely invisible to search engines.

Think: “databases”, “log files”, “spreadsheets”, …
The “Real-Time” Web is Deep Too!

Every:

- Ad impression, click
- Tweet, re-tweet, link
- Wall post, friending, …
- Billing event
- Fast Forward, pause,…
- Server request
- Transaction
- Network message
- Fault
- …

Basically, all the stuff that ends up in:
“databases”, “log files”, “spreadsheets”, …
Deep Real-Time Web
Example: Video/TV Distribution

• Today Nielsen panels 5,000 households nationally and 20,000 in local TV markets for current television viewership.

• Today the U.S. Online Video audience is currently over 150 million viewers.
On-Line Data is More Detailed and Relevant

From playout to reporting.

- Viewer
- Program/Content
- Technology platform
- Minutes viewed
- Program segments viewed
- Ads viewed
- Viewing behaviors (FF, RW)

times 10’s or 100’s M of users, every 5 minutes...
= TBs/day
Giving advertisers the analytic tools to maximize their investment
Identify and leverage trends immediately
Identify operational and delivery problems faster
Challenges of the Deep Real-Time Web

• **It’s BIG**
  • Companies are dealing with many TB per day
  • Used to be just Physicists, Spooks, and Telcos

• **It’s Growing**
  • Machine-generated vs. hand-typed @ 140 char
  • e.g., a video player beacon – 4KB per min per viewer

• **It Gets Stale**
  • Decisions have deadlines; things change

• **It Requires Context**
  • Demographics and history matter

• **Sometimes it Shows Up Late**
  • Seconds, minutes, days, ….
The basic approach is the same for a legacy data warehouse cluster or a brand new Hadoop cluster.
The Latency/Performance Tradeoff: Gone

With Batch: reducing batch size reduces throughput.
With Continuous Analytics this tradeoff goes away.
Taming the Deep Real-Time Web

Key to:

- Monetization
- Personalization
- Behavioral Targeting
- Service Optimization
- New/Improved Products
- New/Improved Business Models

Continuous Analytics provides the required scalability and immediacy.
For more information:

@truviso
www.truviso.com
mfranklin@truviso.com