We Lose

Joe Hellerstein
UC Berkeley
HPTS 2001
History

Generic.com, HPTS 1999

Everyone, et al., HPTS 2001
My Background

• I am a database true believer
• I am an academic
  – Enjoy elegance, beauty, conceptual richness, tweed, etc.
• I am a geek
  – I like to hack
  – I like to play sysadmin (latest fun: OS X)
• Net: I am a intellectual schizophrenic
  – And it’s tearing me apart!
• Been thinking about this last couple years
  – As a result of team-teaching grad students with Eric Brewer
We Have a Beautiful Tradition

• Top-Down design of relational systems, TP
  – Semantics first
    • Codd gives a data model and declarative languages
    • System R crowd defines serializability, etc.
  – Implementation wizardry later
    • And we can rise to any challenge -- just watch us!

• We love to set high goals
  – General-purpose storage/query
  – No sacrifice on performance or availability or distribution or parallelization or load-balancing or….
  – AND still give you beautiful semantics

• Care about the user and their data
But We Lose!

- Grassroots* use Filesystems, not DBs
- Grassroots use App servers, not ORDBs
- Grassroots write Java, PERL, Python, PHP, etc. etc. etc. NOT SQL!
  - or XQuery …

Grassroots: Hackers. But also DBMS engineers, Berkeley grads, Physicists, etc.
Hanging Out with OS Folks

• OS folk have a beautiful tradition too
  – Simple, narrow interfaces and tools
  – Strong ties to the PL world
  – Care about the programmers and their tools

• Bottom-up elegance
  – KISS
  – The art of engineering first. Semantics later.

• [Ahhh… DB folk can hack better than they can!]
  – Not the point!
Why We Lose

• While we’ve been thinking about users…
• The OS/Bottom-Uppers have been targeting programmers

• Without programmers, it’s damn hard to reach users
  – THIS is why we missed the first waves of the Internet
  – THIS is why ORDBMS lost to App Servers
  – Like trying to sell drugs without the Mafia
Why We’re Supposed to Win

• Eventually they’ll come crying to us
  – When they realize they should have had data independence
  – And we give ‘em the best server platform

• Maybe.
  – Maybe that will be too late.
    • Certainly their server platforms have been catching up
  – Databases commoditized and cornered?
    • To slow-moving, evolving, structure-intensive apps that require schema evolution

• Maybe we should reach out more?
Tools and Community

• We need to work on tools
  – Query debuggers
  – Data cleaners
  – Workflow builders/debuggers
  – Right direction: GUIs + logic + a little AI
    • Semi-automatic tools
    • One possibility: our “programmers” are content managers
      (see Stonebraker/Hellerstein SIGMOD ‘01)

• We need to foster a vibrant grassroots community
  – A la the USENIX/OpenSource world
  – How the hell do we do this? Databases are boring!
    • Even open source databases are boring!
    • Far more Linux/BSD buffs than Postgres/Sleepycat/MySQL
Fun With Our Stuff

• Query processing
  – It’s not boring, it’s the coolest thing in computing!
    • Text search, P2P. QP over *their* data, not DB data.
  – Or “continuous” QP
    • Pub/sub for fun (and profit?)

• A toolkit for QP, Cont. QP (pub/sub), etc.
  – Bottom-up: more like query algebra than SQL
  – Dataflow diagrams or “pipe” scripting
  – Plus infrastructure to make it work well over the Internet, etc.
    • You wanted it to be fun, right?
    • Deal with wide-area federation, adaptivity
  – Telegraph project at Berkeley
My Gut

• We lost the cool new Internet enterprise space
• Only the standard enterprise folks will come to us tail-between-legs
• We need to look for a new opportunity
  – Maybe P2P QP
    • Search interface today is just string-match. How to do more?
  – Maybe Internet QP
    • Web search can’t do the most obvious things
    • Needs query/wrapper tools desperately
  – Maybe ubiquitous computing/sensornet QP?
    • See Intel Research tablets @ Berkeley/CMU/Washington
• None of this seems like a business model (yet)
  • But it’s cool – just what we’re missing!
P2P Example: TeleNap
**Web Query Example: Telegraph**

![Image of a web browser window showing a query results page]

### Query Results

<table>
<thead>
<tr>
<th>Name</th>
<th>Occupation</th>
<th>Address</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith, Jimmy</td>
<td>Self employed</td>
<td>Los Angeles, CA</td>
<td>250.00</td>
</tr>
<tr>
<td>Somers, Suzanne</td>
<td>Self</td>
<td>Valencia, CA</td>
<td>1,000.00</td>
</tr>
<tr>
<td>Stamp, Terence</td>
<td>Info Requested</td>
<td>Sanbornville, CA</td>
<td>1,000.00</td>
</tr>
<tr>
<td>Stone, Sharon</td>
<td>Self employed/Actress</td>
<td>Los Angeles, CA</td>
<td>1,000.00</td>
</tr>
<tr>
<td>Streisand, Barbra</td>
<td>Self employed/Singer/Producer</td>
<td>Santa Monica, CA</td>
<td>1,000.00</td>
</tr>
<tr>
<td>Taylor, Elizabeth</td>
<td>Not employed/Homemaker</td>
<td>Tampa, FL 33...</td>
<td>250.00</td>
</tr>
<tr>
<td>Thomas, Heather</td>
<td>CIGNA Healthcare/New Business</td>
<td>Nashville, TN</td>
<td>250.00</td>
</tr>
<tr>
<td>Thomas, Michelle</td>
<td>Washington</td>
<td>Washington, CA</td>
<td>300.00</td>
</tr>
<tr>
<td>Thomas, Olive</td>
<td>National Council of Churches</td>
<td>Maryville, TN</td>
<td>1,000.00</td>
</tr>
<tr>
<td>Thomas, Olive</td>
<td>National Council of Churches</td>
<td>Maryville, TN</td>
<td>1,000.00</td>
</tr>
<tr>
<td>Tomlin, Lily</td>
<td>Self employed/Actress</td>
<td>Los Angeles, CA</td>
<td>250.00</td>
</tr>
<tr>
<td>Triplehorn, Jeanne</td>
<td>Self employed/Actress</td>
<td>Los Angeles, CA</td>
<td>1,000.00</td>
</tr>
<tr>
<td>Wagner, Robert</td>
<td>Self employed/Doctor</td>
<td>McLean, VA 2...</td>
<td>500.00</td>
</tr>
</tbody>
</table>