Thoughts on the Future: Opportunities & Roadblocks

Michael J. Carey
Computer Science Department
UC Irvine

http://isg.ics.uci.edu
Me Predict the Future?

• In 2000:
  – “Oh look – the web – e-commerce – it’s time for me to join an Internet startup!!”

• In 2001:
  – “Google? Search? Come on, Eric Brewer’s already solved that problem with HotBot (a.k.a. Inktomi)!”

• In 2008:
  – “It’s time to rejoin academia – where should I go? Oh, I know – to the UC System!”

(So just do the opposite... 😊)
The Good News

• Arguably the most exciting time in the DB (and information systems) field in 20+ years...
  – Codd gave us the stone tablets with the relational model in 1970
  – In the 70’s and 80’s many of the theoretical and architectural foundations were laid out
  – One might argue that we have been “polishing” since then (albeit in many interesting ways)

But now: **RULES WERE MADE TO BE BROKEN**...! 😊
For Example: Data in the Cloud

• Architecture
  – Shared nothing partitioning (like Teradata)
  – Scalable, self-managing DFS inside (HDFS)
  – Based on a key-value store (e.g., BigTable)

• Data model
  – Relational, sparse relational, semistructured, arrays, key-value based, none, ...

• Data language
  – Basic API, sp-SQL, spj-SQL, HiveQL, full SQL, Pig, Jaql, DryadLINQ, ...

• Consistency model
  – None, ACID, localized-ACID, eventual consistency, other?

All candidates are fair game... (Note: Wild, Wild West!)
A Few Players Own The World

• Where are the world’s most interesting data sets?
  – Google, Facebook, Yahoo!, Twitter, ...

• Where are the Really Big Clusters located?
  – Google, Facebook, Yahoo!, Amazon, ...

• Where do even MS starting salaries have six digits?
  – Yep, you guessed it...

• Therefore not terribly surprising that...
  – Academia is arguably “playing catch-up” nowadays
  – Many top Ph.D. graduates go straight to such places

What are the implications for our field...?
Researchers Want Relevance, But…

• Most potential collaborators are trying to solve a real problem!
  – Have real data, real users, real deliverables
  – And, often have no sense of humor…

• Most of our systems are not complete enough, stable enough, or supported well enough
  – Would have to do the other 80% of the work
  – Might have to pay at least 80% of the salaries

Might also have to write 80% fewer papers…?
Things That Might Help

- Seek industrial collaborations to gain access to real problems and real data (on real clusters)
- (Re)learn to value and admire the production of shared software systems (ex: Postgres, Condor, …)
- Change funding agency behavior somehow?
  - Move away from the current “paper chase” culture
  - Encourage production of open source software
  - Encourage cross-use of software components
  - Fund production of (quality) open source software

Likely to require multiple culture changes…
My Call To You

• Feel free to help define the future!
  – *Do* be inspired by what the Big Players are doing
  – *Don’t* be awestruck and start to follow in places where you can (and should) try to lead

• Work at being relevant!
  – Know how real systems are organized inside
  – Seek challenges and feedback from industrial friends
  – Resist the “papers for papers’ sake” temptation

*Enjoy the exciting times that we’re living in...!*