

# Grand Challenges in Testing Data- intensive Computing Systems

Phyllis Frankl  
Polytechnic Institute of NYU

DBTest Keynote, June 7, 2010

- Sociological problem
  - Understanding risk and investing to prevent low-probability high-risk events
- Technical problems
  - Managing and testing security, privacy, provenance, trust, and regulatory compliance

# Underestimating Risk

- “ For all the criticism BP executives may deserve, they are far from the only people to struggle with such low-probability, high-cost events. Nearly everyone does. ‘These are precisely the kinds of events that are hard for us as humans to get our hands around and react to rationally,’ Robert N. Stavins, an environmental economist at Harvard, says. We make two basic — and opposite — types of mistakes. When an event is difficult to imagine, we tend to underestimate its likelihood. This is the proverbial black swan. Most of the people running Deepwater Horizon probably never had a rig explode on them. So they assumed it would not happen, at least not to them.”
- --Spillonomics: Underestimating Risk, David Leonhardt, NY Times Magazine 6/6/2010
- How can we encourage more realistic assessment of risks from faulty software?

# Related Technical Problem

- Characterize and test low-probability, very costly problems
- Test exceptional behavior, including “fail-safe” backups
- These problems apply to all kinds of software (and more generally, systems) but are likely to be even harder in data-intensive systems
- May be critically important in systems that need to collect and analyze large amounts of data (e.g. from sensor networks)

# security, privacy, provenence, and regulatory compliance

- Need new usable approaches to
  - specify policies and incorporate their enforcement into applications in a manner that's largely transparent to developer
  - Test and/or analyze whether application complies to policy
- May be useful to delegate more application-level policy enforcement to DBMS and other components
  - Need to (transparently) communicate additional data about the data
  - New approach to guarding against injection attacks: complementary character coding and complement aware components