What is PODS
from
the Scientific Perspective?

Phokion G. Kolaitis

University of California Santa Cruz

&

IBM Research - Almaden

What is PODS from the scientific perspective?

The hard queries:

- Can PODS be considered as a purely theoretical conference or as a conference that explores the methods of real data management and analysis?
- What kinds of impact on practice we had in recent years?
- What kinds of impact would we like to have in the coming years?

What is PODS?

- PODS is **not** about

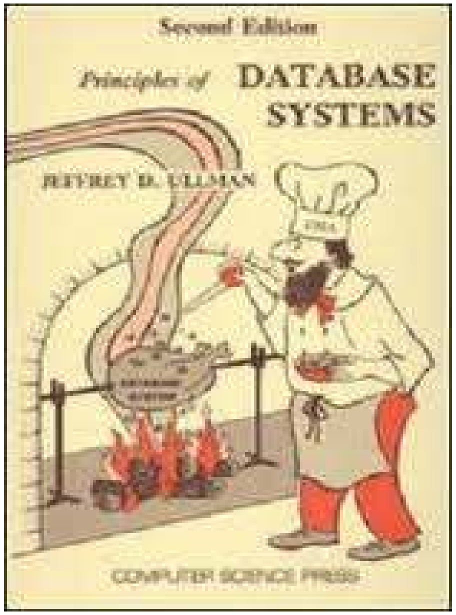
  Portable On Demand Storage

  http://www.pods.com

- PODS is about the

  Principles of Database Systems

http://www.sigmod.org/the-pods-pages



What is PODS?

- Is PODS a purely theoretical conference or a conference that explores the methods of real data management and analysis?

- This is an old, but recurring, question that has been asked since the very early years of PODS.

- In fact, this question has been already addressed in considerable depth in the paper

  Database Theory: Past and Future

  by J.D. Ullman in PODS 1987 (the 6th PODS!)

What is PODS?

- PODS should indeed aspire to “explore the methods of real data management and analysis”.
- The yin and yang duality of database theory and practice

  should be more than just a logo.
- However, in the words of J.D. Ullman (1987):

  “The PODS community is generally regarded as a branch of theoretical Computer Science, it is rarely regarded as a branch of database systems, which is probably sad.”

The Role of (Database) Theory

“Computer theory in the broad sense, is more than justified, possible, or desirable: It is *inevitable*.

…

This is even more true in database research, because abstraction is the essense and *raison d’ être* of databases.”

Database Metatheory: Asking the Big Queries

by C.H. Papadimitriou in PODS 1995.

Successes and Shortcoming of PODS

- Over the years, the PODS community has made significant contributions to the formation and the analysis of the “right” abstractions for database systems research.

- Over the years, the PODS community has had numerous successes in terms of long-lasting impact.

- Yet, why is it that the PODS community
  “is rarely regarded as a branch of database systems”?

Successes and Shortcomings of PODS

- Often times, the PODS community

  "learns far more about the subject than it is necessary."

  J.D. Ullman (1987)

- This is not necessarily bad, but it may lead to a situation in which

  "theoreticians iterate posing and answering their own questions that bring them further and further from the original motivation."

  C.H. Papadimitriou (1995)

Some Missed Opportunities

- The PODS community has failed to develop the principles and methodology for sound ***experimental database research***.
  - To this date, **no** generally accepted framework and set of tools for experimental database research exists.

    In other words:

    What is a “good” database systems experiment?
- The PODS community could have been in the forefront of such an effort.
  - Perhaps there is still time, but certainly an opportunity for significant impact has been missed.

Some Missed Opportunities

- ETL and Data Warehousing
  - Multibillion dollar market ($12.2B in 2012 – Gartner).
  - The “right” abstractions and the “right” models for ETL have yet to be formulated and developed.
- MapReduce
  - What is the “right” computational model for MapReduce?
  - What is the “right” framework for analyzing and comparing MapReduce algorithms?
- By and large, the PODS community has not touched these topics.

Some Things to Keep in Mind

- Maintain a long-term perspective and be patient.

  As Piet Hein once wrote, “Things Take Time”.

- Do **not** rush to declare the demise of an area prematurely.

- The “right” abstractions may fade for a while, but, in the long run, they are here to stay.

A Case in Point

The Rise, Fall, and Rise of Database Dependencies

- Introduced and studied in the 1970s.
- Peaked in the early to mid 1980s.
- Included in the “Last Gasps of the Dying Swans” section of J.D. Ullman’s PODS 1987 paper.

But then database dependencies:

- Made a strong comeback in the past 10-12 years.
- Have been widely used in the formalization of data integration, data exchange, and related data inter-operability tasks.

Another Case in Point

The Rise, Fall, and Rise of Datalog

- Datalog was introduced in PODS 1982 (the first PODS!)
- Between 1982 and 1995,
  Datalog "took the field by storm" (C.H. Papadimitriou – 1995),
  and then faded for quite a while.
- In 1998, M. Stonebraker asserted that
  "No practical applications of recursive query theory have been found to date".

But then:

- Linear Datalog became part of SQL1999 and subsequent standards.
- In the past several years, Datalog has been used in
  "a wide range of practical settings, including security and privacy protocols, program analysis, natural language processing, probabilistic inference, modular robotics, multiplayer games, networking, and distributed systems." (J. Hellerstein – 2010)

Concluding Remarks

- There will always be a place and a reason for theoretical research in database systems.
- “We remind the community that not everything that is implemented is necessary the best that can be done. The theory community has often had its greatest impact when it exposes what the inherent limitations are, concurrency theory and dependency theory offer many results of this nature. Even when our conclusion is that the obvious is the best, we’re upholding the best traditions of science: explore, don’t just hope.”

  J.D. Ullman – PODS 1987