

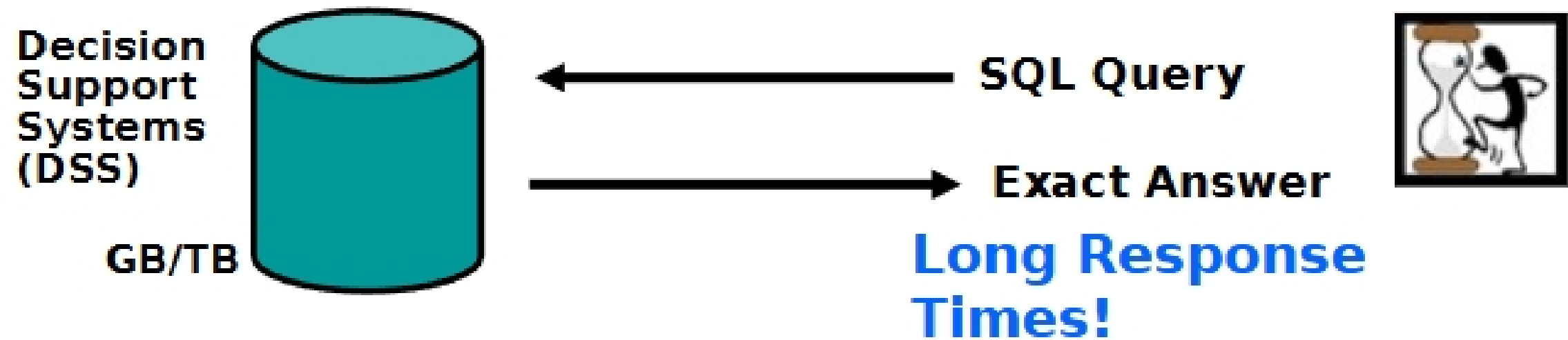
Approximation Techniques for Data Management Systems

***“We are drowning in data but
starved for knowledge”***

John Naisbitt



Traditional Query Processing



- Exact answers **NOT** always required
 - DSS applications usually *exploratory*: early feedback to help identify “interesting” regions
 - *Aggregate queries*: precision to “last decimal” not needed
 - e.g., “What percentage of the US sales are in NJ?”

Fast Approximate Answers

- Primarily for Aggregate queries
- Goal is to quickly report the leading digits of answers
 - In **seconds** instead of minutes or hours
 - Most useful if can provide **error guarantees**

E.g., Average salary

\$59,000 +/- \$500 (with 95% confidence) in 10 seconds
vs. **\$59,152.25** in 10 minutes

- Achieved by answering the query based on *compact synopses* of the data
- Speed-up obtained because synopses are **orders of magnitude smaller** than the original data