DATA INTEGRITY: Threats and Opportunities from Today's Software

DAMA, St. Louis

March 25, 2004

d McGoveran

rnative Technologies

A Graham Hill Rd., Ste 123

Felton, CA   95018

Vmail: 831/338-4621

www.AlternativeTech.com

Why Data Integrity?

By compromising integrity, everything becomes possible…

… and is worth nothing!

Why Data Integrity?

- *Too much data, too little information*

- *You can save it, but do you understand it?*

- *Can you put financial value on it?*

- *Babel Syndrome: Too much data, too many interpretations!*

- *Increasing pressure to certify correctness: our world is <u>semantically fragile</u>.*

What is Data Integrity?

- ***Data** – the basis for making decisions; the result of measurement (quantitative and qualitative)*

- ***Integrity** – soundness, completeness, uncorrupted state*

- ***Syntactics** – the representation*

- ***Semantics** – the meaning*

What Is Data Integrity?

- ***Data uncertainty and confidence***

- ***Security***

- ***Incompatible source and usage contexts***

- ***Redundancy***

Supporting Data Integrity

- *Operational Semantics – usage is everything*

- *Integrity Constraints*

- *Database Transactions – transitions between consistent states*

- *ACID Properties*

Some Dangers

- *Isolation levels – analyze this!*

- *Optimistic concurrency*

- *(pseudo-)Compensation*

- *Unrecognized, untracked dependencies*

- *Ignoring context (source, usage)*

Relational Data Model?

Data, records, files

Vs.

Assertions ("facts"), predicates

Query vs. deductions

Insert vs. affirmation

Delete vs. negation

Why Normalization?

- **Syntactic Integrity**

  Quantifier scope based on dependencies

- **Semantic Integrity**

  Relational operators transition relation variables between consistent values without loss of information

  Quantifier scope (i.e., entity integrity) is neither lost nor modified

Threats: "Dropping ACID"

- ## *Multiple assignment*
  - *CONCERN: SOMEONE MIGHT TRICK YOU INTO USING INTERMEDIATE AND INCONSISTENT RESULTS*

- ## *ACID – criticisms of 'isolation,' 'consistency,' and 'durability'*

Threats: J2EE Application Servers

- *Transaction Assumptions – How optimistic are you?*

- *Bean Managed Persistence*

- *Container Managed Persistence*

- *Transaction Models – Nested? Beans calling Beans calling Beans…*

- *Does your app server play well with others?*

Threats: Business Transaction Protocol

- *What is BTP?*
  - *COHESIONS: DISTRIBUTED TRANSACTIONS IN WHICH RESOURCE MANAGERS NEGOTIATE CONSISTENCY REQUIREMENTS AND OUTCOMES*
  - *WEAK ACID*
  - *COMPENSATION*

- *No way to know what if two BTP transactions are comparable*

- *No semantic integrity*

Threats: EAI

- *Application proliferation*
- *Database proliferation*
- *Data transformation tools are typically syntactic*
- *No concept of application data models*
  - *ARE TRANSFORMATIONS CONSISTENT?*
  - *NO METHODOLOGY FOR COMBINING DATA MODELS*
- *And you want to store that in <u>my</u> database?*

Threats: EII

Enterprise Information Integration

- ## Federated, heterogeneous data access
  - ### UPDATE IS PROMISED

- ## NO Semantics − access is everything

- ## Integration based on simplistic notions of data mapping

Threats: Web Apps

- ## Stateless operations even if we have to:
  - ### IGNORE STATE DEPENDENCIES and REFERENTIAL INTEGRITY
  - ### NEVER USE A SET OPERATION

- ## Table proliferation
  - ### OLD MCDONALD'S ENTITY FARM, NO ANALYSIS
  - ### TEMP AND WORKING TABLES FOR ALL!

Threats: Web Services

- ## Out of context
  - ### WEB SERVICES CAN HIDE JUST ENOUGH INFO TO BE DEADLY
  - ### A DOCTOR IS A DOCTOR, RIGHT?

- ## Choreography

- ## Do you really know what your composite application is doing?

Opportunities: BPMS

Business Process Management System

- **DBMS vs BPMS**

- **Dependency tracking**

- **Transition constraints**

- **Context constraints**

Questions?

BIOGRAPHY

David McGoveran is an industry analyst, and an international management and technology consultant. He is president of Alternative Technologies (Felton, CA), and has been solving difficult relational database, enterprise applications, and business process problems for clients since 1976. He has authored numerous technical articles and co-authored several books (including those with Chris Date), has multiple patents and patents pending, and senior technical editor of the Business Integration Journal.