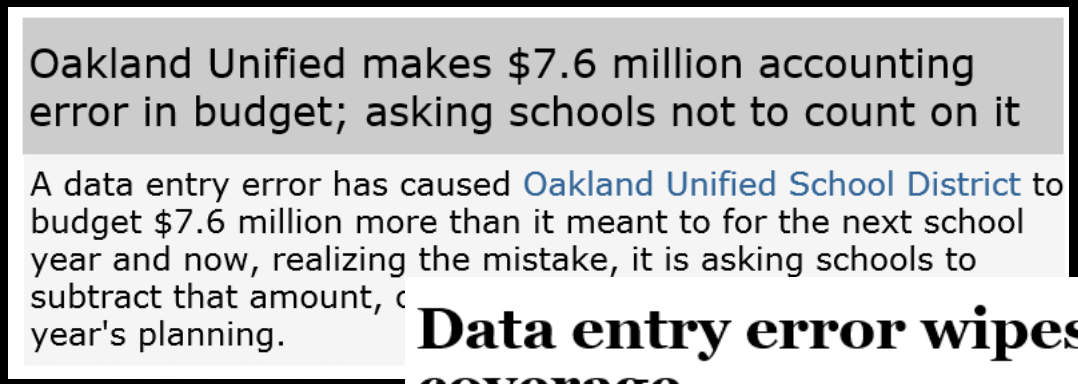
Data Debugging with Continuous Testing

Kıvanç Muşlu♦, Yuriy Brun❄, Alexandra Meliou❄

♦ University of Washington

❄ University of Massachusetts, Amherst

Data Entry Errors



Oakland Unified makes $7.6 million accounting error in budget; asking schools not to count on it

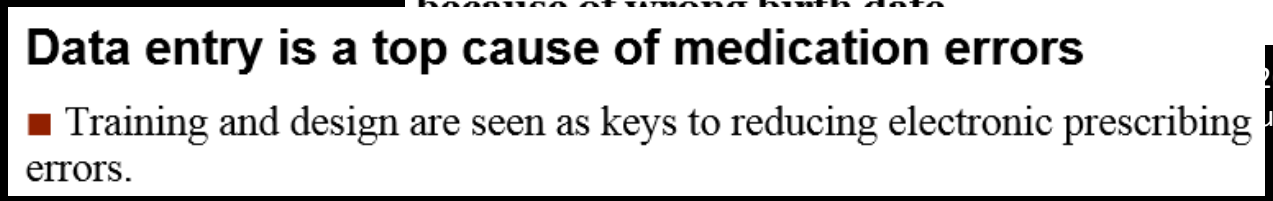
A data entry error has caused Oakland Unified School District to budget $7.6 million more than it meant to for the next school year and now, realizing the mistake, it is asking schools to subtract that amount, c
year's planning.

http://archive.oaklandlocal.c

Data entry error wipes out life insurance coverage

Prudential says woman's policy on son had run out without warning

20101202_1_prudential-employee-
uterol-nebulizer-funeral-procession



Data entry is a top cause of medication errors

■ Training and design are seen as keys to reducing electronic prescribing errors.

http://www.amednews.com/article/20050124/profession/301249959/4

Goal: Detect data entry errors as soon as possible!

Motivation

fname	lname	salary	hbenefit	rbenefit
Alan	Turing	$120,000	$24,000	$20,000
Alonzo	Church	$120,000	$24,000	$21,000
Tim	Berners-Lee	$145,000	$29,000	$26,000
Dennis	Ritchie	$100,000	$20,000	$19,000
Marissa	Mayer	$150,000	$30,000	$26,000
William	Gates	$190,000	$38,000	$30,000

Goal: Reduce each employee's retirement benefit by 10%

Motivation

fname	lname	salary	hbenefit	rbenefit
Alan	Turing	$120,000	$24,000	$18,000
Alonzo	Church	$120,000	$24,000	$18,900
Tim	Berners-Lee	$145,000	$29,000	$23,400
Dennis	Ritchie	$100,000	$20,000	$17,100
Marissa	Mayer	$150,000	$30,000	$23,400
William	Gates	$190,000	$38,000	$27,000

Goal: Reduce each employee's retirement benefit by 10%

Correct query:

```
UPDATE DB SET rbenefit = rbenefit * 0.9;
```

Motivation

fname	lname	salary	hbenefit	rbenefit
Alan	Turing	$120,000	$21,600	$20,000
Alonzo	Church	$120,000	$21,600	$21,000
Tim	Berners-Lee	$145,000	$26,100	$26,000
Dennis	Ritchie	$100,000	$18,000	$19,000
Marissa	Mayer	$150,000	$27,000	$26,000
William	Gates	$190,000	$34,200	$30,000

Goal: Reduce each employee's retirement benefit by 10%

Correct query:

```
UPDATE DB SET rbenefit = rbenefit * 0.9;
```

Data entry error:

```
UPDATE DB SET hbenefit = hbenefit * 0.9;
```

A Couple of Months (and some Valid Updates) Later…

fname	lname	salary	hbenefit	rbenefit
Alan	Turing	$124,000	$24,600	$22,000
Alonzo	Church	$120,000	$21,600	$22,000
Tim	Berners-Lee	$149,000	$27,100	$30,000
Dennis	Ritchie	$103,000	$20,000	$21,000
Marissa	Mayer	$153,000	$31,000	$28,000
William	Gates	$192,000	$36,200	$34,000

William: “My insurance does not pay my hospital bills”

…

- Hard to link the complaint to the data entry error

Goal: Detect Data Entry Errors

Targeted data entry errors

- Data is in valid range: cannot detect with DB primitives
- Data is incorrect: needs application specific knowledge

Related work

- Detection: statistical analysis [BarowyGB 2013]
  - Does not use expert knowledge
  - False positives
- Prevention: challenge-response updates [ChenDLS 2011]
  - Challenge questions can be difficult to answer
  - Difficult to use with automated techniques

Continuous Data Testing

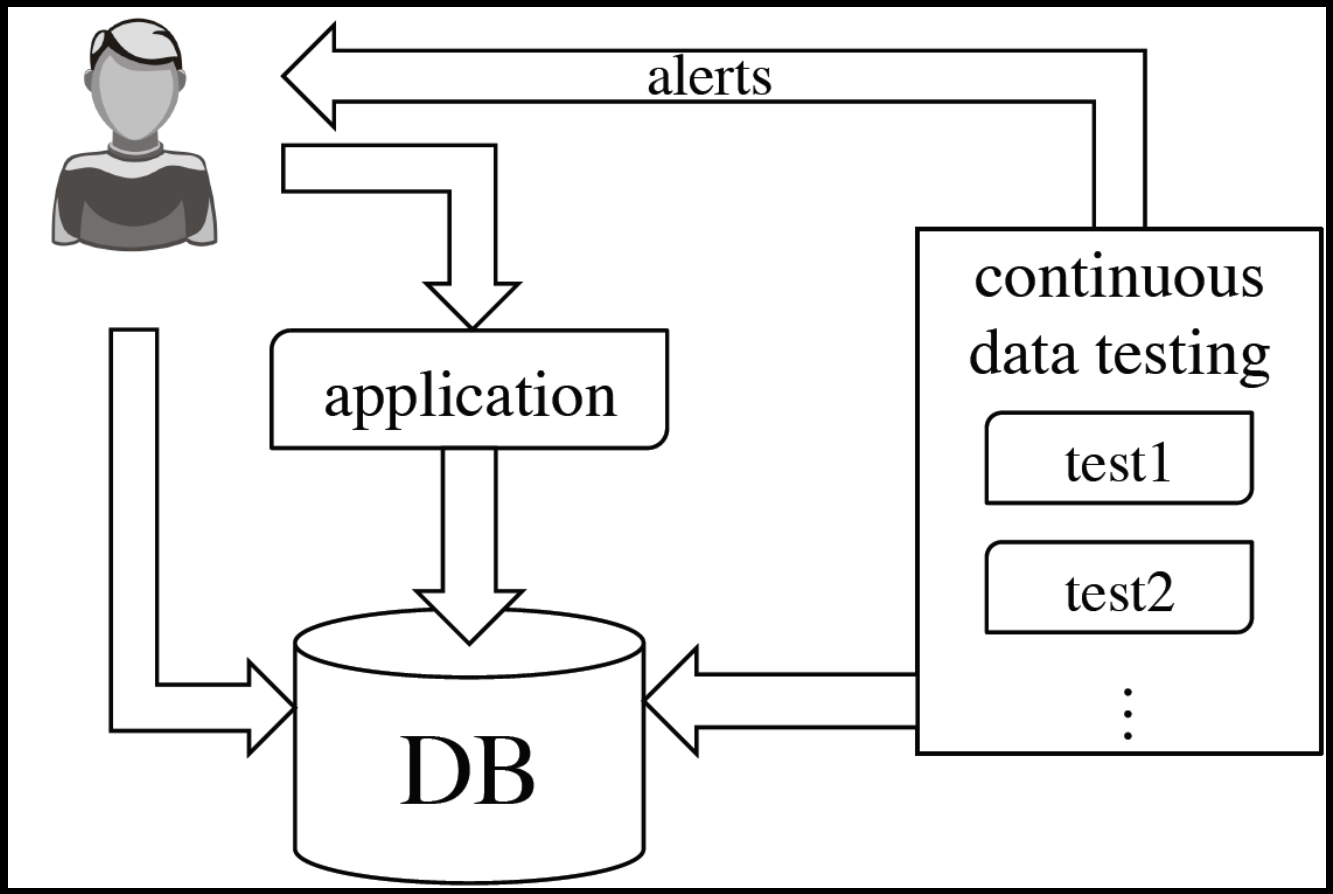
Continuous testing

- Run tests in the background, report failures

Continuous data testing

- Run database tests in the background, report test result changes

Continuous Data Testing Architecture



Test = SQL SELECT representing a regression property

Revisiting Motivation with Continuous Data Testing

fname	lname	salary	hbenefit	rbenefit
Alan	Turing	$120,000	$24,000	$20,000
Alonzo	Church	$120,000	$24,000	$21,000
Tim	Berners-Lee	$145,000	$29,000	$26,000
Dennis	Ritchie	$100,000	$20,000	$19,000
Marissa	Mayer	$150,000	$30,000	$26,000
William	Gates	$190,000	$38,000	$30,000

Test query:

```
SELECT COUNT(*) FROM DB WHERE hbenefit != 0.2*salary
```

Revisiting Motivation with Continuous Data Testing

fname	lname	salary	hbenefit	rbenefit
Alan	Turing	$120,000	$24,000	$18,000
Alonzo	Church	$120,000	$24,000	$18,900
Tim	Berners-Lee	$145,000	$29,000	$23,400
Dennis	Ritchie	$100,000	$20,000	$17,100
Marissa	Mayer	$150,000	$30,000	$23,400
William	Gates	$190,000	$38,000	$27,000

Test query:

```
SELECT COUNT(*) FROM DB WHERE hbenefit != 0.2*salary
```

☑ `UPDATE DB SET rbenefit = rbenefit * 0.9;`

Revisiting Motivation with Continuous Data Testing

fname	lname	salary	hbenefit	rbenefit
Alan	Turing	$120,000	$21,600	$20,000
Alonzo	Church	$120,000	$21,600	$21,000
Tim	Berners-Lee	$145,000	$26,100	$26,000
Dennis	Ritchie	$100,000	$18,000	$19,000
Marissa	Mayer	$150,000	$27,000	$26,000
William	Gates	$190,000	$34,200	$30,000

Test query:

```
SELECT COUNT(*) FROM DB WHERE hbenefit != 0.2*salary
```

☑ `UPDATE DB SET rbenefit = rbenefit * 0.9;`

☒ `UPDATE DB SET hbenefit = hbenefit * 0.9;`

Prototype Implementation

Goals

- Minimize overhead for normal database operations
  12.5% overhead on average

- Catch data entry errors as soon as they occur

When to Test?
Test only when Necessary

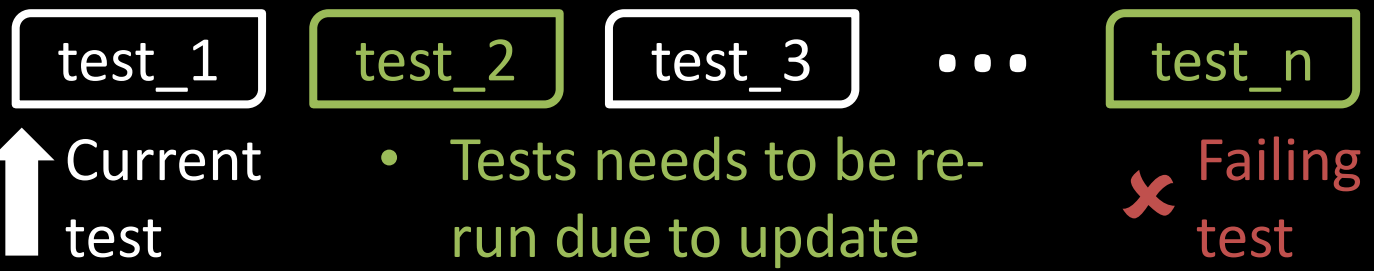
- Naïve approach: Always
  - Unnecessary overhead

➢ Test only when data changes detected by triggers

22.5% → 20.9% overhead on average

What to Test?
Run only the Effected Tests

- Naïve approach: All tests, fixed order
  - Slower failure detection rate:



  - Unnecessary overhead

➢ (Statically) Analyze tests and install triggers

22.5% → 20.9% → 12.5% overhead on average

Challenges

- How to test query performance?
  - Record runtimes for tests, report when delta is big
- UI design: how to make test failures actionable?
  - Create descriptive summaries using causal analysis

Complementary Work

- How to generate tests? Use automated test generation [KhalekK 2011, GauthierMSZ 2012, PanWX 2011]
  - Can be used to suggest tests to the user
- How to test? Consider test results as views
  - Views independent of updates [LevyS 1993]
    - Detect which tests to run more efficiently
  - Incremental view maintenance [GriffinL 1995]
    - Update the test results efficiently and with less overhead
- What to test? Change-impact analysis [RenSTRC 2004]

Contributions

- Continuous data testing:
  a reactive technique to detect data errors
  - Easy to use
  - Unobtrusive: usable with automated methods
  - Generalizes to variety of applications &
    usage models as tests can be application-specific
- Prototype implementation
  - Preliminary experiments suggest low overhead