


# **Belief Databases**

Dan Suciu, University of Washington

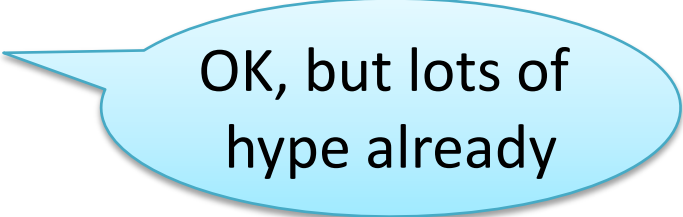
Joint work with Wolfgang Gatterbauer,  
Magda Balazinska, Nodira Khoussainova

# Technology Shifts that Impact Data Management


- Multicore
- Cloud computing:
  - All computation and all data that matters moves to the cloud
- Massive collaborations
  - Wiki's, scientific databases
  - Database as a service
  - *Collaborative Databases*



Sorry, I can't help here



OK, but lots of hype already



Looks exciting to me !

# “Collaborative Databases” (?) = Data contributed by many, stored in cloud

## Data sharing today:

- TableFusion:  
[tables.googlelabs.com/Home](http://tables.googlelabs.com/Home)
- Manyeyes:  
[maneyes.alphaworks.ibm.com/manyeyes](http://maneyes.alphaworks.ibm.com/manyeyes)
- Swivel:  
[www.swivel.com](http://www.swivel.com)
- Dabbledb:  
[dabbledb.com](http://dabbledb.com)

## What they do:

- Upload your “data”
  - Data = spreadsheet
- Allow you to:
  - Share data
  - Visualize data
  - Simple search/filter
- Some allow you to:
  - Annotate

Collaborative Databases need more than this !

# This Talk: Belief Databases

- NatureMapping project
  - Volunteers make observations in the field
  - Data curated by one person – doesn't scale !

## Observations

<u>id</u>	vid	species	date	location	comment
2	Tommy	Crow	06-14-08	Lake Placid	found feathers
53	Freddy	Cougar	06-17-08	Snow Lake	saw footprint
99	Tommy	..	..		
..	..	..	..		

Belief Databases: multiple curators (= users)

# 1. Each user has a set of possible worlds



Alice

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Crow	06-14-08	Lake Placid



Bob

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Raven	06-14-08	Lake Placid

Users/  
Curators

- Each world is consistent
- Inconsistencies across worlds

# 1. Each user has a set of possible worlds



Alice

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Crow	06-14-08	Lake Placid



Bob

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Raven	06-14-08	Lake Placid

## BeliefSQL

```
insert into BELIEF 'Alice' Sightings
values ('s2','Tommy','Crow','06-14-08',
       'Lake Placid')
```

```
insert into BELIEF 'Bob' Sightings
values ('s2','Tommy','Raven','06-14-08',
       'Lake Placid')
```

Q: Who disagrees with Alice and why ?

```
select U2.name, S1.species, S2.species
from Users as U,
     BELIEF 'Alice' Sightings as S1,
     BELIEF U.uid Sightings as S2,
where S1.sid = S2.sid
and S1.species <> S2.species
```

A: {('Bob', 'Crow', 'Raven')}

## 2. Open world assumption



Alice

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Crow	06-14-08	Lake Placid



Bob

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Raven	06-14-08	Lake Placid



Carol

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Crow	06-14-08	Lake Placid
s2	Tommy	Raven	06-14-08	Lake Placid



Repeated key is OK for negative tuples

OWA allows us to give semantics to negative tuples

## 2. Open world assumption



Alice

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Crow	06-14-08	Lake Placid



Bob

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Raven	06-14-08	Lake Placid



Carol

S



<u>sid</u>	vid	species	date	location
s2	Tommy	Crow	06-14-08	Lake Placid
s2	Tommy	Raven	06-14-08	Lake Placid



*I: Carol does not believe that Alice saw a crow nor a raven.*

```
insert into BELIEF 'Carol' not Sightings
values ('s2','Tommy','Crow','06-14-08','Lake Placid')
insert into BELIEF 'Carol' not Sightings
values ('s2','Tommy','Raven','06-14-08','Lake Placid')
```



## 2. Open world assumption



Alice

S



<u>sid</u>	vid	species	date	location
s2	Tommy	Crow	06-14-08	Lake Placid



Bob

S



<u>sid</u>	vid	species	date	location
s2	Tommy	Raven	06-14-08	Lake Placid



Carol

S



<u>sid</u>	vid	species	date	location
s2	Tommy	Crow	06-14-08	Lake Placid
s2	Tommy	Raven	06-14-08	Lake Placid



Q: Who disagrees with Alice and why ?

```
select  U2.name, S1.species
from    Users as U, BELIEF 'Alice' Sightings as S1, BELIEF U.uid Sightings as S2,
where   S1.sid = S2.sid and S1.species <> S2.species
```

A: {( 'Bob', 'Crow' ), ( 'Carol', 'Crow' )}

# 3. Higher-order beliefs



Alice

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Crow	06-14-08	Lake Placid



Bob

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Raven	06-14-08	Lake Placid

C

<u>cid</u>	comment	sid
c1	plain black feathers	s2



Bob

Alice

C

<u>cid</u>	comment	sid
c1	purple-black feathers	s2

Higher order believes: for discussions, explanations

# 3. Higher-order beliefs



Alice

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Crow	06-14-08	Lake Placid



Bob

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Raven	06-14-08	Lake Placid

C

<u>cid</u>	comment	sid
c1	plain black feathers	s2



Bob

Alice

C

<u>cid</u>	comment	sid
c1	purple-black feathers	s2



*I: According to Bob, Alice believes that the feathers of the sighted animal were plain black.*

insert into **BELIEF 'Bob' BELIEF 'Alice'** Comments values ('c1', 'plain black feathers', 's2')

# 3. Higher-order beliefs



Alice

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Crow	06-14-08	Lake Placid



Bob

S

<u>sid</u>	vid	species	date	location
s2	Tommy	Raven	06-14-08	Lake Placid

C

<u>cid</u>	comment	sid
c1	plain black feathers	s2



Bob

Alice

C

<u>cid</u>	comment
c1	purple-

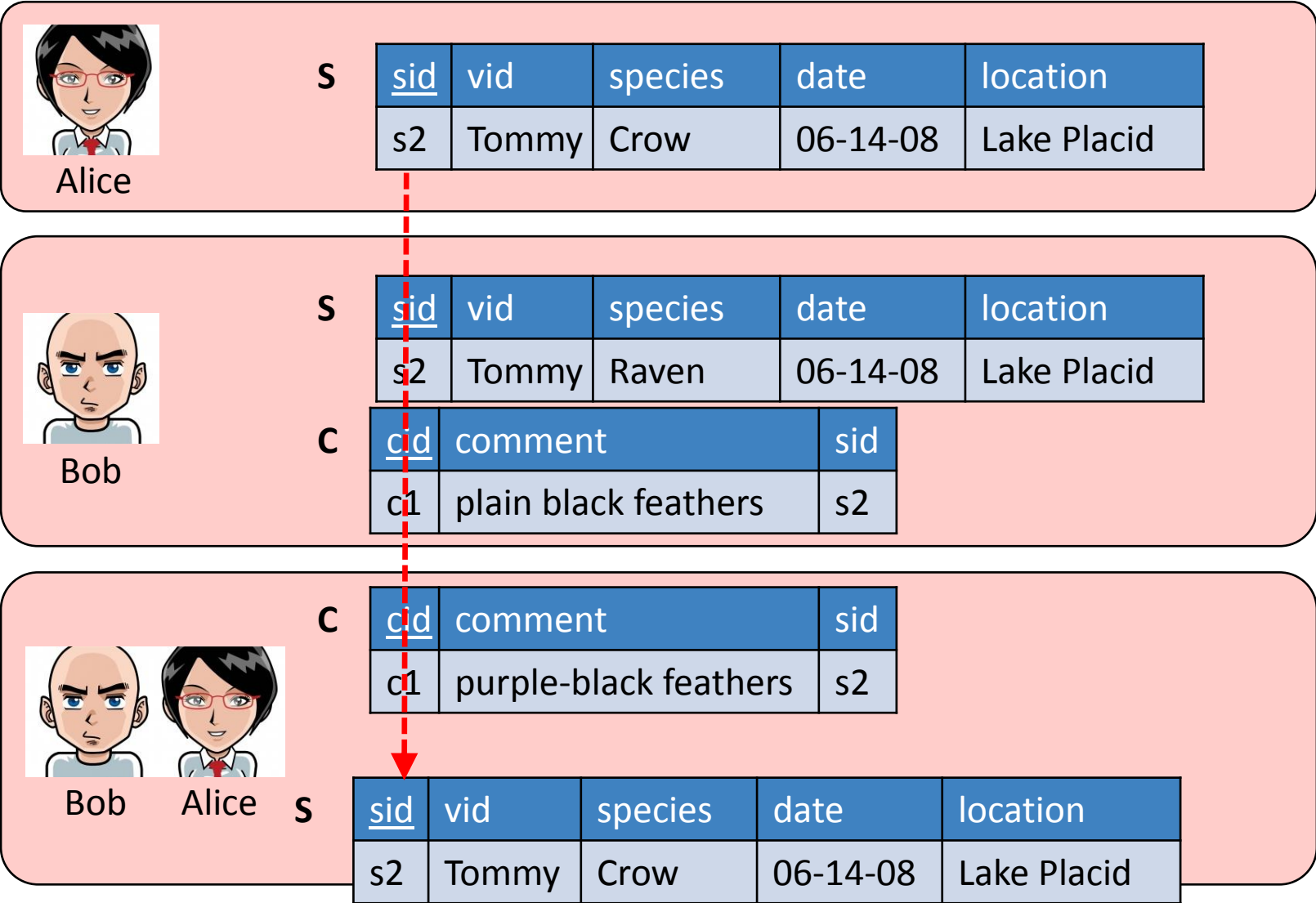


*Q: Which comments does Alice believe according to Bob, which he does not believe himself?*

```
select C1.cid, C1.comment
from BELIEF 'Bob' BELIEF 'Alice' Comments as C1,
BELIEF 'Bob' not Comments as C2
where C1.cid = C2.cid
and C1.comment = C2.comment
and C1.sid = C2.sid
```

*A: {{('c1','plain-black feathers')}}*

# 4. A Default Rule



If a fact is in the database,  
 then each user U believes it unless she explicitly states otherwise

# Open Questions in Belief Databases

- Performance !!!
- Extend/improve/fix the syntax/semantics
- Conflict and trust management:
  - Detection, curation, reputation
- Other modalities and/or axioms
  - “A believes X”, “A knows X”, “A curated X”, ...

# Inspirations and related work (excerpt)

- Annotations
  - Buneman et al. [ICDT 2001 / ICDT 2007]
  - Bhagwat et al. [VLDBJ 2005], Geerts et al. [ICDE 2006]
  - Green & Tannen [PODS 2007]
  - Srivastava & Velegrakis [SIGMOD 2007]
- Modal logic
  - Fagin, Halpern, Moses, Vardi [1995] ←
  - Calvanese et al. [IS 2008], Nguyen [LJ-IGPL 2008]
- Uncertain / incomplete information
  - Sarma et al. [ICDE 2006]
  - Green & Tannen [IEEE Data Eng. 2006]
  - Dalvi & Suciu [PODS 2007]
- Inconsistency / key violations
  - Arenas et al. [PODS 1999]
  - Fuxman et al. [SIGMOD 2005]
- Peer-to-peer computing / collaborative data sharing
  - Bernstein et al. [WebDB 2002]
  - Ives et al. [SIGMOD record 2008]

# Beyond BeliefDBs: Collaborative Databases

- Managing many users
  - Access control, permissions, trust, privacy
- Managing annotations
  - Need semantics of annotations
- Managing inconsistencies/disagreement
  - Cope with them, don't fix them