Synthesizing Third Normal Form Schemata that Minimize Integrity Maintenance and Update Overheads

Parameterizing 3NF by the Numbers of Minimal Keys and Functional Dependencies

Anonymous

1. State of Science
2. Core Idea
3. New Concepts
4. Optimizing Synthesis
5. Experiments
6. Summary

State of Science (SOS?)

Schema Design with Functional Dependencies

- Goal: Maintain data consistency under updates, with a minimum level of effort
- Strategy: Do not admit any redundant data value occurrence in any instance
- Tactics: Transform FDs, that cause redundancy, into keys that prohibit them

Schema Design with Functional Dependencies

- Goal: Maintain data consistency under updates, with a minimum level of effort
- Strategy: Do not admit any redundant data value occurrence in any instance
- Tactics: Transform FDs, that cause redundancy, into keys that prohibit them

Boyce-Codd Normal Form (BCNF): Left-hand side of every non-trivial FD is a key

- Every schema has BCNF decomposition: no redundant data locally on tables
- FDs may be lost during process: they cause redundant data globally across tables

Schema Design with Functional Dependencies

- Goal: Maintain data consistency under updates, with a minimum level of effort
- Strategy: Do not admit any redundant data value occurrence in any instance
- Tactics: Transform FDs, that cause redundancy, into keys that prohibit them

Boyce-Codd Normal Form (BCNF): Left-hand side of every non-trivial FD is a key

- Every schema has BCNF decomposition: no redundant data locally on tables
- FDs may be lost during process: they cause redundant data globally across tables

Third Normal Form (3NF): The left-hand side of every non-trivial FD is a key or every attribute on the right-hand side must be part of some minimal key

- Every schema can be transformed into 3NF without loss of any FD
- Some FDs still cause redundant data locally as they were not morphed into keys
- Number of sources for data redundancy minimized, but
- No a priori bound on the level of data redundancy

State of Science: FDs vs Keys

$C(ity)$, $S(treet)$, $Z(IP)$ with FDs $CS \rightarrow Z$ and $Z \rightarrow C$

FDs $X \rightarrow Y$ and Keys X over Relation Schema R

All records with matching values on X have matching values on Y

Keys are special FDs $X \rightarrow R$

State of Science: FDs vs Keys

$C(ity)$, $S(treet)$, $Z(IP)$ with FDs $CS \to Z$ and $Z \to C$

FDs $X \to Y$ and Keys X over Relation Schema R

All records with matching values on X have matching values on Y
Keys are special FDs $X \to R$

C	S	Z
0	0	0
0	1	1
1	2	2
2	2	3
2	3	3

- Schema CSZ
- 2 keys CS and SZ
- FD $Z \to C$
- in 3NF
- data redundancy
- constraints local

State of Science: FDs vs Keys

C(ity), *S(treet)*, *Z(IP)* with FDs $CS \to Z$ and $Z \to C$

FDs $X \to Y$ and Keys X over Relation Schema R

All records with matching values on X have matching values on Y
Keys are special FDs $X \to R$

C	S	Z
0	0	0
0	1	1
1	2	2
2	2	3
2	3	3

- Schema CSZ
- 2 keys CS and SZ
- FD $Z \to C$
- in 3NF
- data redundancy
- constraints local

C	Z
0	0
0	1
1	2
2	3

S	Z
0	0
1	1
2	2
2	3
3	3

- CZ with key Z
- SZ with key SZ
- both in BCNF
- no data redundancy
- key CS on $CZ \bowtie SZ$

State of Science: Example

Relation Schema R: *E(vent)*, *M(anager)*, *S(tatus)*, *V(enue)*, *T(ime)*

Set $\mathcal{D}$ of Functional Dependencies

$VSE \rightarrow T$, $SET \rightarrow V$, $SME \rightarrow V$, $VS \rightarrow M$, $SME \rightarrow T$, $MT \rightarrow E$, and $ET \rightarrow M$

State of Science: Example

Relation Schema R: *E(vent)*, *M(anager)*, *S(tatus)*, *V(enue)*, *T(ime)*

Set $\mathcal{D}$ of Functional Dependencies

$VSE \to T$, $SET \to V$, $SME \to V$, $VS \to M$, $SME \to T$, $MT \to E$, and $ET \to M$

3NF Synthesis $\mathbb{D}_1$ of $(R, \mathcal{D})$

- $R_1 = ESTV$ and $\mathcal{D}_1$ with 3 keys EST, ESV, and STV
- $R_2 = EMT$ and $\mathcal{D}_2$ with 2 keys ET and MT
- $R_3 = EMSV$ and $\mathcal{D}_3$ with FD $VS \to M$ and 2 keys ESV and EMS

State of Science: Example

Relation Schema R: *E(vent)*, *M(anager)*, *S(tatus)*, *V(enue)*, *T(ime)*

Set $\mathcal{D}$ of Functional Dependencies

$VSE \to T$, $SET \to V$, $SME \to V$, $VS \to M$, $SME \to T$, $MT \to E$, and $ET \to M$

3NF Synthesis $\mathbb{D}_1$ of $(R, \mathcal{D})$

- $R_1 = ESTV$ and $\mathcal{D}_1$ with 3 keys EST, ESV, and STV
- $R_2 = EMT$ and $\mathcal{D}_2$ with 2 keys ET and MT
- $R_3 = EMSV$ and $\mathcal{D}_3$ with FD $VS \to M$ and 2 keys ESV and EMS

3NF Synthesis $\mathbb{D}_2$ of $(R, \mathcal{D})$

- $R_1 = ESTV$ and $\mathcal{D}_1$ with 3 keys EST, ESV, and STV
- $R_4 = EMST$ and $\mathcal{D}_4$ with 2 FDs $MT \to E$, $ET \to M$ & 3 keys EST, EMS, MST
- $R_5 = MSV$ and $\mathcal{D}_5$ with 1 key VS

State of Science: Example

3NF Synthesis $\mathbb{D}_1$ of $(R, \mathcal{D})$

- $R_1 = ESTV$ and $\mathcal{D}_1$ with 3 keys EST, ESV, and STV
- $R_2 = EMT$ and $\mathcal{D}_2$ with 2 keys ET and MT
- $R_3 = EMSV$ and $\mathcal{D}_3$ with FD $VS \rightarrow M$ and 2 keys ESV and EMS

3NF Synthesis $\mathbb{D}_2$ of $(R, \mathcal{D})$

- $R_1 = ESTV$ and $\mathcal{D}_1$ with 3 keys EST, ESV, and STV
- $R_4 = EMST$ and $\mathcal{D}_4$ with 2 FDs $MT \rightarrow E$, $ET \rightarrow M$ & 3 keys EST, EMS, MST
- $R_5 = MSV$ and $\mathcal{D}_5$ with 1 key VS

Which of $\mathbb{D}_1$ and $\mathbb{D}_2$ is better?

State of Science: Example

3NF Synthesis $\mathbb{D}_1$ of $(R, \mathcal{D})$

- $R_1 = ESTV$ and $\mathcal{D}_1$ with 3 keys EST, ESV, and STV
- $R_2 = EMT$ and $\mathcal{D}_2$ with 2 keys ET and MT
- $R_3 = EMSV$ and $\mathcal{D}_3$ with FD $VS \to M$ and 2 keys ESV and EMS

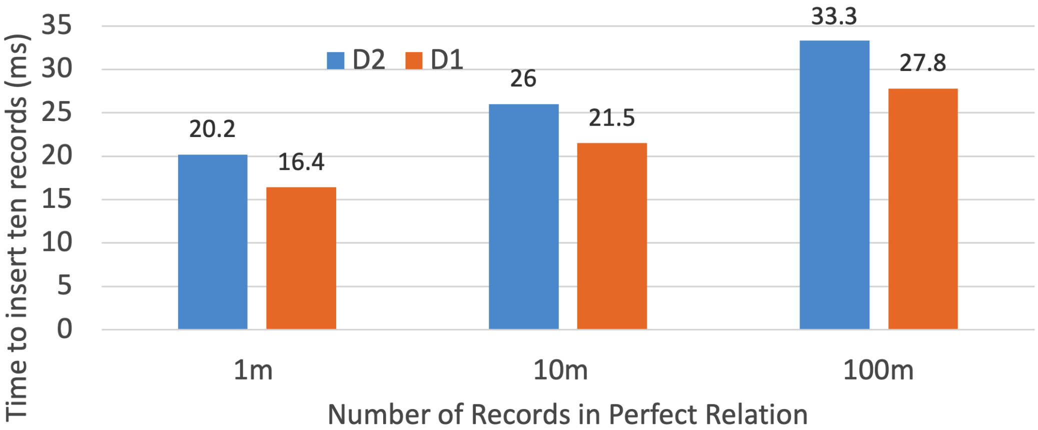
3NF Synthesis $\mathbb{D}_2$ of $(R, \mathcal{D})$

- $R_1 = ESTV$ and $\mathcal{D}_1$ with 3 keys EST, ESV, and STV
- $R_4 = EMST$ and $\mathcal{D}_4$ with 2 FDs $MT \to E$, $ET \to M$ & 3 keys EST, EMS, MST
- $R_5 = MSV$ and $\mathcal{D}_5$ with 1 key VS

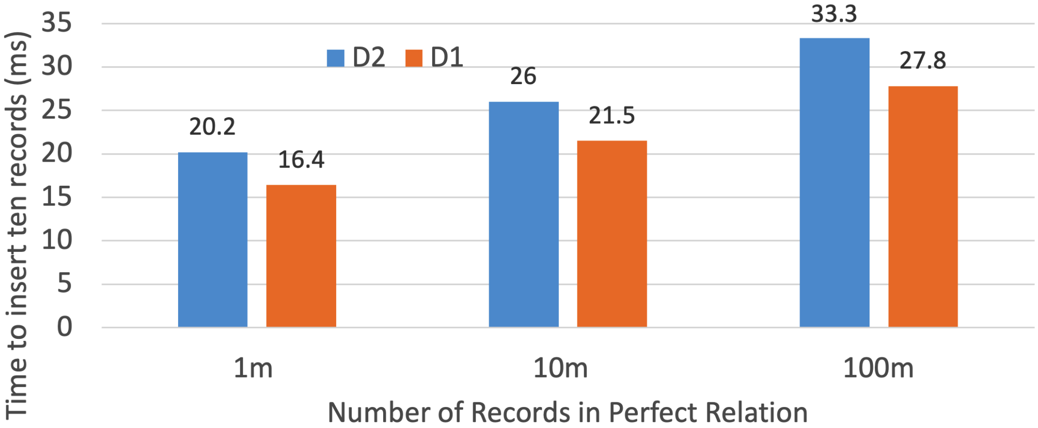
Which of $\mathbb{D}_1$ and $\mathbb{D}_2$ is better?

$\mathbb{D}_1$ better than $\mathbb{D}_2$ if schemata with fewer FDs are favored
$\mathbb{D}_2$ better than $\mathbb{D}_1$ if schemata with more keys are targeted

State of Science: Example



State of Science: Example



Operational performance favors fewer FDs
(these cause the bottleneck in integrity maintenance)

Core Idea

Core Idea

Logical Schema Design that Minimizes Operational Overheads

Opportunity: Classical ties between redundant 3NF schemata

Arbitrary choices when removing redundant schemata during classical 3NF synthesis

Core Idea

Logical Schema Design that Minimizes Operational Overheads

Opportunity: Classical ties between redundant 3NF schemata

Arbitrary choices when removing redundant schemata during classical 3NF synthesis

Parameterize 3NF to decide which

- redundant 3NF schemata to remove
- redundant BCNF schemata to remove

Parameters

- Numbers of (minimal) non-key FDs
- Numbers of (minimal) keys

Core Idea

Logical Schema Design that Minimizes Operational Overheads

Opportunity: Classical ties between redundant 3NF schemata

Arbitrary choices when removing redundant schemata during classical 3NF synthesis

Parameterize 3NF to decide which

- redundant 3NF schemata to remove
- redundant BCNF schemata to remove

Parameters

- Numbers of (minimal) non-key FDs
- Numbers of (minimal) keys

3NF synthesis with a strategy

Combine parameters to break ties between redundant 3NF schemata, for example minimize FD numbers in 3NF schemata & minimize key numbers in BCNF schemata

New Concepts

New Concepts: 3NF Sub-structures

$(R, \mathcal{D})$ denote a relation schema R with a set $\mathcal{D}$ of FDs over R

3NF sub-structure of $(R, \mathcal{D})$

A set $\mathcal{T}$ of keys and non-key prime FDs (FDs with prime attribute on the RHS):

$$\begin{aligned}\mathcal{T} \quad \subseteq \quad & \{X \subseteq R \mid X \to R \in \mathcal{D}^+\} \cup \\ & \{X \to Y \in \mathcal{D}^+ \mid (X \to R \notin \mathcal{D}^+) \wedge (Y - X \subseteq \mathcal{P})\}\end{aligned}$$

where

$$\mathcal{P} = \{A \in R \mid \exists K \to R \in \mathcal{D}^+ \wedge \forall K' \subset K (K' \to R \notin \mathcal{D}^+) \wedge A \in K\}$$

denotes the set of prime attributes for $\mathcal{D}$.

New Concepts: Intransitive Composite Object

$\mathcal{T}$ denotes a 3NF-substructure of $(R, \mathcal{D})$

$\mathcal{T}$ is an intransitive composite object for $(R, \mathcal{D})$ whenver

- 3NF update completeness holds:
  For all relations r over R that satisfy $\mathcal{D}$, for all $t \in dom(R)$, if $r \cup \{t\}$ satisfies $\mathcal{T}$, then $r \cup \{t\}$ satisfies $\mathcal{D}$.

Just validate the keys and FDs in any intransitive composite object for $\mathcal{D}$

New Concepts: Example

$R = \{E, M, S, T\}$ and $\mathcal{D} = \{ET \to MS, M \to E\}$

$(R, \mathcal{D})$ is in 3NF

The set of keys is $\{ET, MT\}$, $\mathcal{P} = EMT$, the only non-prime attribute is $R - \mathcal{P} = S$

New Concepts: Example

$R = \{E, M, S, T\}$ and $\mathcal{D} = \{ET \to MS, M \to E\}$

$(R, \mathcal{D})$ is in 3NF

The set of keys is $\{ET, MT\}$, $\mathcal{P} = EMT$, the only non-prime attribute is $R - \mathcal{P} = S$

$\mathcal{T}' = \{ET, MT, MS \to E\}$ is not an intransitive composite object for $\mathcal{D}$

$\mathcal{T} = \{ET, MT, M \to E\}$ is an intransitive composite object for $\mathcal{D}$

New Concepts: Example

$$R = \{E, M, S, T\} \text{ and } \mathcal{D} = \{ET \to MS, M \to E\}$$

$(R, \mathcal{D})$ is in 3NF

The set of keys is $\{ET, MT\}$, $\mathcal{P} = EMT$, the only non-prime attribute is $R - \mathcal{P} = S$

$\mathcal{T}' = \{ET, MT, MS \to E\}$ is not an intransitive composite object for $\mathcal{D}$

$\mathcal{T} = \{ET, MT, M \to E\}$ is an intransitive composite object for $\mathcal{D}$

$\mathcal{T}'$ is not an intransitive composite object

	Event	*Time*	*Manager*	*Status*
t' :	Workshop	21/11/2024	Sophie	approved
t :	Symposium	19/12/2025	Sophie	declined

$r = \{t'\}$ satisfies $\mathcal{D}$, and $r \cup \{t\}$ satisfies $\mathcal{T}'$ but not $\mathcal{D}$

New Concepts: The 3NF Core

The set of minimal keys implied by $\mathcal{D}$

$$\mathcal{K} = \{X \subseteq R \mid X \to R \in \mathcal{D}^+ \wedge \forall Z \subset X (Z \to R \notin \mathcal{D}^+)\}$$

New Concepts: The 3NF Core

The set of minimal keys implied by $\mathcal{D}$

$$\mathcal{K} = \{X \subseteq R \mid X \to R \in \mathcal{D}^+ \wedge \forall Z \subset X(Z \to R \notin \mathcal{D}^+)\}$$

The set of non-key minimal prime FDs implied by $\mathcal{D}$

$$\mathcal{F} = \{Z \to A \in \mathcal{D}^+ \mid (Z \to R \notin \mathcal{D}^+) \wedge (A \in \mathcal{P} - Z) \wedge (\forall Y \subset Z(Y \to A \notin \mathcal{D}^+))\}$$

New Concepts: The 3NF Core

The set of minimal keys implied by $\mathcal{D}$

$$\mathcal{K} = \{X \subseteq R \mid X \to R \in \mathcal{D}^+ \wedge \forall Z \subset X (Z \to R \notin \mathcal{D}^+)\}$$

The set of non-key minimal prime FDs implied by $\mathcal{D}$

$$\mathcal{F} = \{Z \to A \in \mathcal{D}^+ \mid (Z \to R \notin \mathcal{D}^+) \wedge (A \in \mathcal{P} - Z) \wedge (\forall Y \subset Z (Y \to A \notin \mathcal{D}^+))\}$$

The 3NF-core of $\mathcal{D}$

$$\mathcal{K} \cup \mathcal{F}$$

(provides access to parameters such as the number $k = |\mathcal{K}|$ of minimal keys, and we can minimize $f = |\mathcal{F}|$ by some suitable FD cover, such as a minimal-reduced cover)

New Concepts: Example

$$R = \{E, M, S, T\} \text{ and } \mathcal{D} = \{ET \rightarrow MS, M \rightarrow E\}$$

$\mathcal{T}_c = \mathcal{K} \cup \mathcal{F}$ forms the 3NF-core of $\mathcal{D}$ where $\mathcal{K} = \{ET, MT\}$ and $\mathcal{F} = \{M \rightarrow E\}$

New Concepts: Intransitive Composite Object Normal Form

Intransitive composite object normal form

$(R, \mathcal{D})$ is in *intransitive Composite Object Normal Form* (*iCONF*) if and only if the 3NF-core of $\mathcal{D}$ is an intransitive composite object for $\mathcal{D}$.

New Concepts: Intransitive Composite Object Normal Form

Intransitive composite object normal form

$(R, \mathcal{D})$ is in *intransitive Composite Object Normal Form* (*iCONF*) if and only if the 3NF-core of $\mathcal{D}$ is an intransitive composite object for $\mathcal{D}$.

For every cover $\mathcal{D}'$ of $\mathcal{D}$, $(R, \mathcal{D})$ is in iCONF iff $(R, \mathcal{D}')$ is in iCONF

New Concepts: Intransitive Composite Object Normal Form

Intransitive composite object normal form

$(R, \mathcal{D})$ is in *intransitive Composite Object Normal Form* (*iCONF*) if and only if the 3NF-core of $\mathcal{D}$ is an intransitive composite object for $\mathcal{D}$.

For every cover $\mathcal{D}'$ of $\mathcal{D}$, $(R, \mathcal{D})$ is in iCONF iff $(R, \mathcal{D}')$ is in iCONF

Theorem

$(R, \mathcal{D})$ is in 3NF if and only if $(R, \mathcal{D})$ is in iCONF

The 3NF-core $\mathcal{K} \cup \mathcal{F}$ separates $\mathcal{D}$ into its set $\mathcal{K}$ of minimal keys and its set $\mathcal{F}$ of non-key minimal prime FDs, subject to minimization by minimal-reduced FD covers

New Concepts: Example

- $R_3 = EMSV$ and $\mathcal{D}_3$ with FD $VS \to M$ and 2 keys ESV and EMS
- $R_4 = EMST$ and $\mathcal{D}_4$ with 2 FDs $MT \to E$, $ET \to M$ & 3 keys EST, EMS, MST

Do we prefer $(R_3, \mathcal{D}_3)$ or $(R_4, \mathcal{D}_4)$?

New Concepts: Example

- $R_3 = EMSV$ and $\mathcal{D}_3$ with FD $VS \rightarrow M$ and 2 keys ESV and EMS
- $R_4 = EMST$ and $\mathcal{D}_4$ with 2 FDs $MT \rightarrow E$, $ET \rightarrow M$ & 3 keys EST, EMS, MST

Do we prefer $(R_3, \mathcal{D}_3)$ or $(R_4, \mathcal{D}_4)$?

If we prefer to have fewer FDs, we may pick $(R_3, \mathcal{D}_3)$ over $(R_4, \mathcal{D}_4)$, but if we prefer to have more keys, then we may pick $(R_4, \mathcal{D}_4)$ over $(R_3, \mathcal{D}_3)$

New Concepts: k-CONF subsumed as special case of $(k, 0)$-iCONF

For a schema $(R, \mathcal{D})$ in 3NF, the following are equivalent:

1. The 3NF-core of $\mathcal{D}$ over R is covered by $\mathcal{K}$
2. $(R, \mathcal{D})$ is in BCNF with $k = |\mathcal{K}|$ minimal keys
3. $(R, \mathcal{D})$ is in CONF of level k

Optimizing 3NF Synthesis

Parameterized Third Normal Form

$(R, \mathcal{D})$ is in *(k, f)-3NF* iff $(R, \mathcal{D})$ is in 3NF and there is some minimal-reduced cover $(\mathcal{K}, \mathcal{F})$ for the 3NF-core of $(R, \mathcal{D})$ where $\mathcal{K}$ has cardinality k and $\mathcal{F}$ has cardinality f

Parameterized Third Normal Form

$(R, \mathcal{D})$ is in (k, f)-*3NF* iff $(R, \mathcal{D})$ is in 3NF and there is some minimal-reduced cover $(\mathcal{K}, \mathcal{F})$ for the 3NF-core of $(R, \mathcal{D})$ where $\mathcal{K}$ has cardinality k and $\mathcal{F}$ has cardinality f

- $<_{O'\text{-}BCNF}$: ranking resulting from an order *O'-BCNF* where $f = 0$
- $<_{O''\text{-}3NF}$: ranking resulting from an order *O''-3NF* where $f > 0$
- $<_{O\text{-}BCNF/3NF}$: least preferred of former precedes most preferred of latter

Parameterized Third Normal Form

$(R, \mathcal{D})$ is in *(k, f)-3NF* iff $(R, \mathcal{D})$ is in 3NF and there is some minimal-reduced cover $(\mathcal{K}, \mathcal{F})$ for the 3NF-core of $(R, \mathcal{D})$ where $\mathcal{K}$ has cardinality k and $\mathcal{F}$ has cardinality f

- $<_{O'\text{-}BCNF}$: ranking resulting from an order *O'-BCNF* where $f = 0$
- $<_{O''\text{-}3NF}$: ranking resulting from an order *O''-3NF* where $f > 0$
- $<_{O\text{-}BCNF/3NF}$: least preferred of former precedes most preferred of latter

Example

For *O'-BCNF* $= <_k$ and *O''-3NF* $= (<_f, >_k)$ we obtain merged ranking $<_{O\text{-}BCNF/3NF}$:

$$1 < \ldots < k < (1, k_1) < \ldots < (1, 2) < \ldots < (f, k_f) < \ldots < (f, 2)$$

(critical schemata have at least two different keys)

Parameterized Third Normal Form

$(R, \mathcal{D})$ is in *(k, f)-3NF* iff $(R, \mathcal{D})$ is in 3NF and there is some minimal-reduced cover $(\mathcal{K}, \mathcal{F})$ for the 3NF-core of $(R, \mathcal{D})$ where $\mathcal{K}$ has cardinality k and $\mathcal{F}$ has cardinality f

- $<_{O'\text{-}BCNF}$: ranking resulting from an order *O'-BCNF* where $f = 0$
- $<_{O''\text{-}3NF}$: ranking resulting from an order *O''-3NF* where $f > 0$
- $<_{O\text{-}BCNF/3NF}$: least preferred of former precedes most preferred of latter

Example

For *O'-BCNF* $= <_k$ and *O''-3NF* $=$ $(<_f, >_k)$ we obtain merged ranking $<_{O\text{-}BCNF/3NF}$:

$$1 < \ldots < k < (1, k_1) < \ldots < (1, 2) < \ldots < (f, k_f) < \ldots < (f, 2)$$

(critical schemata have at least two different keys)

Parameterized Third Normal Form

$(R, \mathcal{D})$ is in *(k, f)-3NF* iff $(R, \mathcal{D})$ is in 3NF and there is some minimal-reduced cover $(\mathcal{K}, \mathcal{F})$ for the 3NF-core of $(R, \mathcal{D})$ where $\mathcal{K}$ has cardinality k and $\mathcal{F}$ has cardinality f

- $<_{\textit{O'-BCNF}}$: ranking resulting from an order *O'-BCNF* where $f = 0$
- $<_{\textit{O''-3NF}}$: ranking resulting from an order *O''-3NF* where $f > 0$
- $<_{\textit{O-BCNF/3NF}}$: least preferred of former precedes most preferred of latter

Example

For *O'-BCNF* $= <_k$ and *O''-3NF* $= (<_f, >_k)$ we obtain merged ranking $<_{\textit{O-BCNF/3NF}}$:

$$1 < \ldots < k < (1, k_1) < \ldots < (1, 2) < \ldots < (f, k_f) < \ldots < (f, 2)$$

(critical schemata have at least two different keys)

Parameterized Third Normal Form

$(R, \mathcal{D})$ is in *(k, f)-3NF* iff $(R, \mathcal{D})$ is in 3NF and there is some minimal-reduced cover $(\mathcal{K}, \mathcal{F})$ for the 3NF-core of $(R, \mathcal{D})$ where $\mathcal{K}$ has cardinality k and $\mathcal{F}$ has cardinality f

- $<_{O\text{'-}BCNF}$: ranking resulting from an order *O'-BCNF* where $f = 0$
- $<_{O\text{''-}3NF}$: ranking resulting from an order *O''-3NF* where $f > 0$
- $<_{O\text{-}BCNF/3NF}$: least preferred of former precedes most preferred of latter

Example

For *O'-BCNF* $= <_k$ and *O''-3NF* $= (<_f, >_k)$ we obtain merged ranking $<_{O\text{-}BCNF/3NF}$:

$$1 < \ldots < k < (1, k_1) < \ldots < (1, 2) < \ldots < (f, k_f) < \ldots < (f, 2)$$

(critical schemata have at least two different keys)

Comparing Schemata in Third Normal Form

3NF-rank r_O^R of $(R, \mathcal{D})$

- $(R, \mathcal{D})$ in 3NF
- ranking $<_O$ for a finite order O
- r_O^R is the smallest rank of any (k, f) in $<_O$ for which $(R, \mathcal{D})$ is in (k, f)-3NF

Comparing Schemata in Third Normal Form

3NF-rank r_O^R of $(R, \mathcal{D})$

- $(R, \mathcal{D})$ in 3NF
- ranking $<_O$ for a finite order O
- r_O^R is the smallest rank of any (k, f) in $<_O$ for which $(R, \mathcal{D})$ is in (k, f)-3NF

order $<_O^R$ on 3NF schemata

$$(R, \mathcal{D}) <_O^R (R', \mathcal{D}') \text{ if and only if } r_O^R <_O r_O^{R'}$$

Running Example Formalized

- $R_3 = EMSV$ and $\mathcal{D}_3$ with FD $VS \to M$ and 2 keys ESV and EMS
- $R_4 = EMST$ and $\mathcal{D}_4$ with 2 FDs $MT \to E$, $ET \to M$ & 3 keys EST, EMS, MST

$O = <_f$ with ranking $1 < 2$

- minimizes the number of non-key FDs
- Since $r_O^{R_3} = 1$ and $r_O^{R_4} = 2$, we obtain $(R_3, \mathcal{D}_3) <_O^R (R_4, \mathcal{D}_4)$

Running Example Formalized

- $R_3 = EMSV$ and $\mathcal{D}_3$ with FD $VS \rightarrow M$ and 2 keys ESV and EMS
- $R_4 = EMST$ and $\mathcal{D}_4$ with 2 FDs $MT \rightarrow E$, $ET \rightarrow M$ & 3 keys EST, EMS, MST

$O = <_f$ with ranking $1 < 2$

- minimizes the number of non-key FDs
- Since $r_O^{R_3} = 1$ and $r_O^{R_4} = 2$, we obtain $(R_3, \mathcal{D}_3) <_O^R (R_4, \mathcal{D}_4)$

$O' = >_k$ with ranking $3 < 2$

- maximizes the number of minimal keys
- Since $r_{O'}^{R_3} = 2$ and $r_{O'}^{R_4} = 3$, we obtain $(R_4, \mathcal{D}_4) <_{O'}^R (R_3, \mathcal{D}_3)$

Computational Complexity of Parameterized Synthesis

PARAMETERISED 3NF	
Input:	$(R, \mathcal{D})$, non-negative integers k and f
Problem:	Decide if $(R, \mathcal{D})$ is in (k, f)-3NF

PARAMETERISED 3NF WITH KEYS	
Input:	$(R, \mathcal{D})$, non-negative integer f
	Set $\mathcal{K}$ of minimal keys for $\mathcal{D}$
Problem:	Decide if $(R, \mathcal{D})$ is in (k, f)-3NF

PARAMETERISED 3NF DESIGN	
Input:	$(R, \mathcal{D})$, non-negative integers k and f
	Set $S \subseteq R$
Problem:	Decide whether $(S, \mathcal{D}[S])$ is in (k, f)-3NF

Computational Complexity of Parameterized Synthesis

PARAMETERISED 3NF	
Input:	$(R, \mathcal{D})$, non-negative integers k and f
Problem:	Decide if $(R, \mathcal{D})$ is in (k, f)-3NF

PARAMETERISED 3NF WITH KEYS	
Input:	$(R, \mathcal{D})$, non-negative integer f
	Set $\mathcal{K}$ of minimal keys for $\mathcal{D}$
Problem:	Decide if $(R, \mathcal{D})$ is in (k, f)-3NF

PARAMETERISED 3NF DESIGN	
Input:	$(R, \mathcal{D})$, non-negative integers k and f
	Set $S \subseteq R$
Problem:	Decide whether $(S, \mathcal{D}[S])$ is in (k, f)-3NF

Theorem

1. PARAMETERISED 3NF *is NP-complete*
2. PARAMETERISED 3NF WITH KEYS *is polynomial*
3. PARAMETERISED 3NF DESIGN *is NP-complete* $\square$

Comparing 3NF Decompositions

D a set of lossless, dependency-preserving 3NF decompositions of $(R, \mathcal{D})$ in 3NF

Comparing 3NF Decompositions

D a set of lossless, dependency-preserving 3NF decompositions of $(R, \mathcal{D})$ in 3NF

- ranking $<_{O\text{-}BCNF/3NF}$ spans all values of parameters that occur in some $\mathbb{D} \in \mathbf{D}$
- for $\mathbb{D} \in \mathbf{D}$, and for rank r in $<_{O\text{-}BCNF/3NF}$

$$\mathcal{S}_r^{\mathbb{D}} = \{(S, \mathcal{D}[S]) \in \mathbb{D} \mid (S, \mathcal{D}[S]) \text{ has rank } r_O^S = r \text{ in } <_{O\text{-}BCNF/3NF}\}$$

Comparing 3NF Decompositions

D a set of lossless, dependency-preserving 3NF decompositions of $(R, \mathcal{D})$ in 3NF

- ranking $<_{O\text{-}BCNF/3NF}$ spans all values of parameters that occur in some $\mathbb{D} \in \mathbf{D}$
- for $\mathbb{D} \in \mathbf{D}$, and for rank r in $<_{O\text{-}BCNF/3NF}$

$$\mathcal{S}_r^{\mathbb{D}} = \{(S, \mathcal{D}[S]) \in \mathbb{D} \mid (S, \mathcal{D}[S]) \text{ has rank } r_O^S = r \text{ in } <_{O\text{-}BCNF/3NF}\}$$

For $\mathbb{D}', \mathbb{D}'' \in \mathbf{D}$, $\mathbb{D}'$ is *D-better* than $\mathbb{D}''$ ($\mathbb{D}' <^{D}_{O\text{-}BCNF/3NF} \mathbb{D}''$) if and only if for the worst rank r in $<_{O\text{-}BCNF/3NF}$ where $\mathcal{S}_r^{\mathbb{D}'} \neq \mathcal{S}_r^{\mathbb{D}''}$, $\mathcal{S}_r^{\mathbb{D}'} = \emptyset$

Running Example

3NF Synthesis $\mathbb{D}_1$ of $(R, \mathcal{D})$

- R_1 and $\mathcal{D}_1$ with 3 keys
- R_2 and $\mathcal{D}_2$ with 2 keys
- R_3 and $\mathcal{D}_3$ with 1 FD and 2 keys

3NF Synthesis $\mathbb{D}_2$ of $(R, \mathcal{D})$

- R_1 and $\mathcal{D}_1$ with 3 keys
- R_4 and $\mathcal{D}_4$ with 2 FDs and 3 keys
- R_5 and $\mathcal{D}_5$ with 1 key

$O\text{-}BCNF = <_k$ and $O\text{-}3NF = <_f$ with ranking $<_{O\text{-}BCNF/3NF}$: $1 < 2 < 3 << 1 < 2$

$\mathbb{D}$	$\mathcal{S}_1^{\mathbb{D}}$	$\mathcal{S}_2^{\mathbb{D}}$	$\mathcal{S}_3^{\mathbb{D}}$	$\mathcal{S}_4^{\mathbb{D}}$	$\mathcal{S}_5^{\mathbb{D}}$
$\mathbb{D}_1$	$\emptyset$	$\{R_2\}$	$\{R_1\}$	$\{R_3\}$	$\emptyset$
$\mathbb{D}_2$	$\{R_5\}$	$\emptyset$	$\{R_1\}$	$\emptyset$	$\{R_4\}$

- worst rank on which $\mathbb{D}_1$ and $\mathbb{D}_2$ have different schemata is rank 5
- as $|\mathcal{S}_5^{\mathbb{D}_1}| \neq |\mathcal{S}_5^{\mathbb{D}_2}|$ and $\mathcal{S}_5^{\mathbb{D}_1} = \emptyset$, we have $\mathbb{D}_1 <^{D}_{O\text{-}BCNF/3NF} \mathbb{D}_2$

Parameterized Synthesis

Require: $(R, \mathcal{D})$ with FD set $\mathcal{D}$ over schema R, 3NF-order *O-3NF*, BCNF-order *O-BCNF*
Ensure: Lossless, FD-preserving 3NF decomposition $\mathbb{D}$ of $(R, \mathcal{D})$ that is $<^{D}_{\textit{O-BCNF/3NF}}$-optimal

1. Compute f and k for all critical schemata (those in 3NF but not BCNF)
2. In reverse $<_{\textit{O-3NF}}$-ranks, remove the critical schema if it is redundant, otherwise add to $\mathbb{D}$
3. Compute k for all BCNF schemata
4. In reverse $<_{\textit{O-BCNF}}$-ranks, remove the BCNF schema if it is redundant, otherwise add to $\mathbb{D}$
5. Remove any schemata in $\mathbb{D}$ if they are subsumed by others
6. Add a schema that contains some global minimal key
7. **Return**($\mathbb{D}$)

Parameterized Synthesis

Require: $(R, \mathcal{D})$ with FD set $\mathcal{D}$ over schema R, 3NF-order *O-3NF*, BCNF-order *O-BCNF*
Ensure: Lossless, FD-preserving 3NF decomposition $\mathbb{D}$ of $(R, \mathcal{D})$ that is $<^{D}_{\textit{O-BCNF/3NF}}$-optimal

1: Compute f and k for all critical schemata (those in 3NF but not BCNF)
2: In reverse $<_{\textit{O-3NF}}$-ranks, remove the critical schema if it is redundant, otherwise add to $\mathbb{D}$
3: Compute k for all BCNF schemata
4: In reverse $<_{\textit{O-BCNF}}$-ranks, remove the BCNF schema if it is redundant, otherwise add to $\mathbb{D}$
5: Remove any schemata in $\mathbb{D}$ if they are subsumed by others
6: Add a schema that contains some global minimal key
7: **Return**($\mathbb{D}$)

Theorem

On input $(R, \mathcal{D}, \textit{O-3NF}, \textit{O-BCNF})$, the algorithm returns a lossless, dependency-preserving decomposition into 3NF that is $<^{D}_{\textit{O-BCNF/3NF}}$-optimal. □

Running Example

Input

$VSE \to T$, $SET \to V$, $SME \to V$, $VS \to M$, $SME \to T$, $MT \to E$, and $ET \to M$
$O_{3NF} = <_f$ and $O_{BCNF} = <_k$

Atomic cover

add $MST \to V$, $STV \to E$

Schemata generated

- $R_1 = ESTV$ and $\mathcal{D}_1$ with 3 keys EST, ESV, and STV
- $R_2 = EMT$ and $\mathcal{D}_2$ with 2 keys ET and MT
- $R_3 = EMSV$ and $\mathcal{D}_3$ with 1 FD $VS \to M$ and 2 keys ESV and EMS
- $R_4 = EMST$ and $\mathcal{D}_4$ with 2 FDs $MT \to E$, $ET \to M$ & 3 keys EST, EMS, MST
- $R_5 = MSV$ and $\mathcal{D}_5$ with 1 key VS
- $R_6 = MSTV$ and $\mathcal{D}_6$ with 1 FD $SV \to M$ and 2 keys STV and MST

Example continued

Schemata generated

- $R_3 = EMSV$ and $\mathcal{D}_3$ with 1 FD $VS \to M$ and 2 keys ESV and EMS
- $R_4 = EMST$ and $\mathcal{D}_4$ with 2 FDs $MT \to E$, $ET \to M$ & 3 keys EST, EMS, MST
- $R_6 = MSTV$ and $\mathcal{D}_6$ with 1 FD $SV \to M$ and 2 keys STV and MST

Critical Schemata

$(R_3, \mathcal{D}_3) =_f^R (R_6, \mathcal{D}_6) <_f^R (R_4, \mathcal{D}_4)$

- R_4-generating FD $EMS \to T$ is redundant, so $(R_4, \mathcal{D}_4)$ is not required
- R_3-generating FD $EMS \to V$ is not redundant now, so the schema $(R_3, \mathcal{D}_3)$ is added to the decomposition.
- R_6-generating FD $MST \to V$ is still redundant, so the schema $(R_6, \mathcal{D}_6)$ is not required.

Schemata generated

- $R_1 = ESTV$ and $\mathcal{D}_1$ with 3 keys EST, ESV, and STV
- $R_2 = EMT$ and $\mathcal{D}_2$ with 2 keys ET and MT
- $R_5 = MSV$ and $\mathcal{D}_5$ with 1 key VS

BCNF Schemata

$$(R_5, \mathcal{D}_5) <_k^R (R_2, \mathcal{D}_2) <_k^R (R_1, \mathcal{D}_1)$$

- $(R_1, \mathcal{D}_1)$, $(R_2, \mathcal{D}_2)$, $(R_5, \mathcal{D}_5)$ are all added to the decomposition
- $R_5 \subseteq R_3$, so $(R_5, \mathcal{D}_5)$ is removed

$\mathbb{D}_1 = \{(R_1, \mathcal{D}_1), (R_2, \mathcal{D}_2), (R_3, \mathcal{D}_3)\}$ is $<^D_{O\text{-}3NF/BCNF}$-optimal

Experiments

Experimental Setup

Algorithms

Implemented in Java, Version 17.0.7, and run on a 12th Gen Intel(R) Core(TM) i7-12700, 2.10GHz, with 128GB RAM, 1TB SSD, and Windows 10

Data Sets

- FD sets mined from 12 real-world benchmark data plus TPC-H [a]
- Benchmarks for mining keys and FDs, but also used in previous work
- Compare to state-of-the-art and analyze on various FDs to test scalability

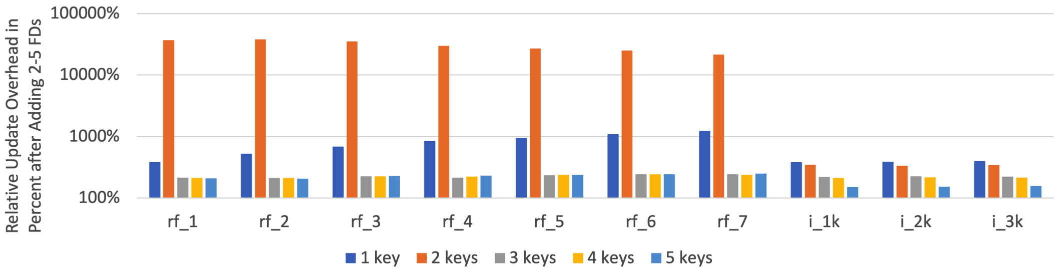
[a] `hpi.de/naumann/projects/repeatability/data-profiling/fds.html`

Comparison between...

- *iConf-fk*: $(<_f, >_k)$ and $<_k$
- *iConf-f*: $<_f$ and $<_k$
- *Conf*: only *O-BCNF*$= <_k$
- *BC-Cover*: Computes BCNF whenever possible
- *Synthesis*: Classical 3NF

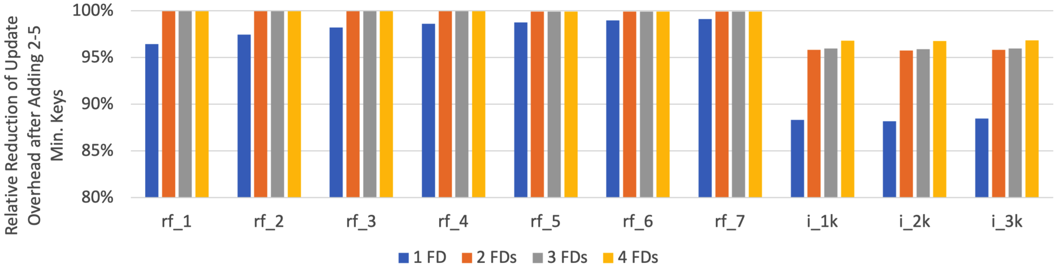
1. How do keys and FDs affect update and query performance?

Update Overheads From Adding More FDs



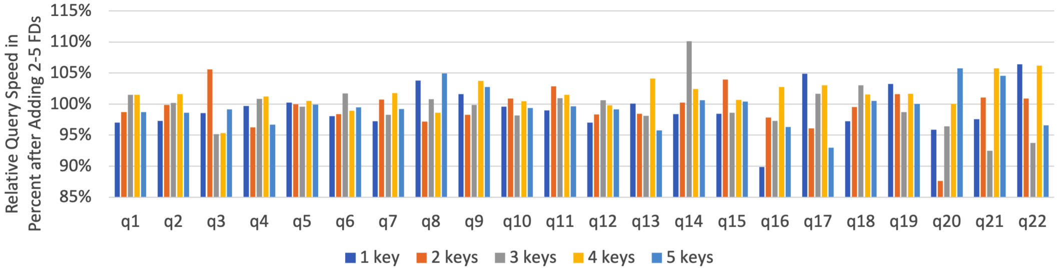
- The average overhead across the 7 refresh operations and all constraint sets is more than 6400%, and across the 3 inserts it is more than 264%.
- Having sufficiently many keys does scale update performance when more non-key FDs are present.

Reducing Update Overheads by Adding More Keys



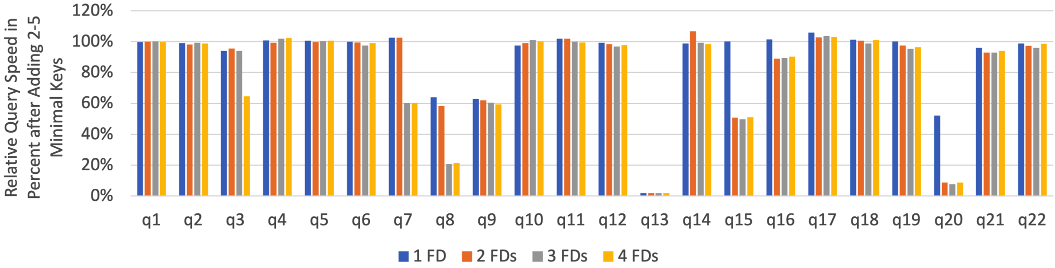
- The average reduction across the 7 refresh operations and all constraint sets is more than 99.4%, and across the 3 inserts it is more than 94%.
- Adding a few more FDs incurs huge update overheads, but having a few more minimal keys can scale integrity maintenance well.

Relative Query Speed after Adding More FDs



- The average speed across the 22 queries is just below 99.8%.
- While some queries are affected, on average there is little impact on query performance resulting from FDs.

Relative Query Speed after Adding More Keys



- The average speed is just below 83.6%, so a speed up of over 14.4%.
- UNIQUE indices resulting from arbitrary selections of minimal keys show impact

Takeaways

Update performance

- Slowed down significantly by FDs

  Infeasible to rely on FDs only for integrity maintenance

  Parameter f is a valuable parameter to minimize
- Scaled up significantly by keys

  Smaller k, fewer `UNIQUE` indices to maintain

  Need sufficient k on critical schemata

Query performance

- Unaffected by FDs

  Query optimization with FDs perhaps not well implemented yet
- Improved significantly by keys and their `UNIQUE` indices

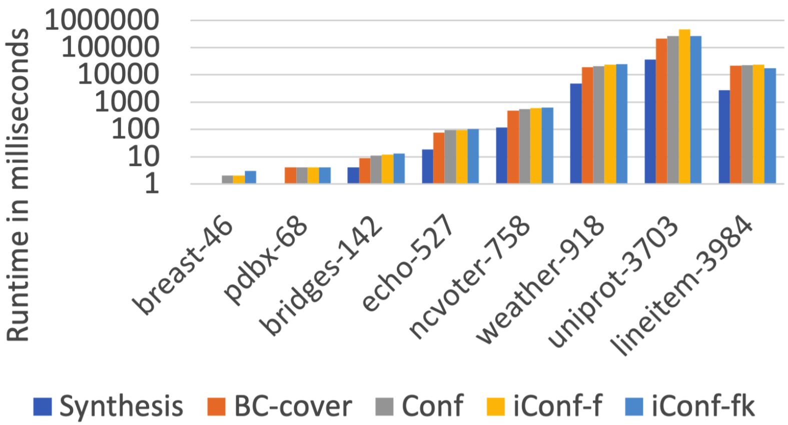
  Larger k, more options for query optimization

2. How well do our algorithms perform?

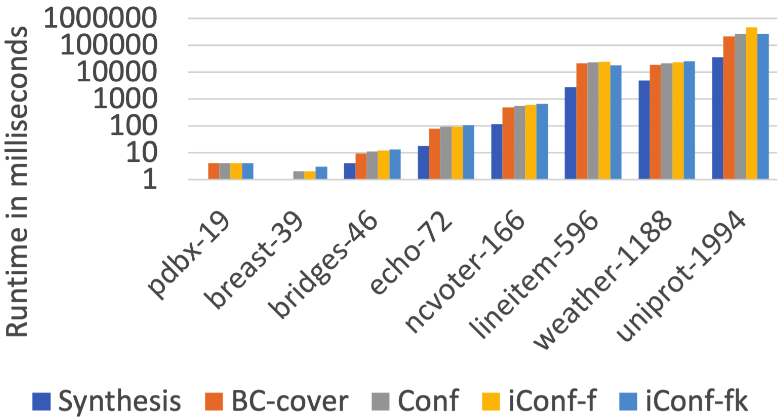
Runtime of Synthesis Algorithms

	Characteristics			Time of Algorithms (in ms)				
Data set	*#R*	*#C*	*#FD*	*Synthesis*	*BC-Cover*	*Conf*	*iConf-f*	*iConf-fk*
abalone	4,177	9	137	3	9	11	17	53
adult	48,842	14	78	2	4	5	5	6
breast	699	11	46	1	1	2	2	3
bridges	108	13	142	4	9	11	12	13
echo	132	13	527	18	77	93	94	105
hepatitis	155	20	8,250	2064	11,797	13,134	14,551	15,865
letter	20,000	17	61	4	6	7	7	8
lineitem	6,001,215	16	3,984	2,698	21,269	23,056	23,696	17,364
ncvoter	1,000	19	758	115	489	547	595	640
pdbx	17,305,799	13	68	1	4	4	4	4
uniprot	512,000	30	3,703	36,238	213,958	266,825	468,059	266,257
weather	262,920	18	918	4,796	18,824	20,925	23,184	25,140

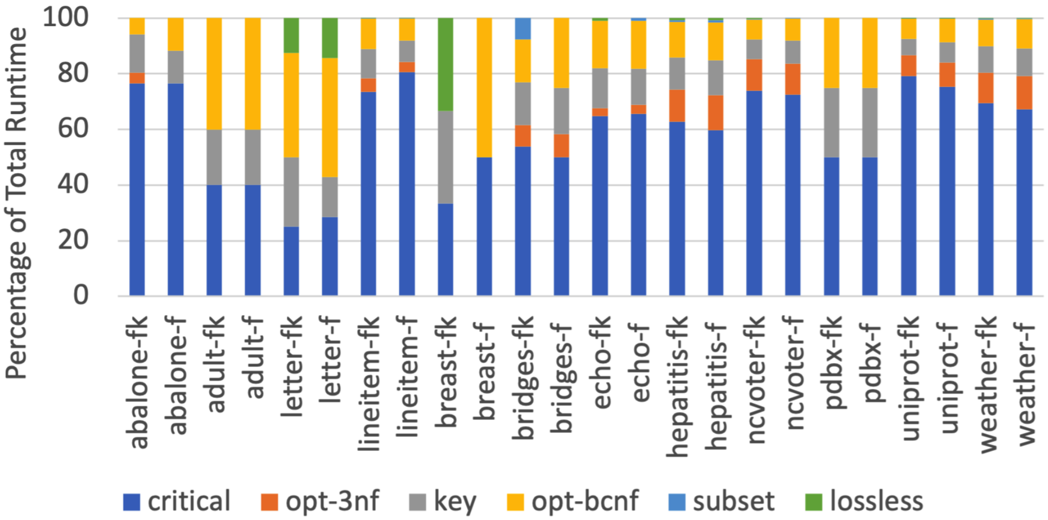
Runtime in Number of FDs



Runtime in Size of Output



Breakdown of Total Runtime



Output Analysis of Algorithm

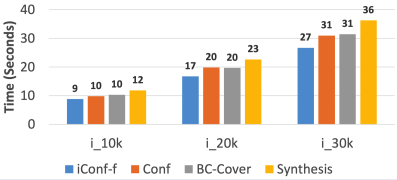
		Decomposition			Schema in BCNF		Schema in 3NF	
Data set	Alg	Size	BCNF	3NF	#Keys	Distribution	#FDs	Distribution
abalone	iConf-fk	26	22	4	1.64	[3:1,2:12,1:9]	2	[4:1,2:1,1:2]
	iConf-f	23	18	5	1.61	[2:11,1:7]	1.8	[4:1,2:1,1:3]
	Conf	21	16	5	1.81	[3:2,2:9,1:5]	2.4	[4:2,2:1,1:2]
	BC-Cover	20	15	5	2.07	[5:1,3:2,2:8,1:4]	2.4	[4:2,2:1,1:2]
	Synthesis	21	14	7	1.93	[3:2,2:9,1:3]	2.29	[4:2,3:1,2:1,1:3]
ncvoter	iConf-fk	166	145	21	1.19	[2:27,1:118]	1.19	[3:1,2:2,1:18]
	iConf-f	166	145	21	1.2	[3:1,2:27,1:117]	1.19	[3:1,2:2,1:18]
	Conf	168	147	21	1.2	[3:1,2:28,1:118]	1.29	[4:1,2:3,1:17]
	BC-Cover	162	141	21	1.28	[4:2,3:5,2:23,1:111]	1.29	[4:1,2:3,1:17]
	Synthesis	154	123	31	1.24	[4:1,3:3,2:20,1:99]	1.35	[4:1,2:8,1:22]
lineitem	iConf-fk	590	560	30	1.39	[15:1,10:1,6:3,5:3,4:5,3:23,2:105,1:419]	1.4	[4:1,2:9,1:20]
	iConf-f	596	567	29	1.39	[15:1,10:1,6:4,5:3,4:5,3:23,2:105,1:425]	1.41	[4:1,2:9,1:19]
	Conf	587	558	29	1.38	[15:1,10:1,6:3,5:3,4:5,3:22,2:104,1:419]	2.62	[10:2,9:1,5:1,4:1,3:2,2:10,1:12]
	BC-Cover	562	533	29	2.32	[15:1,11:1,10:2,9:9,8:9,7:19,6:26,5:26,4:26,3:26,2:46,1:342]	2.62	[10:2,9:1,5:1,4:1,3:2,2:10,1:12]
	Synthesis	531	466	65	2.29	[11:1,10:1,9:8,8:9,7:19,6:21,5:25,4:24,3:18,2:30,1:310]	2.28	[10:3,9:1,5:1,4:4,3:6,2:20,1:30]

3. How well do optimizations transcend operationally?

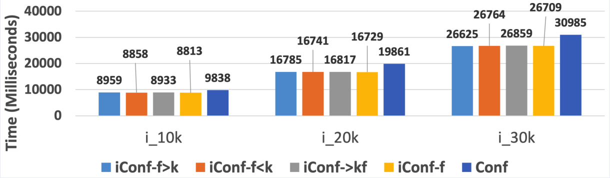
Abalone

		Decomposition			Schema in BCNF		Schema in 3NF	
Data set	Alg	Size	BCNF	3NF	#Keys	Distribution	#FDs	Distribution
abalone	iConf-fk	26	22	4	1.64	[3:1,2:12,1:9]	2	[4:1,2:1,1:2]
	iConf-f	23	18	5	1.61	[2:11,1:7]	1.8	[4:1,2:1,1:3]
	Conf	21	16	5	1.81	[3:2,2:9,1:5]	2.4	[4:2,2:1,1:2]
	BC-Cover	20	15	5	2.07	[5:1,3:2,2:8,1:4]	2.4	[4:2,2:1,1:2]
	Synthesis	21	14	7	1.93	[3:2,2:9,1:3]	2.29	[4:2,3:1,2:1,1:3]

abalone



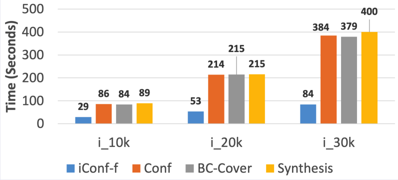
abalone (optimizations)



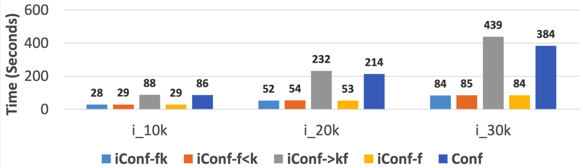
NCVoter

		Decomposition			Schema in BCNF		Schema in 3NF	
Data set	Alg	Size	BCNF	3NF	#Keys	Distribution	#FDs	Distribution
ncvoter	iConf-fk	166	145	21	1.19	[2:27,1:118]	1.19	[3:1,2:2,1:18]
	iConf-f	166	145	21	1.2	[3:1,2:27,1:117]	1.19	[3:1,2:2,1:18]
	Conf	168	147	21	1.2	[3:1,2:28,1:118]	1.29	[4:1,2:3,1:17]
	BC-Cover	162	141	21	1.28	[4:2,3:5,2:23,1:111]	1.29	[4:1,2:3,1:17]
	Synthesis	154	123	31	1.24	[4:1,3:3,2:20,1:99]	1.35	[4:1,2:8,1:22]

ncvoter



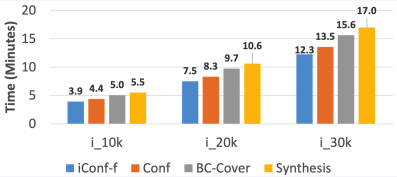
ncvoter (optimizations)



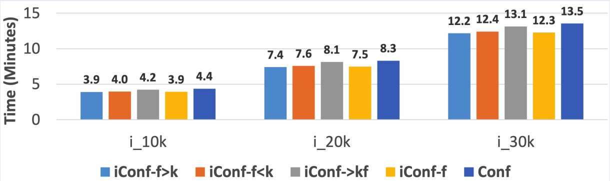
Lineitem

		Decomposition			Schema in BCNF		Schema in 3NF	
Data set	Alg	Size	BCNF	3NF	#Keys	Distribution	#FDs	Distribution
lineitem	iConf-fk	590	560	30	1.39	[15:1,10:1,6:3,5:3,4:5,3:23,2:105,1:419]	1.4	[4:1,2:9,1:20]
	iConf-f	596	567	29	1.39	[15:1,10:1,6:4,5:3,4:5,3:23,2:105,1:425]	1.41	[4:1,2:9,1:19]
	Conf	587	558	29	1.38	[15:1,10:1,6:3,5:3,4:5,3:22,2:104,1:419]	2.62	[10:2,9:1,5:1,4:1,3:2,2:10,1:12]
	BC-Cover	562	533	29	2.32	[15:1,11:1,10:2,9:9,8:9,7:19,6:26,5:26,4:26,3:26,2:46,1:342]	2.62	[10:2,9:1,5:1,4:1,3:2,2:10,1:12]
	Synthesis	531	466	65	2.29	[11:1,10:1,9:8,8:9,7:19,6:21,5:25,4:24,3:18,2:30,1:310]	2.28	[10:3,9:1,5:1,4:4,3:6,2:20,1:30]

lineitem



lineitem (optimizations)



Average reduction of overheads across algorithms

Comparison	total	per schema
iConf-f over *Conf*	20.0%	23.5%
Conf over *BC-Cover*	3.0%	6.2%
BC-Cover over *Synthesis*	5.7%	8.7%

Summary

Summary: Conclusion

Measure the Effort of Integrity Maintenance at the Logical Level

- Classify 3NF-schemata by the numbers k of keys and f of FDs they exhibit
- 3NF: all integrity constraints can be enforced by keys or prime FDs
- (k, f)-3NF: all integrity constraints can be enforced by k keys and f prime FDs

Summary: Conclusion

Measure the Effort of Integrity Maintenance at the Logical Level

- Classify 3NF-schemata by the numbers k of keys and f of FDs they exhibit
- 3NF: all integrity constraints can be enforced by keys or prime FDs
- (k, f)-3NF: all integrity constraints can be enforced by k keys and f prime FDs

Optimize Synthesis Strategically by Utilizing Parameters

- Improve 3NF Synthesis by breaking ties between redundant 3NF schemata
- Use combination of parameters k and f to break ties

Summary: Conclusion

Measure the Effort of Integrity Maintenance at the Logical Level

- Classify 3NF-schemata by the numbers k of keys and f of FDs they exhibit
- 3NF: all integrity constraints can be enforced by keys or prime FDs
- (k, f)-3NF: all integrity constraints can be enforced by k keys and f prime FDs

Optimize Synthesis Strategically by Utilizing Parameters

- Improve 3NF Synthesis by breaking ties between redundant 3NF schemata
- Use combination of parameters k and f to break ties

Logical Schema Design that Minimizes Operational Overheads

- Optimizations on logical level transcend to operational level
- Bottleneck of integrity maintenance is minimized
- Achieved by separating non-key FDs from keys

Summary: Future Work

Use the Size of Keys and FDs

- Total number of attribute occurrences in keys and FDs
- Requires optimal instead of minimal-reduced covers
- Optimal covers potentially inefficient but much better (prize of optimality)

Summary: Future Work

Use the Size of Keys and FDs

- Total number of attribute occurrences in keys and FDs
- Requires optimal instead of minimal-reduced covers
- Optimal covers potentially inefficient but much better (prize of optimality)

Interaction with Other Constraints

- MVDs (4NF) and JDs (5NF): What are parameters and notions of covers here?
- CCs: Bounded synthesis to minimize worst-case levels of data redundancy
- INDs (IDNF): What is the impact of referential integrity?

Summary: Future Work

Use the Size of Keys and FDs

- Total number of attribute occurrences in keys and FDs
- Requires optimal instead of minimal-reduced covers
- Optimal covers potentially inefficient but much better (prize of optimality)

Interaction with Other Constraints

- MVDs (4NF) and JDs (5NF): What are parameters and notions of covers here?
- CCs: Bounded synthesis to minimize worst-case levels of data redundancy
- INDs (IDNF): What is the impact of referential integrity?

Other Dimensions of Data Quality and Data Models

- Variety, Veracity, Velocity
- JSON, Graphs, Object-Stores

Summary: References

1. Marcelo Arenas: Normalization theory for XML. SIGMOD Rec. 35(4): 57-64 (2006)
2. Philip A. Bernstein: Synthesizing Third Normal Form Relations from Functional Dependencies. ACM Trans. Database Syst. 1(4): 277-298 (1976)
3. Philip A. Bernstein, Nathan Goodman: What does Boyce-Codd Normal Form Do? VLDB 1980: 245-259
4. Joachim Biskup, Umeshwar Dayal, Philip A. Bernstein: Synthesizing Independent Database Schemas. SIGMOD Conference 1979: 143-151
5. Joachim Biskup: Boyce-Codd and Object Normal Forms. Inf. Process. Lett. 32(1):29–33 (1989)
6. E. F. Codd: Recent Investigations in Relational Data Base Systems. IFIP Congress 1974: 1017-1021
7. Ronald Fagin: A Normal Form for Relational Databases That Is Based on Domains and Keys. ACM Trans. Database Syst. 6(3): 387-415 (1981)
8. I. J. Heath: Unacceptable File Operations in a Relational Data Base. SIGFIDET Workshop 1971: 19-33
9. Solmaz Kolahi, Leonid Libkin: An information-theoretic analysis of worst-case redundancy in database design. ACM Trans. Database Syst. 35(1): 5:1-5:32 (2010)
10. C. Lucchesi, S. Osborn: Candidate Keys for Relations. J. Comput. Syst. Sci. 17(2): 270-279 (1978)
11. Sylvia Osborn: Testing for Existence of a Covering BCNF. Inf. Process. Lett. 8(1): 11-14 (1979)
12. Zhuoxing Zhang, Wu Chen, Sebastian Link: Composite Object Normal Forms - Parameterizing BCNF by the Number of Minimal Keys. Proc. ACM Manag. Data 1(1): 13:1-13:25 (2023)