ELSEVIER

Information Processing Letters 62 (1997) 255–258

Information Processing Letters

On keys and normal forms

Wai Yin Mok *

Brigham Young University, 590 Wymount Terrace, Provo, UT 84604, USA

Received 18 July 1995; revised 28 March 1996
Communicated by G.R. Andrews

Abstract

In this paper, we give a necessary and sufficient condition for a Boyce–Codd Normal Form (BCNF) relation scheme to be in Fourth Normal Form (4NF). We also give a necessary and sufficient condition for a 4NF relation scheme to be in Projection-Join Normal Form (PJNF). From these results, we derive necessary and sufficient conditions for a BCNF relation scheme to be in PJNF. In addition, we give a less stringent condition for a Third Normal Form (3NF) relation scheme to be in BCNF. Finally, we show that our theorems generalize the results by Date and Fagin (1992). © 1997 Published by Elsevier Science B.V.

Keywords: Databases; Relational databases; Normalization; Third Normal Form; Boyce–Codd Normal Form; Fourth Normal Form; Projection-Join Normal Form

1. Introduction

In [1], Date and Fagin derived some results that give the database designer sufficient conditions, defined in terms of *functional dependencies* (FDs) alone, that guarantee that a relation scheme being designed is automatically in a higher normal form. However, their conditions are too stringent. To find less stringent sufficient conditions for a lower normal form to be in a higher normal form is the purpose of this paper. In particular, we consider 3NF, BCNF, 4NF, and PJNF relation schemes and give conditions defined in terms of keys to guarantee that a lower normal form is in a higher normal form. In addition, for BCNF, 4NF, and PJNF relation schemes, the conditions that are sufficient for a lower normal form to be in a higher normal form are also necessary.

We present our arguments as follows. In Section 2, we give the basic definitions and results needed in this paper. These definitions and results are common and most can be found in [2]. In Section 3, we prove our theorems, and in Section 4, we show that the lemmas and theorems in [1] follow from our theorems. We conclude in Section 5.

2. Preliminaries

A *relation scheme* R is a finite set of attributes. An FD $X \to Y$ of R is an FD where $XY \subseteq R$. An FD $X \to Y$ is *nontrivial* if Y is not a subset of X. A *join dependency* (JD) $J = \bowtie\{R_1, \ldots, R_m\}$ of R is a

* Present address: Department of Mathematical Sciences, The University of Akron, Akron, OH 44325-4002, USA. Email: wmok@cs.uakron.edu.

PII S0020-0190(97)00076-8

JD where $\bigcup_{i=1}^{m} R_i = R$. Each R_i of J is a *component* of J. A *nontrivial* JD has no component equal to R. An *multivalued dependency* (MVD) is a special JD that has only two components. An MVD $\bowtie\{R_1, R_2\}$ can also be written as $(R_1 \cap R_2) \twoheadrightarrow (R_2 - R_1)$.

A *key* K of a relation scheme R is a subset of R such that $K \to R$ and there is no proper subset K' of K such that $K' \to R$. A key is *simple* if it consists of only one attribute. An attribute A is a *key attribute* if A is part of a key.

A relation scheme R is in 3NF if whenever $X \to A$ is a nontrivial FD of R, $X \to R$ or A is a key attribute. R is in BCNF if whenever $X \to A$ is a nontrivial FD of R, $X \to R$. R is in 4NF if whenever $X \twoheadrightarrow Y$ is a nontrivial MVD of R, $X \to R$. It has also been shown that R is in 4NF if and only if every MVD of R follows from the key dependencies of R [1]. R is in PJNF if every JD of R follows from the key dependencies of R.

A JD $J = \bowtie\{R_1, \ldots, R_m\}$ implies many MVDs. However, it is sufficient to consider the MVDs in $MVD(J)$ where $MVD(J)$ is the set of MVDs of the form $\bowtie\{\bigcup_{i \in S_1} R_i, \bigcup_{i \in S_2} R_i\}$ where $S_1 \cup S_2 = \{1, \ldots, m\}$, and $S_1 \cap S_2 = \emptyset$ [2].

We use Theorem 13.2 in [2], which states that the following conditions on a JD J are equivalent:

(1) J is acyclic.

(2) $MVD(J)$ implies J.

Note that if there is a component of J equal to R, then J is trivial and thus J is acyclic since $MVD(J)$ implies J vacuously.

To test whether the key dependencies of a relation scheme R imply a JD of R, we use the following Membership Algorithm [1]. The input is a JD J and the keys $K_1, \ldots, K_n$ of R. Initialize a JD J' as J. Apply the following rule until it can be no longer applied: if key K_i of R is a subset of $Y \cap Z$ for some components Y and Z of J', remove the components Y and Z from J' and add a new component YZ to J'. The final J' is the output of the Membership Algorithm and is denoted by $OMA(R, J)$. The key dependencies of R imply J if and only if there is a component of $OMA(R, J)$ equal to R. Note that the key dependencies of R and $OMA(R, J)$ together always imply J [2].

We shall use an inference rule, which is called C2 in [2], in this paper.

(C2) If $X \twoheadrightarrow Y$ and $Z \to W$, where $W \subseteq Y$ and $Y \cap Z = \emptyset$, then $X \to W$.

3. The main results

In this section, we prove several theorems. In the next section, we show that the lemmas and theorems in [1] follow from these theorems.

Theorem 1. *If R is a* 3NF *relation scheme, R is in* BCNF *if for every pair of distinct keys K and K' of R, $K \cap K' = \emptyset$.*

Proof. This theorem is the contrapositive of Theorem 3 in [3]. However, we simplify their proof as follows. If R is not in BCNF, then there is a nontrivial FD $X \to A$ of R such that $X \not\twoheadrightarrow R$. Since R is in 3NF and $X \not\twoheadrightarrow R$, $A \in K$ where K is a key of R. $X(K - A) \to K$ and thus $X(K - A) \to R$. Therefore, there is a key K' of R such that $K' \subseteq X(K - A)$. $K \neq K'$ since $A \in K$ but $A \notin K'$. K' must contain an attribute B in $K - A$; otherwise $K' \not\twoheadrightarrow R$ since $X \not\twoheadrightarrow R$. Therefore, $K \cap K' \neq \emptyset$. □

The converse of Theorem 1 is not true. Consider, for example, $R = ABC$ with keys AB and BC and no other FDs. R is in BCNF, but the two keys of R overlap.

Theorem 2. *If R is a* BCNF *relation scheme, R is in* 4NF *if and only if for every MVD M of R, there is a key of R that is a subset of a component of M.*

Proof. Let $M = \bowtie\{R_1, R_2\}$ be a nontrivial MVD of R. Without loss of generality, we assume there is a key K of R such that $K \subseteq R_1$. Since K is a key, $K \to (R_2 - R_1)$. Since $(R_1 \cap R_2) \twoheadrightarrow (R_2 - R_1)$, $K \subseteq R_1$, and since $K \to (R_2 - R_1)$, by C2, $(R_1 \cap R_2) \to (R_2 - R_1)$. Since M is nontrivial, $(R_2 - R_1)$ is not empty and thus $(R_1 \cap R_2) \to (R_2 - R_1)$ is nontrivial. Since R is in BCNF, $(R_1 \cap R_2) \to R$ and therefore, R is in 4NF. If each key K of R is neither a subset of R_1 nor R_2, then, $K \cap (R_2 - R_1) \neq \emptyset$ and $K \cap (R_1 - R_2) \neq \emptyset$. Consider a relation r on R with two tuples t_1 and t_2 where t_1 has 1's in all the columns and t_2 has 1's in all the $(R_1 \cap R_2)$-columns and 2's in the rest of the columns. Now, r satisfies every key dependency of R; however, r does not satisfy M and therefore the key dependencies of R do not imply M. Hence, R is not in 4NF. □

Theorem 3. *If R is a* 4NF *relation scheme, R is in* PJNF *if and only if for every JD J of R, $OMA(R, J)$ is acyclic.*

Proof. Let J be a JD of R. Since R is in 4NF, the key dependencies of R imply $MVD(J)$. Since every component of $OMA(R, J)$ is the union of some components of J, $MVD(OMA(R, J)) \subseteq MVD(J)$ and thus the key dependencies of R imply $MVD(OMA(R, J))$. If $OMA(R, J)$ is acyclic, by Theorem 13.2 in [2], $MVD(OMA(R, J))$ implies $OMA(R, J)$ and thus the key dependencies of R imply $OMA(R, J)$. Since the key dependencies of R and $OMA(R, J)$ together imply J, the key dependencies of R imply J. Hence, R is in PJNF. If R is in PJNF, there is a component of $OMA(R, J)$ equal to R. Therefore, $OMA(R, J)$ is acyclic. □

Lemma 4. *If R is a* BCNF *relation scheme, R is in* 4NF *if for every JD J of R, there is a key of R that is a subset of a component of J.*

Proof. Let J be a JD of R. It is sufficient to consider the MVDs in $MVD(J)$. If there is a key K of R that is a subset of a component of J, by the definition of $MVD(J)$, for every MVD M in $MVD(J)$, K is a subset of a component of M. Therefore, by Theorem 2, R is in 4NF. □

The converse of Lemma 4 is not true. Consider, for example, $R = ABC$ and $J = \bowtie\{AB, BC, AC\}$. $MVD(J)$ contains only trivial MVDs and thus R is in 4NF. However, the only key ABC of R is not a subset of any component of J.

Theorem 5. *If R is a* BCNF *relation scheme, R is in* PJNF *if and only if for every JD J of R, there is a key of R that is a subset of a component of J and $OMA(R, J)$ is acyclic.*

Proof. Let J be a nontrivial JD of R. If there is a key of R that is a subset of a component of J, by Lemma 4, R is in 4NF. Therefore, if $OMA(R, J)$ is acyclic, by Theorem 3, R is in PJNF. Assume R is in PJNF. If there is not a key of R that is a subset of a component of J, then $OMA(R, J)$ equals J. Since J is nontrivial, R is not in PJNF. Hence, there is a key of R that is a subset of a component of J. By Lemma 4, R is in 4NF. Therefore, since R is in PJNF, by Theorem 3, $OMA(R, J)$ is acyclic. □

Note that Theorem 1 and Theorem 5 together give sufficient conditions for a 3NF relation scheme to be in PJNF.

4. Related work

In this section, we relate the results in this paper and the results in [1]. In particular, we show that the lemmas and theorems in [1] follow from our theorems.

Corollary 6 (Lemma 4.2 in [1]). *If R is a* 3NF *relation scheme, R is in* BCNF *if every key of R is simple.*

Proof. If R is a 3NF relation scheme and every key of R is simple, then for every pair of distinct keys K and K' of R, $K \cap K' = \emptyset$. Therefore, by Theorem 1, R is in BCNF. □

Corollary 7 (Lemma 4.4 in [1]). *If R is a* BCNF *relation scheme, R is in* PJNF *if every key of R is simple.*

Proof. If R is a BCNF relation scheme and every key of R is simple, then for every JD J of R, every key of R is contained in a component of J. Now, by using a similar argument for Theorem 2, we can show that $OMA(R, J)$ is acyclic. Therefore, by Theorem 5, R is in PJNF. □

Corollary 8 (Theorem 4.1 in [1]). *If R is a* 3NF *relation scheme, R is in* PJNF *if every key of R is simple.*

Proof. This corollary follows immediately from Corollaries 6 and 7. □

Corollary 9 (Theorem 5.2 in [1]). *If R is a* BCNF *relation scheme, R is in* 4NF *if some key of R is simple.*

Proof. If R is a BCNF relation scheme and some key of R is simple, then for every MVD M of R, there is a key of R that is a subset of a component of M. Therefore, by Theorem 2, R is in 4NF. □

5. Concluding remarks

We have generalized the results in [1]. In particular, we have given less stringent conditions for a lower normal form to be in a higher normal form. Theorem 1 is less stringent than Lemma 4.2 in [1] because keys need not be single attributes. In fact, we only need keys not to overlap. Theorem 2 is less stringent than Theorem 5.2 in [1] because we do not need a key to be simple. We simply need a key to be a subset of a component of every MVD. Finally, Theorem 5 is less stringent than Lemma 4.4 in [1] because we do not need keys to be simple. We only need a key to be a subset of a component of every JD and the output of the Membership Algorithm is acyclic.

Acknowledgments

I like to thank Professor David W. Embley of Brigham Young University and the anonymous referees for many helpful suggestions.

References

[1] C.J. Date and R. Fagin, Simple conditions for guaranteeing higher normal forms in relational databases, *ACM Trans. Database Systems* 17 (3) (1992) 465–476.

[2] D. Maier, *The Theory of Relational Databases* (Computer Science Press, Rockville, MD, 1983).

[3] M.W. Vincent and B. Srinivasan, A note on relation schemes which are in 3NF but not in BCNF, *Inform. Process. Lett.* 48 (1993) 281–283.