

WIKIPEDIA

The Free Encyclopedia

Chase (algorithm)

The chase is a simple fixed-point algorithm testing and enforcing implication of data dependencies in database systems. It plays important roles in database theory as well as in practice. It is used, directly or indirectly, on an everyday basis by people who design databases, and it is used in commercial systems to reason about the consistency and correctness of a data design. New applications of the chase in meta-data management and data exchange are still being discovered.

The chase has its origins in two seminal papers of 1979, one by Alfred V. Aho, Catriel Beeri, and Jeffrey D. Ullman^[1] and the other by David Maier, Alberto O. Mendelzon, and Yehoshua Sagiv.^[2]

In its simplest application the chase is used for testing whether the projection of a relation schema constrained by some functional dependencies onto a given decomposition can be recovered by rejoining the projections. Let t be a tuple in $\pi_{S_1}(R) \bowtie \pi_{S_2}(R) \bowtie \dots \bowtie \pi_{S_k}(R)$ where R is a relation and F is a set of functional dependencies (FD). If tuples in R are represented as t_1, \dots, t_k , the join of the projections of each t_i should agree with t on $\pi_{S_i}(R)$ where $i = 1, 2, \dots, k$. If t_i is not on $\pi_{S_i}(R)$, the value is unknown.

The chase can be done by drawing a tableau (which is the same formalism used in tableau query). Suppose R has attributes A, B, \dots and components of t are a, b, \dots . For t_i use the same letter as t in the components that are in S_i but subscript the letter with i if the component is not in S_i . Then, t_i will agree with t if it is in S_i and will have a unique value otherwise.

The chase process is confluent. There exist implementations of the chase algorithm,^[3] some of them are also open-source.^[4]

Example

Let $R(A, B, C, D)$ be a relation schema known to obey the set of functional dependencies $F = \{A \rightarrow B, B \rightarrow C, CD \rightarrow A\}$. Suppose R is decomposed into three relation schemas $S_1 = \{A, D\}$, $S_2 = \{A, C\}$ and $S_3 = \{B, C, D\}$. Determining whether this decomposition is lossless can be done by performing a chase as shown below.

The initial tableau for this decomposition is:

A	B	C	D
a	b ₁	c ₁	d
a	b ₂	c	d ₂
a ₃	b	c	d

The first row represents S_1 . The components for attributes A and D are unsubscripted and those for attributes B and C are subscripted with $i = 1$. The second and third rows are filled in the same manner with S_2 and S_3 respectively.

The goal for this test is to use the given F to prove that $t = (a, b, c, d)$ is really in R . To do so, the tableau can be chased by applying the FDs in F to equate symbols in the tableau. A final tableau with a row that is the same as t implies that any tuple t in the join of the projections is actually a tuple of R .

To perform the chase test, first **decompose all FDs in F so each FD has a single attribute on the right hand side** of the "arrow". (In this example, F remains unchanged because all of its FDs already have a single attribute on the right hand side: $F = \{A \rightarrow B, B \rightarrow C, CD \rightarrow A\}$.)

When equating two symbols, if one of them is unsubscripted, make the other be the same so that the final tableau can have a row that is exactly the same as $t = (a, b, c, d)$. If both have their own subscript, change either to be the other. However, to avoid confusion, all of the occurrences should be changed.

First, apply $A \rightarrow B$ to the tableau. The first row is (a, b_1, c_1, d) where a is unsubscripted and b_1 is subscripted with 1. Comparing the first row with the second one, change b_2 to b_1 . Since the third row has a_3 , b in the third row stays the same. The resulting tableau is:

A	B	C	D
a	b_1	c_1	d
a	b_1	c	d_2
a_3	b	c	d

Then consider $B \rightarrow C$. Both first and second rows have b_1 and notice that the second row has an unsubscripted c . Therefore, the first row changes to (a, b_1, c, d) . Then the resulting tableau is:

A	B	C	D
a	b_1	c	d
a	b_1	c	d_2
a_3	b	c	d

Now consider $CD \rightarrow A$. The first row has an unsubscripted c and an unsubscripted d , which is the same as in third row. This means that the A value for row one and three must be the same as well. Hence, change a_3 in the third row to a . The resulting tableau is:

A	B	C	D
a	b_1	c	d
a	b_1	c	d_2
a	b	c	d

At this point, notice that the third row is (a, b, c, d) which is the same as t . Therefore, this is the final tableau for the chase test with given R and F . Hence, whenever R is projected onto S_1 , S_2 and S_3 and rejoined, the result is in R . Particularly, the resulting tuple is the same as the tuple of R that is projected onto $\{B, C, D\}$.

References

1. Alfred V. Aho, Catriel Beeri, and Jeffrey D. Ullman: "The Theory of Joins in Relational Databases", *ACM Trans. Datab. Syst.* 4(3):297-314, 1979.
 2. David Maier, Alberto O. Mendelzon, and Yehoshua Sagiv: "Testing Implications of Data Dependencies". *ACM Trans. Datab. Syst.* 4(4):455-469, 1979.
 3. Michael Benedikt, George Konstantinidis, Giansalvatore Mecca, Boris Motik, Paolo Papotti, Donatello Santoro, Efthymia Tsamoura: *Benchmarking the Chase*. In *Proc. of PODS*, 2017.
 4. "The Llunatic Mapping and Cleaning Chase Engine" (<https://github.com/donatellosantoro/Llunatic>). 6 April 2021.
- Serge Abiteboul, Richard B. Hull, Victor Vianu: *Foundations of Databases*. Addison-Wesley, 1995.
 - A. V. Aho, C. Beeri, and J. D. Ullman: *The Theory of Joins in Relational Databases*. *ACM Transactions on Database Systems* 4(3): 297-314, 1979.
 - J. D. Ullman: *Principles of Database and Knowledge-Base Systems, Volume I*. Computer Science Press, New York, 1988.
 - J. D. Ullman, J. Widom: *A First Course in Database Systems* (3rd ed.). pp. 96–99. Pearson Prentice Hall, 2008.
 - Michael Benedikt, George Konstantinidis, Giansalvatore Mecca, Boris Motik, Paolo Papotti, Donatello Santoro, Efthymia Tsamoura: *Benchmarking the Chase*. In *Proc. of PODS*, 2017.

Further reading

- Sergio Greco; Francesca Spezzano; Cristian Molinaro (2012). *Incomplete Data and Data Dependencies in Relational Databases*. Morgan & Claypool Publishers. ISBN 978-1-60845-926-1.

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Chase_\(algorithm\)&oldid=1046631670](https://en.wikipedia.org/w/index.php?title=Chase_(algorithm)&oldid=1046631670)"

■