

Subscribe to Our Blog

Blog Home (<https://www.sisense.com/blog/>)

Sign up to get the latest news and developments in business analytics, data analysis and Sisense.

Business Perspectives (<https://www.sisense.com/blog/category/business-perspectives/>)

BI Best Practices (<https://www.sisense.com/blog/category/bi-best-practices/>)

Subscribe Tech Talk (<https://www.sisense.com/blog/category/tech-talk/>)

Sisense News (<https://www.sisense.com/blog/category/sisensenews/>)

Related Posts

- How to Calculate Year Over Year Growth
(<https://www.sisense.com/blog/tech-talk/how-to-calculate-year-year-growth/>) | 7 min read

- SQL Cheat Sheet: Retrieving Column Description in SQL Server
(<https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/>)

- Dashboard Design Best Practices – 4 Key Principles
(<https://www.sisense.com/blog/4-design-principles-creating-better-dashboards/>)

Watch a Demo

- As those of you who have watched my recent webinar Data Modeling Fundamentals With Sisense ElastiCube dashboard (https://pages.sisense.com/WBWalkthroughforyourfirstEC-Datamodelingfundamentals_Website.html) might recall, a **primary key** is a unique identifier given to a record in our database, which we can use when querying the database in order to
- The Beginner's Guide to MySQL

WHEN (AND HOW) TO USE SURROGATE KEYS



Hila Kantor (<https://www.sisense.com/blog/author/hila/>)
February 19, 2018

Storage Engines


<https://www.sisense.com/blog/beginners-guide-to-mysql-storage-engines/>

join multiple sources. This article will discuss the concept of **surrogate keys** and show some examples of when and how to apply them using simple SQL.

Subscribe to Our Blog

Sign up to get the latest news and

developments in business analytics, data

Before we dive into natural vs surrogate keys, let's recall four important rules to follow when selecting a primary key for your data analysis and Sisense model:

Subscribe

1. **The primary key must be unique for each record.** A primary key with duplicates will lead to inaccurate queries with duplicate records and totals. If two customers are assigned the primary key, their sales activity will be unintentionally blended together. If the customer is accidentally duplicated, their sales activity will also be duplicated. Database architects refer to this

- [How to Calculate Year-to-Date Variance](#)

2. **The primary key must apply uniform rules for all records.** Whether your key is strictly numeric, alphanumeric, or a random system-generated value, the key must be programmed in a consistent format. This format must exist despite whatever requirements there are in the business requirements. An inconsistent format can lead to difficult data analysis, especially in

- [parent/child data relationships](#)

- [SQL Cheat Sheet: Retrieving](#)

3. **The primary key must stand the test of time.** A key based off of contextual data

(<https://www.sisense.com/glossary/contextual-data/>) at the present time, may not have the same contextual meaning later. For

example, if a customer ID key is based on customer name, what happens when a customer is acquired or reorganized?

(<https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/>) Changing key formats should be avoided at all costs. Changing keys will require changing all stored procedures referencing the new key in any JOINS or WHERE clauses, as well as UPDATES to all existing references to the old key in all of your database tables.

- [Dashboard Design Best](#)

4. **The primary key must be read-only.** In order to stand the test of time, primary keys should never be edited. Edited primary

keys can have typos (123123 vs 132123), varying formats based on the user's preference (1 vs 000001), and allow for

(<https://www.sisense.com/blog/4-overwriting-a-previously-deleted-record>) Never allow anyone to edit the value of primary keys.

[design-principles-creating-better-](#)

[dashboards/](#)

Selecting a Primary Key: Surrogate vs Natural Keys

- [The Beginner's Guide to MySQL](#)

Storage Engines

First, let's go over the difference between these two forms of primary keys:

<https://www.sisense.com/blog/beginners-guide-to-mysql-storage-engines/>

A **natural key** is a key that has **contextual or business meaning** (for example, in a table containing STORE, SALES and DATE we might use the DATE field as a natural key when joining with another table detailing inventory).

Sign up to get the latest news and

developments in business analytics, data

A natural key can be system-generated, but natural keys are at least partially determined by a manual process. Some natural keys are totally manually generated. One of the most widely recognized uses of a natural key is a stock ticker symbol – i.e. MSFT, APPL, and GOOGL. Natural keys serve as a great primary key when contextual meaning is important.

Subscribe

A **surrogate key** is a key which **does not have any contextual or business meaning**. It is manufactured “artificially” and only for the purposes of data analysis. The most frequently used version of a surrogate key is an increasing sequential integer or “counter” value (i.e. 1, 2, 3). Surrogate keys can also include the current system date/time stamp, or a random alphanumeric string.

Related Posts

- How to Calculate Year Over Year

Growth

[\(https://www.sisense.com/blog/calculate-](https://www.sisense.com/blog/calculate-year-over-year-growth/)

year-over-growth/)

Want to master data modeling? Watch our on demand webinar

([https://www.sisense.com/webinars/data-modeling-fundamentals-sisense-](https://www.sisense.com/webinars/data-modeling-fundamentals-sisense-elasticube/)

elasticube/) and learn the fundamental skills every analyst should have.

Server

[\(https://www.sisense.com/blog/sql-](https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/)

cheat-sheet-retrieving-column-

description-sql-server/)

When should you stick to natural keys in your data model?

- Dashboard Design Best

Practices – 4 Key Principles

The main advantage of natural keys is in their simplicity and in the fact that the data maintains its original context. They will often be

(relatively) easy to recognize to people viewing the data, and relying on natural keys reduces the need to enrich the data using

custom SQL. Additionally,

dashboards/)

See Sisense In Action (/Get/Watch-

SHARES

- The Beginner's Guide to MySQL

Storage Engines



Natural keys are great for multiple data types in the database. Natural keys allow the user to easily identify the data type from the key, even when multiple data types use similar key formats. Financial databases frequently format their keys using a natural and sequential key together.

Subscribe to Our Blog

Sign up to get the latest news and developments in business analytics, data analysis and Sisense

Expense Report: ER-123

Timesheets: TS-123

AR Invoices: INV-123

Subscribe

Related Posts

Even though all three records contain a sequential ID of 123, the natural key prefix allows the user to immediately identify

- How to Calculate Year Over Year Growth

Natural keys work well when connecting two systems with two different primary key formats. Thus for example, we can use year-year-growth/)

- SQL Cheat Sheet: Retrieving Salesforce Customer ID: 001D000000IRt53

Column Description in SQL

Server

Quickbooks Customer ID: MSFT

(https://www.sisense.com/blog/sql-

cheat-sheet-retrieving-column-

description-sql-server/)

- Dashboard Design Best

Practices – 4 Key Principles

Natural key Customer ID: 001D000000IRt53_MSFT

(https://www.sisense.com/blog/4-

design-principles-creating-better-

dashboards/)

• **Natural keys make for a more easy-to-understand GUI.** A customer ID such as GOOGL is easy for a user to recognize (for

- The beginning of the stock ticker symbol is for Google. Easier recognition also allows for easier search.

Storage Engines

Drawbacks of Using Natural Keys

<https://www.sisense.com/blog/10-techniques-to-boost-your-data-modeling/>
<https://www.sisense.com/blog/10-techniques-to-boost-your-data-modeling/>

While it might be tempting and initially easier to rely on existing natural keys, this could prove problematic when scaling the data model (<https://www.sisense.com/blog/10-techniques-to-boost-your-data-modeling/>), or in a more complex environment, which we

will demonstrate using an example of stock tickers:

developments in business analytics, data

analysis and Sisense.

- **Natural keys do not apply uniform rules for each record.** Designators or variables in the natural key make the key difficult to query and understand after the fact. For example, stock ticker symbols of preferred shares have a multitude of designators, including P, PR, and /PR. Trying to query for the designator P (`SELECT * FROM stock_quotes WHERE stock_ticker_symbol like %P`) would return all results where the stock ticker symbol ends in P, regardless if the symbol is actually preferred stock or not.

Related Posts

- **Natural keys do not stand the test of time.** Symbols which might have been business meaning could become meaningless, or bear a different meaning in the future. Thus, for example, the symbols GOOG and GOOGL do not accurately represent the reorganization of the company from Google to Alphabet.
- **Natural keys can be easily confused with each other.** Sticking with the previous example – when Twitter was ready to launch their IPO under the ticker TWTR, many investors bought from a defunct electronics company named Tweeter, trading under the ticker TWTRQ. Because TWTR and TWTRQ contain the same first four letters, many investors unintentionally invested in the wrong stock. Tweeter later changed their ticker symbol to THEGQ, which could also be misconstrued with GQ Magazine (a privately-held company under Conde Nast).
- SQL Cheat Sheet: Retrieving Column Description in SQL Server

Advantages of Using Surrogate Keys

As mentioned, a surrogate key sacrifice some of the original context of the data. However, it can be extremely useful for analytical purposes for the following reasons:

- Dashboard Design Best Practices – 4 Key Principles

Watch a Demo

- **Surrogate keys are unique.** Because surrogate keys are system-generated, it is impossible for the system to create and store a duplicate value.
- **Surrogate keys apply uniform rules to all records.** The surrogate key value is the result of a program, which creates the surrogate key value for each record.
- The surrogate key value is a SQL key created as a result of a program will apply uniform rules for each record.



Because surrogate keys lack any context or business meaning, there will be no need

- **Surrogate keys allow for unlimited values.** Sequential, timestamp, and random keys have no practical limits to unique

Subscribe to Our Blog

Sign up to get the latest news and

Combining Natural and Surrogate Keys

development in business analytics, data analysis and Sisense.

Certain business scenarios might require keeping the natural key intact as a means for users to interact with the database. In these cases... [Subscribe](#)

Related Posts

- **When a key is recommended, use a surrogate key field as the primary key, and a natural key as a foreign key.**

While users may interact with the natural key, the database can still have surrogate keys outside of the users' view, with no

- [How to Calculate Year Over Year Growth](#)

- **If a natural key must be used without an additional surrogate key, be sure to combine it with a surrogate key**

[element.](https://www.sisense.com/blog/calculate-element/)

In our financial database example, Expense Reports (ER-123) have a natural key is used in conjunction with a surrogate sequential key. This format prevents many of the natural key side effects listed above.

- [SQL Cheat Sheet: Retrieving Column Description in SQL Server](#)

An Example of Adding a Surrogate Key Using Custom SQL

In the following example, we will look at a table containing historical data about product prices. By using a custom SQL expression in the Sisense Elasticube Manager (/product/), we create the surrogate key ProdDate_Key, which in this case is created by combining the other fields into a single, unique identifier that can easily be queried later.

- [Dashboard Design Best Practices — 4 Key Principles](#)

<https://www.sisense.com/blog/4-design-principles-creating-better-dashboards/>

See Sisense In Action (/Get/Watch-

- [The Beginner's Guide to MySQL](#)

Storage Engines

Product List Price History - Table Preview
(https://www.sisense.com/blog/beginners-guide-to-mysql-storage-engines/)

Output Errors

StartDate	EndDate	ProductID	ListPrice
7/1/2005 12:00:00 AM	6/30/2006 12:00:00 AM	707	33.6442
7/1/2006 12:00:00 AM	6/30/2007 12:00:00 AM	707	33.6442
7/1/2007 12:00:00 AM	12/31/2008 12:00:00 AM	707	34.99
7/1/2005 12:00:00 AM	6/30/2006 12:00:00 AM	708	33.6442
7/1/2006 12:00:00 AM	6/30/2007 12:00:00 AM	708	33.6442
7/1/2007 12:00:00 AM	12/31/2008 12:00:00 AM	708	34.99
7/1/2005 12:00:00 AM	6/30/2006 12:00:00 AM	709	9.5
7/1/2005 12:00:00 AM	6/30/2006 12:00:00 AM	710	9.5
7/1/2005 12:00:00 AM	6/30/2006 12:00:00 AM	711	33.6442
7/1/2006 12:00:00 AM	6/30/2007 12:00:00 AM	711	33.6442
7/1/2007 12:00:00 AM	12/31/2008 12:00:00 AM	711	34.99
7/1/2005 12:00:00 AM	6/30/2006 12:00:00 AM	712	8.6442
7/1/2006 12:00:00 AM	6/30/2007 12:00:00 AM	712	8.6442
7/1/2007 12:00:00 AM	12/31/2008 12:00:00 AM	712	8.99
7/1/2005 12:00:00 AM	6/30/2006 12:00:00 AM	713	48.0673
7/1/2006 12:00:00 AM	6/30/2007 12:00:00 AM	713	48.0673
7/1/2007 12:00:00 AM	12/31/2008 12:00:00 AM	713	49.99
7/1/2005 12:00:00 AM	6/30/2006 12:00:00 AM	714	48.0673
7/1/2006 12:00:00 AM	6/30/2007 12:00:00 AM	714	48.0673
7/1/2007 12:00:00 AM	12/31/2008 12:00:00 AM	714	49.99
7/1/2005 12:00:00 AM	6/30/2006 12:00:00 AM	715	48.0673
7/1/2006 12:00:00 AM	6/30/2007 12:00:00 AM	715	48.0673
7/1/2007 12:00:00 AM	12/31/2008 12:00:00 AM	715	49.99
7/1/2005 12:00:00 AM	6/30/2006 12:00:00 AM	716	48.0673

1 2 3 4

Subscribe to Our Blog

Sign up to get the latest news and developments in business analytics, data analysis and Sisense.

Related Posts

- [How to Calculate Year Over Year Growth](https://www.sisense.com/blog/calculate-year-over-year-growth/)
(https://www.sisense.com/blog/calculate-year-over-year-growth/)
- [SQL Cheat Sheet: Retrieving Column Description in SQL Server](https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/)
(https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/)

- [Dashboard Design Best Practices – 4 Key Principles](#)

(http://www.sisense.com/blog/4-design-principles-creating-better-dashboards/)

See Sisense In Action (/Get/Watch-Demo)

SQL used to add surrogate key:

- [The Beginner's Guide to MySQL](#)

Storage Engines
 SELECT DISTINCT
 (https://www.sisense.com/blog/beginners-
 guide-to-mysql-storage-engines/)
 Pro_Date_Key, Date,

Subscribe to Our Blog

PH.ListPrice

Sign up to get the latest news and

FROM [ProductListPriceHistory] PH JOIN [AllDates] ON Date between PH.StartDate AND PH.EndDate

analysis and Sisense.

Result:

Subscribe

Related Posts

- How to Calculate Year Over Year Growth
 (https://www.sisense.com/blog/calculate-year-year-growth/)
- SQL Cheat Sheet: Retrieving Column Description in SQL Server
 (https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/)
- Dashboard Design Best Practices – 4 Key Principles
 (https://www.sisense.com/blog/4-design-principles-creating-better-dashboards/)
- The Beginner's Guide to MySQL

Watch a Demo

See Sisense In Action (/Get/Watch-

SHARES

Demos

Storage Engines

ProductDailyPrice - Table Preview

https://www.sisense.com/blog/beginners-guide-to-mysql-storage-engines/

Subscribe to Our Blog

Sign up to get the latest news and developments in business analytics, data analysis and Sisense.

Related Posts

- How to Calculate Year Over Year Growth
(https://www.sisense.com/blog/calculate-year-year-growth/)
- SQL Cheat Sheet: Retrieving Column Description in SQL Server
(https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/)

ProdDate_Key	Date	ProductID	ListPrice
707_2006-3-24	3/24/2006 12:00:00 AM	707	33.6442
708_2006-3-24	3/24/2006 12:00:00 AM	708	33.6442
709_2006-3-24	3/24/2006 12:00:00 AM	709	9.5
710_2006-3-24	3/24/2006 12:00:00 AM	710	9.5
711_2006-3-24	3/24/2006 12:00:00 AM	711	33.6442
712_2006-3-24	3/24/2006 12:00:00 AM	712	8.6442
713_2006-3-24	3/24/2006 12:00:00 AM	713	48.0673
714_2006-3-24	3/24/2006 12:00:00 AM	714	48.0673
715_2006-3-24	3/24/2006 12:00:00 AM	715	48.0673
716_2006-3-24	3/24/2006 12:00:00 AM	716	48.0673
717_2006-3-24	3/24/2006 12:00:00 AM	717	1263.4598
718_2006-3-24	3/24/2006 12:00:00 AM	718	1263.4598
719_2006-3-24	3/24/2006 12:00:00 AM	719	1263.4598
720_2006-3-24	3/24/2006 12:00:00 AM	720	1263.4598
721_2006-3-24	3/24/2006 12:00:00 AM	721	1263.4598
722_2006-3-24	3/24/2006 12:00:00 AM	722	297.6346
723_2006-3-24	3/24/2006 12:00:00 AM	723	297.6346
724_2006-3-24	3/24/2006 12:00:00 AM	724	297.6346
725_2006-3-24	3/24/2006 12:00:00 AM	725	306.5636
726_2006-3-24	3/24/2006 12:00:00 AM	726	306.5636
727_2006-3-24	3/24/2006 12:00:00 AM	727	306.5636

150199 Rows

- Dashboard Design Best

Practices – 4 Key Principles

Want to master data modeling? Watch our on demand webinar

(https://www.sisense.com/webinars/data-modeling-fundamentals-sisense-elasticube/) and learn the fundamental skills every analyst should have.

- The Beginner's Guide to MySQL

Storage Engines



Tags: [beginners guide to mysql storage engines](https://www.sisense.com/blog/beginners-guide-to-mysql-storage-engines/) (<https://www.sisense.com/blog/beginners-guide-to-mysql-storage-engines/>) | [Data modeling](https://www.sisense.com/blog/tag/data-modeling/) (<https://www.sisense.com/blog/tag/data-modeling/>) | [how to's](https://www.sisense.com/blog/tag/how-tos/) (<https://www.sisense.com/blog/tag/how-tos/>) | [SQL](https://www.sisense.com/blog/tag/sql/) (<https://www.sisense.com/blog/tag/sql/>)

Subscribe to Our Blog

Sign up to get the latest news and developments in business analytics, data analysis and Sisense.

Subscribe

Related Posts

- [How to Calculate Year Over Year Growth](https://www.sisense.com/blog/calculate-year-year-growth/) (<https://www.sisense.com/blog/calculate-year-year-growth/>)
- [SQL Cheat Sheet: Retrieving Column Description in SQL Server](https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/) (<https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/>)
- [Dashboard Design Best Practices – 4 Key Principles](https://www.sisense.com/blog/4-design-principles-creating-better-dashboards/) (<https://www.sisense.com/blog/4-design-principles-creating-better-dashboards/>)

Watch a Demo

- [See Sisense In Action \(/Get/Watch-Demo\)](#)
- [The Beginner's Guide to MySQL](#)

SHARES

Storage Engines


<https://www.sisense.com/blog/beginners-guide-to-mysql-storage-engines/>

Subscribe to Our Blog

Sign up to get the latest news and developments in business analytics, data analysis and Sisense.

Explore Other Resources

Related Posts

- How to Calculate Year Over Year Growth
(<https://www.sisense.com/blog/calculate-year-year-growth/>)
- SQL Cheat Sheet: Retrieving Column Description in SQL Server
(<https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/>)
- Dashboard Design Best Practices – 4 Key Principles
(<https://www.sisense.com/blog/4-design-principles-creating-better-dashboards/>)
- See Sisense In Action (/Get/Watch-Demo)
- The Beginner's Guide to MySQL
(<https://www.sisense.com/blog/when-and-how-to-use-surrogate-keys/>)



Storage Engines
 [\(https://www.sisense.com/blog/beginners-guide-to-mysql-storage-engines/\)](https://www.sisense.com/blog/beginners-guide-to-mysql-storage-engines/)

Subscribe to Our Blog

Sign up to get the latest news and developments in business analytics, data analysis and Sisense.

Subscribe

Related Posts

- How to Calculate Year Over Year Growth
(<https://www.sisense.com/blog/calculate-year-year-growth/>)
- SQL Cheat Sheet: Retrieving Column Description in SQL Server
(<https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/>)

- Dashboard Design Best Practices — 4 Key Principles
(<https://www.sisense.com/blog/4-design-principles-creating-better-dashboards/>)

Watch a Demo

Subscribe

- The Beginner's Guide to MySQL



(https://www.sisense.com/whitepapers/bi-v-utm_source=trendemon&utm_medium=co)
7 Tips for Selecting a Business Intelligence Solution
Learn More



(https://www.sisense.com/blog/much-data-utm_source=trendemon&utm_medium=co)
How Much Data Will You Have in 3 Years? | Sisense Blog
Learn More

SUBSCRIBE TO OUR BLOG

Sign up to get the latest news and insights.

Storage Engines


<https://www.sisense.com/blog/beginners-guide-to-mysql-storage-engines/>

Subscribe to Our Blog

Sign up to get the latest news and developments in business analytics, data analysis and Sisense.

Resources

Subscribe

Related Posts

- How to Calculate Year Over Year Growth
(<https://www.sisense.com/blog/calculate-year-year-growth/>)
- SQL Cheat Sheet: Retrieving Column Description in SQL Server
(<https://www.sisense.com/blog/sql-cheat-sheet-retrieving-column-description-sql-server/>)
- Dashboard Design Best Practices – 4 Key Principles
(<https://www.sisense.com/blog/4-design-principles-creating-better-dashboards/>)
- The Beginner's Guide to MySQL

SHARES

FOLLOW US

(<https://www.facebook.com/Sisense/>) (<https://twitter.com/sisense/>) (<https://www.linkedin.com/company/sisense/>)

Copyright © 2018 Sisense Inc.

All rights reserved.