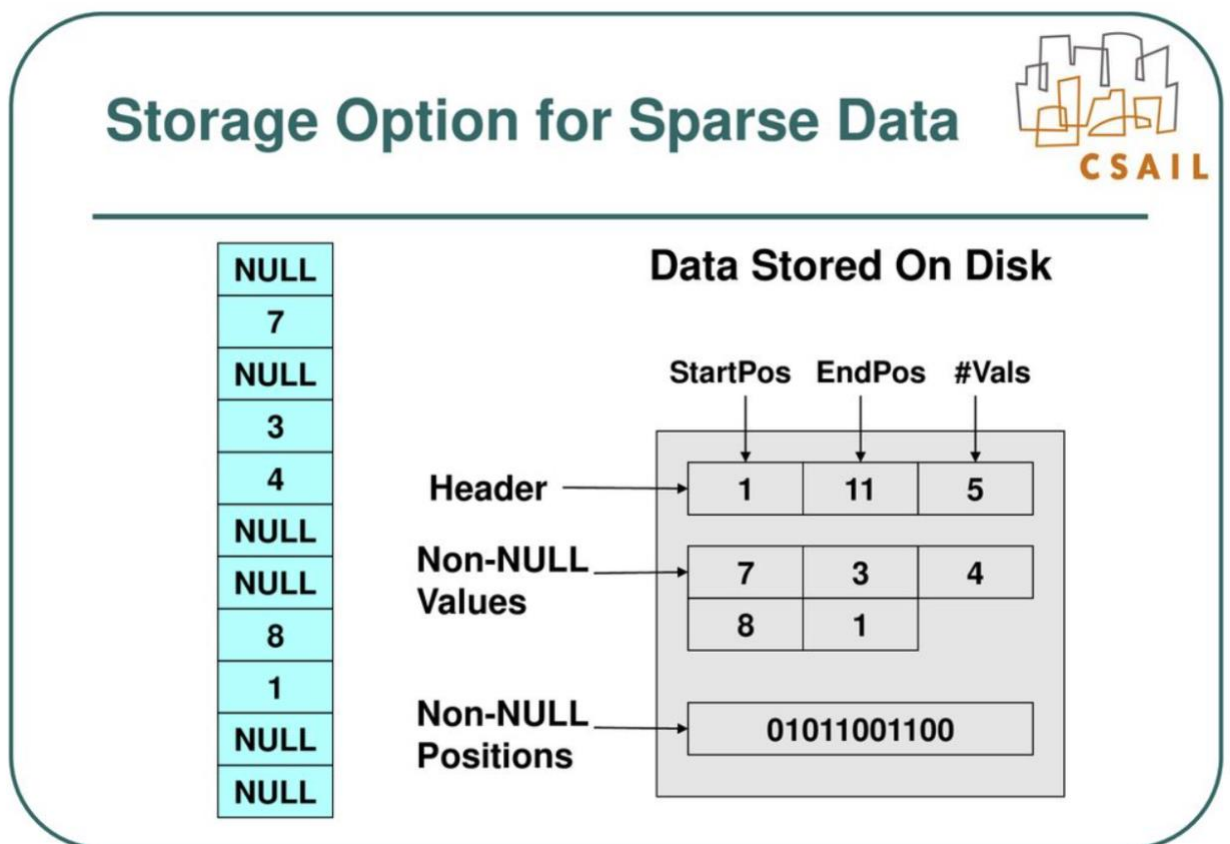


Sparse Data

Sparse in reference to data stores “means that there are many gaps present in the data being recorded.” (Harshita, 2018). This means that there will be spaces where there’s no data. So, a sparse value is basically a null value.

Sparse columns are special columns that can contain null values. These special columns are designated as sparse to save space. This concept and capability to create such columns can be applied to many databases. Some include SQL Server, Azure SQL Database, and Azure SQL Managed Instance (Microsoft, 2021).

The image below describes a column that contains sparse data:



Additionally, recent versions of RDBMS' also have the sparse option. One RDBMS that has this option is MS SQL Server. This is a very popular option.

The advantages of the sparse option include saving space. The SPARSE keyword is used to designate a column as a sparse column. Then, when there are null values, they don't take up any space. However, there are many specifics of how much space can be saved depending on the data type:

Estimated space savings by data type

Sparse columns require more storage space for non-NULL values than the space required for identical data that is not marked SPARSE. The following tables show the space usage for each data type. The **NULL Percentage** column indicates what percent of the data must be NULL for a net space savings of 40 percent.

Fixed-Length Data Types

Data type	Nonsparse bytes	Sparse bytes	NULL percentage
bit	0.125	5	98%
tinyint	1	5	86%
smallint	2	6	76%
int	4	8	64%
bigint	8	12	52%
real	4	8	64%
float	8	12	52%

References

Harshita, Srivastava. (2018). *What Is Sparse Data?* Retrieved from <https://magoosh.com/data-science/what-is-sparse-data/>

Microsoft. (2021). *Use Sparse Columns*. Retrieved from <https://docs.microsoft.com/en-us/sql/relational-databases/tables/use-sparse-columns?view=sql-server-ver15>

First Image: Abadi, Daniel. (2018). *Column Stores for Wide and Sparse Data*. Retrieved from <https://slideplayer.com/slide/14613848/>

Second Image: Microsoft. (2021). *Use Sparse Columns*. Retrieved from <https://docs.microsoft.com/en-us/sql/relational-databases/tables/use-sparse-columns?view=sql-server-ver15>