# 数据治理- NOT NULL 和 NULL 是否影响存储空间大小？

**1、MySQL**

在 MySQL 中，**NOT NULL 和 NULL** 对于字段所**占用**的**存储空间没有直接的影响**。字段在数据库中占用的**存储空间主要取决于**字段的**数据类型**和**长度**，而**不是**字段是否允许存储 NULL 值。    在 MySQL 中，常用的数据类型的存储空间大小如下（以示例数据类型 VARCHAR 为例）：

**CHAR：**    固定长度字符串，以实际存储的字符长度计算，例如 CHAR(10) 实际存储 "hello" 和 "world" 都占用 10 个字符的存储空间。

**VARCHAR：**    **可变长度字符串**，以实际存储的字符长度加上 1 或 2 个字节的存储空间计算，额外的字节用于存储字符串长度信息。

**INT：**   **4 字节整数**，存储范围为 -2^31 到 2^31-1 （约 -2.1 亿 到 2.1 亿），如果用 UNSIGNED 修饰则存储范围为 0 到 2^32-1。

以上仅是一些常见数据类型的存储空间情况，不同数据类型和设置还会影响存储空间的使用情况。因此，字段被设置为 NOT NULL 或 NULL 并不会直接影响其在数据库中的存储空间大小。

**2、Apache Kudu（列式存储）**

    在 Apache Kudu 中，与传统的关系型数据库 MySQL 不同，NOT NULL 和 NULL 在存储空间大小上有一定的区别。在 Kudu 中，字段是否允许存储 NULL 值会影响到数据在磁盘上的存储方式，从而间接影响到存储空间的大小和性能。    在 Kudu 中，当字段被设置为 NOT NULL 时，Kudu 在存储数据时只存储实际的数据值，而不会分配额外的空间来存储 NULL 值的标记。因此，对于 NOT NULL 的字段，节省了存储空间，并且可以提高性能。    当字段可以存储 NULL 值时，Kudu 需要额外的空间来处理 NULL 值的标记。这导致存储空间的一定增加，并且在处理包含 NULL 值的数据时可能会影响查询性能。    因此在 Kudu 中使用 NOT NULL 可以节省存储空间并提高性能，但在实际应用中，应根据业务需求和数据的特点来选择合适的字段约束。

**3、Apache HBase（列式存储）**

在 Apache HBase 中，与传统的关系型数据库不同，**HBase 是一个分布式列式存储系统**，不直接支持像 MySQL 中的 NOT NULL 和 NULL 约束。在 HBase 中，存储空间的大小主要取决于数据的实际值、列族和版本数等因素，而不是字段是否允许存储 NULL 值。    在 HBase 中，数据存储的基本单位是单元格（cell），单元格包含行键（row key）、列族（column family）、列限定符（column qualifier）、时间戳（timestamp）和值（value）等信息。在 HBase 中，单元格只会存储有值的数据，如果某个单元格没有数据，则该单元格不会被存储，也不会占用存储空间。    因此，在 HBase 中，存储空间大小受到数据的实际值以及数据模型的设计影响。而对于是否允许存储 NULL 值，HBase 并不会额外分配空间来存储 NULL 值的标记，**因为单元格只会存储有值的数据**。   **因此在 HBase 中，存储空间大小主要受数据实际值和数据设计的影响，而不受字段是否允许存储 NULL 值的影响。**

**4、StarRocks**    在 StarRocks 中，与传统的关系型数据库不同，StarRocks 是一种亚秒级 SQL 计算引擎，采用分布式数据库架构，其存储引擎不同于传统的行存储数据库。在 StarRocks 中，存储空间大小取决于数据类型、数据分布、数据压缩等因素，而与字段是否允许存储 NULL 值的约束关系比较小。    StarRocks 的存储模型是以列式存储为基础的，与传统的关系型数据库不同，它是按照列族和分区进行列式存储的。在 StarRocks 中，对于列族中的每一列，会根据存储类型和数据特性进行数据压缩处理，以节省存储空间。    对于字段是否允许存储 NULL 值，在 StarRocks 中，NULL 值会被存储，并且会占用一些额外的存储空间。StarRocks 会使用位图索引来表示 NULL 值，这可能会占用一些额外的存储空间，但相对于整个数据集来说，这个额外的占用通常比较小，且会随着压缩算法和存储优化的使用而减小。    因此在 StarRocks 中，存储空间大小主要受到数据类型、数据特性、压缩算法等因素的影响，相对而言，NULL 值的存储对整体存储空间的影响相对较小。因此，在实际应用中，可以根据业务需求和数据特点来选择字段是否允许存储 NULL 值。