**小型数据中心RAID方案选择**

**一.背景**

RAID0：简单的、无数据校验的数据条带化技术，只需要单数量硬盘，具有高读写性能，但没有数据冗余，一旦数据损坏，将无法恢复。

RAID1：将数据完全一致地分别写到工作磁盘和镜像磁盘，它的磁盘空间利用率为50%，但同时意味着需要双倍数量的磁盘，拥有完全容错的能力，实现成本高。

RAID5：校验数据分布在阵列中的所有磁盘上，而没有采用专门的校验磁盘。当一个数据盘损坏时，系统可以根据同一条带的其他数据块和对应的校验数据来重建损坏的数据。需要1.1倍磁盘数量，兼顾兼顾存储性能、数据安全和存储成本等各方面因素，它可以理解为 RAID0 和 RAID1 的折中方案。

RAID01：综合0和1，先做条带化再作镜像，本质是对物理磁盘实现镜像；

RAID10：综合0和1，先做镜像再作条带化，是对虚拟磁盘实现镜像。

原意于建立一个小型数据中心，保存多源、异构的非结构化数据，供未来人工智能学习。

则该数据中心具有一定数量的数据，且该数据用于人工智能学习，需要保证数据的拓展性及正确性。

**二.方案**

1. 选用RAID0。RAID0的最大优势是只需要单倍数量（即1D）的磁盘，意味着该方案的成本最低，但劣势在于RAID0没有冗余保护，以为着内部数据一旦发生损坏，将没有任何修复的可能性。
2. 选用RAID1。RAID1的最大优势是将原磁盘的数据完全镜像备份，因此，若原数据发生了损坏，可以很快在镜像中复制更改，但劣势在于该方案需要双倍数量（即2D）的磁盘，增加了一倍的成本，并且数据量只能达到50%。
3. 选用RAID01和RAID10。二者是上述两种方案的结合，同样拥有RAID0和RAID1的优点，但劣势在于因为采用了RAID1，最终的数据量上限也只能达到50%。
4. 选用RAID5：优势在于，相比于上述三种方案，RAID5只需要1.1倍数量（即1.1D）的磁盘，同时其可用数据量为(n-1)/n，兼顾了性能与成本。

**三.结论**

综上，当选用RAID5时，若磁盘数量为2，其数据量等同于方案三的50%，但当磁盘数量大于2时，其数据量将大于50%，并且随着磁盘增多其数据量同样增加。

综合数据使用量和磁盘数量成本，故认为RAID5是最佳选择方案。