02/22下午预答辩问题总结

对于本次答辩态度不够端正，题目都没有，答辩内容中废话过多，逻辑不清晰，语句也不够简练，实施方案也不够具体，考虑不够全面。

背景部分进行大幅度修改，删除前面部分的废话，准备改从数据量增长，SSD存储容量有限这个点引入压缩技术。然后介绍压缩技术发生的三个层次和现在压缩下移的趋势过渡到SSD内部压缩。然后介绍SSD内部压缩的优势和问题，借此引出课题的主题——基于透明压缩的数据管理方案。

存储硬件压缩的优势：

1 相比于主机端压缩不会占用CPU资源，同时还能保持一定量的带宽。

2 相比于文件系统的压缩，由于块接口限制了数据块的大小是固定的，因此压缩之后的变长数据块可能导致性能下降（具体怎么导致下降得进行更深入调研，目前考虑对块重新组织成符合接口协议大小的块而引入的额外的开销以及压缩/解压开销过大）。

方案设计方面，目前看来有点不够具体，也不够充分。打算后续跟师兄们进行进一步的沟通交流。

压缩模块对于数据管理模块是前置模块，也是紧紧相关的一个模块。高压缩率能提供更好的性能，减少写操作，但是也会增加读取数据的难度和延迟。对于读取性能，初步构思说跟写操作一样维护一个读取用的缓冲区来保存热数据，或者对于压缩模块增加选择压缩的功能。

数据管理模块简单构思了一种方案，基本思想是将一页page分割成几块小的部分，例如4KB分成2个1KB、2个512B、4个256KB共八个部分。一个压缩块存储时可以使用空闲的小块组合成一个具有较小的浪费空间来存储一个压缩页。仅仅是简单构思，没有理论推导也没有具体测试。在后续的工作中或者与师兄的交流中再思考可行性。

最后，本课题针对的应是数据具有一定压缩比、小文件居多的负载而开展的。（后面进行测试时候顺便测试一下大文件多、压缩比小的负载）。