Practical 2: Analyze impact of storage format

GAHAN SARAIYA, 18MCEC10

18mcec10@nirmauni.ac.in

I. Introduction

The data can not be always considered to be fit in one size and the goal of this article is to deduce the impact of various storage format.

Data needs to be stored in persistence storage, the focus is always to retrieve data efficiently. To achieve that we need to have optimal storage mechanism/schemes that allows to access the data efficiently.

II. Static data placement on disk pages

I. N-ary Storage Model (NSM)

Also known as Slotted Page.

- record are stored sequentially on data pages

- Accesses full record (all attributes of record)

✓ attributes of a same record are stored together

✗ Doesn't work well on modern memory hierarchies

Flow The whole page from disk is loaded in to main memory and then block(s) are fetched CPU cache
Impact

✓ Best results if full access to record is require

✗ Partial record access is slow. Due to Fixed page layout it wastes I/O bandwidth

✗ At CPU cache low spatial locality

II. Decomposition Storage Model (DSM)

Original Table is partitioned in to single attribute sub-tables. Each sub-table is stored in separate NSM page

- N-attribute table stored in to N sub-tables

✓ Saves I/O (eliminates unnecessary I/O)

✗ Hard to modify/construct record afterwards

Flow   Only required NSM page(s) (single attributed sub-table) is/are loaded in to main memory and blocks are fetched from loaded page **Impact**

✓ Best results if partial access to record is require

✗ full record access require reconstruction which may lead to unnecessary I/O and performance degraded compare to NSM

III.   Partition Attribute Across (PAX)

In single page it partitions data for spacial locality by attribute

Flow   Likewise in NSM it loads whole PAX page from disk in to main memory. partitioned data are stored in to contiguous block and only require block(s) needs to be fetched **Impact**

✓ Optimizes communication/transfer between Main memory and CPU cache memory (Increases Hit Ratio)

✓ Optimal solution for both full and partial record access

III.   SUMMARY

Table 1: Comparison of NSM, DSM and PAX

Performance	Cache↔Memory		Memory↔Disk	
Page Layout	**full record access**	**partial record access**	**full-record access**	**partial record access**
NSM	✔	✗	✔	✗
DSM	✗	✔	✗	✔
PAX	✔	✔	✔	✗