Flash Memory Summit
SNIA STORAGE SECURITY SUMMIT
Wednesday, May 11, 2022 • Virtual

# Bring Compression to Postgres at Zero Performance Cost

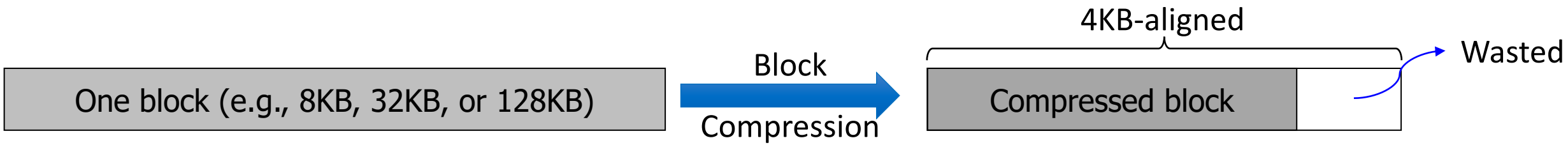
Tong Zhang, ScaleFlux
tong.zhang@scaleflux.com

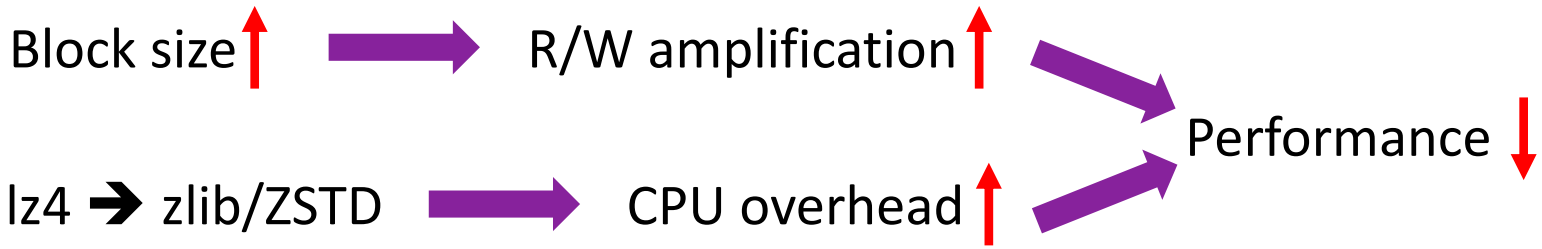
# Data Compression




Filesystem w/ transparent compression (ZFS, Btrfs, …)




Improve compression ratio &
reduce storage space waste



# Transparent Compression



PostgreSQL

Journaling filesystem (ext4, XFS, …)

Block layer w/ transparent compression (VDO, …)


Improve compression ratio & reduce storage space waste

lz4 → zlib/ZSTD → CPU overhead ↑ → Performance ↓

# Transparent Compression



PostgreSQL

Journaling filesystem (ext4, XFS, …)

Normal block layer


Computational Storage Drive (CSD) with Built-in Transparent Compression

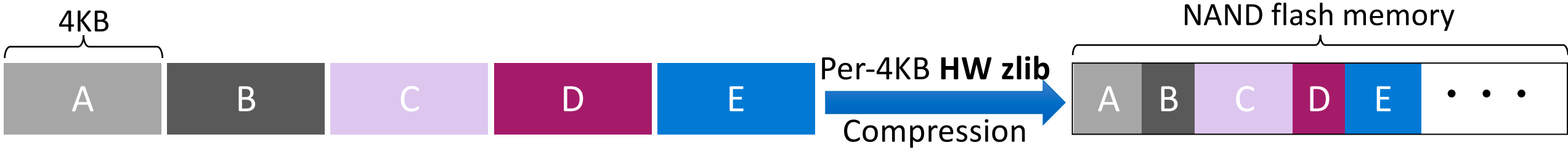
 

# CSD w/ Built-in Transparent Compression


Block-layer transparent compression (Linux VDO, VMware, …)

VS

ScaleFlux CSD 3000: Built-in transparent compression




HW zlib compression → Much higher compression ratio @ zero CPU cost

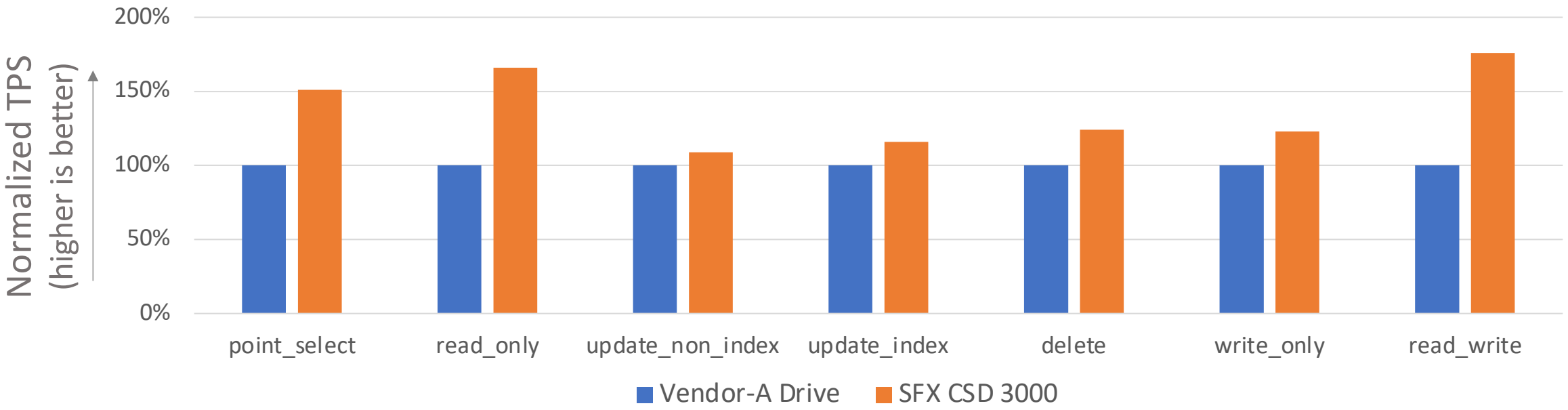
Tight packing → Zero space waste

# Straightforward Usage in Postgres

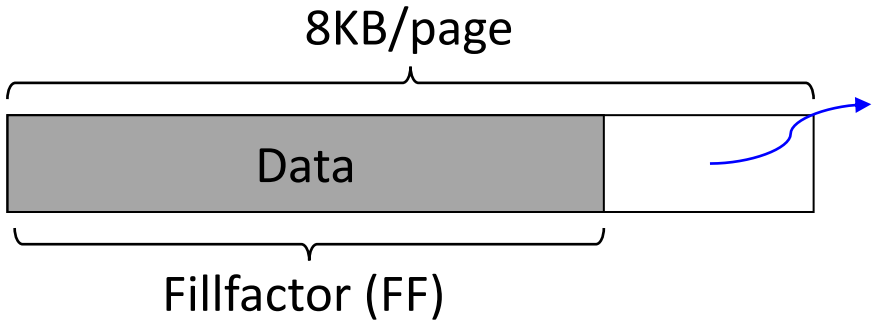


- Xeon E5-2667 v4 32-core @ 3.2GHz, 256GB DRAM
- CentOS 7.5.1804, Postgres 14.4, Sysbench 1.1.0 (64 threads)
- 3.84TB vendor-A NVMe drive vs. 3.84TB ScaleFlux CSD 3000 (both PCIe Gen4x4)
- Transparent compression: **1TB** Postgres dataset ➔ **262GB**

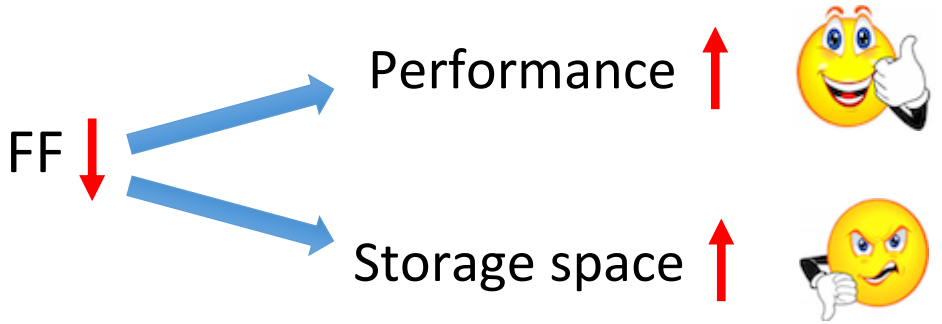


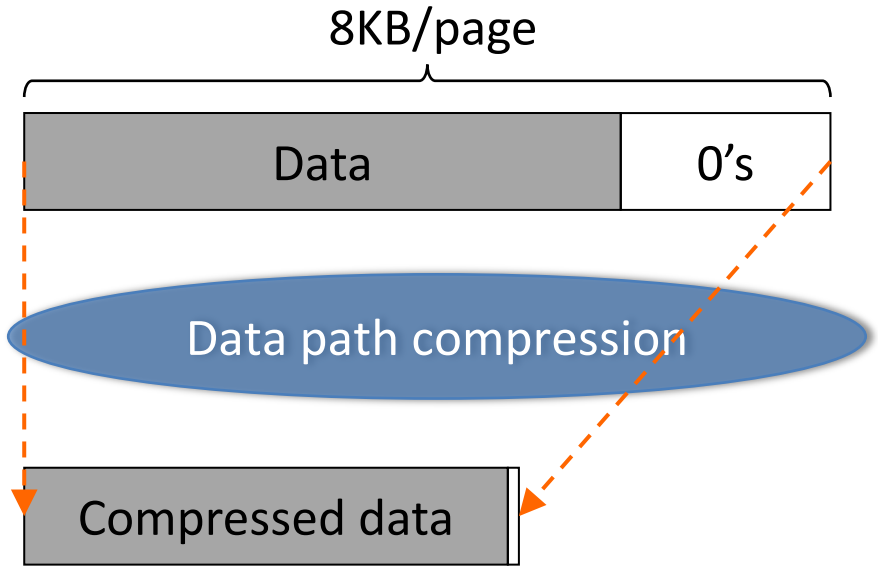

 

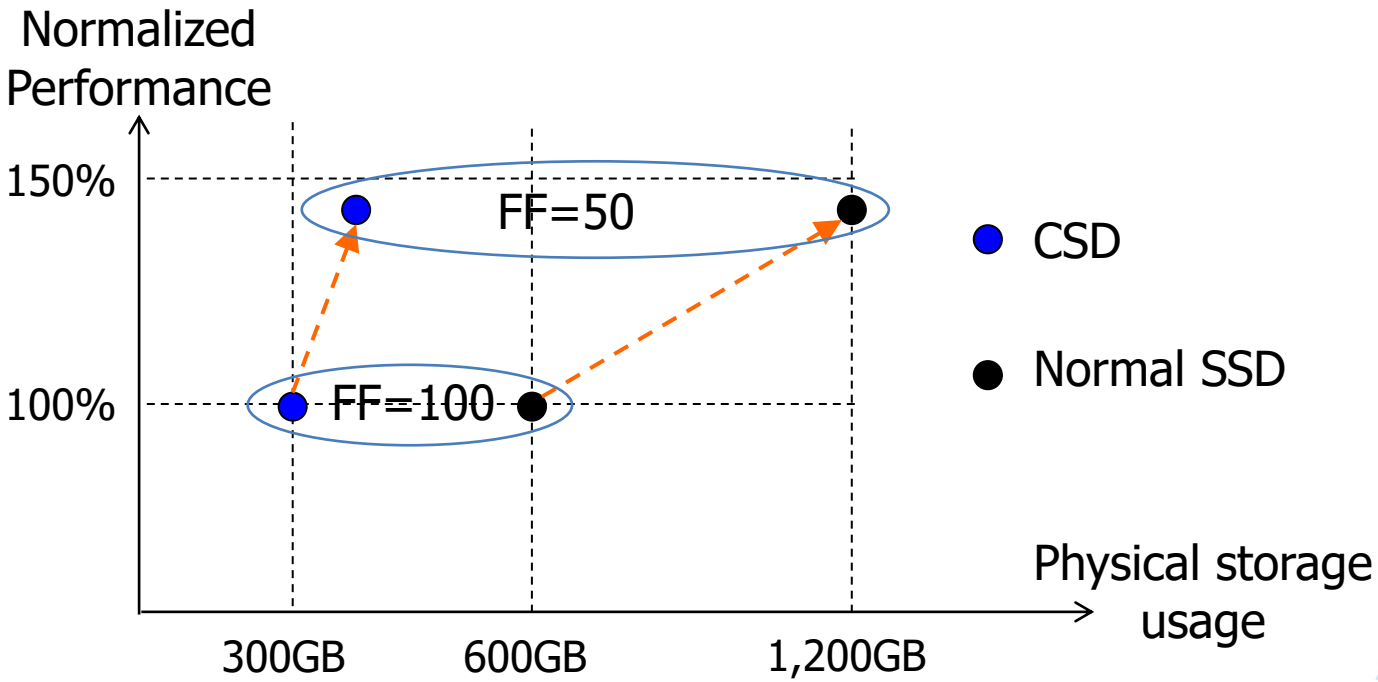
# One Step Further




Reserved for future update



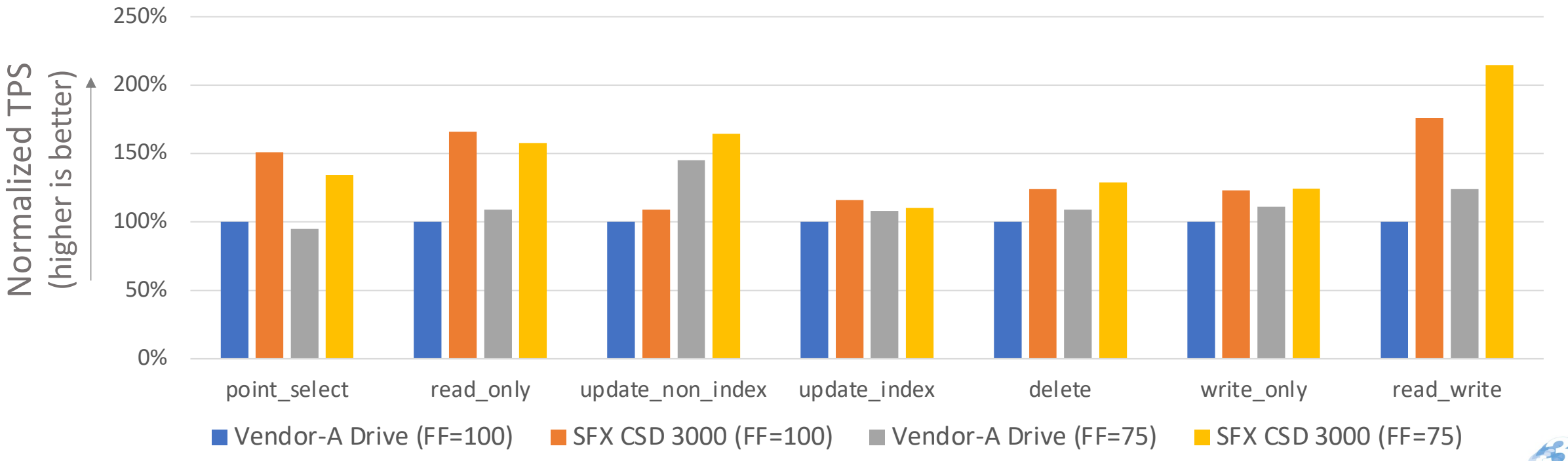



# One Step Further



| Fillfactor | Drive | Logical size (GB) | Physical size (GB) | Compression Ratio |
|---|---|---|---|---|
| 100 | Vendor-A | 1,000 | 1,000 | 1.00 |
| | SFX CSD 3000 | | **262** | 3.91 |
| 75 | Vendor-A | 1,600 | 1,600 | 1.00 |
| | SFX CSD 3000 | | **349** | 4.58 |

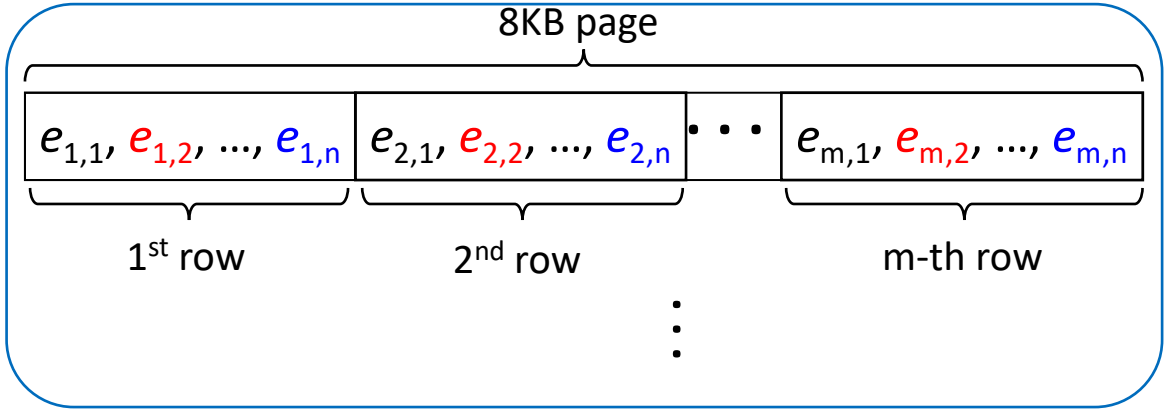


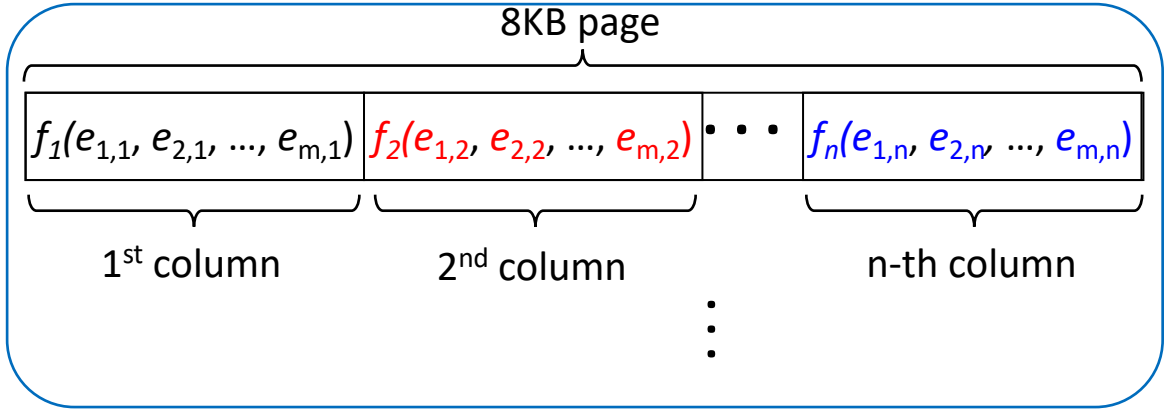

 

# More Innovation Opportunities



**In-memory** row-based pages ⟷ Row-column conv./trans. ⟷ **On-disk** column-based pages


- ❑ In memory: row-based format to best serve OLTP workloads
- ❑ On storage: column-based format to improve data compressibility ➔ lower storage cost

# Conclusion



PostgreSQL

Journaling filesystem (ext4, XFS, …)

Normal block layer


**In-storage transparent compression: The perfect match with Postgres**