25 – Oracle In-Memory Databases

1. Real-Time enterprises
    A. efficient, agile, data driven
    B. Hardware Trends
        i. Larger, Faster disk and memory
    C. Technology
        i. TimesTen-In-Memory Database
        ii. Database In-Memory
        iii. In-Memory on Exadata storage
2. Background
    A. Dual format
        i. Row format : slower for OLAP, faster for OLTP
        ii. Column format : faster for OLAP, slower for OLTP
        iii. Since both format has pros and cons,
        So make row format at buffer cache and column format at in memory
        txn uses buffer cache, and analytics uses in-memory
    B. In-memory column format
        i. In-Memory Compression Unit
            1. Dictionary compress and bitmap them
            2. RLE compression (store data just once and map with count of contiguous same data)
            3. OZIP compression
            (find pattern of RLE compression and compress it)
        ii. SIMD vector Processing
            A. Only process column that is needed

3. Top-5 Oracle Database In-Memory Innovation
    A. Dual Format Architecture
        i. If some rows are modified, just set a bit to make invalid row.
        analytics will ignore invalid row by seeing invalid bit.
        ii. Fast Background Repopulation
            1. Double buffering : make new IMCU, if complete, switch it.
            2. Incremental Repopulation :
            when build new IMCU, use metadata to allow quick formatting
            3. Column level invalidations :
            modifying data mostly touches some column, not all column
            So column level invalidation is possible
        iii. Accelerates Mixed Workloads
    B. Vectorized Analytics
        i. SIMD vector Processing
        ii. In-Memory Joins
        iii. In-Memory Aggregates
    C. In-Memory + Exadata
        i. Exadata
            1. Store nodes : do some filter and send data to compute nodes
            2. Compute nodes : doing higher level SQL operation
        ii. Fault tolerance
            1. Each column duplicated across 2 or more nodes
    D. Intelligent Automation
        i. Observe access patterns with “heatmap”
        (which is hot, which is cold, etc)
        ii. Classify data, Populate data with this classification

E. Converged Analytics

4. What is it?

A. Detailed logic of IMCU

B. Detailed logic of OZIP compression

C. Hot to decide which data is cold or hot