



# Multi-Temperature Data Warehousing:

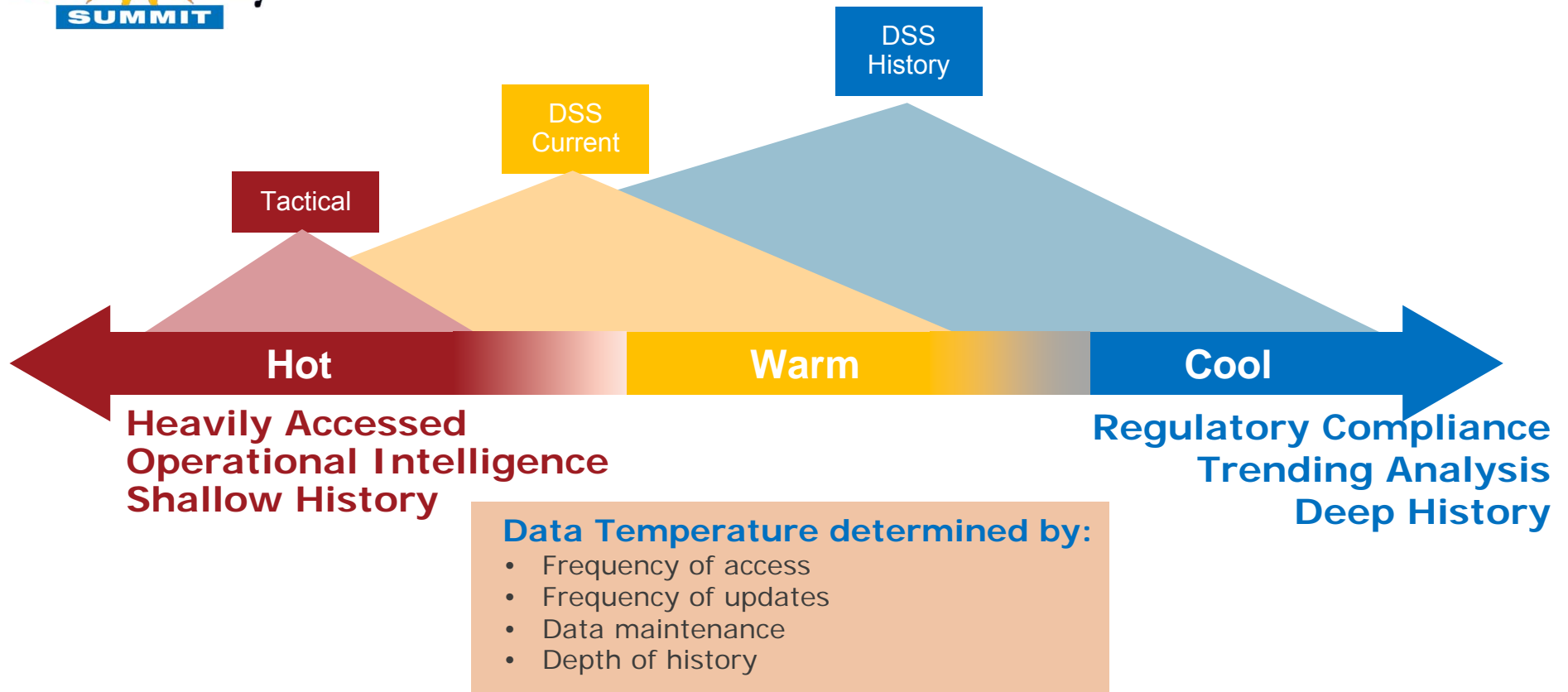
## Managing Data According to Frequency of Use

Rob Kubo

Director, Platform Product Management

Teradata Corporation

# What Is Multi-Temperature Warehousing?



**A Multi-Temperature Warehouse has the ability to prioritize the use of system resources based on business rules while maximizing utilization of storage with ever increasing capacity**

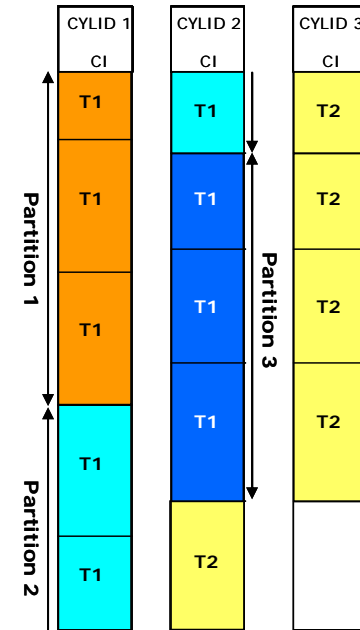
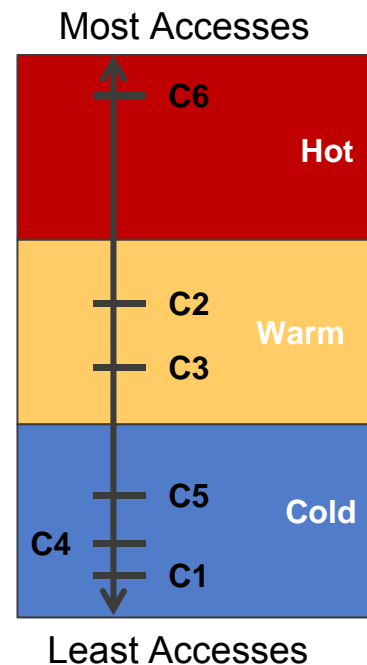


# Temperature Management

## – Metric Collection –

- Data block access frequency is tracked and aggregated up to the allocation unit (cylinder) level;
- Temperature is used to represent frequency of access;
  - Internally, TVS manages temperature at a very granular basis but externalizes temperature as follows:
    - HOT** – Most frequently accessed
    - WARM** – Moderate access
    - COLD** – Very light access

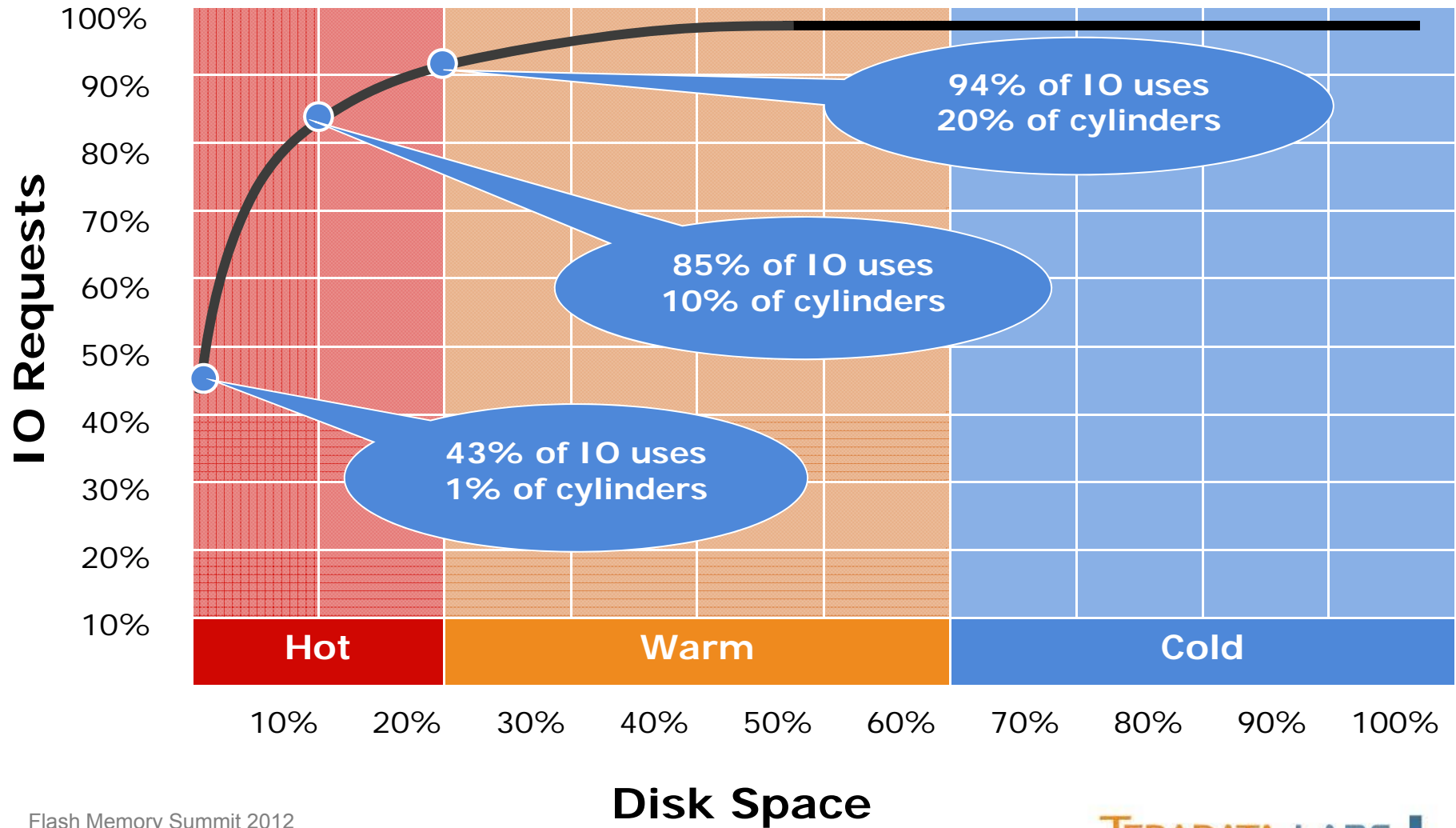
Cylinder #	Accesses
1	20
2	400
3	300
4	39
5	100
6	1000



- Cylinder = unit of storage allocation which is contiguous space on disk
  - > Two sizes ~2MB and ~12 MB
- Rows are stored in data blocks in RowID order;
- Data blocks are stored in cylinders in TableID / RowID order;



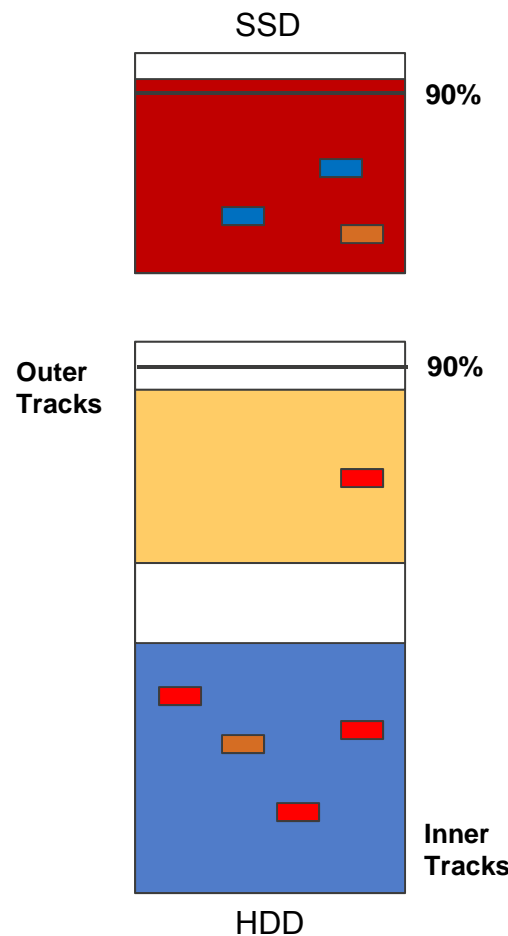
# Data Temperature Tool: 7 Day Trace





# Automated Data Migration

- TVS will automatically perform one migration operation per Parallel Unit every 5-minutes
  - Move a HOT cylinder in SLOW storage to a faster location;
  - If no space exists in the faster storage, a COLD or WARM cylinder will first be migrated out of the way;
- Requires no DBA or user involvement – Completely automated;
- Consumes no more than 1-2% of system resources (CPU and I/O)



Automated threshold based migration

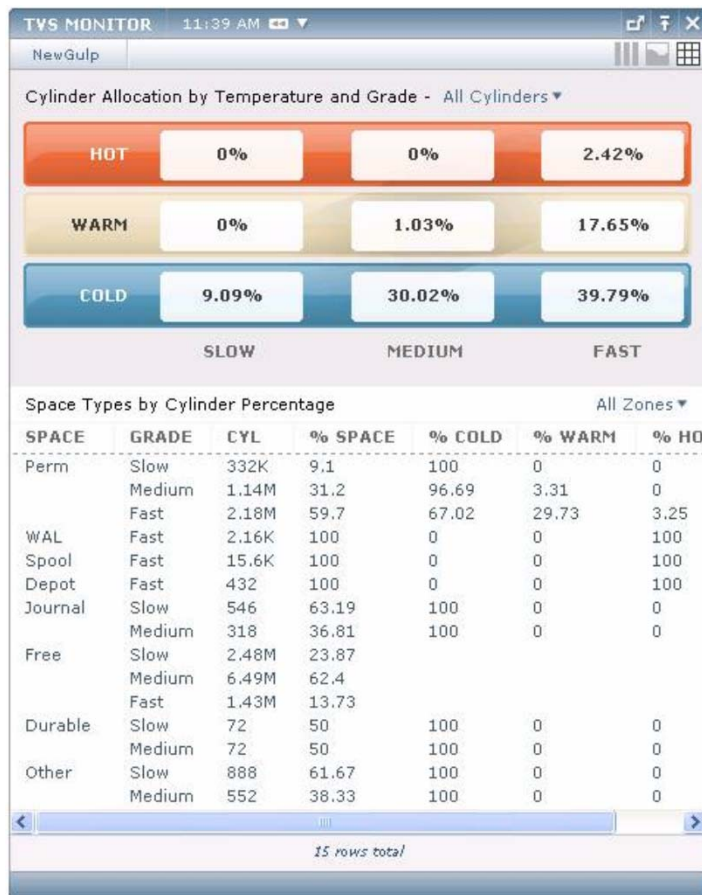
- Invoked when FAST or MEDIUM storage exceed 90% capacity;
- Migrates COLD and WARM data out of FAST, and COLD data out of WARM until both FAST and WARM drop below their 90% capacity mark.
- Background migration of colder data out of the faster locations ensures there is space for new HOT allocations;
- Overhead is included in the 1-2% of system resources previously mentioned



# GUI Monitoring – Viewpoint Portlet

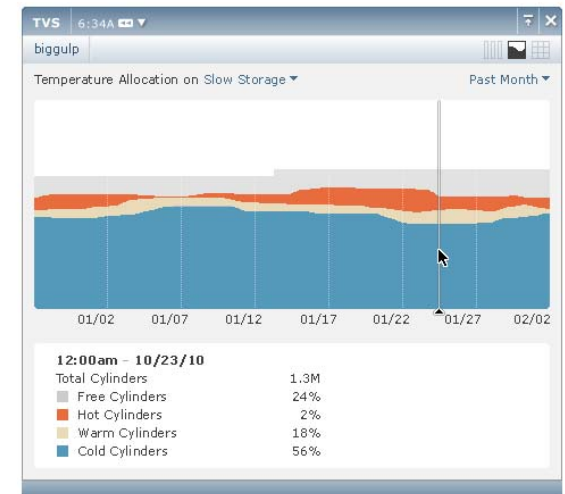
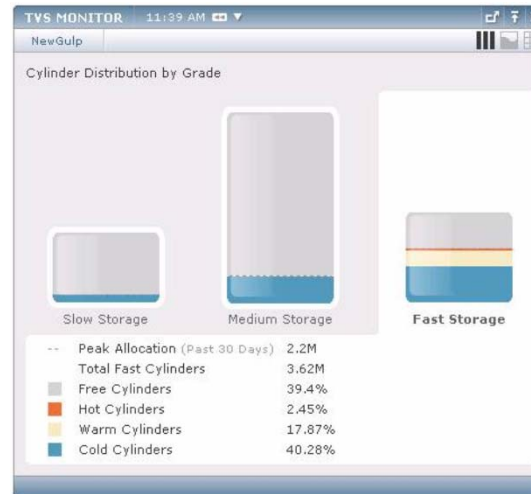
Display and filter cylinders by temperature and use (e.g. spool)

- Available with Teradata Management Portlets
- Monitor TVS activity, cylinder data temperatures, detailed usage



View cylinder distribution

View hot, warm and cold cylinder migration trends

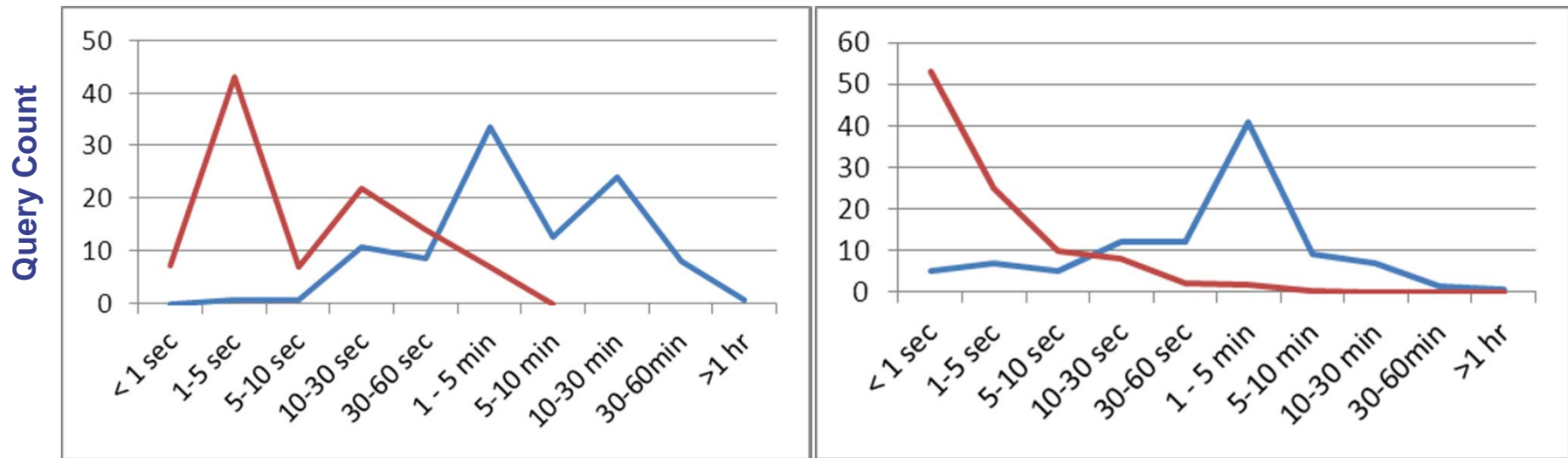




# Hybrid Storage @ Work – Examples

- Health Care Financial Provider -

### High Priority “Standard Short” Queries



HDD ONLY — (blue line)  
HYBRID — (red line)



# Hybrid Storage @ Work

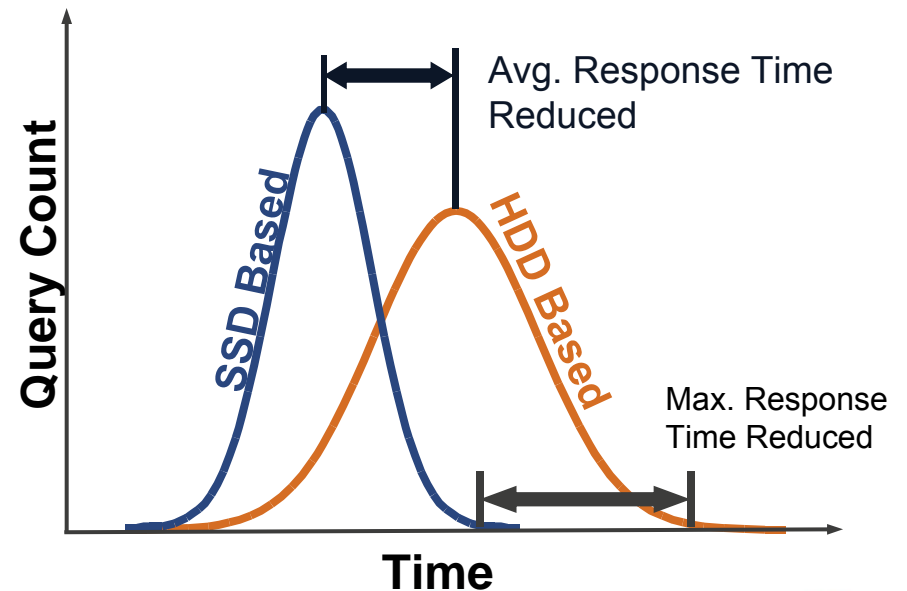
- Large Food Retailer -

## Early production performance results

- Migration of two systems to Hybrid storage
  - “A”: 5550 (HDD) to 6690 (HYBRID)
  - “B”: 5550 (HDD) to 6690 (HYBRID)
- 6690 (HYBRID) storage config:
  - 15 SSD (400GB) & 60 HDD (300GB) per node
    - 25% FAST storage data space
- Microstrategy queries - average response times
  - 22% reduction on System A
    - Up to 45% reduction of peak time response
  - 50% reduction on System B
- Microstrategy reports
  - 6 different reports – run times of six (6) min to one (1) hour
  - Run times reduced by 3% to 60%

	HDD	Hybrid
Response Time - All Jobs	0:00:52	0:00:39
Response Time - Prime Time	0:00:35	0:00:29
Response Time - Non Prime Time	0:01:37	0:01:07

## Faster and More Consistent Query Response Times



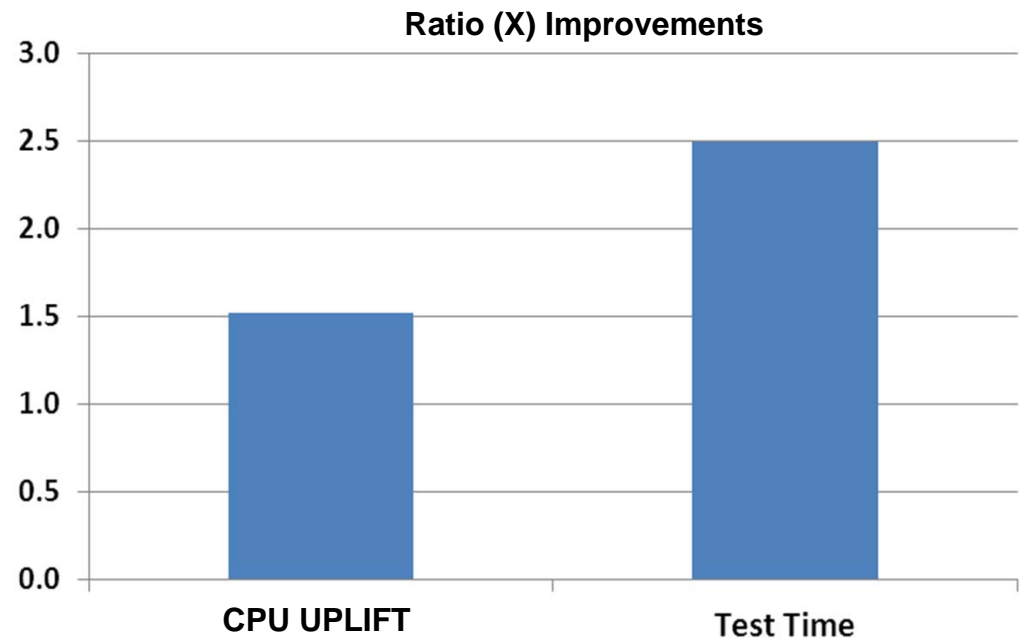




# Hybrid Storage @ Work

*- Manufacturing Industrial -*

- Migrated to Hybrid Storage
  - From 12N 5550 (HDD) to 6N 6680 (HYBRID)
  - 1.5X performance capability (CPU Uplift Based)
  - 18 SSD/36 HDD per node
- Stress test - 7,692 SQL statements
  - Simulate 52 distinct functional users
  - WAS: 14 h 28 m
  - NOW: 5 h 50m





## Hybrid Storage @ Work - U.S. Based Retailer -

- Migrated from 1N 5550 (HDD) to 1N 6680 (HYBRID)
  - 50% increase in CPU Uplift
  - 16 SSD, 46 HDD
- Microstrategy reports
  - Run at 7AM every Sunday.
  - WAS: 10-11 hours; NOW: 2-3 hours
- Long running ETL jobs
  - WAS: 57 Min; NOW: 7 Min

