INTRODUCING SYBASE IQ 15.3

WITH PlexQ® TECHNOLOGY
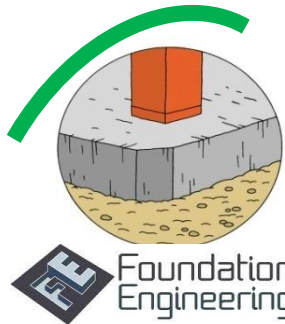
Joydeep Das, Sybase Analytics Product Management

August 22, 2011
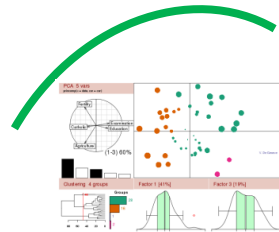
SYBASE IQ 15.x

A platform (r)evolution

v15.0 — Foundation Engineering — **March, 2009**

v15.1 — In-Database Analytics — **July, 2009**

v15.2 — **June, 2010**

v15.3 — Introducing Sybase IQ PlexQ® Technology — **June, 2011**

SYBASE® | An SAP Company
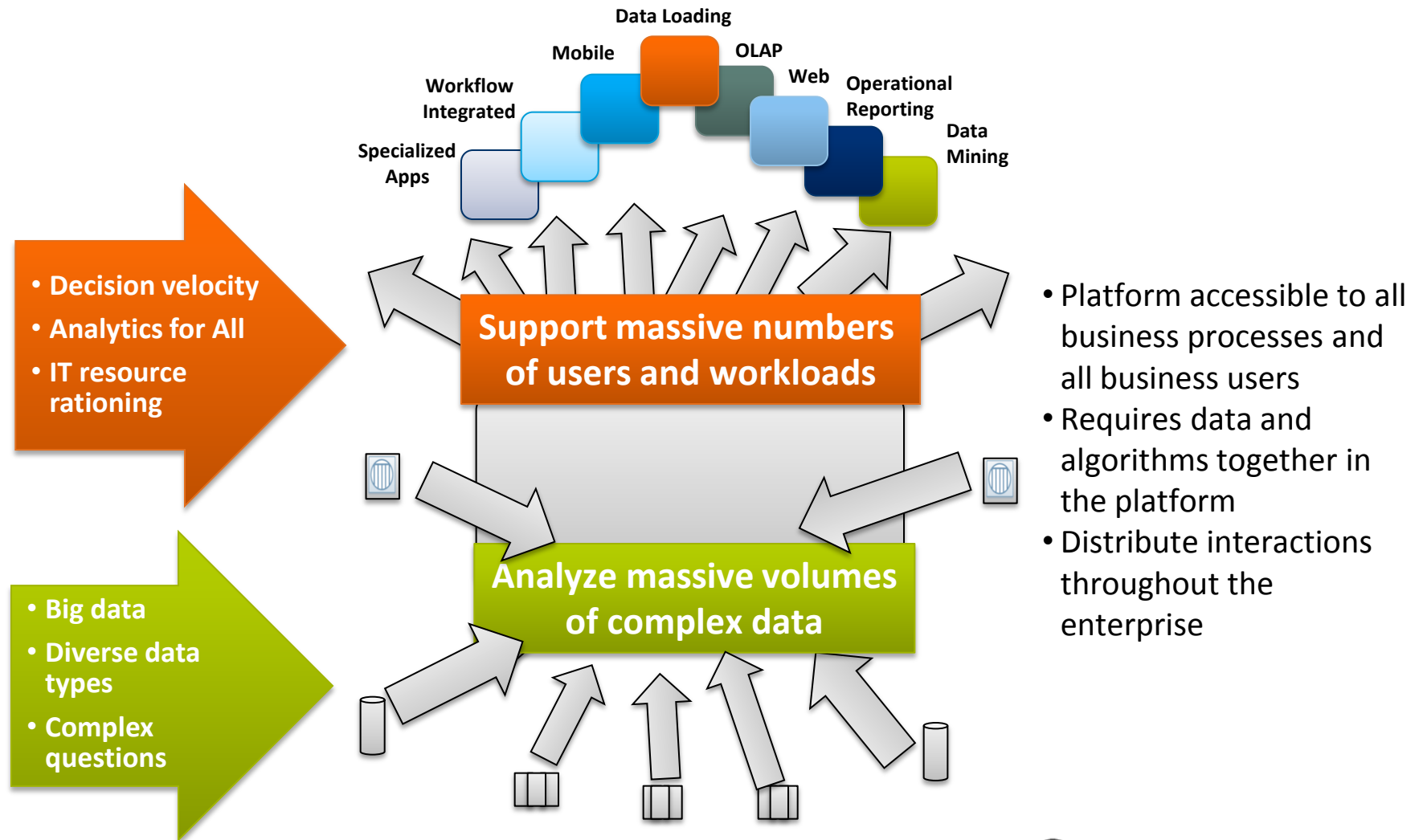
SYBASE IQ 15.3 MOTIVATIONS

Business analytics challenges to overcome

- Big data
- Diverse data types
- Complex questions
- Decision velocity
- Analytics for all
- It resource rationing

SYBASE IQ 15.3 MOTIVATIONS

Business analytics needs a new platform

- **Decision velocity**
- **Analytics for All**
- **IT resource rationing**

- **Big data**
- **Diverse data types**
- **Complex questions**

Support massive numbers of users and workloads

Analyze massive volumes of complex data

Data Loading
Mobile
OLAP
Workflow Integrated
Web
Operational Reporting
Specialized Apps
Data Mining

- Platform accessible to all business processes and all business users
- Requires data and algorithms together in the platform
- Distribute interactions throughout the enterprise

SYBASE® | An SAP Company

INTRODUCING SYBASE IQ 15.3 PlexQ®

Massively intelligent, intelligence for everyone

SYBASE IQ 15 *Redefined*

MASSIVELY INTELLIGENT, INTELLIGENCE FOR EVERYONE

- Performance
- Flexibility
- Scalability
- Data Management

DRIVES BUSINESS TRANSFORMATION

Build analytics into **Applications & Processes**

Consolidate BI projects with **Virtual Data Marts**

Rationalize EDW with **Elastic PlexQ® Grid**

Drive insight with **Predictive Analytics**

Empower users with **Self-Service Analytics**

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ® KEY FEATURES

Massively intelligent, intelligence for everyone

Massively Parallel Processing (MPP) – Shared Everything

Distributed Parallel Query Operators

Extract, Load, Transform Jobs (InfoPrimer)

Logical Servers

Application + Workload Partitioning

Elastic Resource Provisioning

Sybase IQ 15.3 PlexQ®

Smart Applications

Web Services API

Ruby On Rails Driver
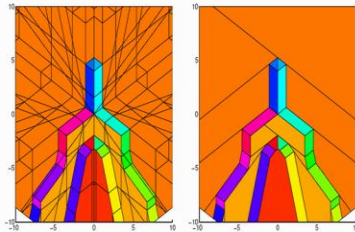
Multimedia Analytics Plug-In

Productivity Tools

Web based Administration & Monitoring

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ® MPP

Why bother?

Challenges

Very
Large
Data
Bases

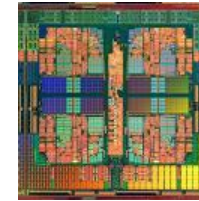**Data Size**   **Complexity**

Expectations

**High Performance,
Cost effective**

Opportunities

**Multi-core
Chips**   **Cost Effective
HW**

Sample problems to solve

- Scenario planning simulations
- Aggregation of large data sets
- Matrix computations

Expectations from infrastructure

- Consistent high performance for same workload
- High performance across a spectrum of workload
- Judicious, cost effective use of resources
- Scale out without tuning
- High resiliency in event of hardware failures

SYBASE IQ 15.3 PlexQ® MPP

An evolution in scalability for data, complex workloads, and users
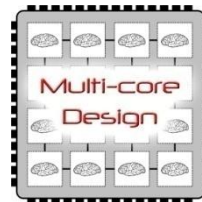
Sybase IQ 12.x
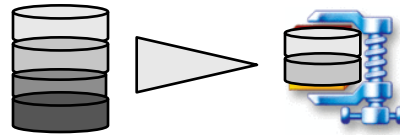
Storage Scalability with Compression

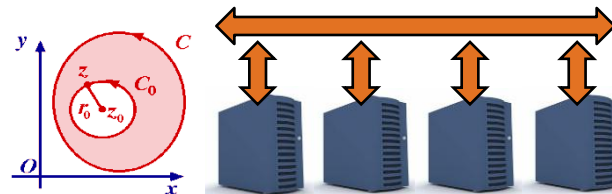User Scalability with Multiplex

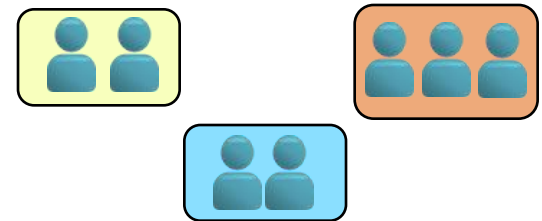**Sybase IQ 15.0**

Multi-core Load / Query parallelism

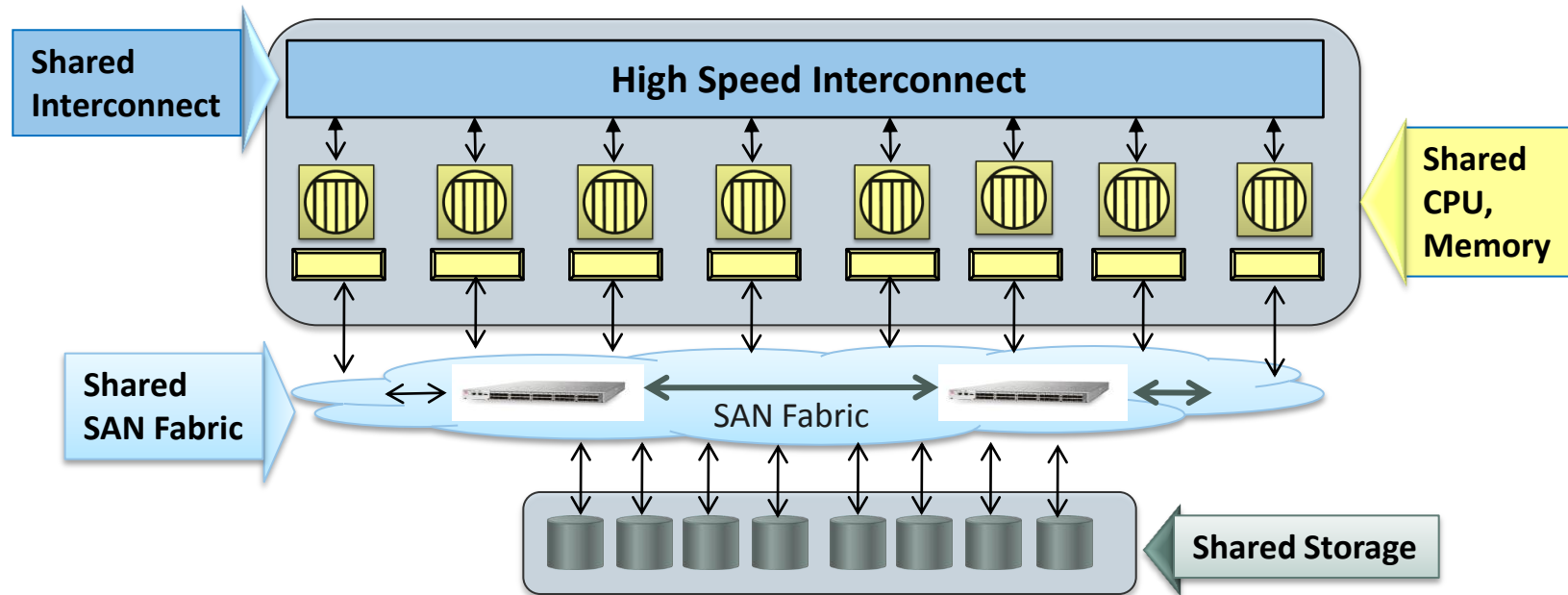Grid based multi-table parallel loads

Sybase IQ 15.3

Scale query processing across grid

Create and scale user communities

SYBASE® | An **SAP** Company

SYBASE IQ 15.3 PlexQ® MPP

Shared Everything with Distributed Query Processing (DQP)

Shared Interconnect

High Speed Interconnect

Shared CPU, Memory

Shared SAN Fabric

SAN Fabric

Shared Storage

Characteristics

- Compute and storage resources are decoupled
- Single query can span multiple nodes, disks
- Dynamic, versatile resource allocation
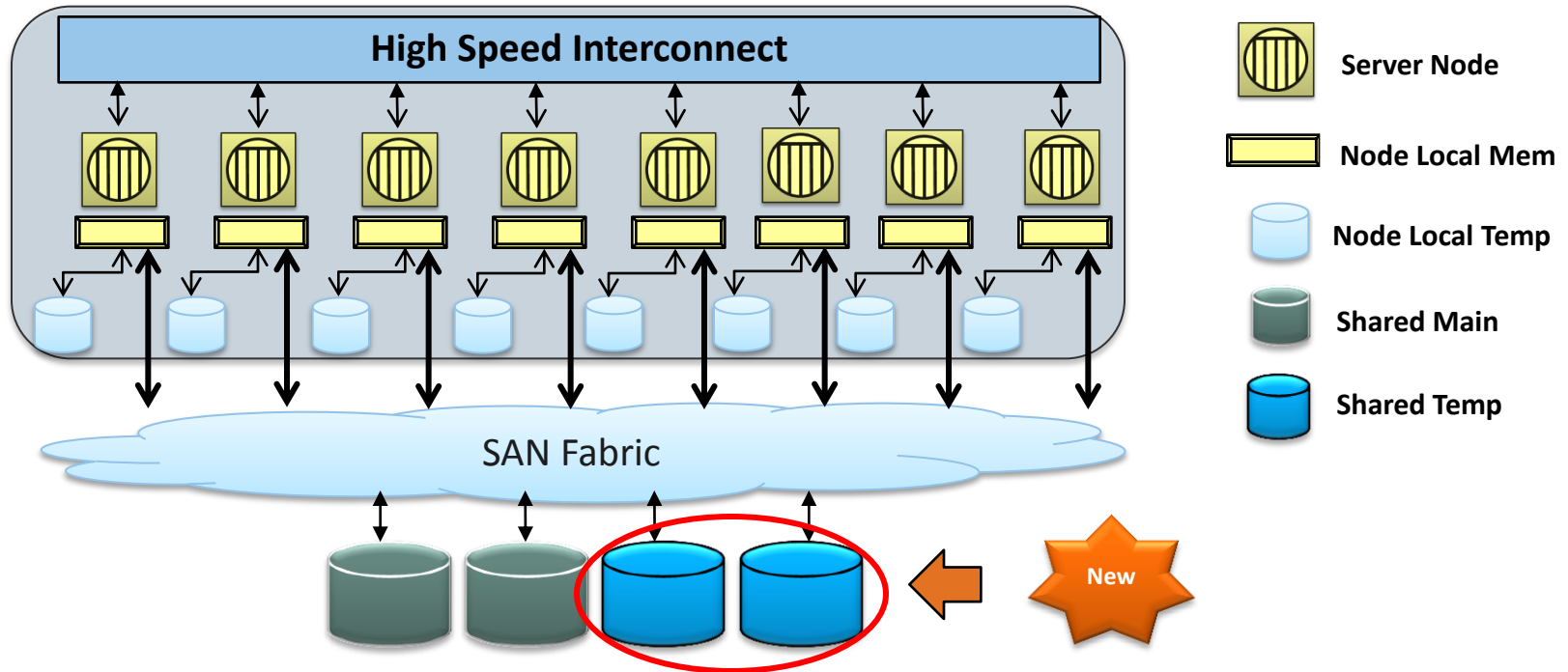 - Fully shared or private

Advantages

- Economical query scale out across low cost HW units
- Independent scale out of compute and storage
- Simple, non-partitioned data placement
- Scales across spectrum: batch, ad-hoc, concurrent

Addresses: big data, complex queries, analytics for all

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ® MPP

Shared Everything — shared temporary storage for DQP

High Speed Interconnect

SAN Fabric

New

Server Node

Node Local Mem

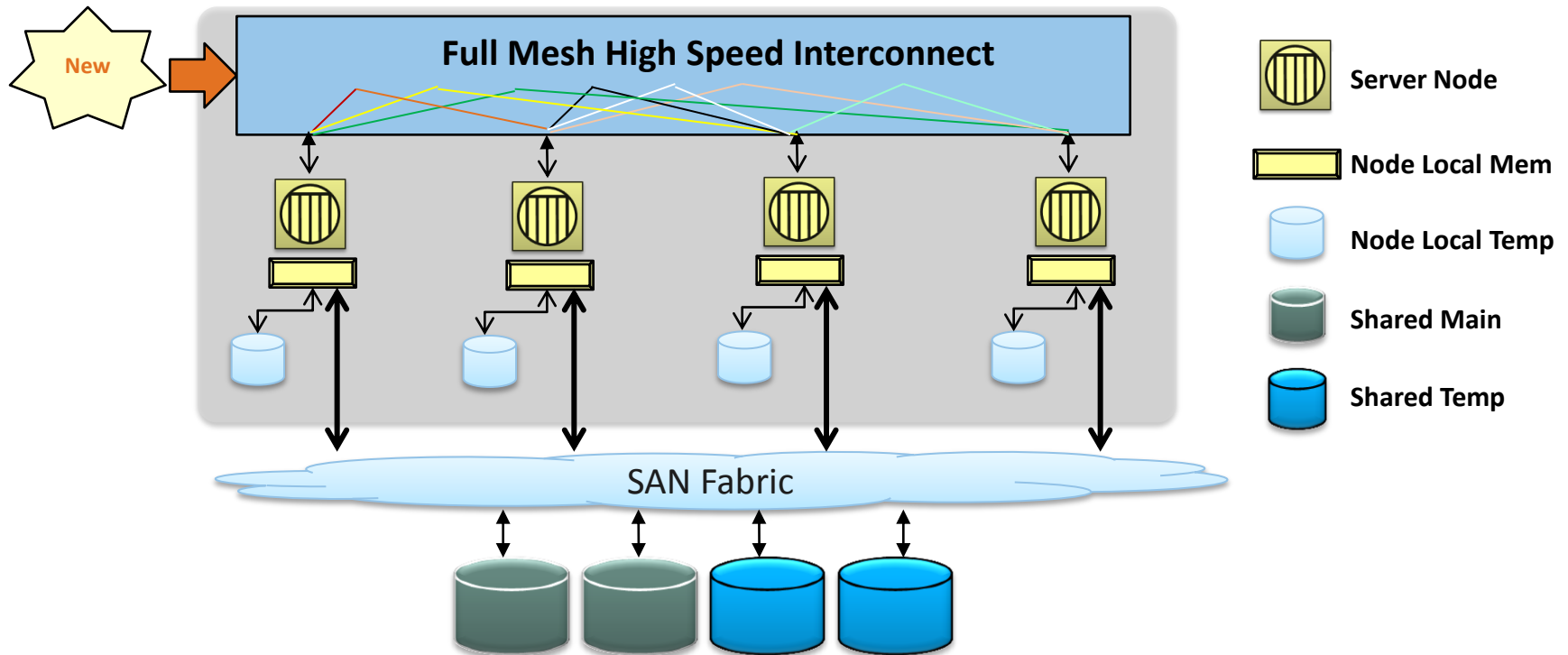Node Local Temp

Shared Main

Shared Temp

Shared temporary space

- DQP heavily uses new dbspace IQ_SYSTEM_SHARED_TEMP for sharing temporary data
    - More efficient than local temporary store — write once, consume multiple times
- New buffer cache manager for IQ_SYSTEM_SHARED_TEMP
    - Will keep information for both local and shared temporary spaces

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ® MPP

Shared Everything — shared temporary storage for DQP

New

Full Mesh High Speed Interconnect

SAN Fabric

Server Node

Node Local Mem
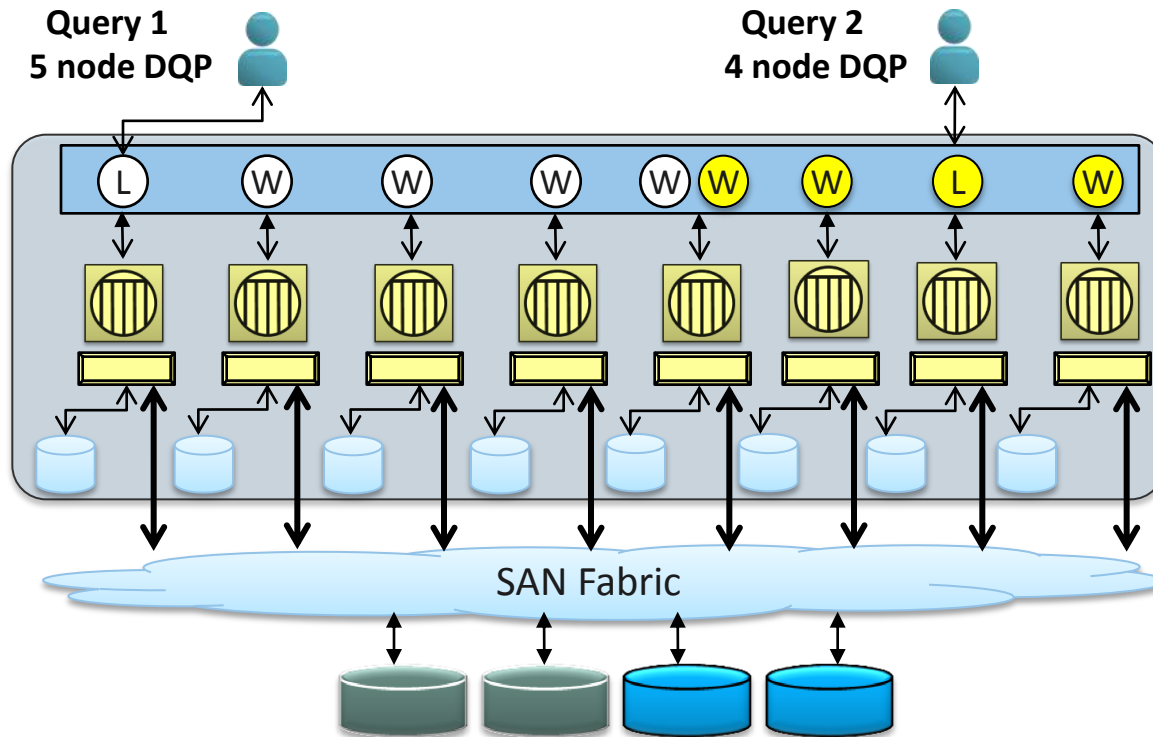
Node Local Temp

Shared Main

Shared Temp

Full mesh interconnect
- Inter node process communication framework
- Bi-directional peer-to-peer full mesh for DQP
  - At most two connections active per node
  - Small binary payloads
  - Secure and resilient

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ® MPP
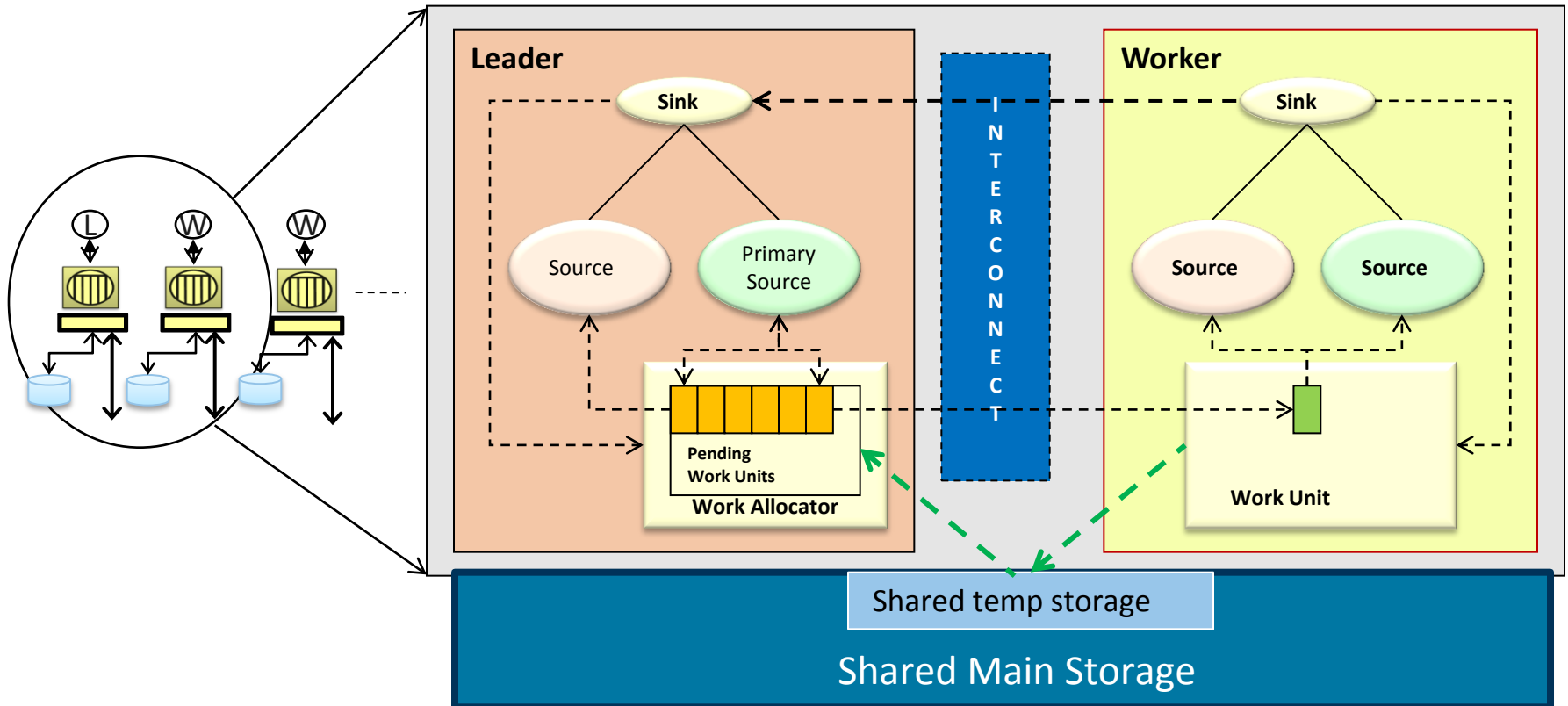
Shared Everything DQP fundamentals

Execution setup
- **Leader node:** receives and initiates query
    - Any node can be a leader, one leader per query, many concurrent Leaders possible
- **Worker node:** nodes picking up work units from leader
    - Many worker nodes per query, same worker node can serve multiple queries

SYBASE IQ 15.3 PlexQ® MPP

Shared Everything DQP workflow

Leader

Sink

Source

Primary Source

Pending Work Units

Work Allocator

INTERCONNECT

Worker

Sink

Source

Source

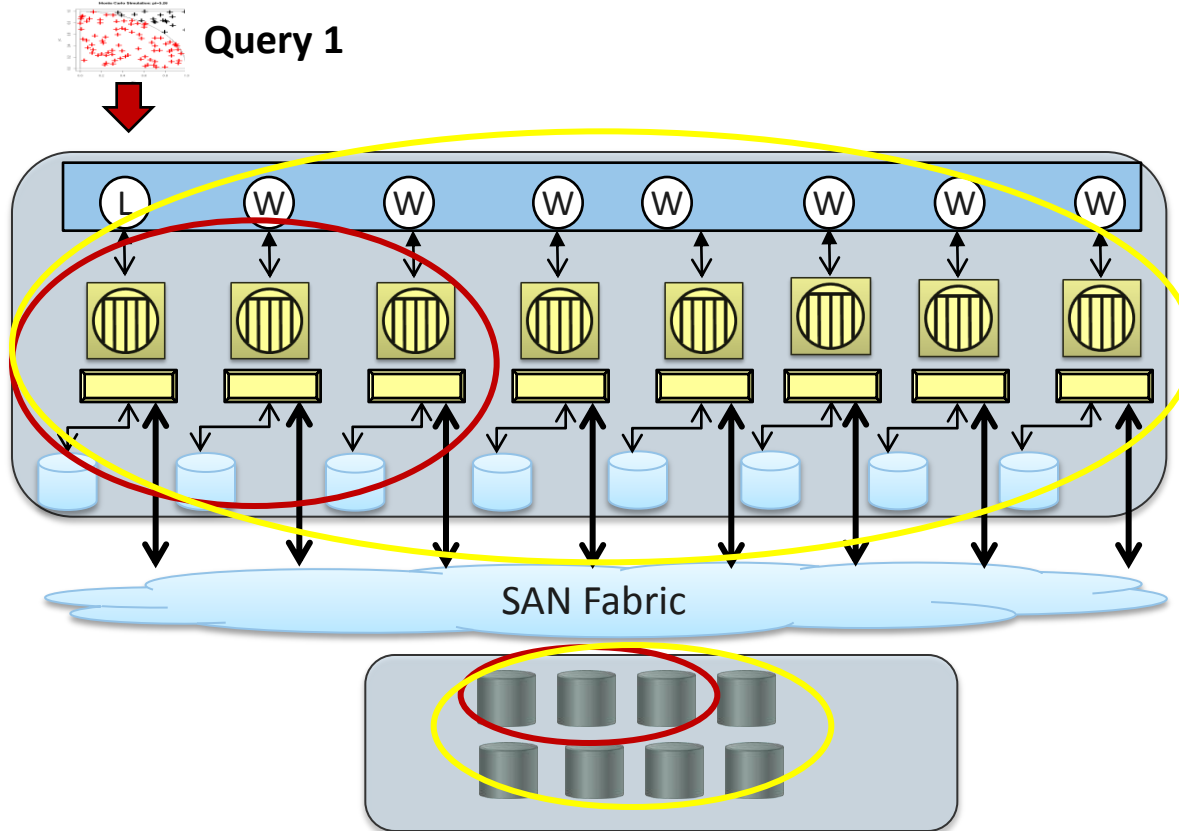Work Unit

Shared temp storage

Shared Main Storage

DQP execution: democratic & collaborative

- **Leader** multicasts workmap with workunits
- **Workers** receive, prepare, pickup work units for execution
- **Automatically** workload balanced via "pull" model

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ® MPP

Shared Everything DQP resource usage scenarios
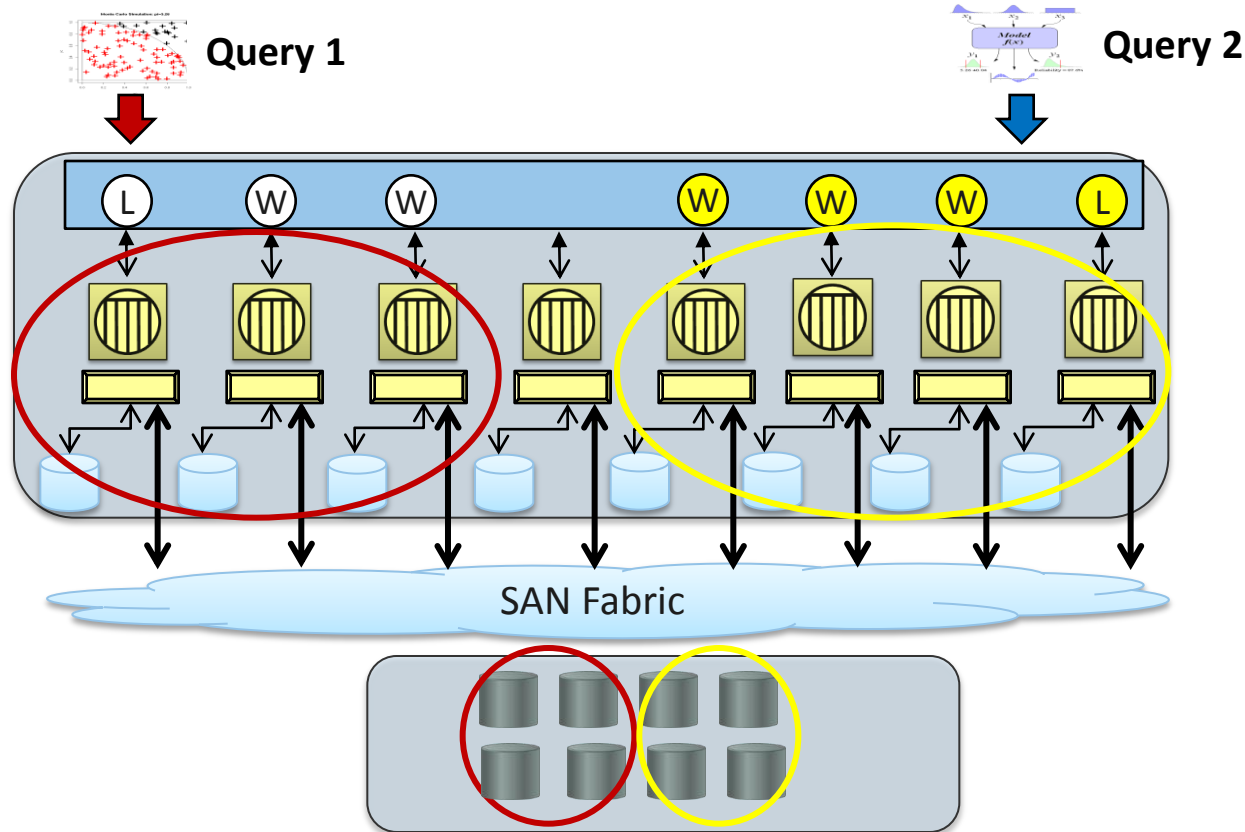
Query 1

SAN Fabric

Scenario I: single query across some or all resources
- **Some resources:** Monthly bank revenue grouped by every month in the quarter
- **All resources:** Daily retail inventory across all stores in the chain

SYBASE IQ 15.3 PlexQ® MPP

Shared Everything DQP resource usage scenarios

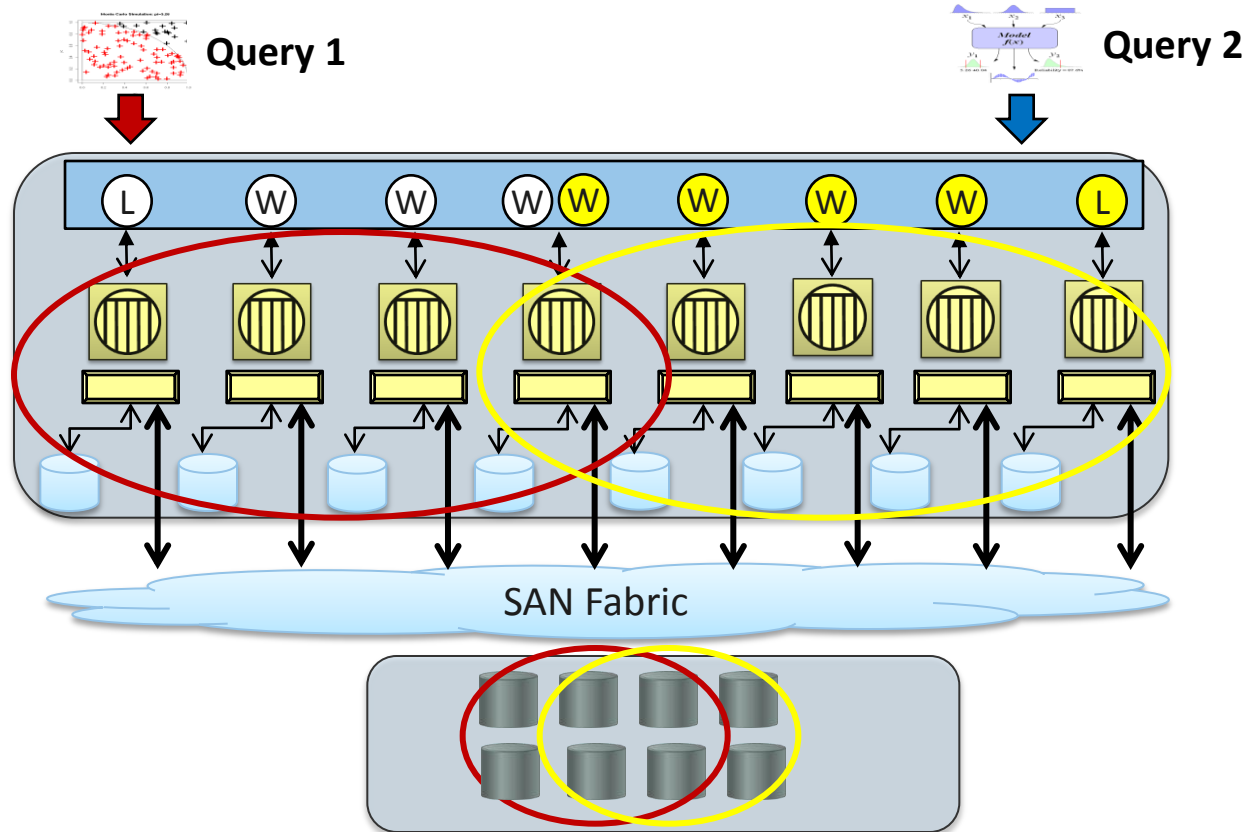

Query 1

Query 2

SAN Fabric

Scenario II: concurrent queries mutually exclusive resources
- **Query 1:** Finance queries across revenue, margin data
- **Query 2:** Marketing queries across campaign response data

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ® MPP

Shared Everything DQP resource usage scenarios



Query 1

Query 2

SAN Fabric

Scenario III: concurrent queries overlapping storage resources, overlapping compute resources
- **Query 1:** Marketing queries across campaign response AND Sales data at Peak Time
- **Query 2:** Sales queries across geographic sales data AND marketing campaign data at Peak Time

SYBASE IQ 15.3 PlexQ®

With Virtual Data Marts (VDM) — a deeper look

Full Mesh High Speed Interconnect

SAN Fabric

VDM1
VDM2
Shared
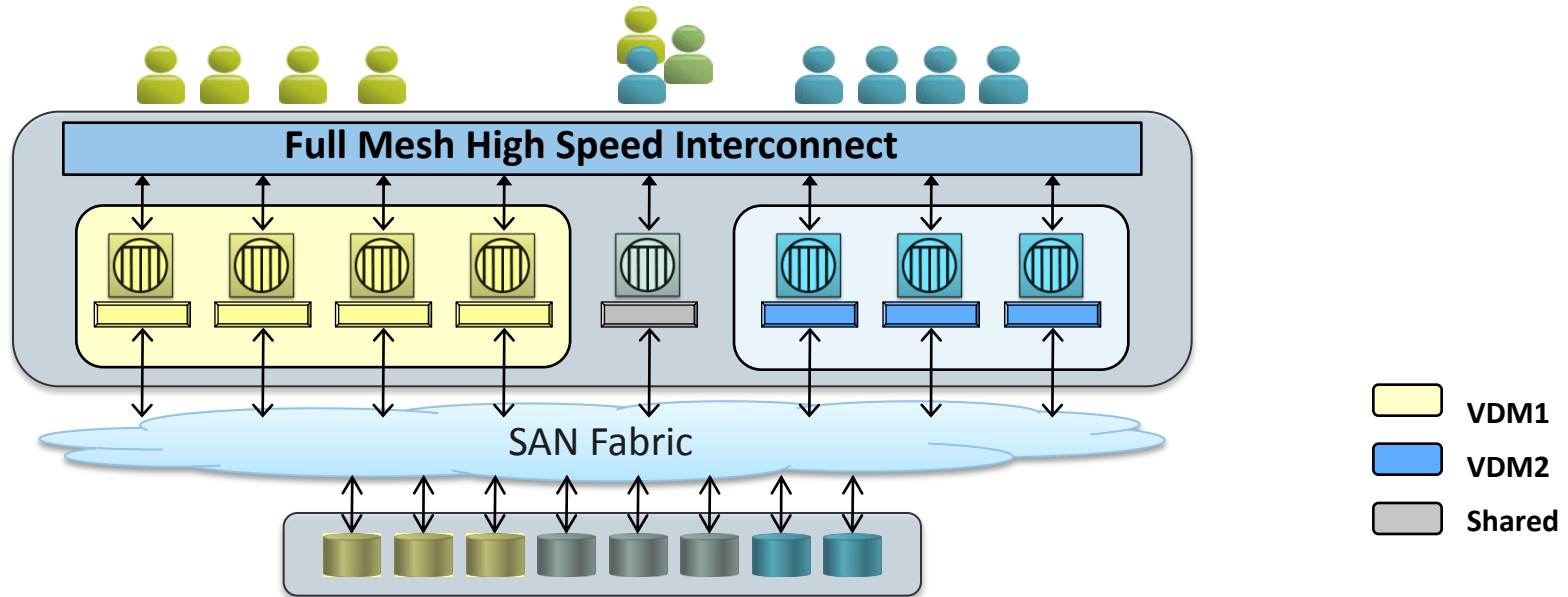
Characteristics

- VDM and works via login permission control
- VDM can isolate applications, workload, users
- DQP within VDM boundaries only
- Dynamic/scheduled resource reassignment b/w VDM

Advantages

- Effective user group assignment to resources
- Effective workload isolation and partitioning
- Elastic, flexible resource provisioning/utilization
- Economical scalability
- Ideal for MPP AND concurrent workload

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ®

Virtual Data Marts (VDM) — applicability

Full Mesh High Speed Interconnect

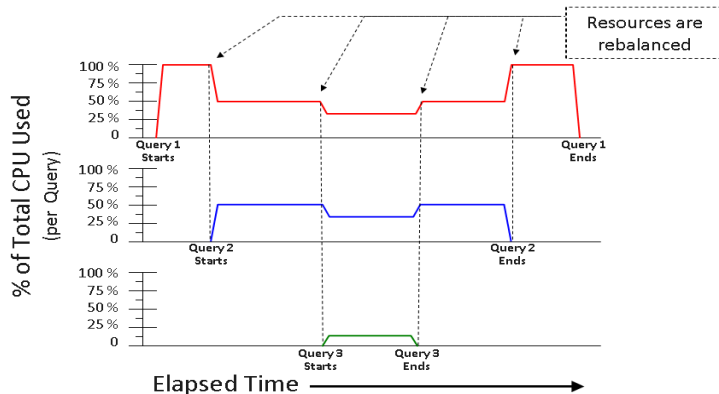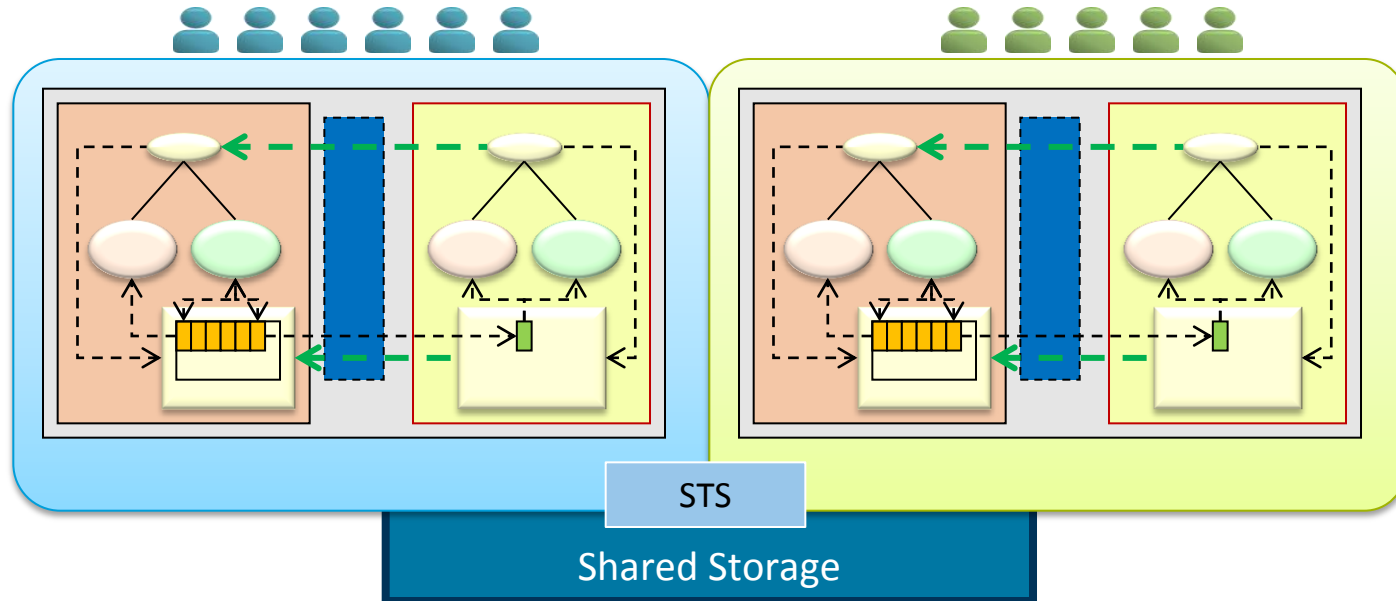**SAN Fabric**
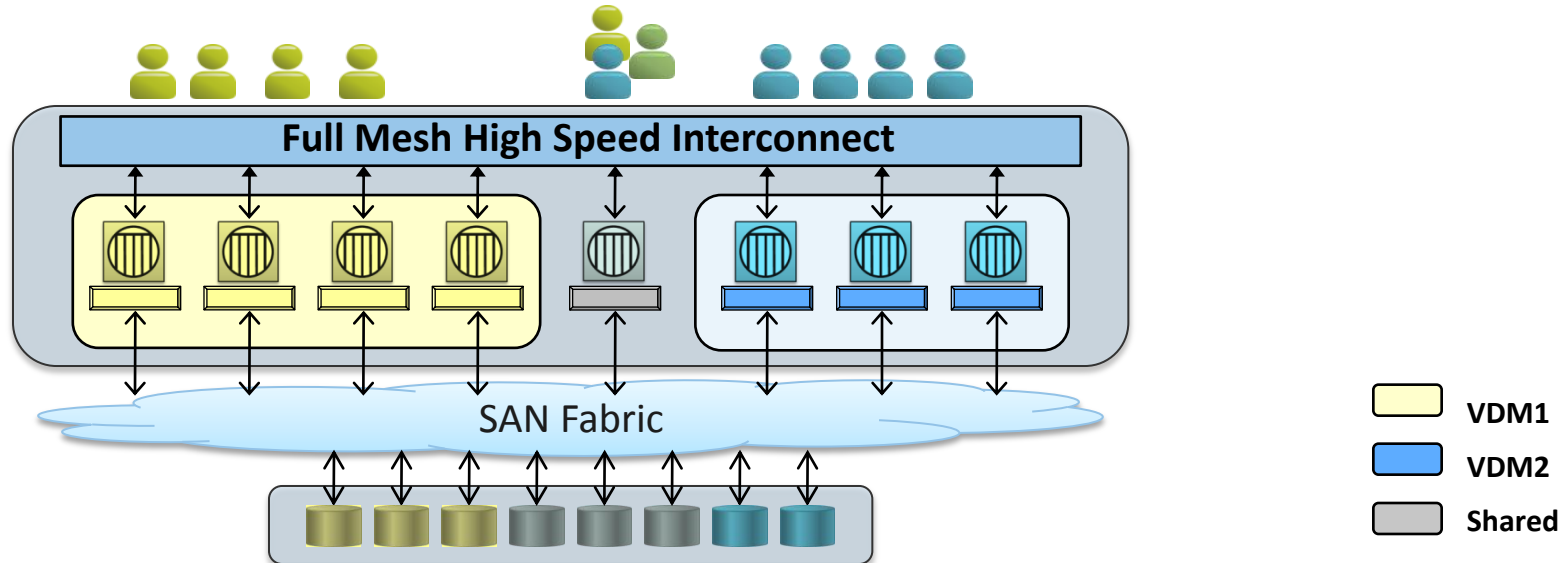
VDM1
VDM2
Shared

Use case scenarios

- **Workload isolation**
  - VDM 1: Data mining; VDM 2: reporting
  - VDM 1: Loading; VDM 2: queries
- **User isolation**
  - VDM 1: Finance; VDM 2: marketing
- **Private cloud provisioning**
  - Move 1 node from LS1 to LS2 at 5 pm every day
  - Add 1 node to LS2 at peak load time

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ®

MPP Shared Everything DQP concurrent workload balancing

STS

Shared Storage

Resources are rebalanced

% of Total CPU Used (per Query)

Query 1 Starts — Query 1 Ends
Query 2 Starts — Query 2 Ends
Query 3 Starts — Query 3 Ends

Elapsed Time

DQP execution: auto concurrent workload balance

- Automatic workload balancing
  - Any node can be a leader node
  - Every job is self throttling (even loads)
    - Shadow parallelism active for run time adjustment
  - Further workload isolation via VDM (Logical Server)

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ®

MPP Shared Everything — best of both worlds
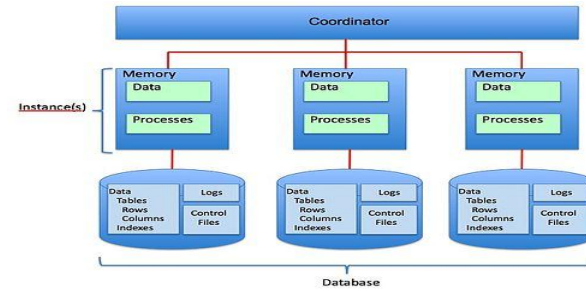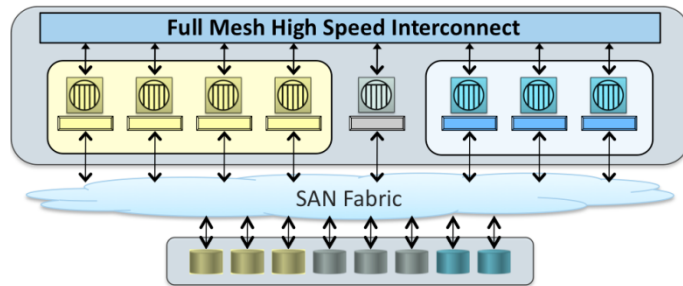
## PlexQ: MPP Shared Everything

- Massive independent, linear scale out
  - Compute OR memory OR storage
- Tremendous resource optimization
  - Dynamic resource provisioning
- No forcible data partitioning
- Well behaved, full spectrum support
  - Batch, ad-hoc, concurrent
- Mature DR support
- Less data duplication
- No single "choke-point" leader node
- SAN affordable and reliable

PlexQ: MPP Shared Nothing Contenders

! Massive homogeneous unit scale out
  x Compute + memory + storage in chunks
! Sub-optimal resource utilization
  x Share nothing !
! Forced data (re)partitioning
! Poorly behaved full spectrum support
  ✓ Good for complex queries (planned)
  x Struggle with ad-hoc, concurrent queries
- Complicated DR support but better fault tolerance
! Significant data duplication across nodes
! Master "coordinator" a "choke-point"
- DAS very affordable but more redundancy needed

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ®

**In-DB processing: extract, load, transform with Sybase IQ InfoPrimer**
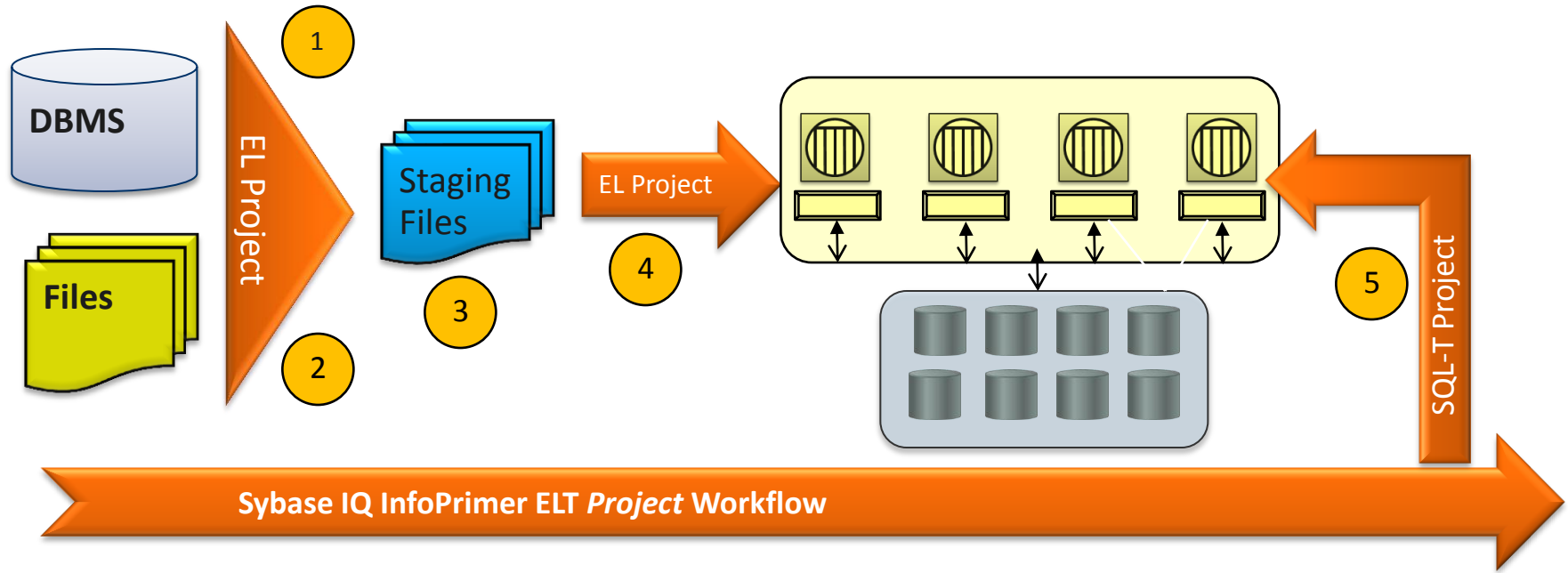
PlexQ InfoPrimer ELT: addressing decision velocity

1. EL: Fast and quick extract from DB sources (with low click count)
2. EL: Extract data from files (with low click count)
3. EL: Stage data into temporary files and optionally apply preprocessing steps, such as binary encoding
4. EL: High speed bulk load staged and source files into IQ
5. SQL-T: Transform loaded data to merge into tables – execute in Sybase IQ directly with error control
6. ETL: Path still possible, deploy transformation at the right point (ETLT)

Addresses: big data, decision velocity

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ®

Built-in web services with Soap API
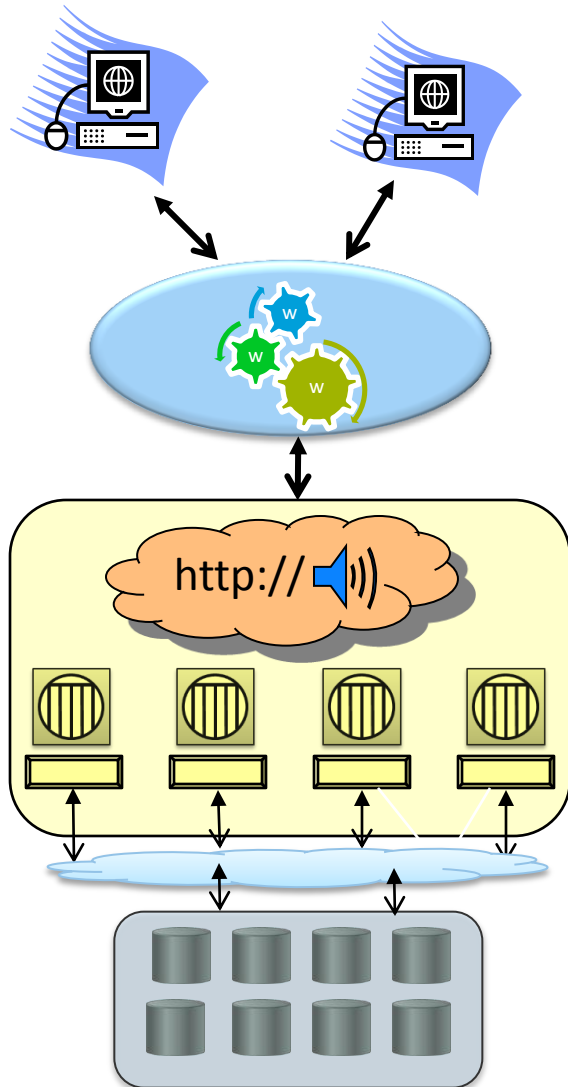
Sybase IQ PlexQ web services

- Web services is the most common orchestration backbone
- Robust bi-directional http listener — can accept client side requests & initiate server side requests
- Web services stored in Sybase IQ
    - Define which URLs are valid and what they do
    - SOAP 1.1, DISH, HTML, XML, JSON, RAW
- Simple commands to set up web services
    - Start Sybase IQ with http port enabled
    - Create procedures, functions, …
    - Create service, alter service, …
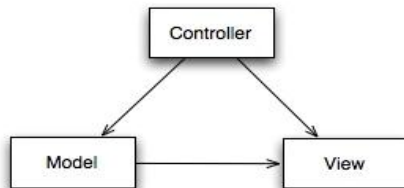
Addresses: Analytics For All, IT resource rationing

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ®

Web 2.0 integration with Ruby on Rails
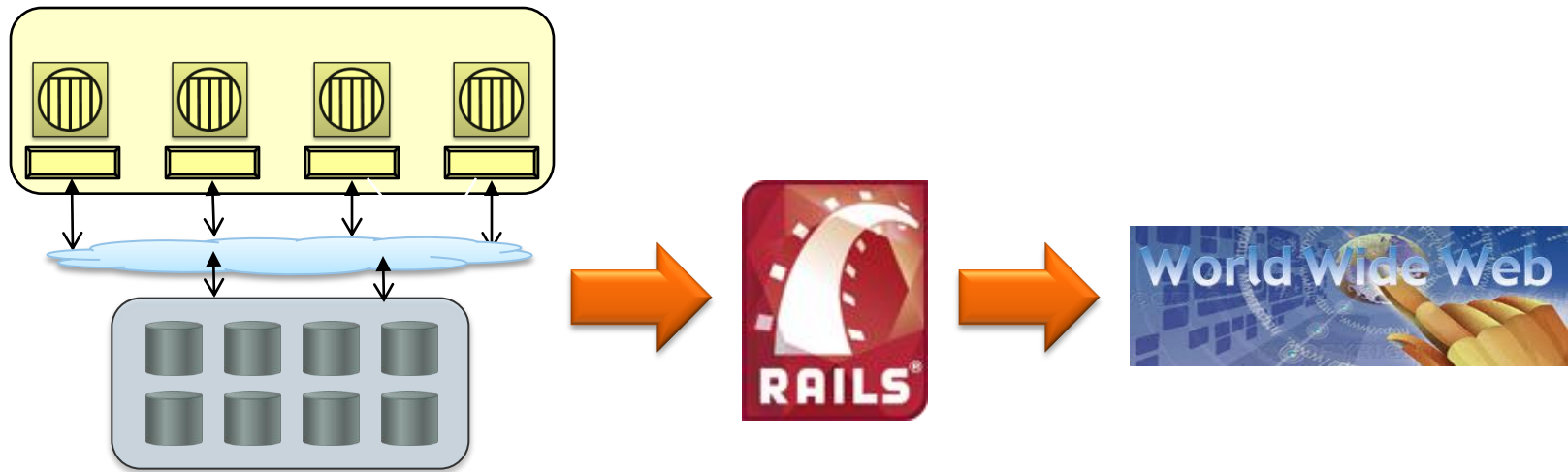
Ruby On Rails Fundamentals

- Open source web app framework for Ruby Language
- Agile development with Model (OR), view (Render), controller (workflow) with dynamic run time focus

The Controller is repsonsible for the control flow of the application.

The Model contains business data and behaviour.

The View renders Models to the user, triggered by the controller.
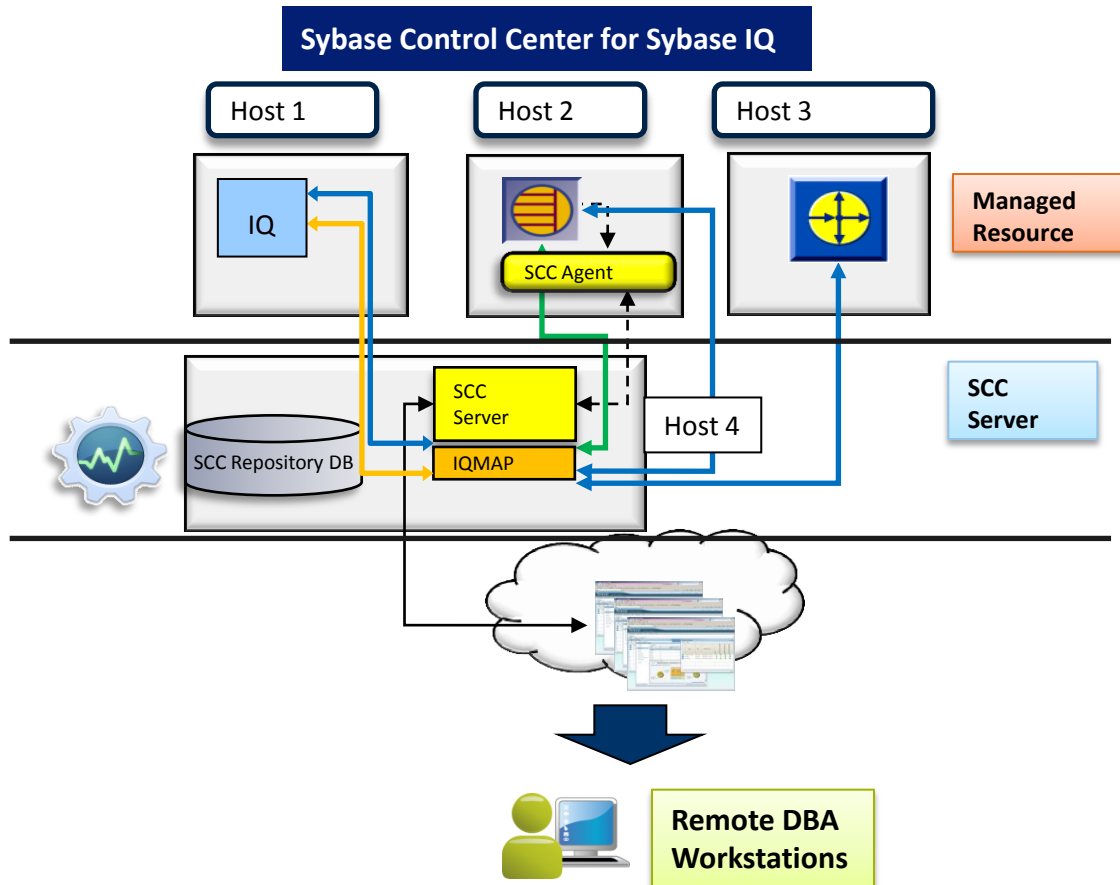
Sybase IQ PlexQ Ruby support

- Allow Ruby programs access to large, high performance data access / analysis
- Sybase IQ specific Ruby driver bundle
    - Native Ruby driver: access to Sybase IQ
    - Active record adapter: Ruby ObjRel map to Sybase IQ
    - Ruby DBI: database interface to Sybase IQ

Addresses: Diverse Data Types, Analytics For All

SYBASE IQ 15.3 PlexQ®

Enhancing administration & monitoring experience with SCC

Sybase Control Center for Sybase IQ

Host 1 | Host 2 | Host 3

IQ

SCC Agent

Managed Resource

SCC Server

SCC Repository DB

IQMAP

Host 4

SCC Server

Remote DBA Workstations

Sybase IQ PlexQ SCC

Integrated, Browser Based Administration & Monitoring

- Several Sybase Central operations migrated
- All v15.3 operations added
- Sybase Central will co-exist for a while

Addresses: IT resource rationing

SYBASE IQ 15.3

Taking on business analytics challenges

- Big data

- Diverse data types

- Complex questions

- Decision velocity

- Analytics for all

- It resource rationing
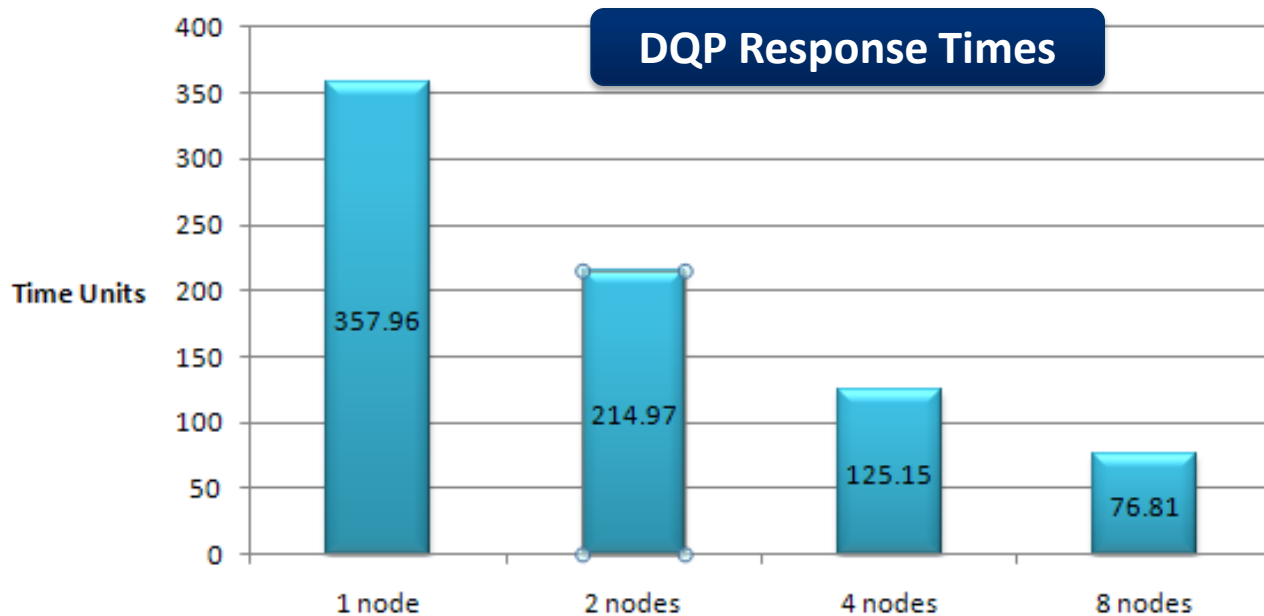
EXPERIENCES WITH SYBASE IQ 15.3

SYBASE® | An SAP Company

SYBASE IQ 15.3 PlexQ®

Beta program findings

Application intent: S curve calculation
Performance characteristics: distributed deep nested joins, groups; ~33% per node

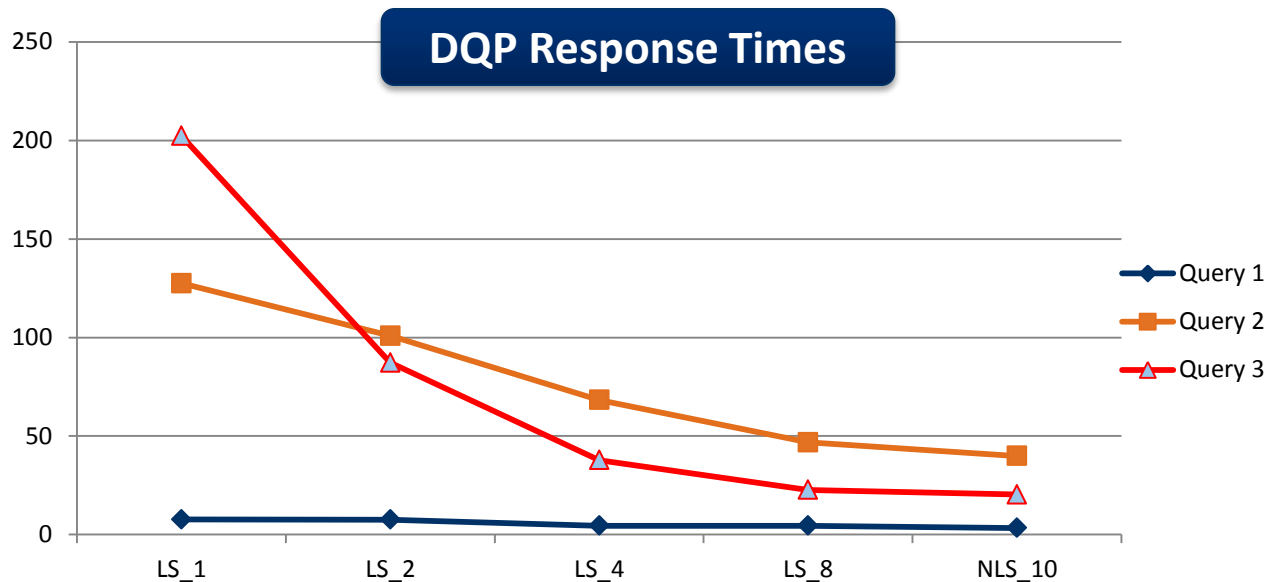**DQP Response Times**

Configuration: 8 machines with 8 cores each

SYBASE IQ 15.3 PlexQ®

Beta program findings

Application intent: measuring audience website visitation patterns
Execution characteristics: not all queries benefit; grouping, sorting, joins benefit most

DQP Response Times

Configuration: 8 machines with 8 cores each and 2 machines of 4 cores each

SYBASE® | An **SAP** Company

SYBASE IQ 15.3 PlexQ®

Foundation for great things to come

- Measures up to a true business analytics platform
  - Addresses key challenges in business analytics
- Sets the bar for MPP architectures relevant for full spectrum workloads
- Application enabled through web services, ruby, and multi-media UDF API
- Built for IT satisfaction: ease of administration and monitoring
- Stage is set for innovative applications to take root
  - MapReduce, in-database analytics, Web 2.0 participation