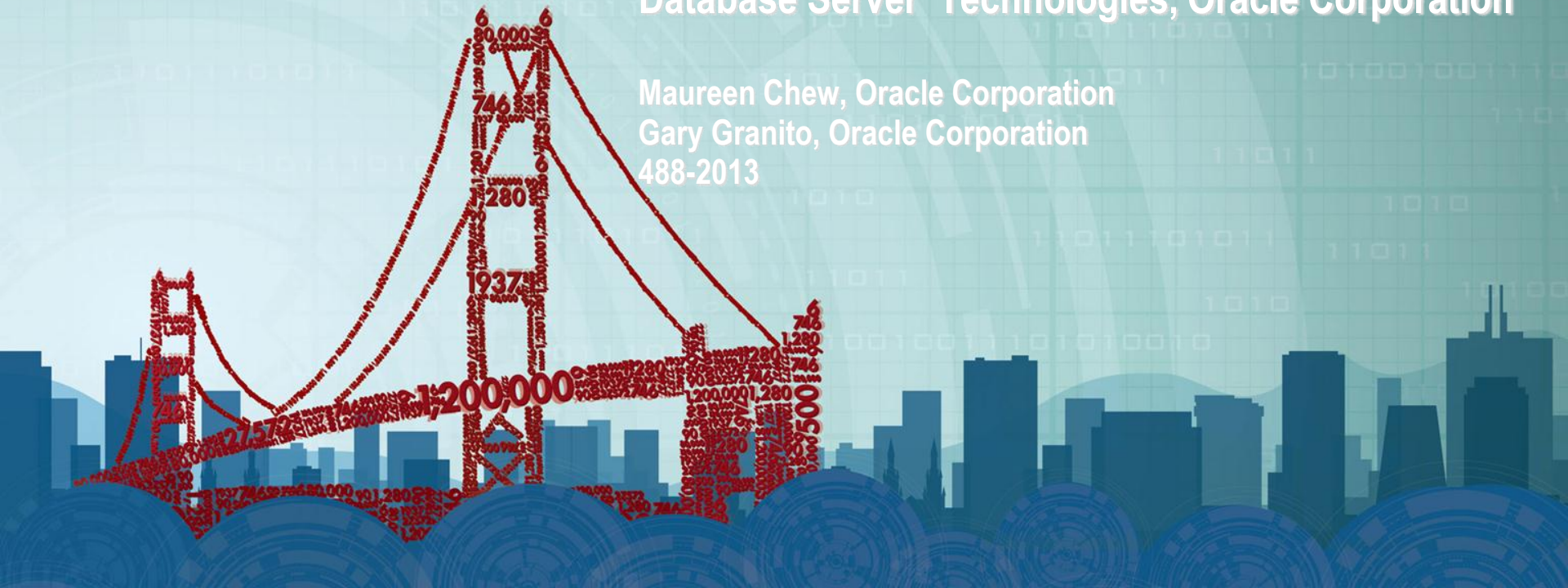


Big Data – Dream IT. Build IT. Realize IT

Paul Kent, Vice President, Big Data, SAS
Andy Mendelsohn, Senior Vice President,
Database Server Technologies, Oracle Corporation

Maureen Chew, Oracle Corporation
Gary Granito, Oracle Corporation
488-2013



Agenda

- Introduction
- SAS Visual Analytics, SAS High Performance Analytics on Oracle Engineered Systems
- Oracle & SAS Collaboration
 - Setting the Stage for Big Data
 - Next Generation Database Technology
- Comprehensive Platforms for SAS - Complete
 - Sun ZFS Storage Appliance for SAS Grid – case study
 - SAS on Current and Next Generation of SPARC servers
 - » Bureau of Labor and Statistics SAS Case Study
 - » SAS on SPARC M5-32 testing highlights



Paul Kent, Vice President, Big Data, SAS

SAS Visual Analytics and SAS High Performance Analytics on Oracle Engineered Systems



SAS and Oracle

- *Reflection* on a stronger partnership than ever
- SAS High-Performance Analytics and SAS Visual Analytics on Oracle Engineered Systems
- Extensive engineering collaboration
 - Sizing, configuration guidance and best practices for deployment
 - Support for POVs
- A strong technology and business alliance to develop solutions and products brings tremendous value and confidence



Big Data is RELATIVE not ABSOLUTE

BIG DATA

- When volume, velocity and variety of data exceeds an organization's storage or compute capacity for accurate and timely decision-making
-
-
-
-

ANALYTICS

- The process surrounding the development, interpretation, and useful application of statistics to solve a problem.
- Analytics applied to data provides the 4th V = Value
- Three types: Descriptive, Predictive, Prescriptive
-
-
-

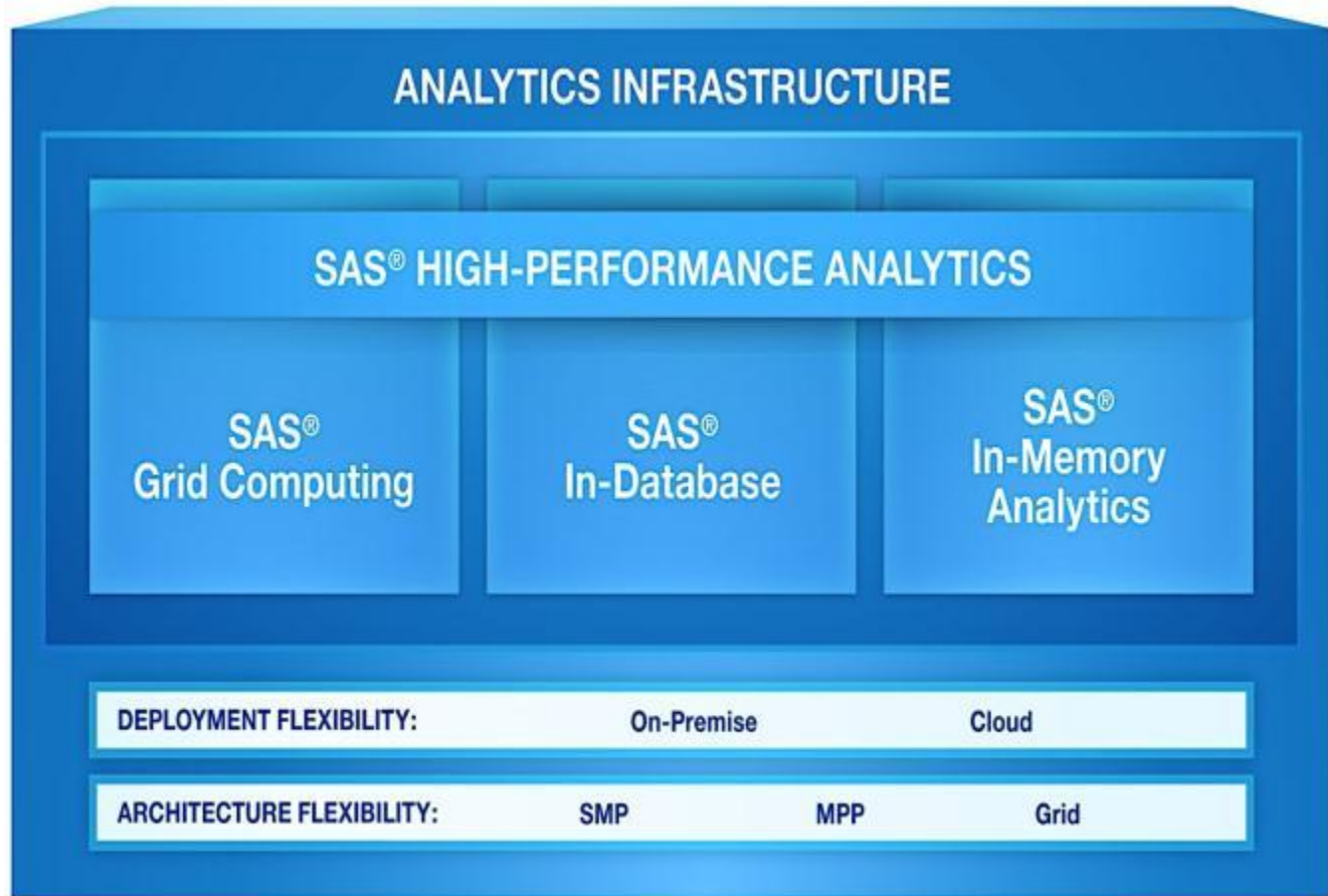
BIG ANALYTICS

- The combination of using **ANALYTICS on BIG DATA**
- **AND/OR** the capability to run advanced or complex analytics on any size data.
-
-
-



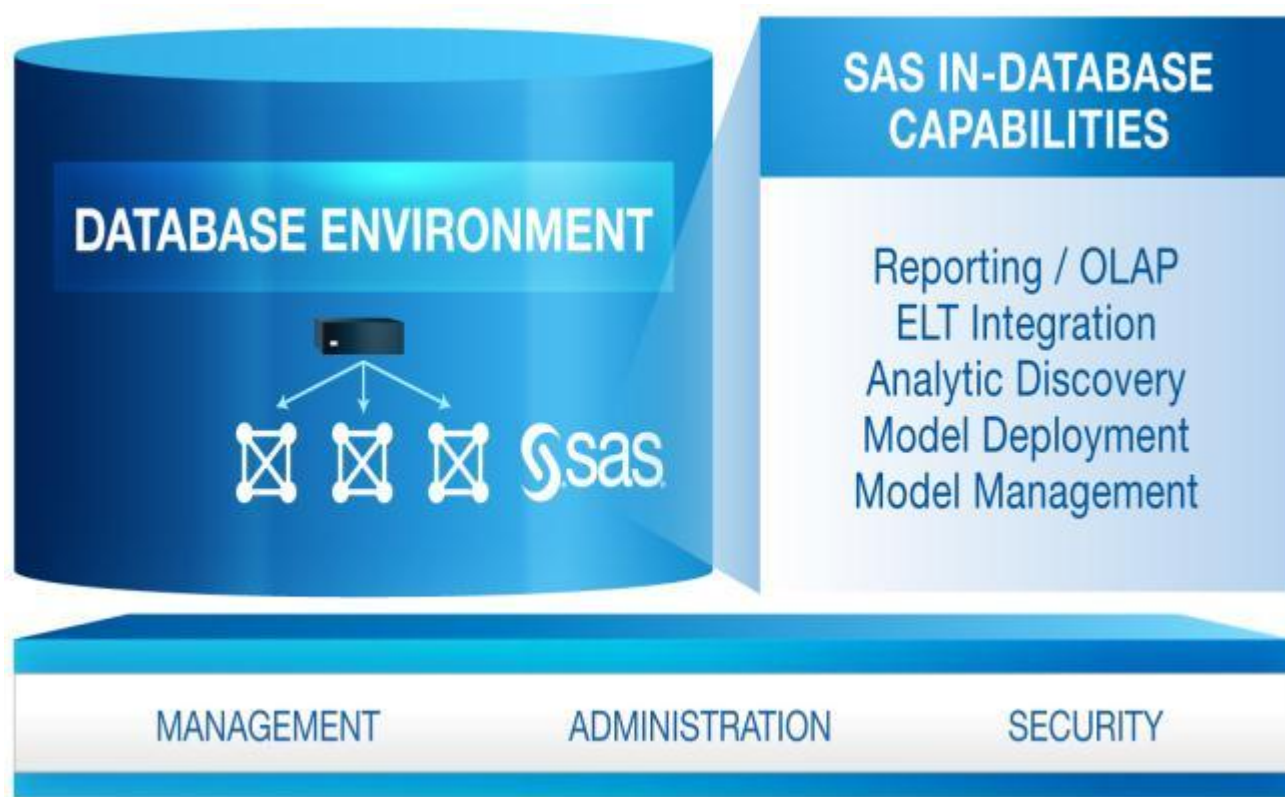
SAS® HIGH-PERFORMANCE ANALYTICS

KEY COMPONENTS



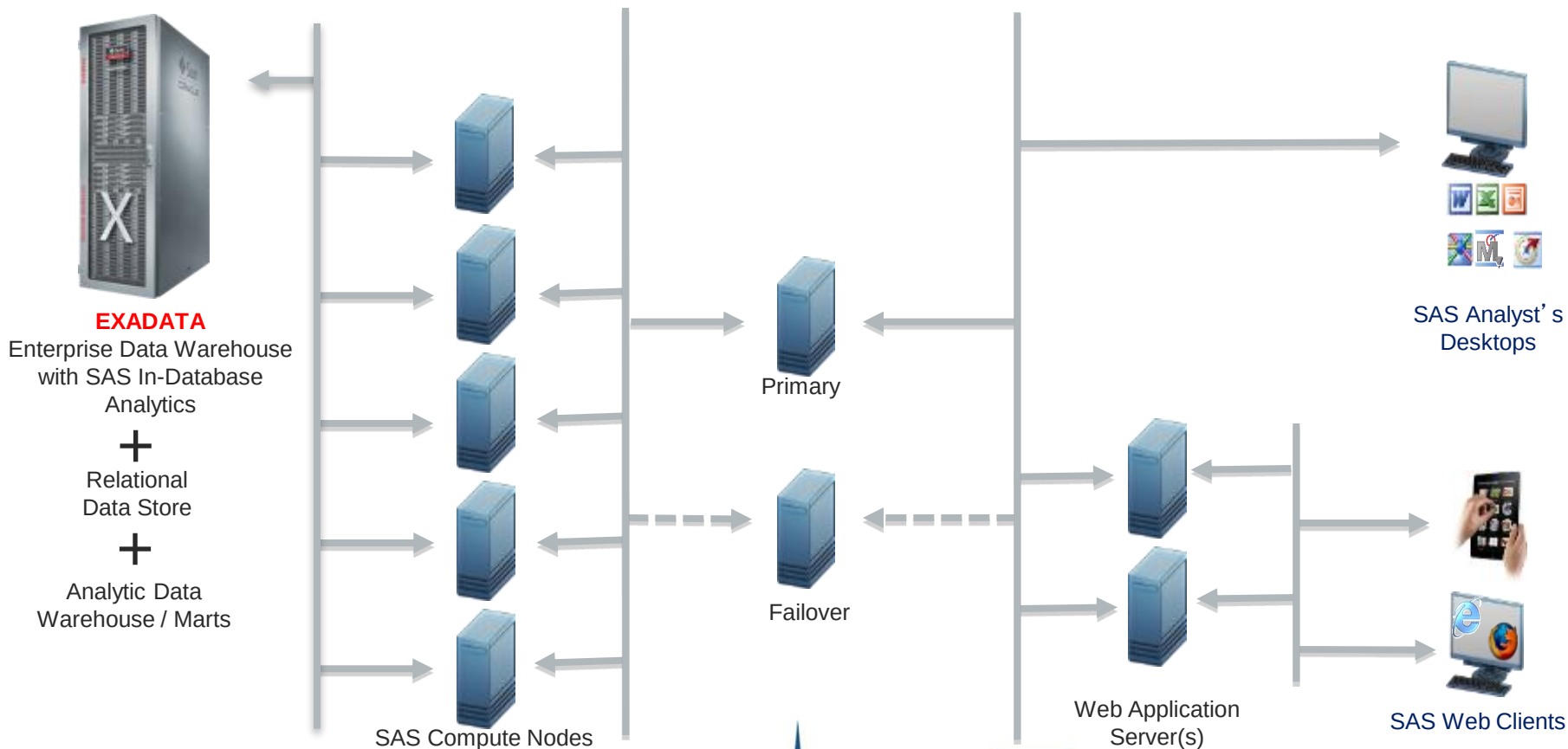
SAS® HIGH- PERFORMANCE ANALYTICS

SAS® IN-DATABASE



SAS BUSINESS ANALYTICS FRAMEWORK (ADD IN-DATABASE)

SAS Scoring Accelerator on Oracle Database and Exadata
Available since June 2012



Data Tier
Data Tier

Server Tier

Metadata Tier

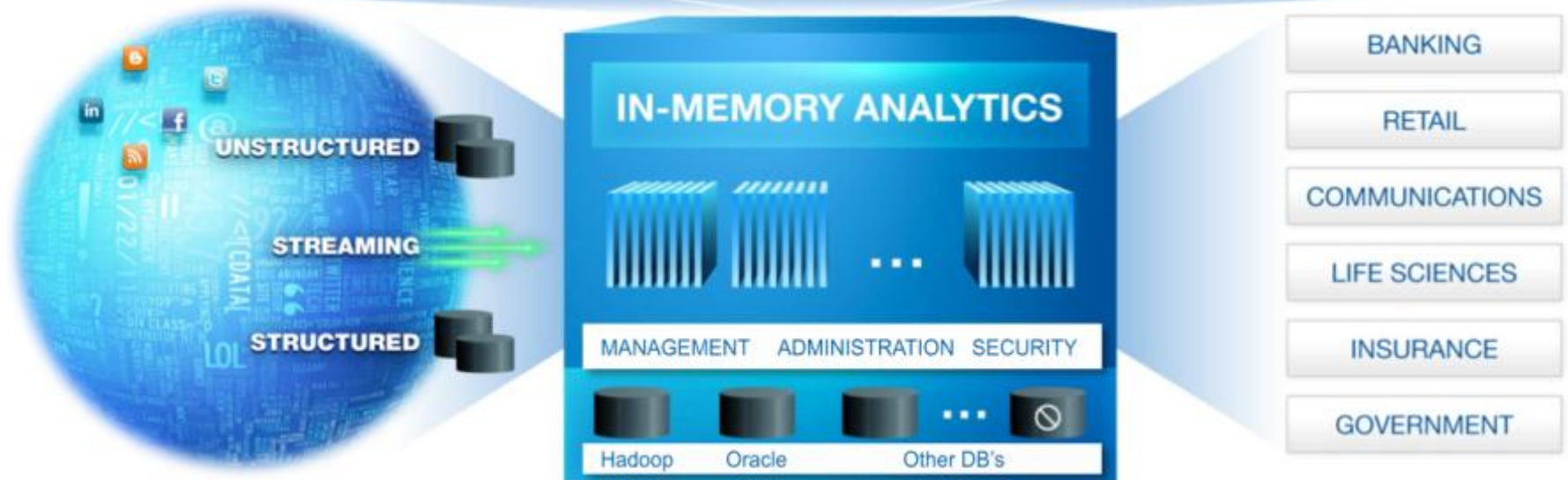
Web Tier

Client Tier



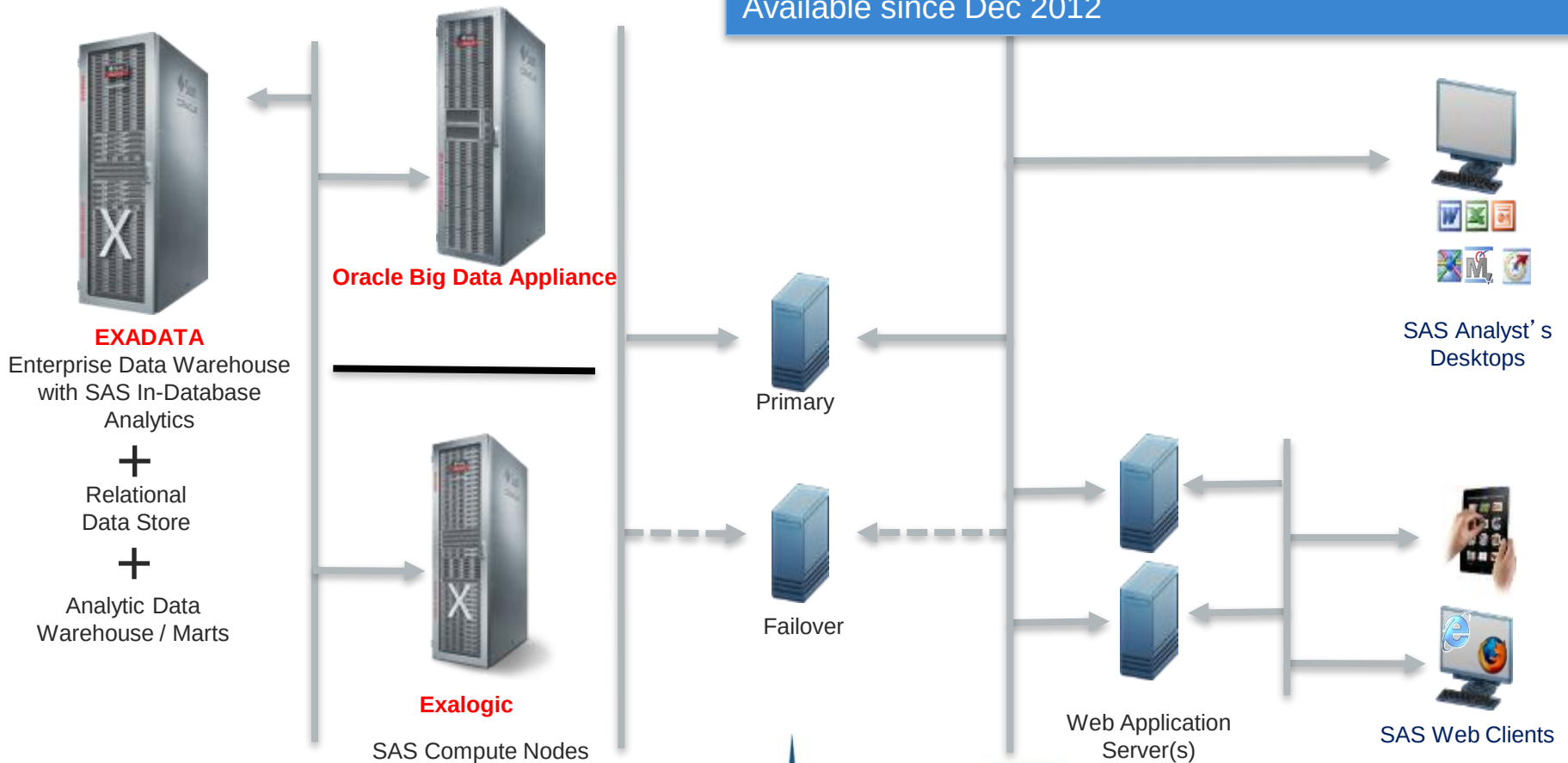
SAS® HIGH-PERFORMANCE ANALYTICS

SAS® IN-MEMORY ANALYTICS



SAS BUSINESS ANALYTICS FRAMEWORK (ADD IN-MEMORY)

SAS Visual Analytics, SAS High-Performance Analytics Server for Oracle Engineered Systems
Available since Dec 2012



Data Tier

Server Tier

Metadata Tier

Web Tier

Client Tier



HIGH-PERFORMANCE ANALYTICS

Using Different Data and Computing Appliances with Asymmetric HPA

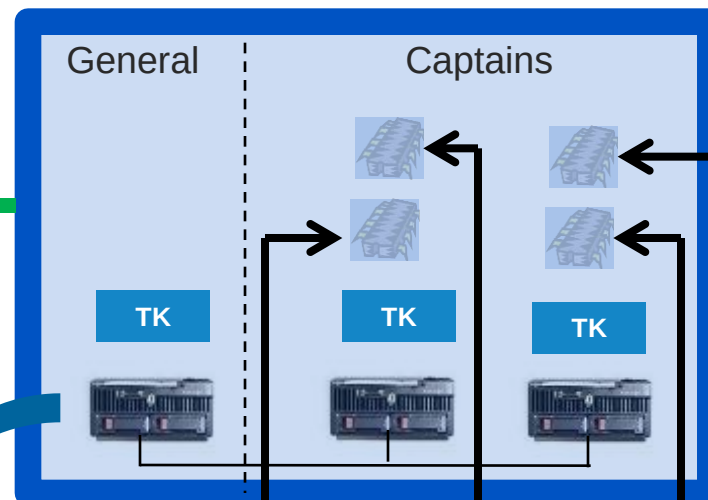
SAS Server



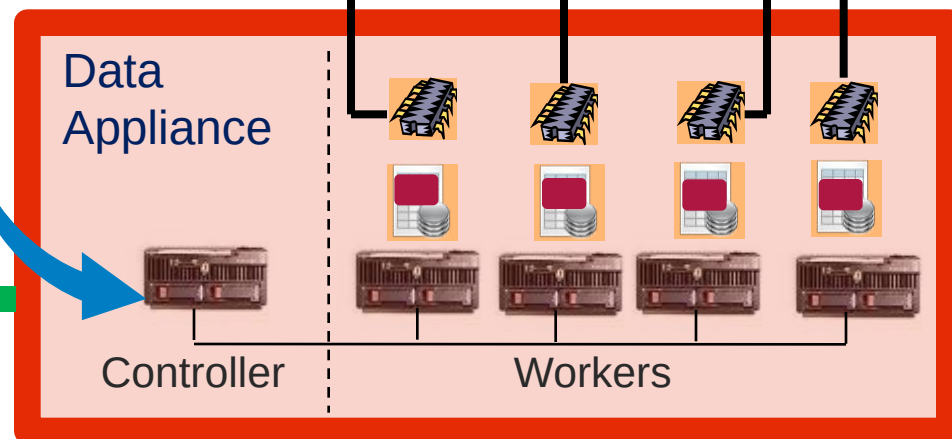
```
libname a oracle  
  server="dataAppliance";  
  
proc hpcorr data=a.flights;  
  performance  
    mode=asym  
    host="computingAppliance";  
run;
```

TKGrid

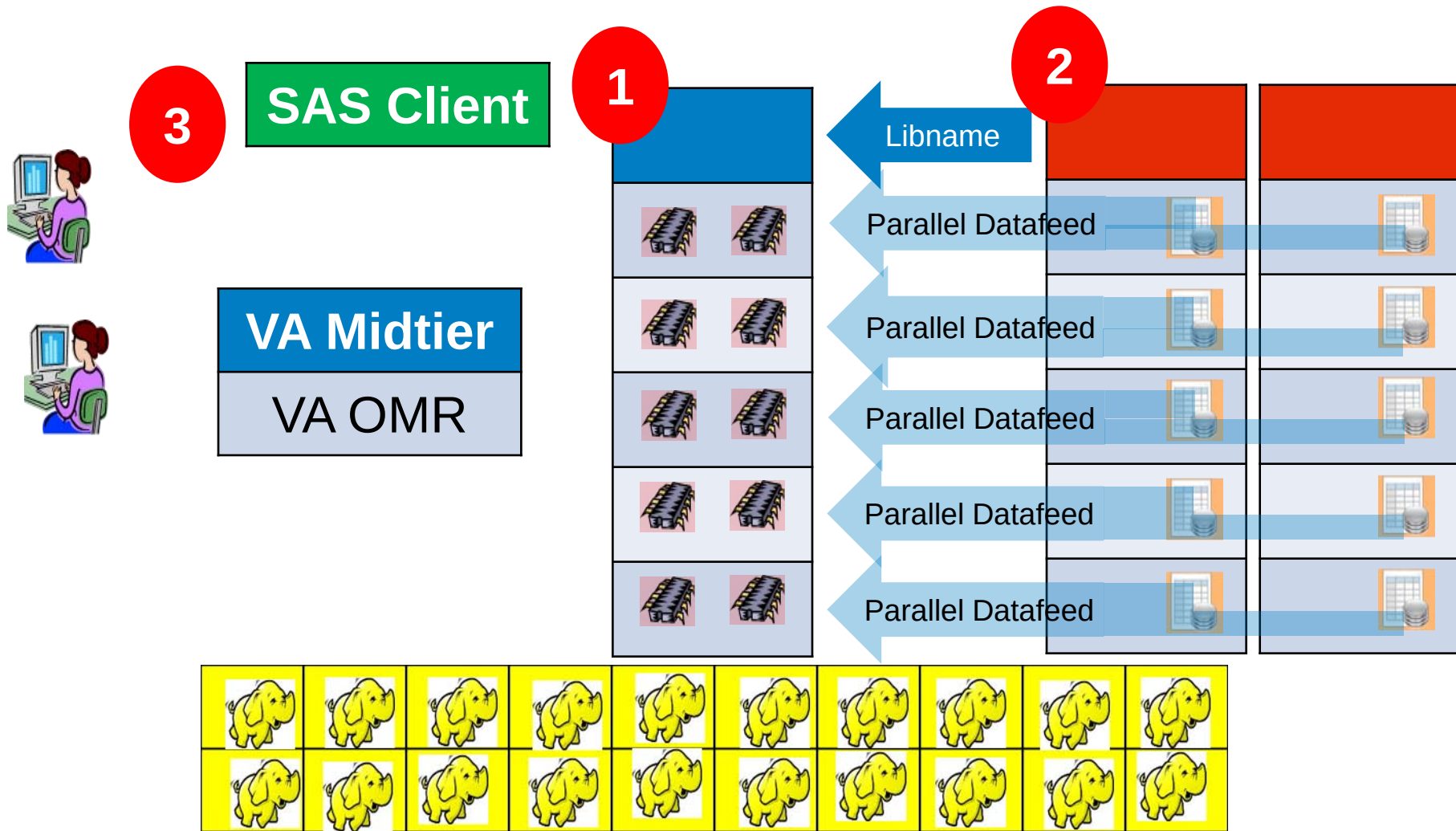
Computing Appliance



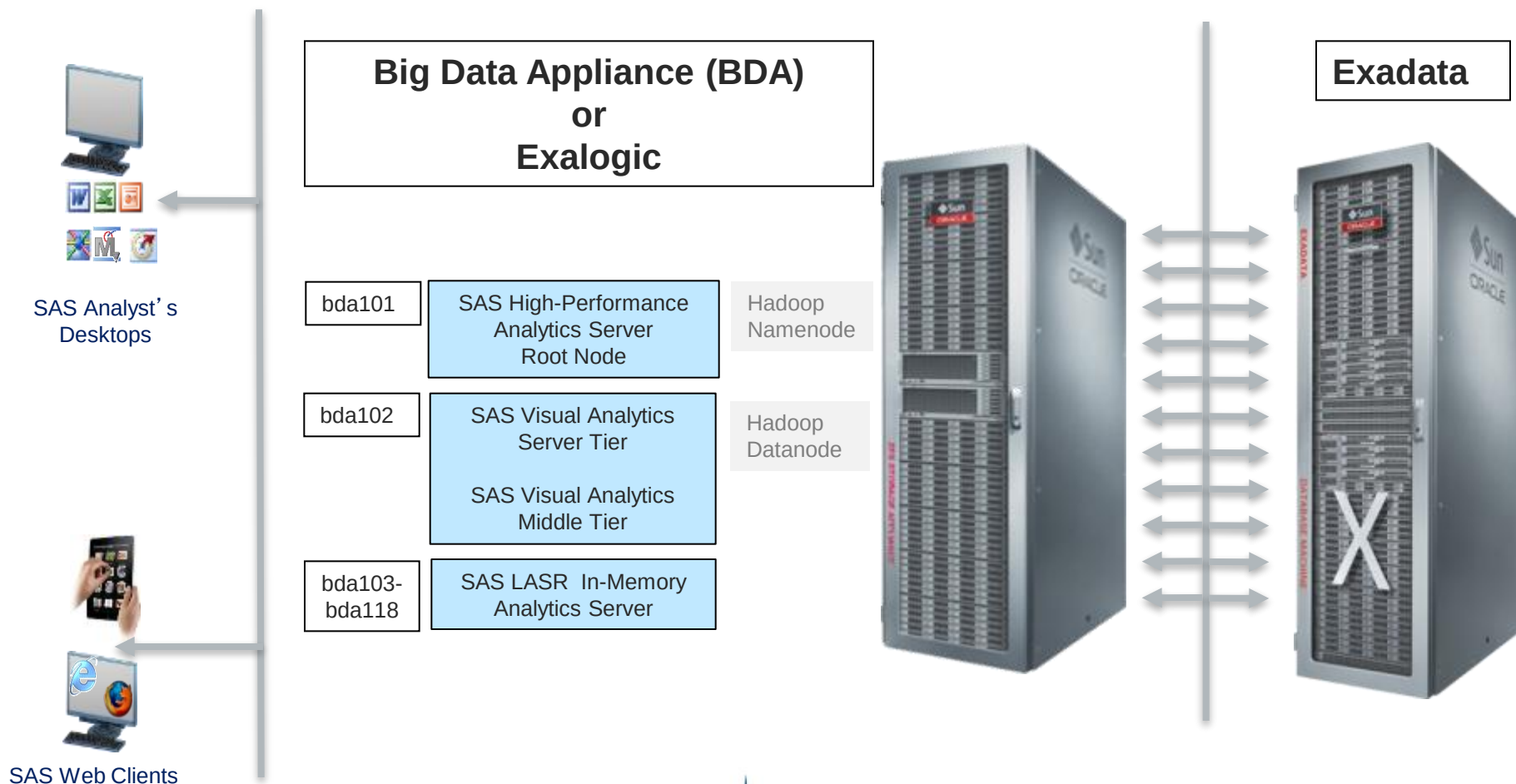
**Access
Engine**



Vanilla LASR Server – with a Database, other sources



SAS High Performance Analytics, SAS Visual Analytics on Oracle Engineered Systems



Oracle Linux support

- SAS 9.3 supports Oracle Linux 5.5
 - <http://support.sas.com/resources/sysreq/hosts/unix>
- And higher...
 - http://support.sas.com/resources/thirdpartysupport/baseline_plus.html

Operating Systems

Beginning with the minimum system requirements for that release of SAS software, SAS will support both major and minor versions, as defined by the Independent Software Vendor (ISV), unless the operating-system upgrade also necessitates an unsupported third-party product change. SAS will also support vendor-supplied patches to any supported version of the operating system.

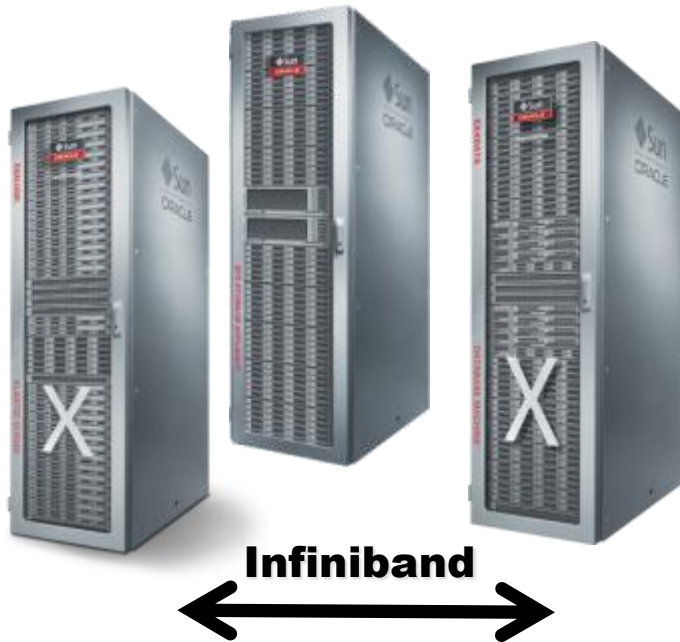
For example, if the baseline for SAS >version x is Solaris 8, SAS version x will support Solaris 9 and 10 under this policy. However, if a SAS software product or component requires version y of a third-party product but version y is not yet supported on Solaris 10, then that SAS product or component will not be supported yet on Solaris 10.



SAS High-Performance Analytics - Choice

Distributed

- Exalogic or BDA
- Oracle Linux



SMP (SAS 9.4)

- SPARC M5-32, Solaris 11.1
- Single domain test – 48 cores, 2TB RAM
- SMP – In-Memory Analytic Server (LASR)
 - Lift 100GB table from Exadata to LASR
 - -> “hp” PROCS running in multi-threaded fashion



SAS High-Performance Analytics

- SAS High Performance Analytics on Oracle Big Data Appliance – LIVE at the Oracle Booth
 - High-Performance PROCs (ie: hpreg, hplogistic, hpreduce,) running in parallel nodes
 - Demonstrating multiple high-performance data sources
 - » Parallel data feeder for Exadata
 - » Hadoop Distributed File System (HDFS)
 - » In-Memory, Distributed Analytical Server (LASR)
 - » Data can flow bi-directionally between sources above



Andy Mendelsohn, Senior Vice President, Database Server Technologies, Oracle

Oracle and SAS Collaboration: Setting the Stage for Big Data



Oracle and SAS Collaboration

- Stronger partnership than ever
- Joint engineering Proof of Value
- Where Best of Breed Solutions Converge
- Where we are today and a look ahead

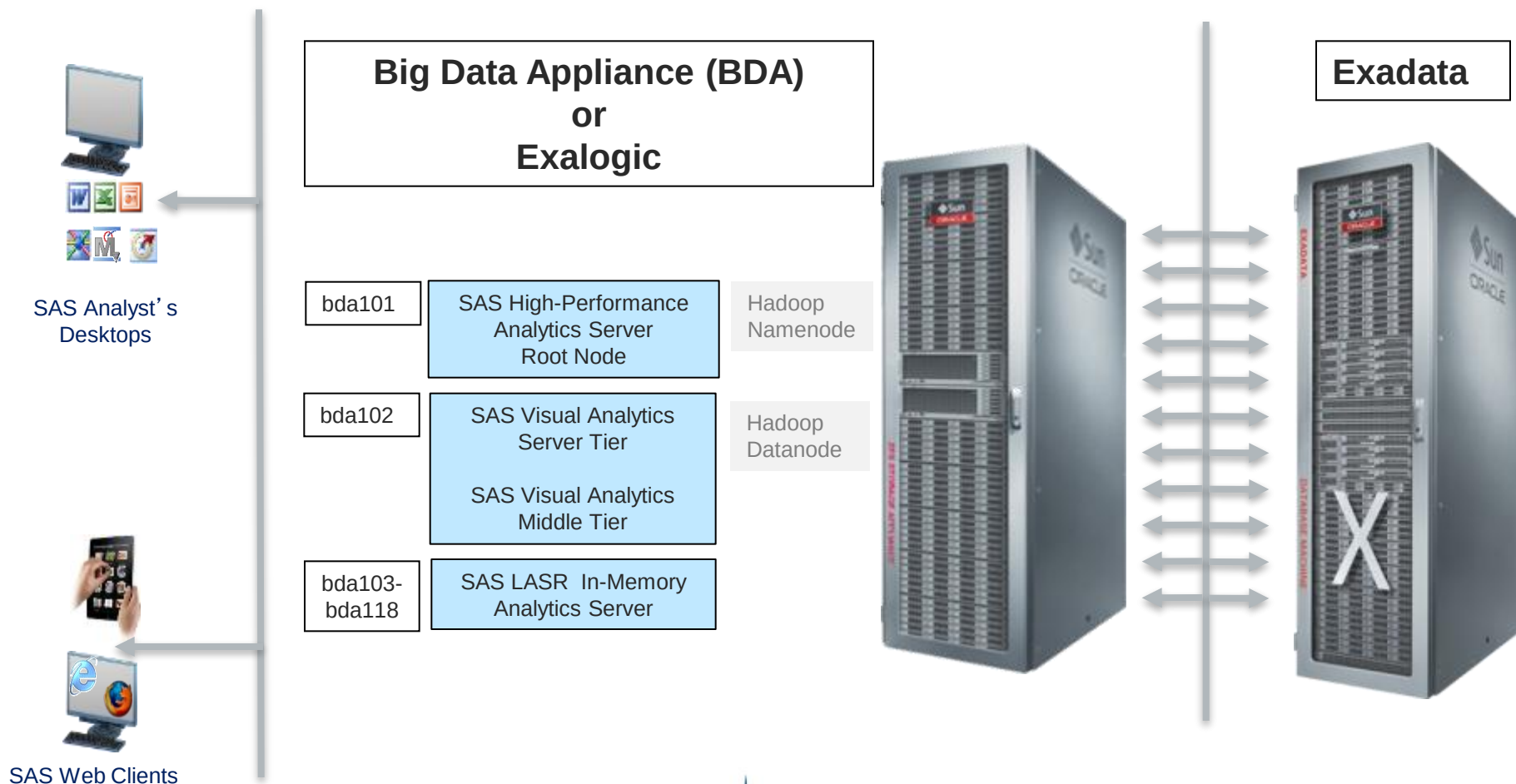


Oracle and SAS

- *Reflection* on a stronger partnership than ever
- Oracle Engineered Systems for SAS High-Performance Analytics and SAS Visual Analytics
- Extensive engineering collaboration
 - Sizing, configuration guidance and best practices for deployment
 - Support for POVs
- A strong technology and business alliance to develop solutions and products brings tremendous value and confidence



SAS High Performance Analytics, SAS Visual Analytics on Oracle Engineered Systems



Where Best of Breed Solutions Converge

Oracle: Premiere Database Platform for SAS

- Synergy between SAS and Oracle Database



THE
POWER
TO KNOW...

requires data, we provide the data platform

- Proven in the market
 - Vast Majority of SAS database users depend on Oracle
- Proven at SAS
 - **SAS R&D invests** in Exadata for development, test, and field readiness of joint SAS/Oracle solution
 - **SAS MIS chose** several Exadata Database Machines and SPARC T5 servers for internal business critical applications
 - **SAS Solutions on Demand chose** several Exadata Database Machines for premiere hosting of SAS cloud services



SAS OnDemand – Runs on Exadata

Key Capabilities/Deliverables

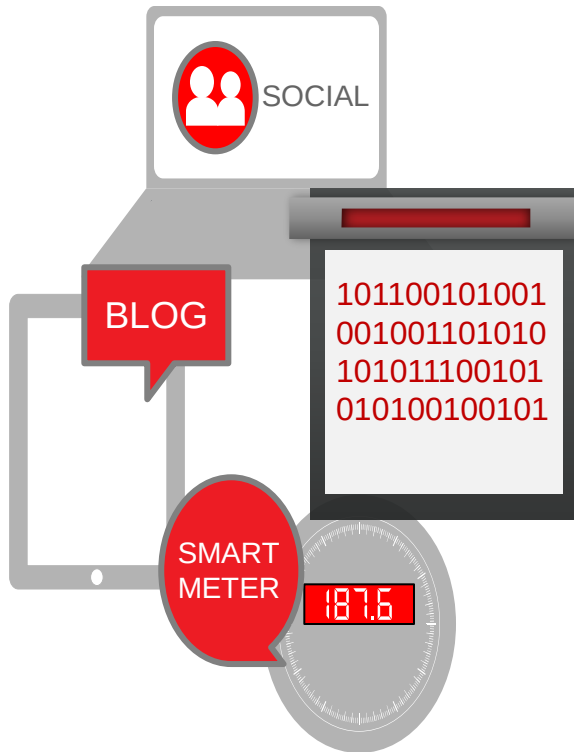
- **Decreased Time to Production for new customers** - (no need to take Exadata down to expand HW or SW)
- **Best of Breed:**
 - Quality and quantity of tools for management, monitoring, development and diagnosis via Oracle Enterprise Manager
 - Resource Management/Optimization through Exadata's IO Resource Management and Database Resource Manager.
 - Best of Breed Security(HW to SW) capabilities through Oracle's Advanced Security Option
- **Business Continuity:**
 - High Availability SLA's >99%
 - Superior Backup, restore, and recovery
- **Oracle DB License Consolidation** - management efficiencies and savings



Next Generation Database Technology



Major Database Focus Areas



BIG
DATA



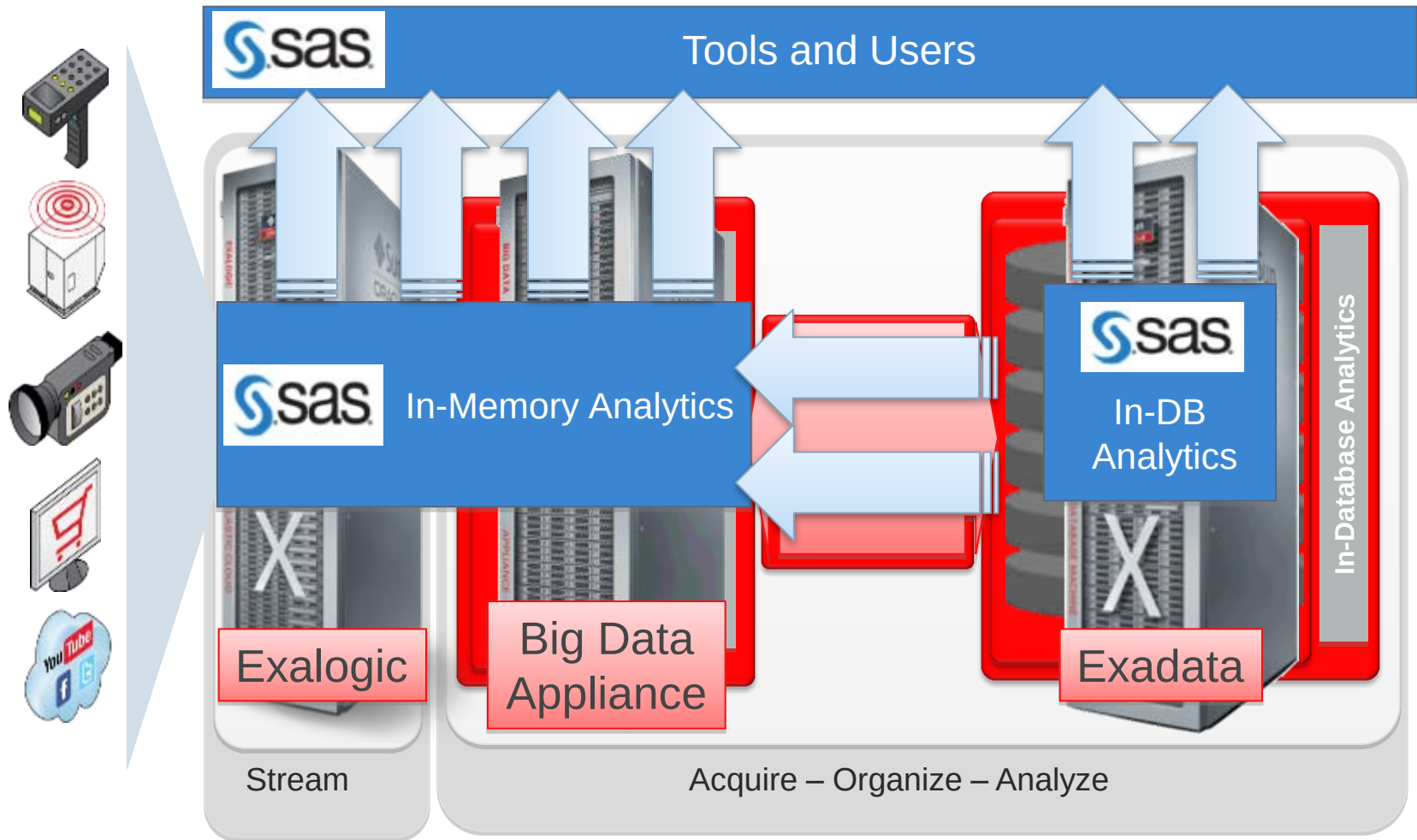
ENGINEERED
SYSTEMS



CLOUD
COMPUTING

Oracle/SAS Big Data Solution

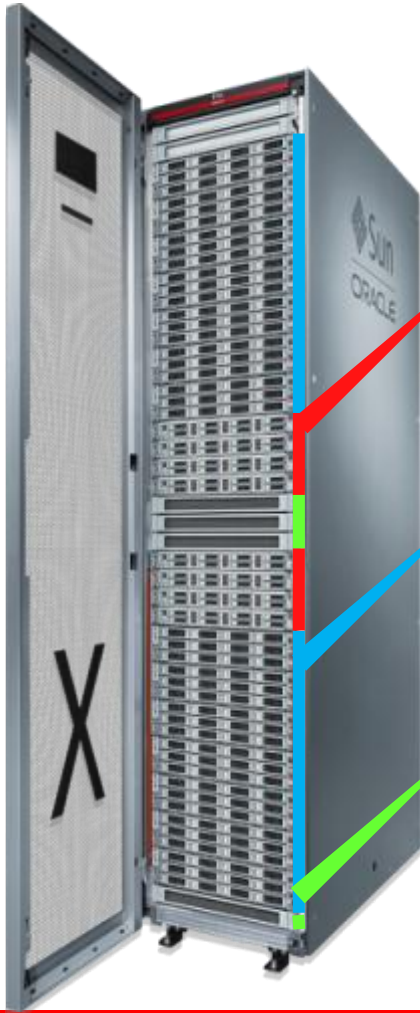
Comprehensive end-to-end solution



ORACLE

Oracle Exadata Database Machine

For OLTP, Data Warehousing & Consolidated Workloads



Scale-Out Database Servers

- Industry standard 2 or 8-socket servers
- Oracle Database 11g, RAC, ASM, EM12c
- Compatible with all 11g databases, app's, tools

Scale-Out Intelligent Storage Servers

- Industry standard 2-socket storage servers
- *Exadata Storage Server Software*

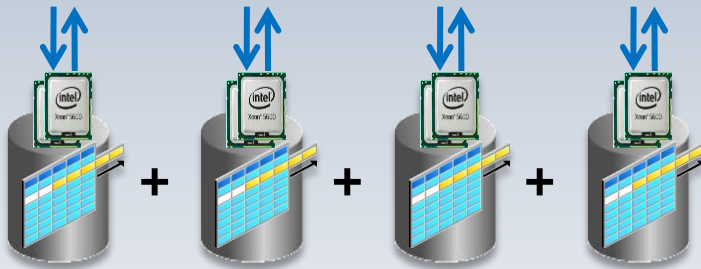
InfiniBand Network

- Unified internal connectivity (40 Gb/sec)

Exadata Innovations

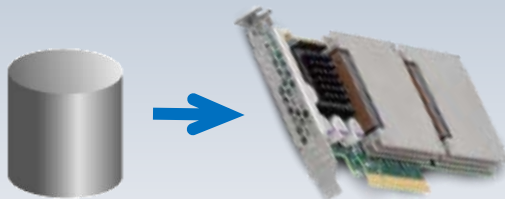
Intelligent storage

- Scale-out InfiniBand storage
- Smart Scan query offload



Smart PCI Flash Cache

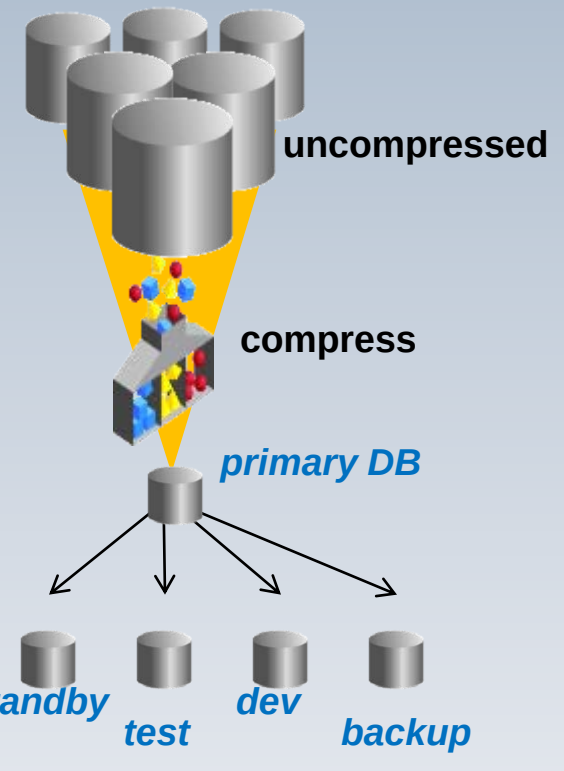
- Accelerates random I/O up to 30x
- Triples data scan rate



**Data
remains
compressed
for scans
and in Flash**

Hybrid Columnar Compression

- 10x compression for warehouses
- 15x compression for archives

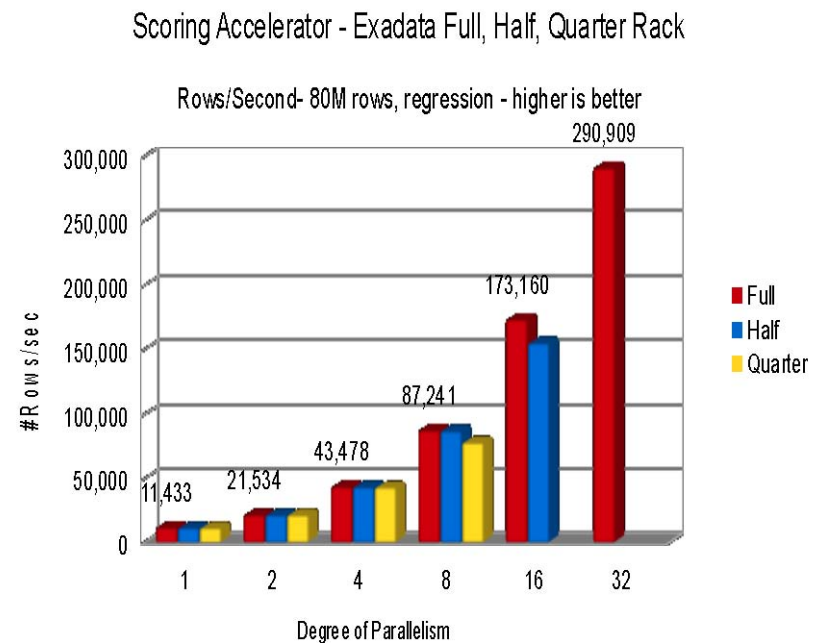


ORACLE

SAS Scoring Accelerator on Exadata

Performance Evaluation

- Performance test on quarter, half, and full rack of Exadata proved linear scalability
 - Doubling the parallelism cuts runtime in half
- Identical behavior for all configurations up to saturation of resources
- Extreme performance gains
 - 80M rows: 2hrs down to 4 min
 - 250M rows: 6 hrs down to 14 min



Introducing

ORACLE®

DATABASE

12^c

Application Development

Big Data

Consolidation

Data Optimization

Data Warehousing

High Availability

In-Memory

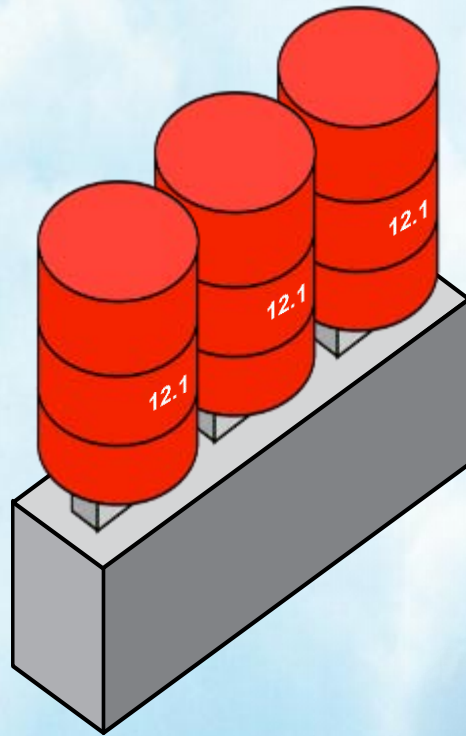
Performance & Scalability

Security & Compliance

ORACLE®

Oracle Database 12c Pluggable Databases

New architecture for consolidating databases on Clouds



No application changes

Isolation and multitenancy

Fast provisioning and cloning

Secure and highly available

Lower IT costs

Manage many as one

Greater resource utilization

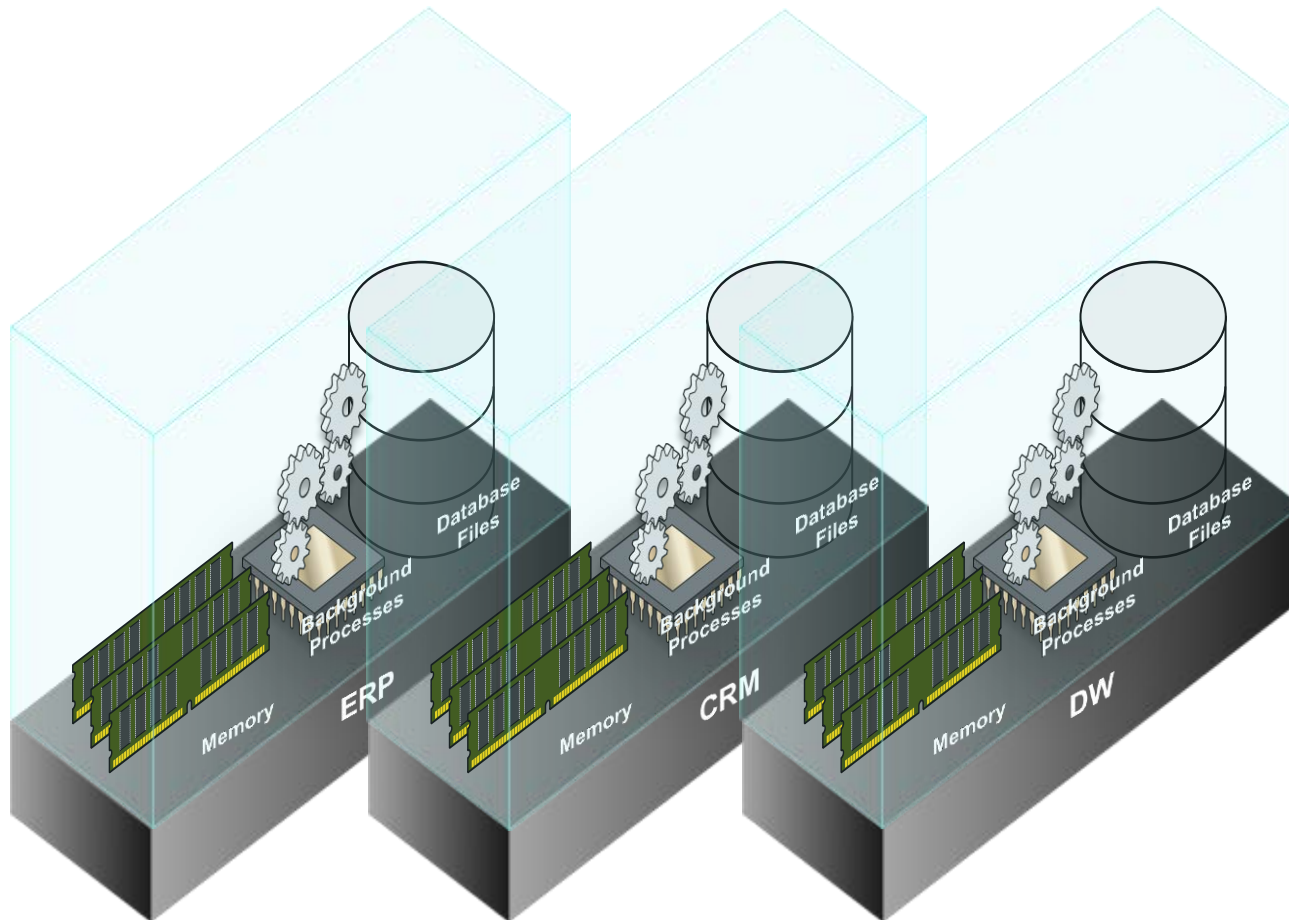
Performant and scalable

ORACLE

Oracle Database Architecture

Requires memory, processes and database files

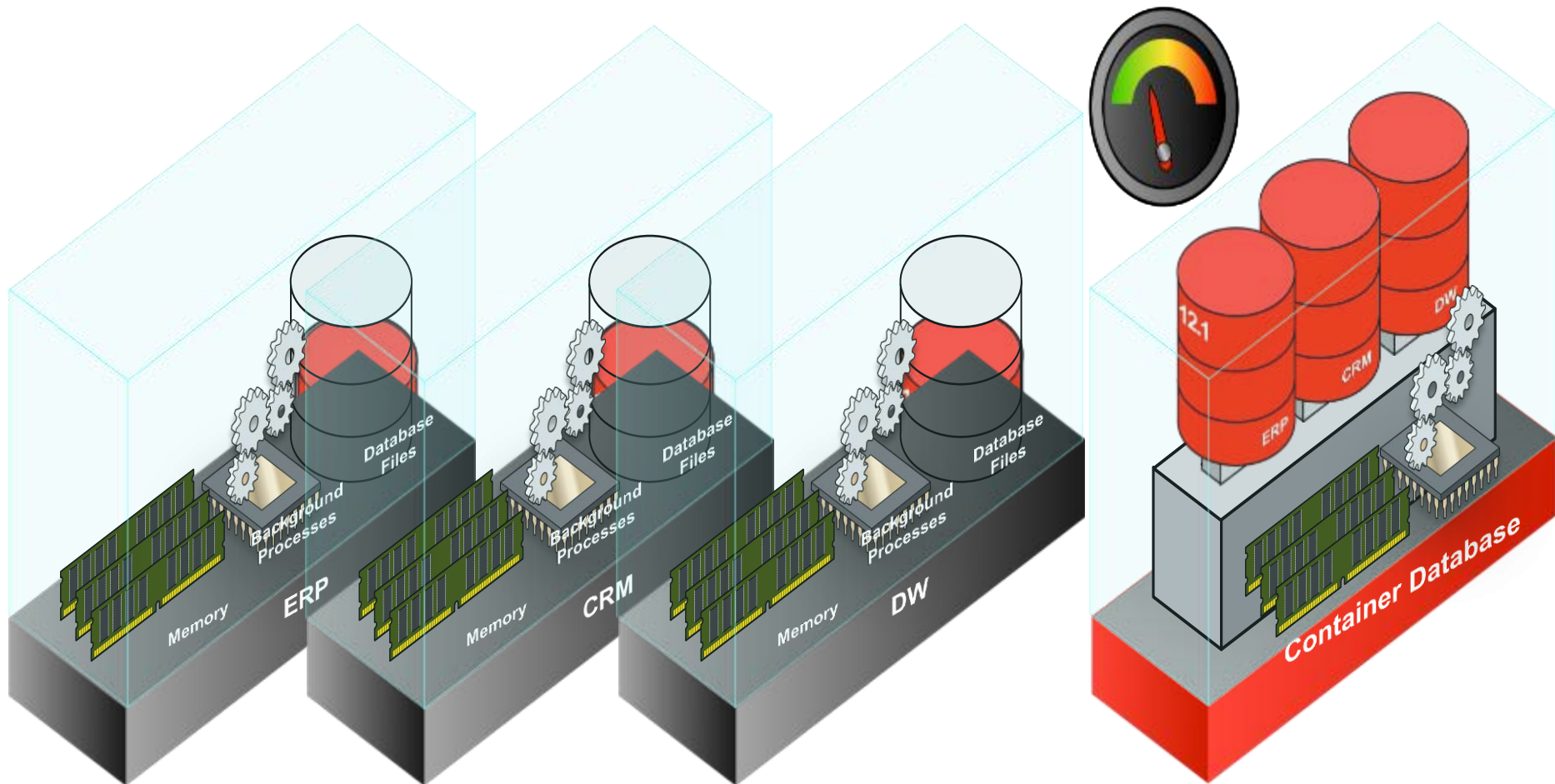
System Resources



New Pluggable Databases Architecture

Memory and processes required at container level only

System Resources

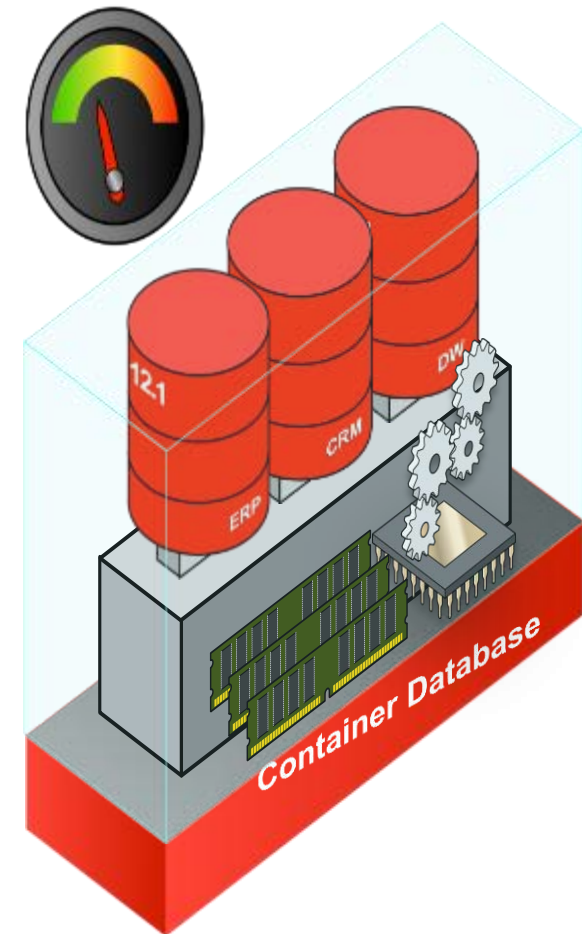


ORACLE

New Pluggable Databases Architecture

Memory and processes required at container level only

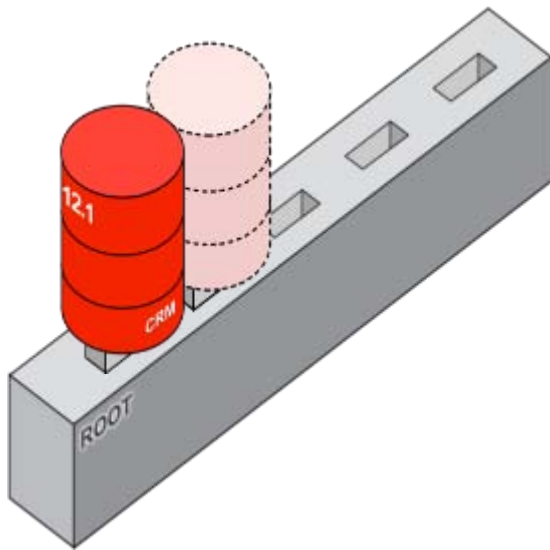
System Resources



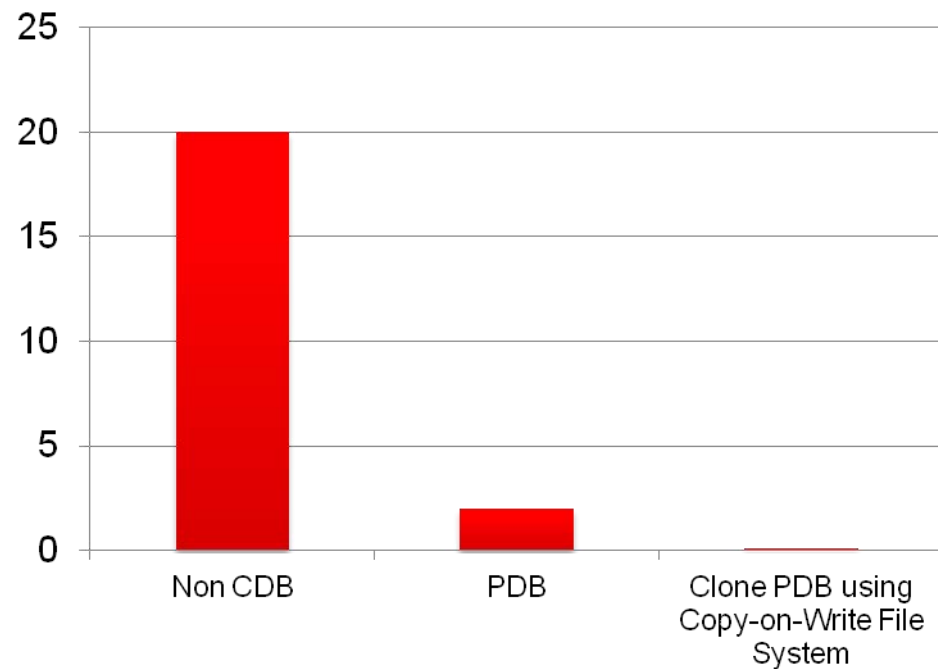
ORACLE

Pluggable Databases for Fast Provisioning

Pluggable Databases can be quickly provisioned from seed

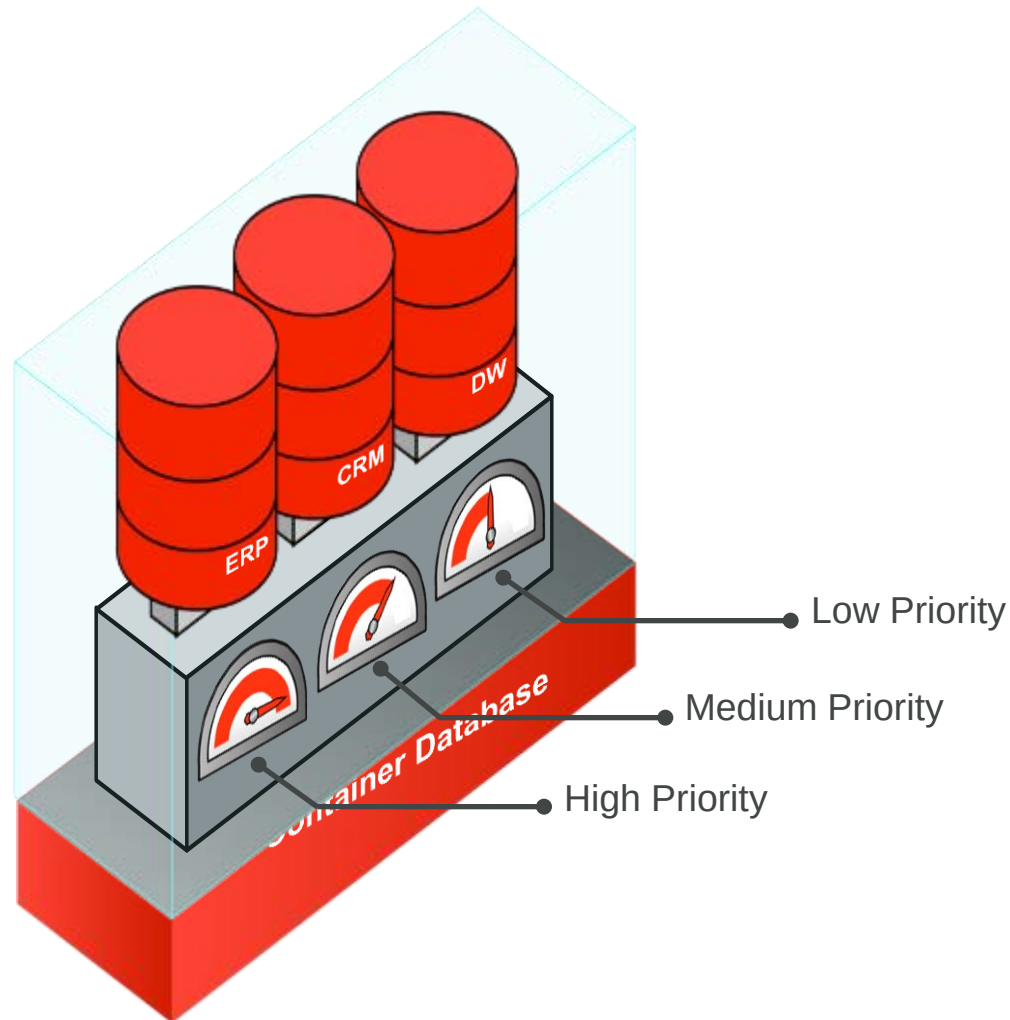


Time Taken to Provision New Database



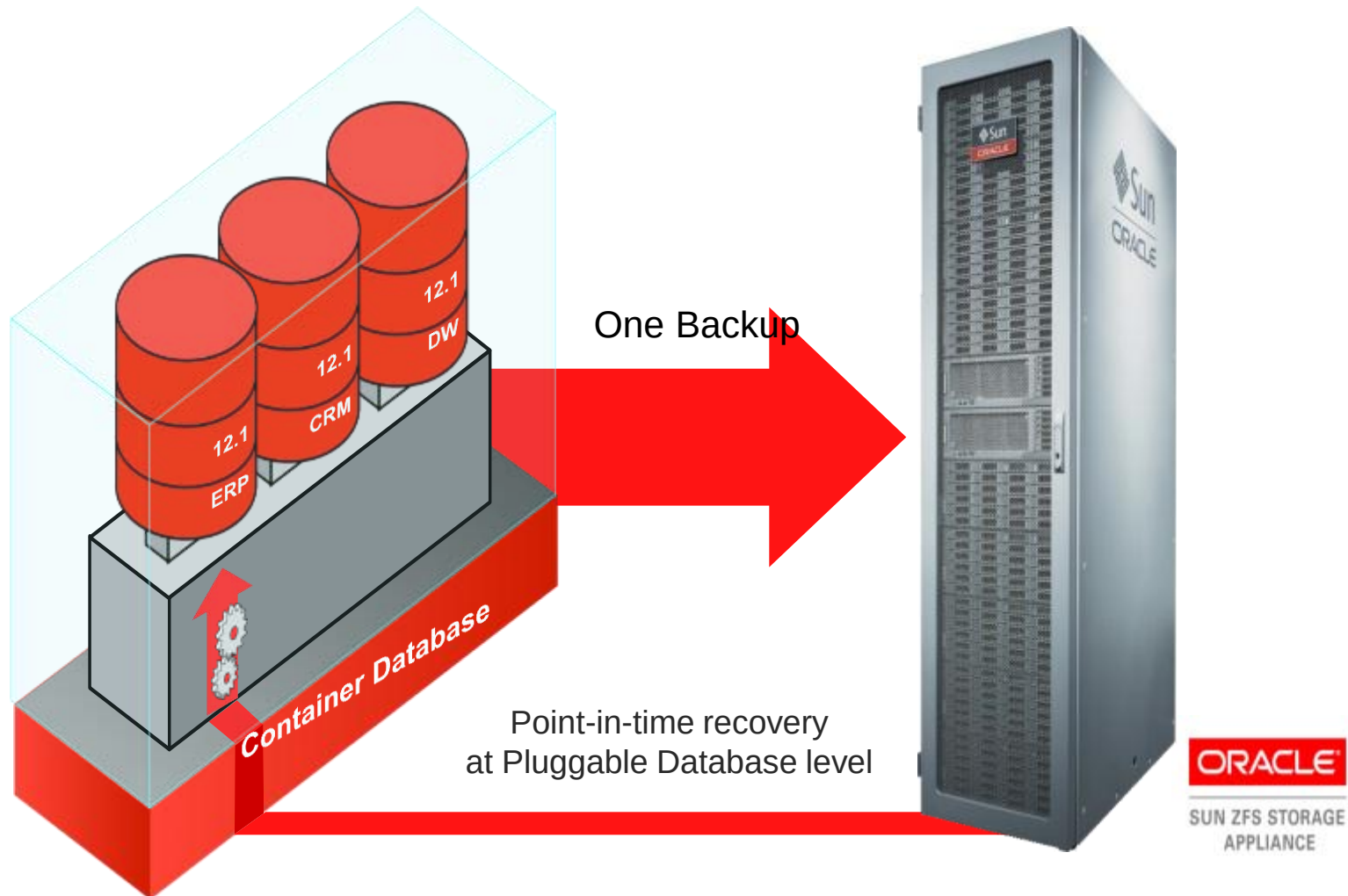
Managing Shared Resources

Resource management for Pluggable Databases



Pluggable Databases. Manage Many as One

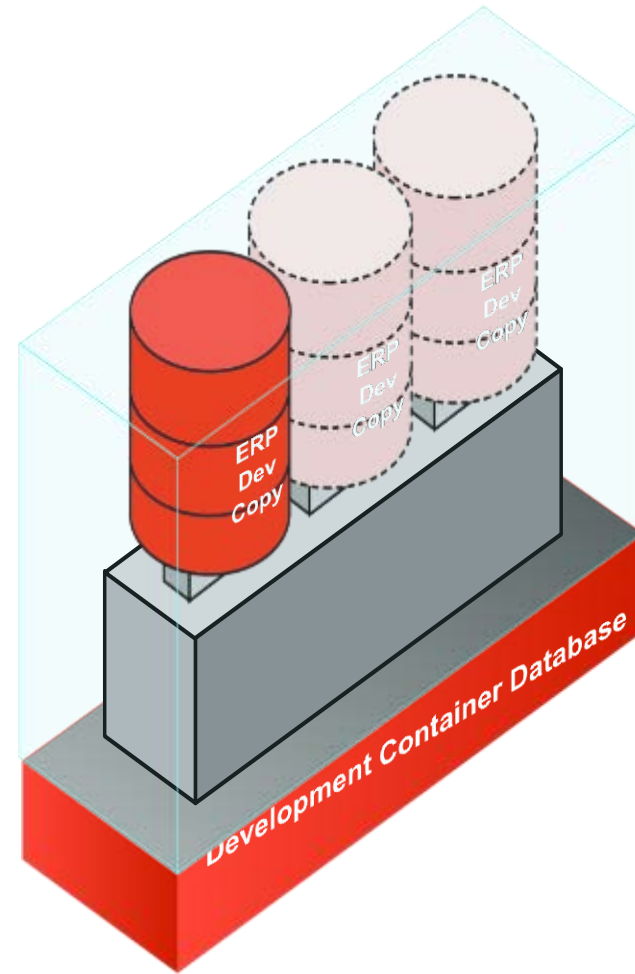
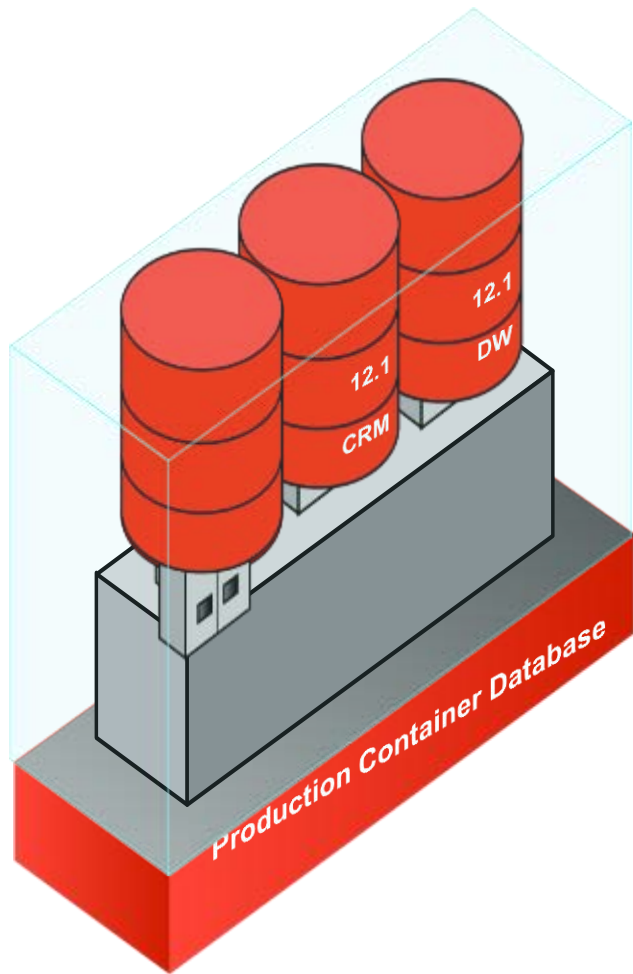
Backup Databases as One; Recover at Pluggable Database Level



ORACLE

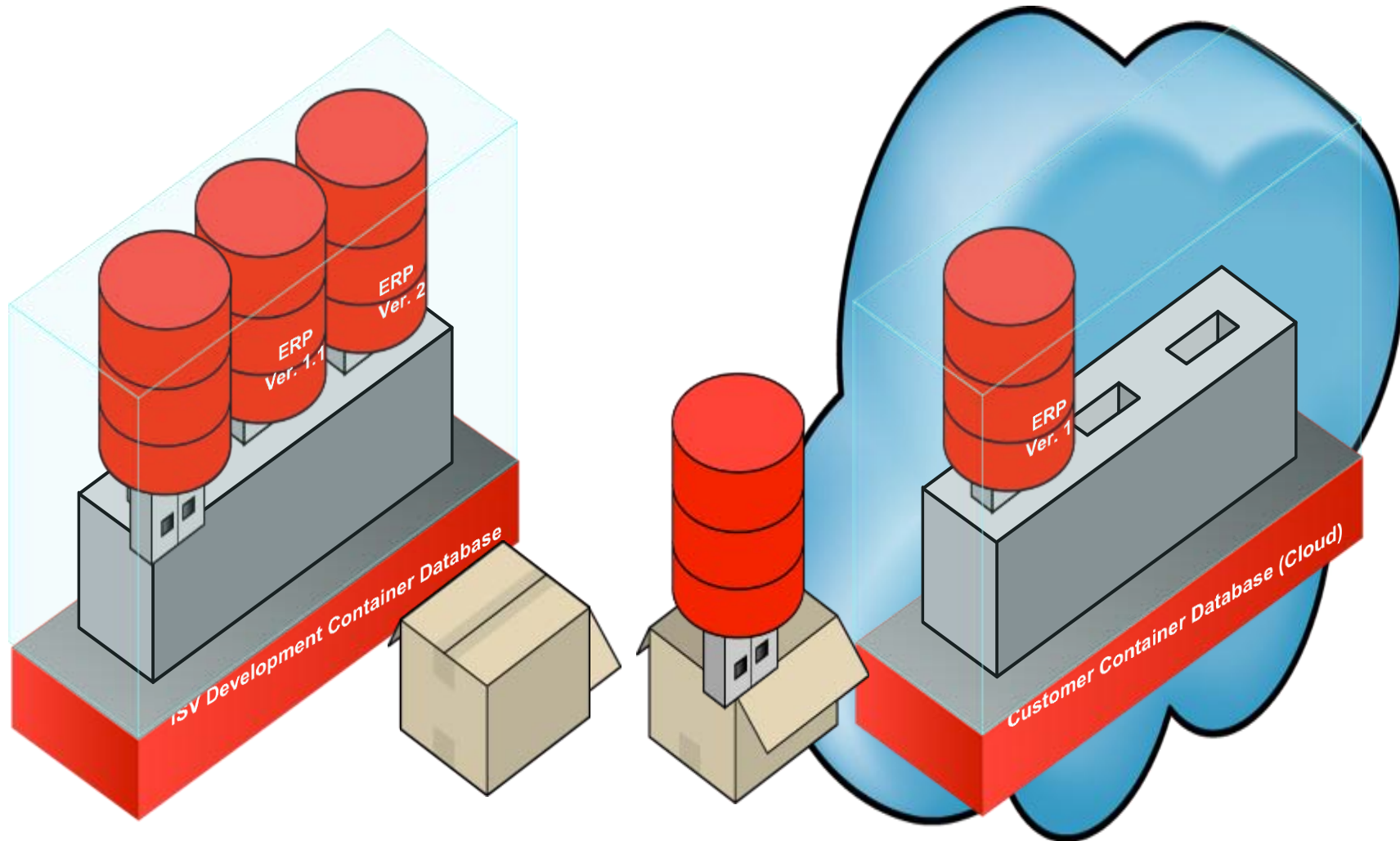
Pluggable Databases for Test and Development

Fast, flexible copy and snapshot of pluggable databases



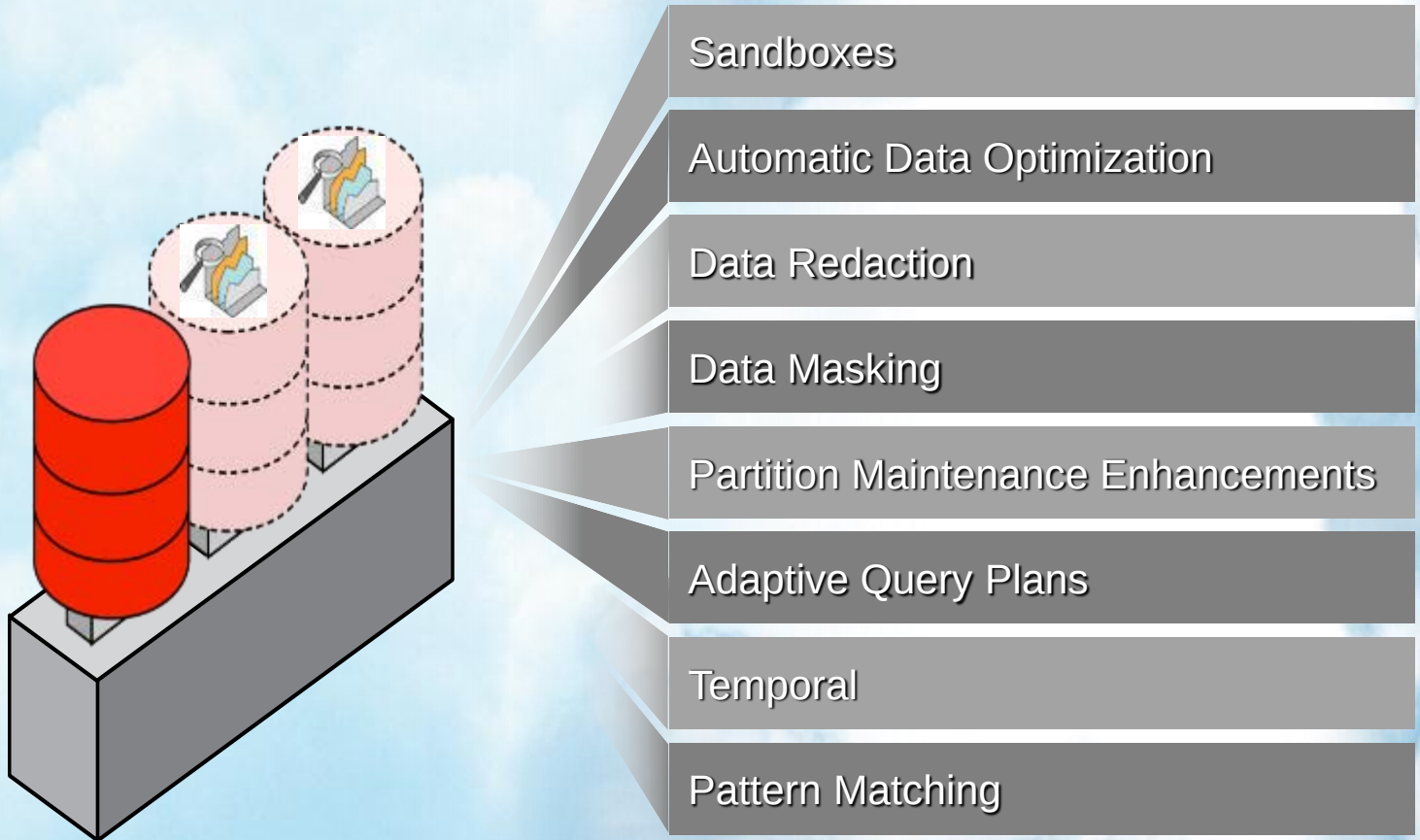
Pluggable Databases. Perfect for ISVs.

Packaged apps and reference data are easily distributed



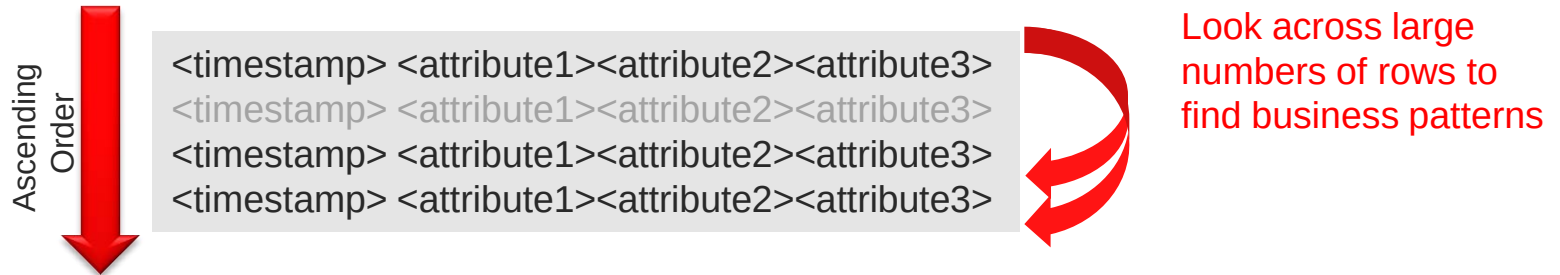
Oracle Database 12c

Enabling Data Scientists



Pattern Matching

Simplified Analysis of Big Data



- Scalable discovery of business event sequences
 - Clickstream logs: sessionization, search behaviour
 - Financial transactions: fraud detection, double bottom (“W”) stock analysis
 - Telco: dropped calls
 - Medical sensors: automated medical observations and detections

Pattern Matching

Finding Double Bottom (W)

```
}  
  
private boolean lt(String a, String b) {  
    if (a.isEmpty() || b.isEmpty()) {  
        return false;  
    }  
    return Double.parseDouble(a) < Double.parseDouble(b);  
}  
  
public String getState() {  
    return this.state;  
}  
}  
BagFactory bagFactory = BagFactory.getInstance();  
  
@Override  
public Tuple exec(Tuple input) throws IOException {  
  
    long c = 0;  
    String line = "";  
    String pbkey = "";  
    V0Line nextLine;  
    V0Line thisLine;  
    V0Line processLine;  
    V0Line evalLine = null;  
    V0Line prevLine;  
    boolean noMoreValues = false;  
    String matchList = "";  
    ArrayList<V0Line> lineFifo = new ArrayList<V0Line>();  
    boolean finished = false;  
  
    DataBag output = bagFactory.newDefaultBag();  
  
    if (input == null) {  
        return null;  
    }  
    if (input.size() == 0) {  
        return null;  
    }  
    Object o = input.get(0);  
    if (o == null) {  
        return null;  
    }  
  
    //Object o = input.get(0);  
    if (!(o instanceof DataBag)) {  
        int errCode = 2114;  
        String msg = "Expected input to be DataBag, but"  
            + " got " + o.getClass().getName();  
    }  
}
```

```
SELECT first_x, last_z  
FROM ticker MATCH_RECOGNIZE (  
    PARTITION BY name ORDER BY time  
    MEASURES FIRST(x.time) AS first_x,  
              LAST(z.time) AS last_z  
    ONE ROW PER MATCH  
    PATTERN (X+ Y+ W+ Z+)  
    DEFINE X AS (price < PREV(price)),  
           Y AS (price > PREV(price)),  
           W AS (price < PREV(price)),  
           Z AS (price > PREV(price) AND  
                z.time - FIRST(x.time) <= 7 ) )
```

250+ Lines of Java and PIG

12 Lines of SQL

20x less code, 5x faster

Plug into the Cloud

ORACLE®
DATABASE

12^c

Application Development

Big Data

Consolidation

Data Optimization

Data Warehousing

High Availability

In-Memory

Performance & Scalability

Security & Compliance

ORACLE®

Where we are today and a look ahead

- SAS High-Performance Analytics, SAS Visual Analytics ready to Deploy on Oracle Engineered Systems
 - Oracle Exadata
 - Oracle BDA
 - Oracle Exalogic
- SAS Dev running Oracle Database 12c beta
 - SAS/ACCESS to Oracle ✓
 - SAS Scoring Accelerator ✓
- Even more benefits for SAS/Oracle customers coming with Oracle Database 12c



Oracle and SAS

- A Stronger partnership than ever
- Oracle Engineered Systems for SAS High-Performance Analytics and SAS Visual Analytics
- Joint engineering and field engagement
 - R&D, Real World Performance Engineering, Partner Technology Solutions, Product Management
- Joint external engagements
 - Sizing, configuration guidance and best practices for deployment
 - Support for POCs
- Fast time-to-market
 - Certification of Oracle Database 12c



Maureen Chew, Principle Software Engineer, Oracle Corporation
Gary Granito, Solution Architect, Oracle Corporation

Comprehensive Platforms for SAS – Complete



Agenda

- Introduction
- SAS Visual Analytics, SAS High Performance Analytics on Oracle Engineered Systems
- Oracle & SAS Collaboration
 - Setting the Stage for Big Data
 - Next Generation Database Technology
- Comprehensive Platforms for SAS - Complete
 - Sun ZFS Storage Appliance for SAS Grid – case study
 - SAS on Current and Next Generation of SPARC servers
 - » Bureau of Labor and Statistics SAS Case Study
 - » SAS on SPARC M5-32 testing highlights



Comprehensive Platforms for SAS - Complete

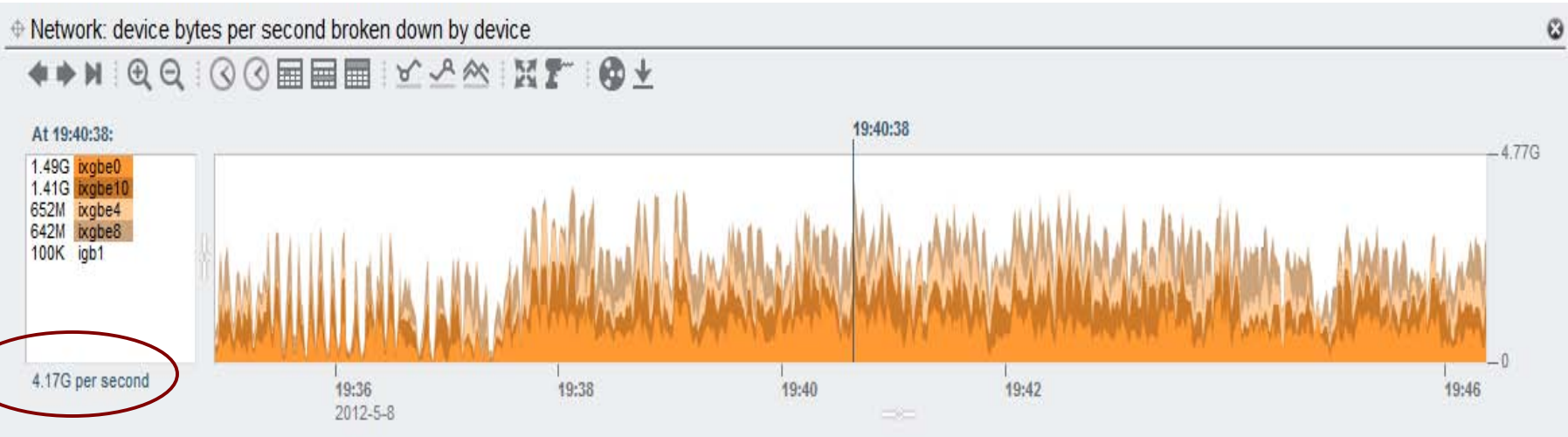
- Sun ZFS Storage Appliance for SAS Grid
 - » Customer POV – SAS Grid, ZFS SA for shared storage, 4 trays – peak 4.17GB/sec
- SPARC T4 momentum – Bureau of Labor Statistics
 - » SAS landscape used to produce monthly metrics
 - » SPARC T4 evaluation metric
- SPARC [M5@SAS](#), product launch last month
 - » Showcasing scale-up – concurrency testing, oversubscribe cores by 2X still shows excellent performance
 - » Large memory brings new meaning for “In-Memory” paradigms
 - » SAS Analytic Server (LASR)
 - » Using tmpfs for SASWORK



SAS Mixed Grid Analytic Workload Results

Customer POV, SAS Grid, Sun ZFS storage appliance

9 Nodes 6 Jobs	SAS Run 5/8/12 19:30:13
Peak GB/s	4.17
Job Completion Time	0:43:32



SAS Grid POC Modeling Results

Oracle / Customer X / SAS POC	2 Trays	2 Trays	2 Trays	4 Trays
Grid Servers (12 Core each)	4	6	8	9
Total Cores	48	72	96	108
Throughput Required (50 MB/s per core)	2400	3600	4800	5400
45% ZFS Load Requirement MB/s	1080	1620	2160	2430
Observed Throughput Load MB/s	2830	3550	3710	4170
Avg NFS Latency to Host ms	1.86	2.46	4	7.87
Max Avg NFS Latency ms (Peak Throughput)	3.23	4.85	5.34	10.23
% Disk Busy Average	4.24%	5.09%	9.45%	22.44%
% Disk Busy Average (Peak Throughput)	10.50%	13.90%	17.02%	35.17%
Avg CPU	13%	13%	18%	24%
Max CPU	84%	86%	89%	90%
Compression	4.1x	4.3x	4.37x	1.12x
Job Run Time	1:17:43	1:06:30	0:51:37	43.32



Comprehensive Platforms for SAS - Complete

- Sun ZFS Storage Appliance for SAS Grid
 - » Customer POV – SAS Grid, ZFS SA for shared storage, 4 trays – peak 4.17GB/sec
- SPARC T4 momentum – Bureau of Labor Statistics
 - » SAS landscape used to produce monthly metrics
 - » SPARC T4 evaluation metric
- SPARC [M5@SAS](#), product launch last month
 - » Showcasing scale-up – concurrency testing, oversubscribe cores by 2X still shows excellent performance
 - » Large memory brings new meaning for “In-Memory” paradigms
 - » SAS Analytic Server (LASR)
 - » Using tmpfs for SASWORK



Bureau of Labor Statistics



NEWS RELEASE



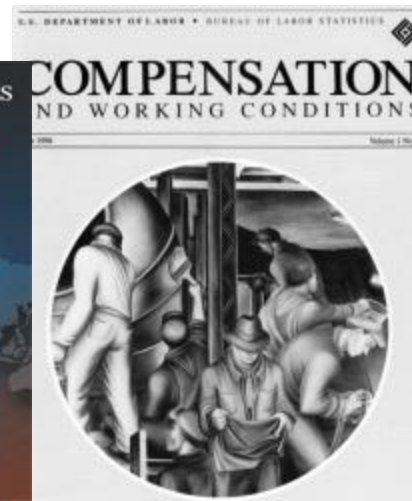
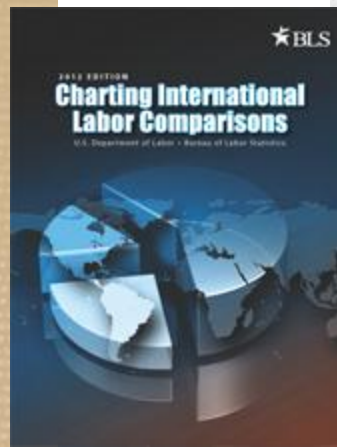
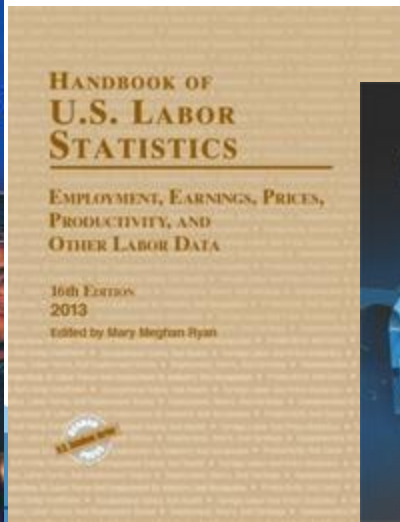
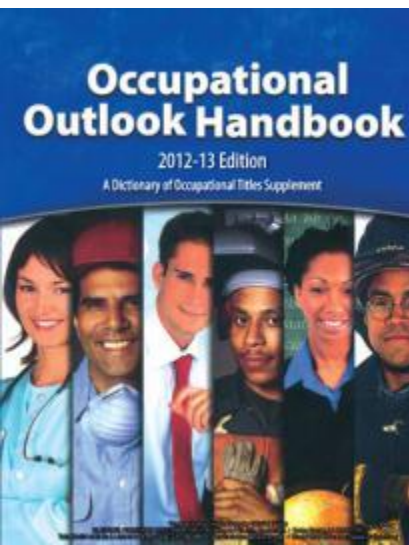
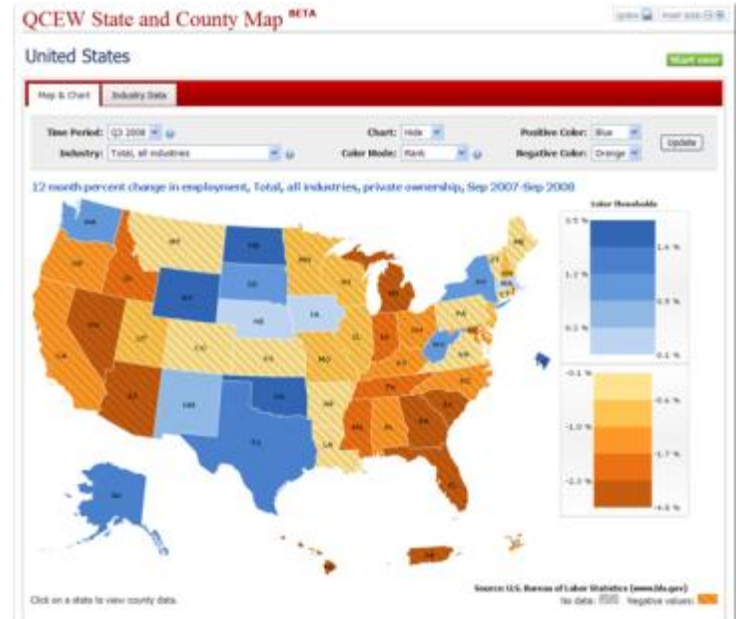
Transmission of material in this release is embargoed until
8:30 a.m. (EDT) Tuesday, April 16, 2013

USDL-13-0668

Technical information: (202) 691-7000 • Reed.Steve@bls.gov • www.bls.gov/cpi
Media Contact: (202) 691-5902 • PressOffice@bls.gov

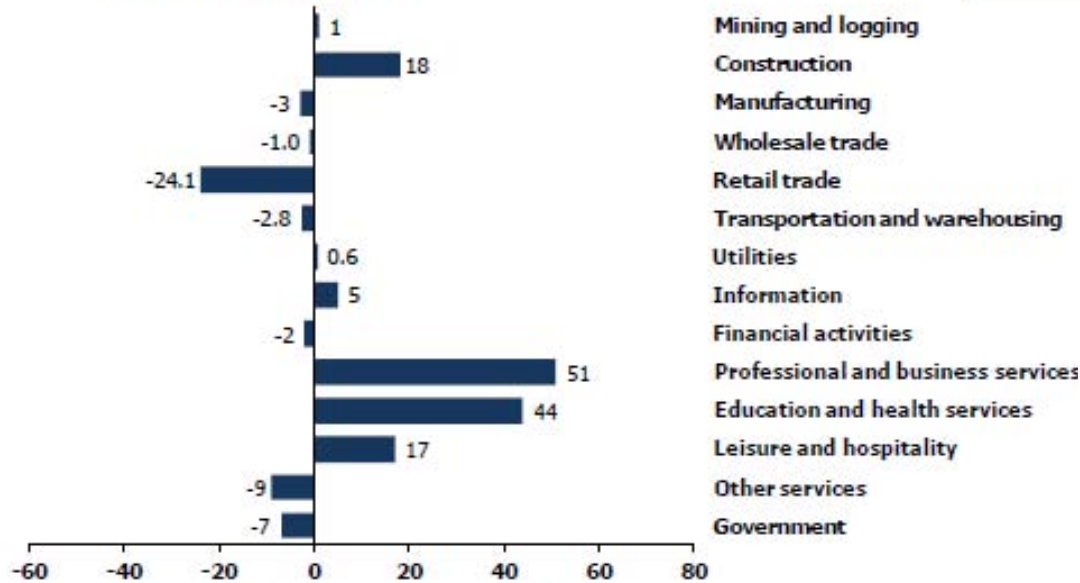
CONSUMER PRICE INDEX – MARCH 2013

The Consumer Price Index for All Urban Consumers (CPI-U) decreased 0.2 percent in March on a seasonally adjusted basis, the U.S. Bureau of Labor Statistics reported today. Over the last 12 months, the all items index increased 1.5 percent before seasonal adjustment.



Industry Employment

Employment in total nonfarm Over-the-month change, March 2013 Seasonally adjusted, in thousands

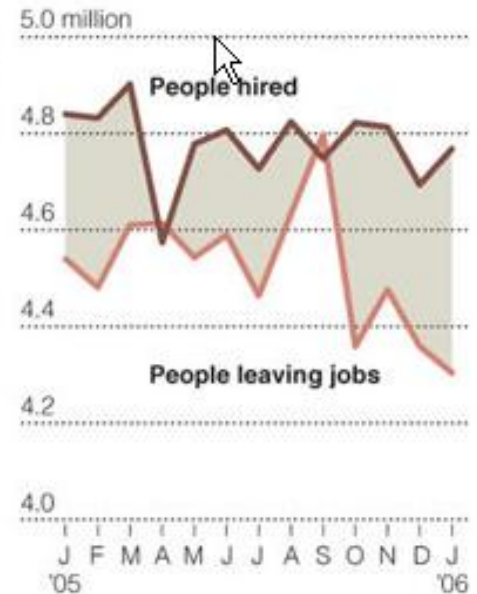


Source: Bureau of Labor Statistics, Current Employment Statistics survey, April 05, 2013.
Note: Data are preliminary.

Total nonfarm
88

Less Job Movement

While hiring has held steady, the number of people leaving jobs has declined.



Source: Bureau of Labor Statistics

The New York Times

Chart 1. Green goods and services private sector employment, 2010–11 annual averages

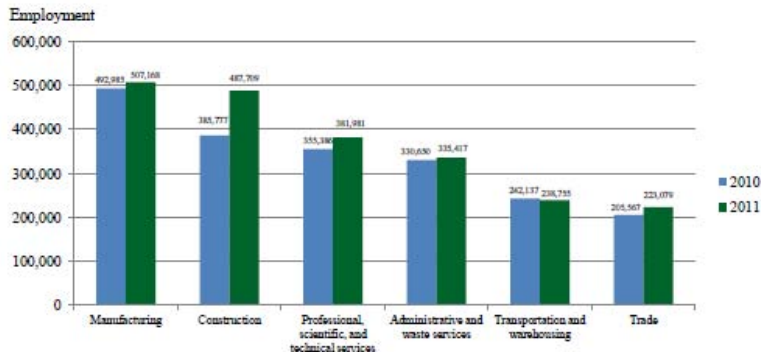
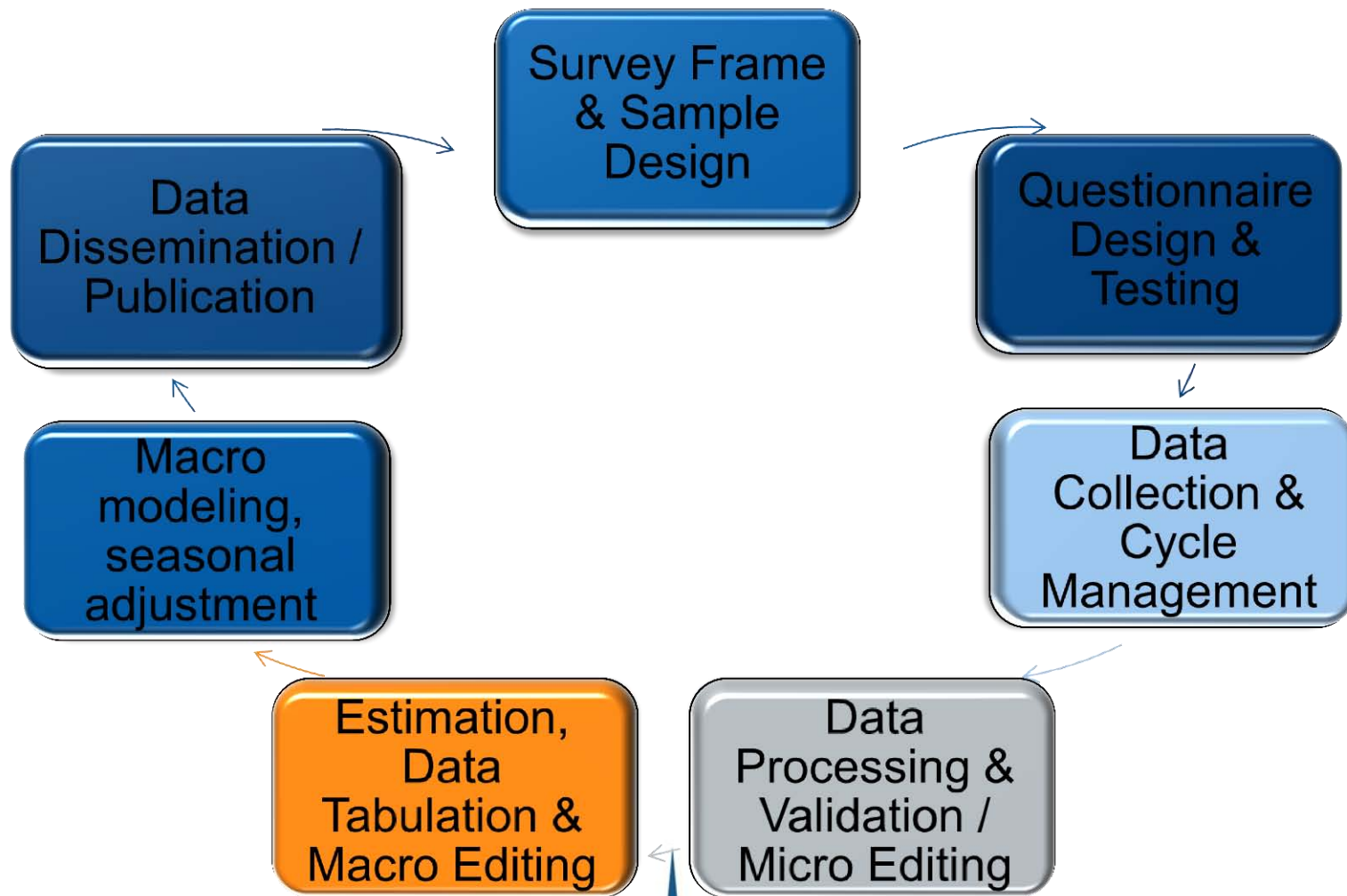


Table 3. Average hourly earnings, 2009 annual averages

Industry	All employees	Production and nonsupervisory employees	Percent difference
Total Private	\$22.21	\$18.62	19.3
Goods-producing	23.77	19.90	19.4
Mining and logging	27.30	23.29	17.2
Construction	24.84	22.67	9.6
Manufacturing	23.03	18.23	26.3
Durable goods	24.51	19.35	26.7
Nondurable goods	20.64	16.56	24.6

UM

Operation and Business Process





SAS/STAT

Microsoft Office Integration

Others



Microsoft®

Office



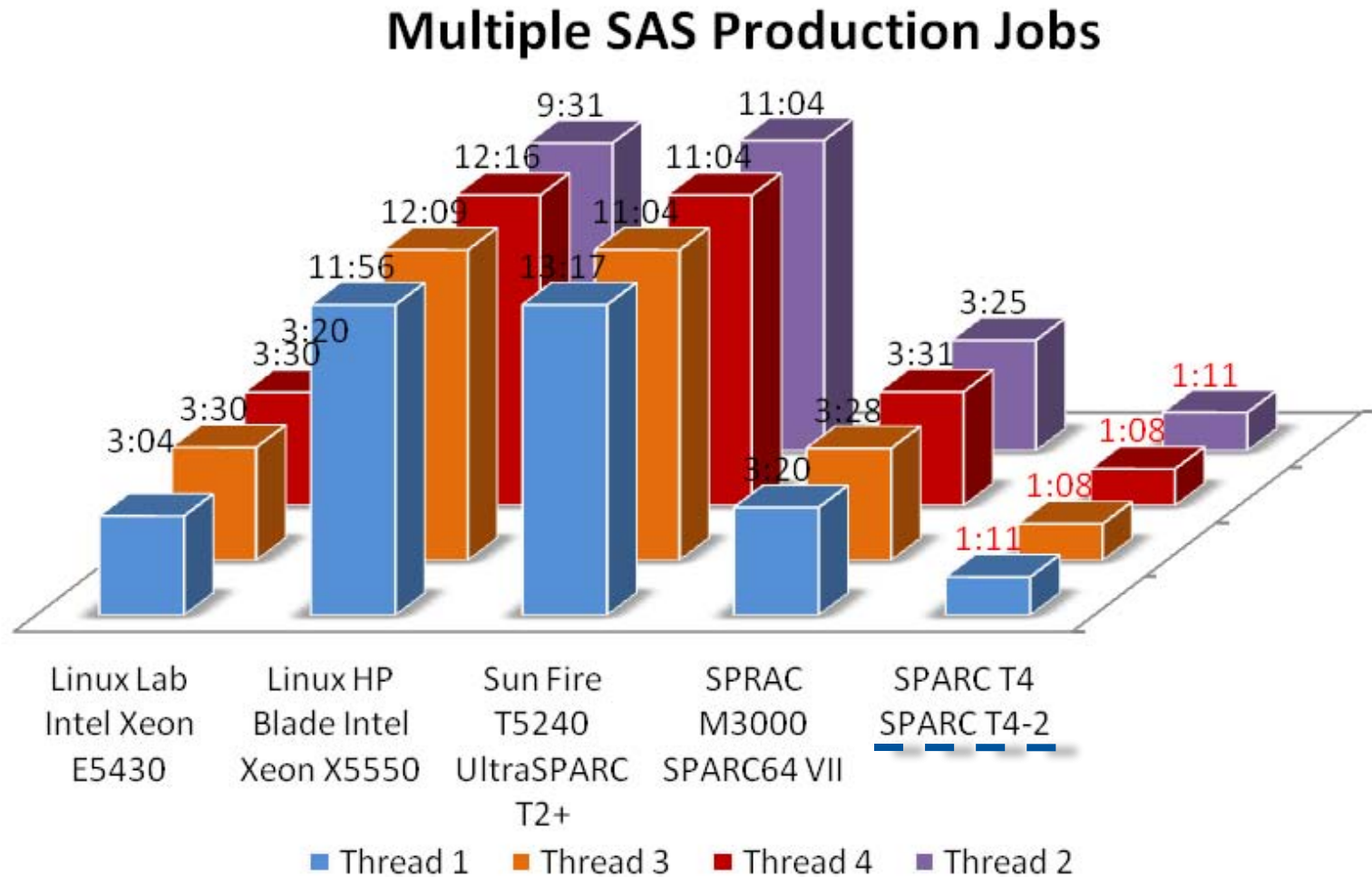
ASCI

SYBASE



```
<html>
<title>HTML</title>
<body>
This is HTML!
</body>
</html>
```

Performance evaluation metric for SPARC T4 @ BLS



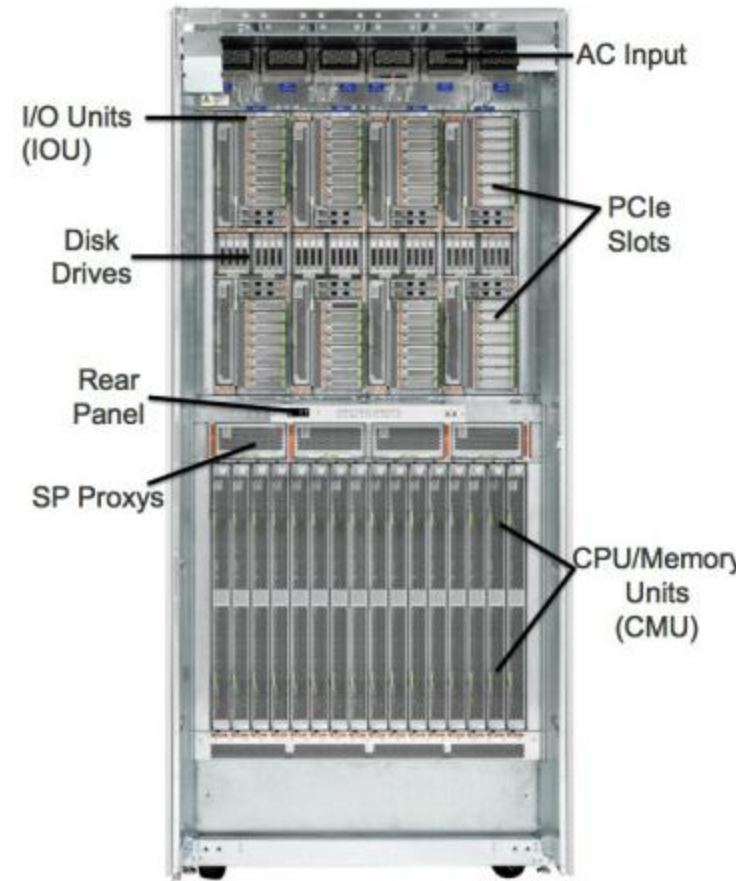
Comprehensive Platforms for SAS - Complete

- Sun ZFS Storage Appliance for SAS Grid
 - » Customer POV – SAS Grid, ZFS SA for shared storage, 4 trays – peak 4.17GB/sec
- SPARC T4 momentum – Bureau of Labor Statistics
 - » SAS landscape used to produce monthly metrics
 - » SPARC T4 evaluation metric
- SPARC [M5@SAS](#), product launch last month
 - » Showcasing scale-up – concurrency testing, oversubscribe cores by 2X still shows excellent performance
 - » Large memory brings new meaning for “In-Memory” paradigms
 - » SAS Analytic Server (LASR)
 - » Using tmpfs for SASWORK



SPARC M5-32 Configuration

- Test Server Details
 - » 32x3.6GHz SPARC M5 (192 cores total)
 - » 8 TB RAM
 - » Solaris 11.1
 - » 4 Physical Domains (Pdom)
 - » Each pdom: 48 cores, 2TB RAM
- I/O
 - » Solaris ZFS
 - » 3 zpools – 6/8 internal disks (/apps), 4 F40 Flash drive partitions (/f0), SAN (3PAR, multipathed, 16 LUNs avail to 4 PDoms)



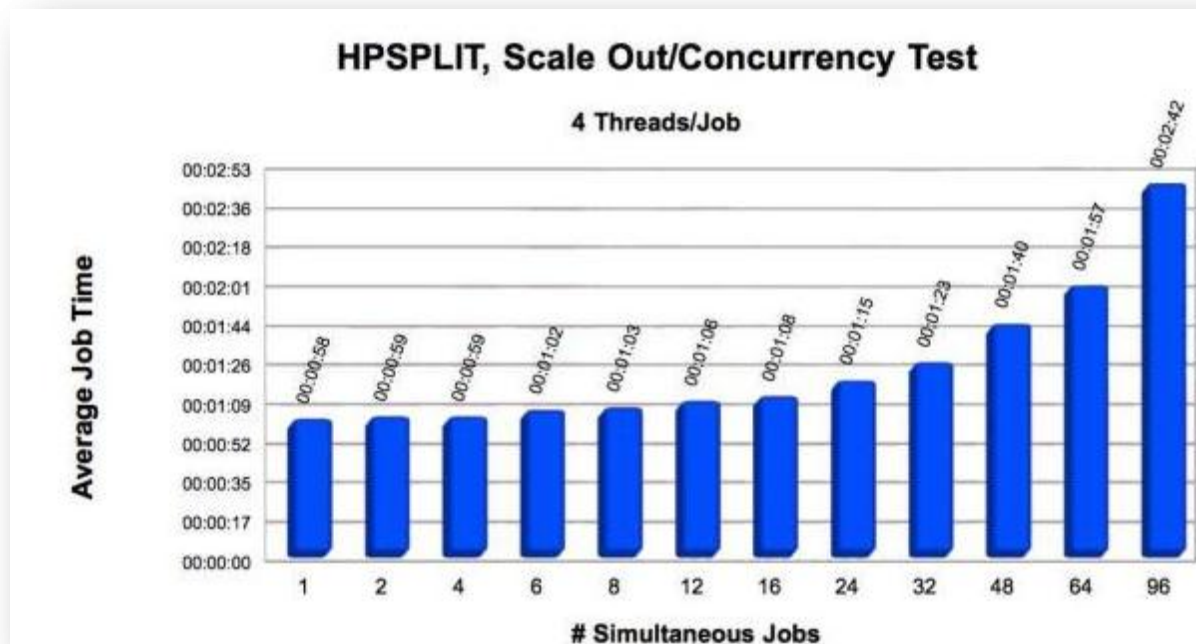
Chronicle of a Beta Test

- Jan 2013
 - » T-minus 3 months to launch
 - » M5 arrives at SAS – from pallet to raised floor



Scalability that Stands the Test

- Concurrency Testing
 - » 96 hpsplit jobs (@CPUCOUNT=4) with 48 cores
- Excellent performance for average job time
 - » Despite system being oversubscribed by 2X
 - » ~2.75min avg job time @concurrency=96 vs ~1min at single job concurrency



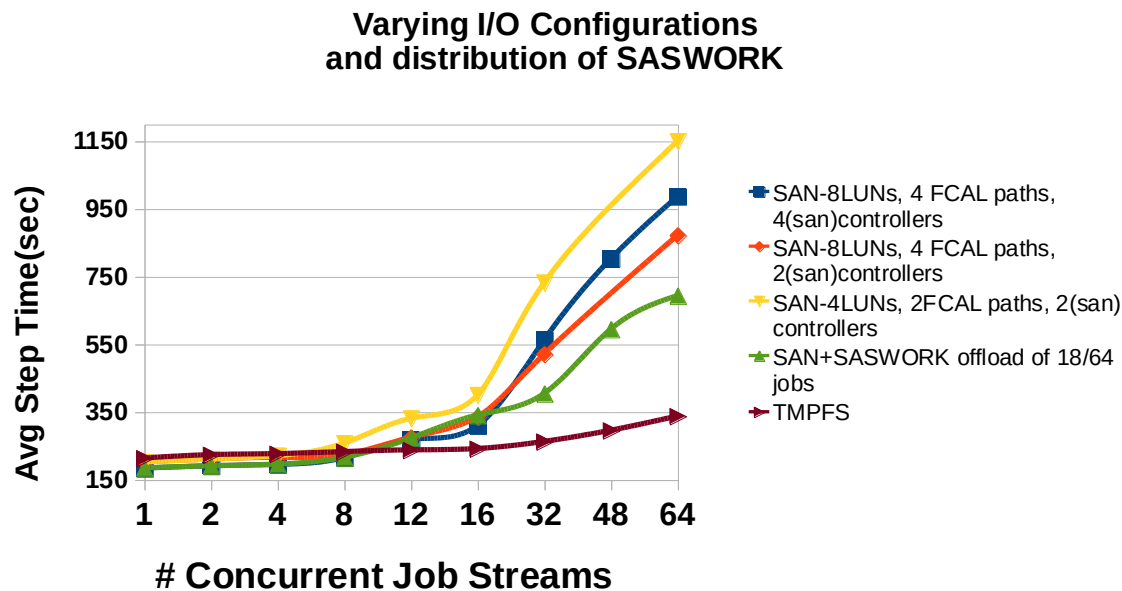
SAS In-Memory Analytic Server (LASR)

- LASR running in SMP fashion
- Single Physical Domain
 - » 48 cores, 2TB RAM
 - » Solaris 11.1
- Lift 100GB table from Exadata to memory
- New 9.4 “hp” PROCs running in multi-threaded fashion from LASR data source



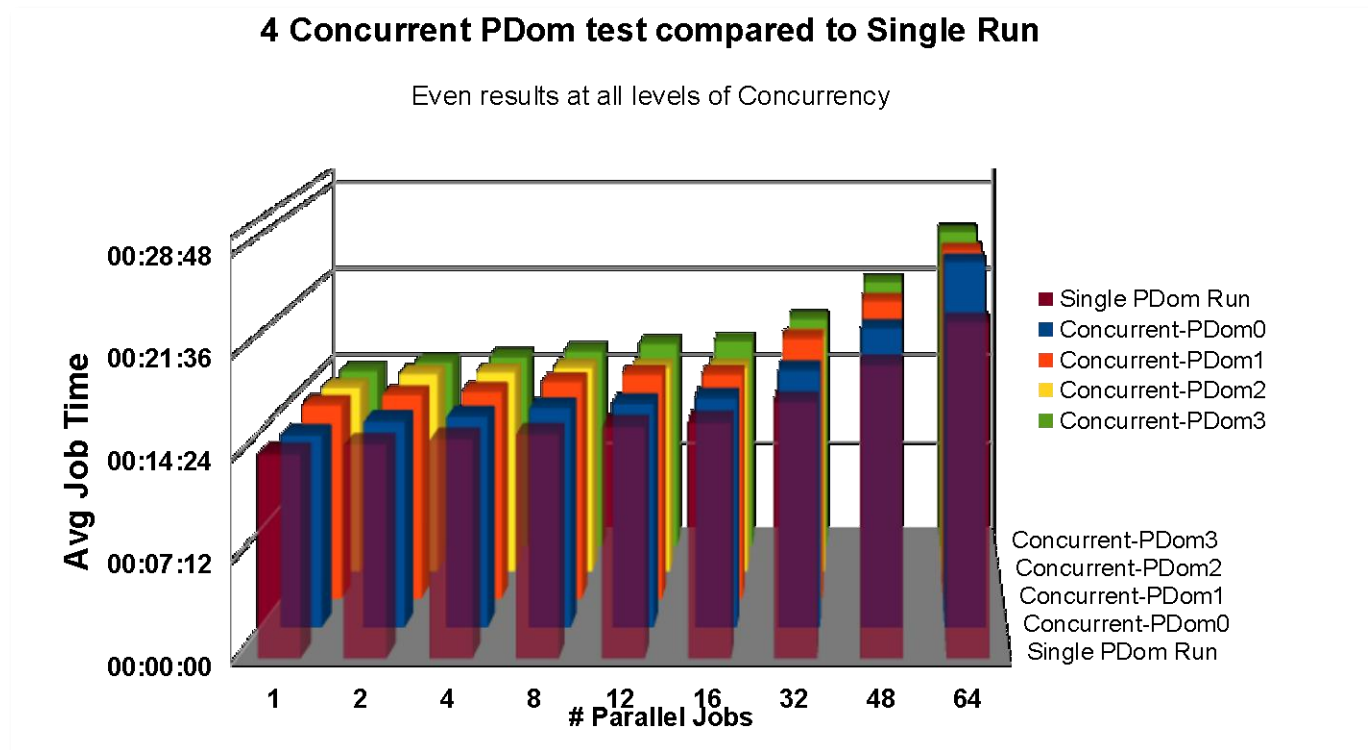
Real World SAS Global Trade Application

- Complex multi-step sas application scalability profile: running identical streams in parallel to model performance at different I/O configurations
- Single PDom, scale up test with SAS 9.4
- Since SAN becomes oversubscribed, move 18 jobs from SASWORK on SAN to internal (internal disks and flash) pools (green)
- TMPFS (red) moves I/O to “in-memory”



Real World SAS Global Trade Application

- Scale up test and run same workload simultaneously across 4 Pdoms
- Use all TMPFS for SASWORK to eliminate SAN and to stress memory subsystem
- At peak, 256 CPU/memory intensive applications are running simultaneously. Each app utilities 750MB memory (MEMSIZE) as well as 80+GB I/O going to TMPFS



Thank You!





San Francisco, CA
April 28–May 1, 2013

