DB2 Now! – DB2 V11.1新功能介绍

张光业
gyzhang@cn.ibm.com
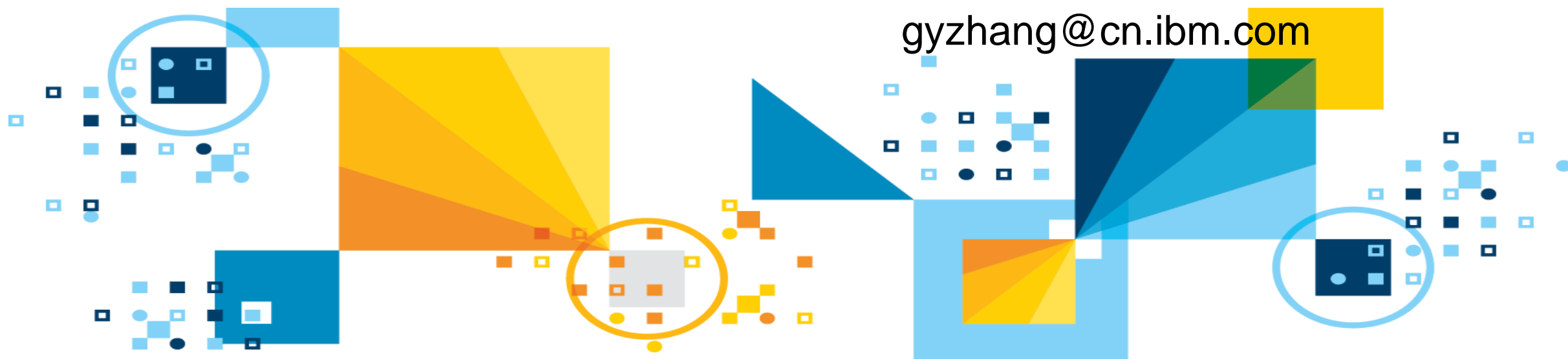
DB2 V11.1 新功能概览

核心业务:
扩展DB2的领导地位

数据仓库类分析:
用途最广、扩展最强的内存计算数据仓库平台

全面的企业级安全

企业级加密
• 集中式密钥管理(KMIP)

仅次于DB2 for zOS的高可用性

简单快速部署
• 几个小时就可以装好运行
更高的可用性
• 采用HADR容灾方案实现零数据丢失
• 更多的在线管理
更多的平台支持
• Power Linux (LE)
• RDMA (x86) 虚拟化

显著的核心数据库改进

极大数据库性能
• 更高的业务吞吐量
简单、快速、更多的在线升级
• 更快、无需离线备份
• 流线型HADR 升级
• DB2 9.7 直接升级到 11.1

大规模扩展数据仓库，内存计算的性能

MPP BLU 扩展能力
• PB 级内存计算数据仓库

下一代内存计算性能, 功能和负载

• **更快的ELT/ETL 性能**
• **优化了更多的查询负载**
• **支持了更多的功能**
  • 派生列
  • RCAC
  • OLAP + BLU性能

增强的兼容性

多语言SQL 支持
• PostgresSQL
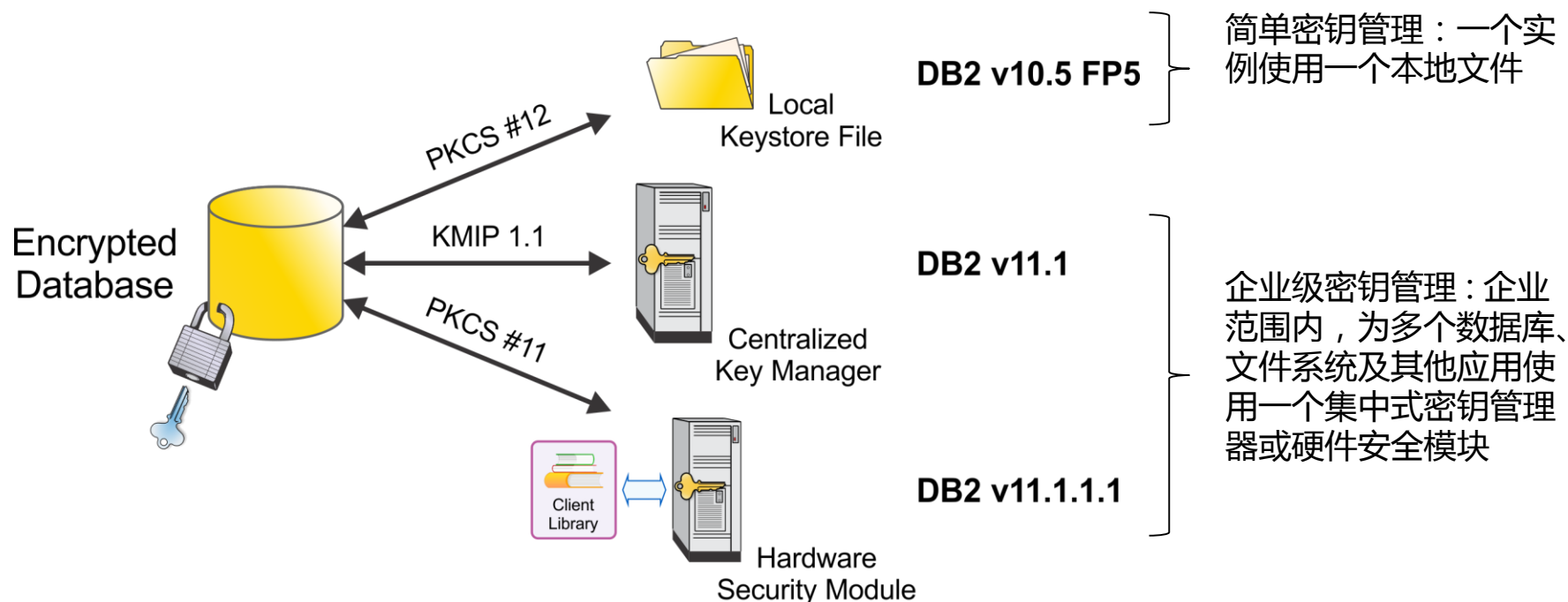**支持欧洲语言**
• 代码页819

DB2 原生加密功能支持硬件安全模块（HSM）

- **DB2 V11.1 GA版本支持KMIP 1.1 集中式密钥管理**
  - IBM Security Key Lifecycle Manager (ISKLM), SafeNet KeySecure 等

- **DB2 11.1.1.1 支持硬件安全模块**
  - Gemalto Safenet HSM (之前的Luna) 版本6.1及以上
  - Thales nShield HSM, Security World Software version 11.50

简单密钥管理：一个实例使用一个本地文件

企业级密钥管理：企业范围内，为多个数据库、文件系统及其他应用使用一个集中式密钥管理器或硬件安全模块

DB2 pureScale 增强

- 简化了安装部署
- SYNC 和NEARSYNC HADR 模式支持零数据丢失容灾
- GDPC 4交换机支持
- Power Linux 小端支持
- Linux RDMA (x86) 虚拟化支持
  - VMware
- 全文检索支持
- 增强性能监控和排错
- 集群健康检查工具
  - db2cluser –verify
  - Validations performed include, but are not limited to, the following:
    - Configuration settings in peer domain and GPFS cluster
    - Communications between members and CFs
    - Replication setting for each file system
   - Status of each disk in the file system

DB2 pureScale HADR 支持SYNC 和NEARSYNC 模式

▪ **SYNC 和 NEARSYNC 模式增加到 pureScale HADR中**

  – 该项改进将把 pureScale HADR强壮的容灾能力和 pureScale的持续可用性结合在一起，提供了零数据丢失（RPO=0）的容灾能力

 – HADR peer 窗口(hadr_peer_window) 不受支持

▪ **pureScale HADR 现在支持的功能包括:**

  – SYNC, NEARSYNC, ASYNC 和 SUPERASYNC 四种模式

  – 延时更新、日志缓冲

  – 非强制接管（角色切换）和强制接管（故障切换）

Primary Cluster

HADR

Standby DR Cluster
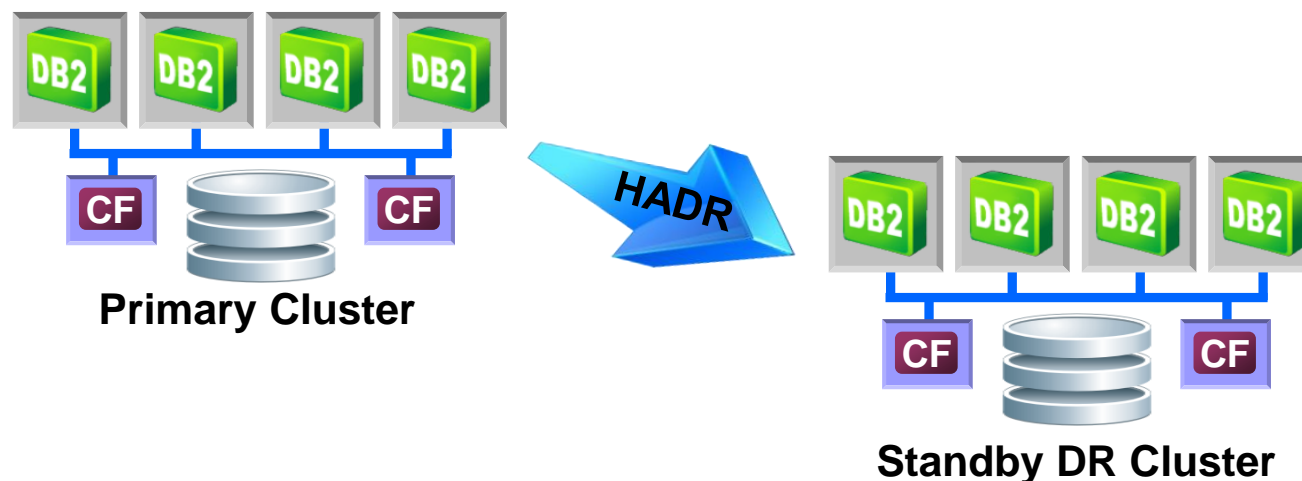
Linux 虚拟化的增强

- **RDMA over Converged Ethernet (RoCE) support added in VMWare**
  - RoCE SR-IOV for RHEL 7.2 only
  - Allows a single adapter to be shared across multiple partitions
- **Single-Root I/O Virtualization (SR-IOV)**
  - Standard that enables one PCI Express (PCIe) adapter to be presented as multiple separate logical devices (Virtual Functions) to virtual machines
  - Allow the virtual machines to run native RoCE and achieve near wire speed performance.
  - Can be enabled on Mellanox ConnectX-3/ConnectX-3 Pro/Connect X-3 VPI adapters for Ethernet
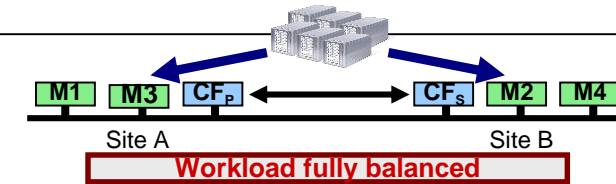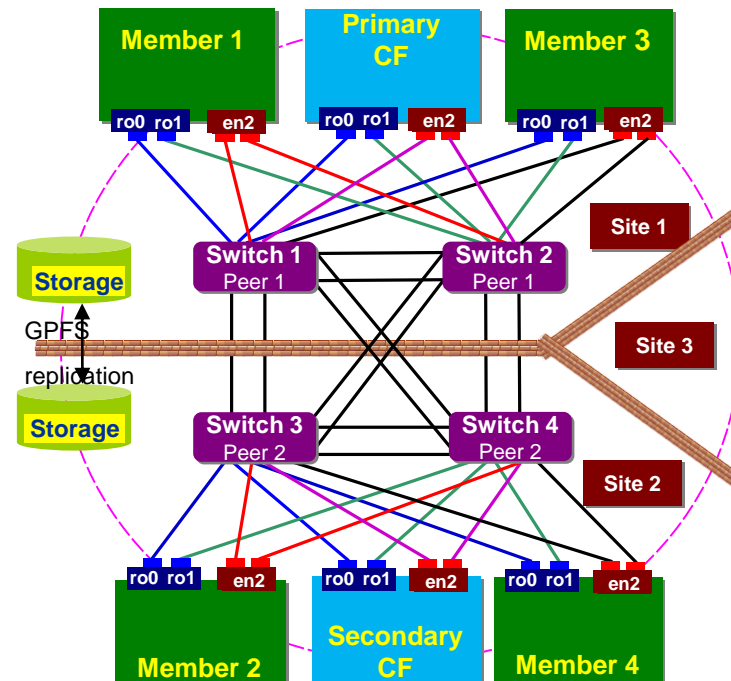
VMWare RoCE 卡共享举例
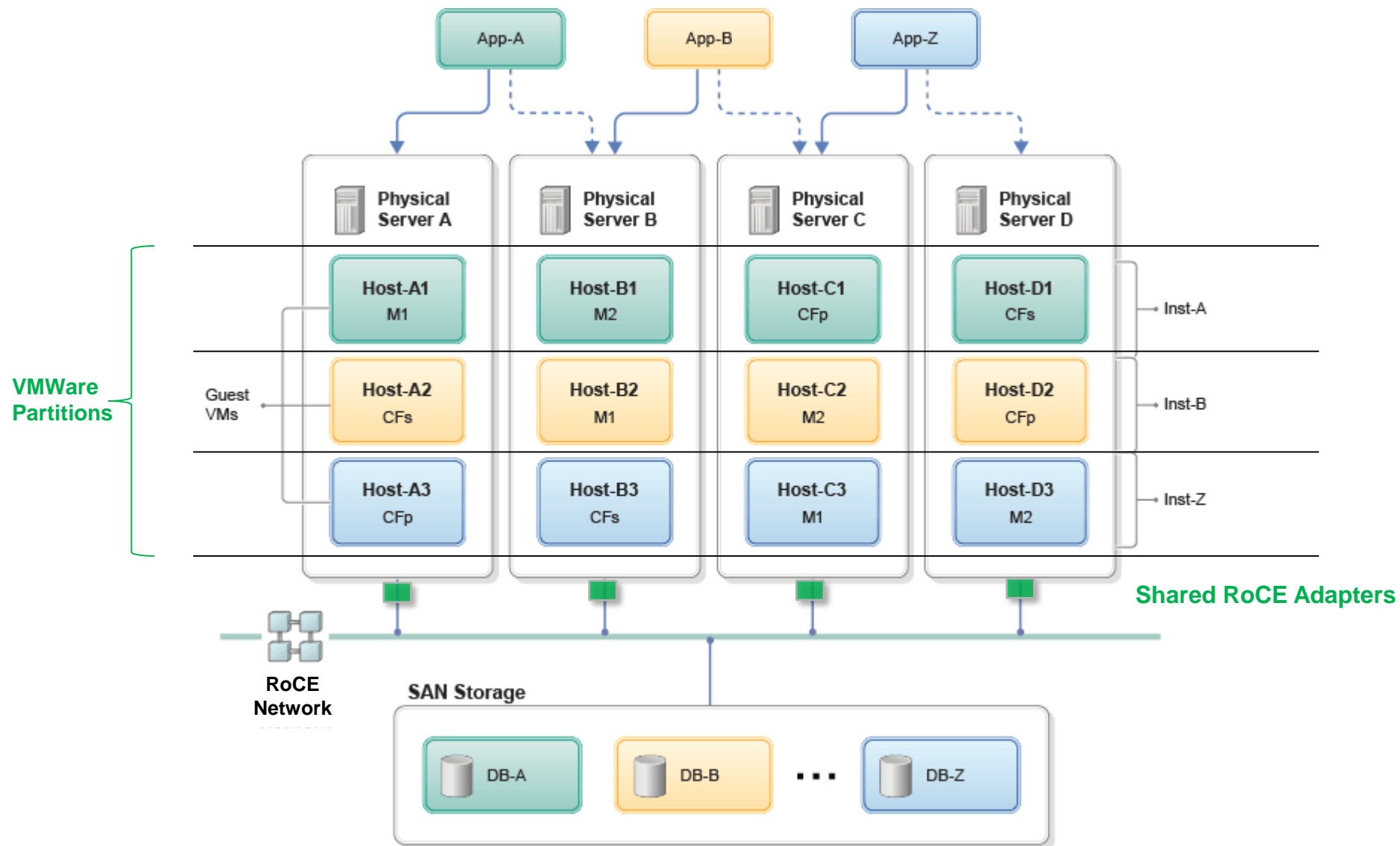
GDPC 支持的增强

- **DB2 V11 adds improved high availability for Geographically dispersed DB2 pureScale clusters (GDPC) for both RoCE & TCP/IP**

  - Multiple adapter ports per member and CF to support higher bandwidth and improved redundancy at the adapter level

  - Dual switches can be configured at each site to eliminate the switch as a site-specific single point of failure (i.e. 4-switch configuration)
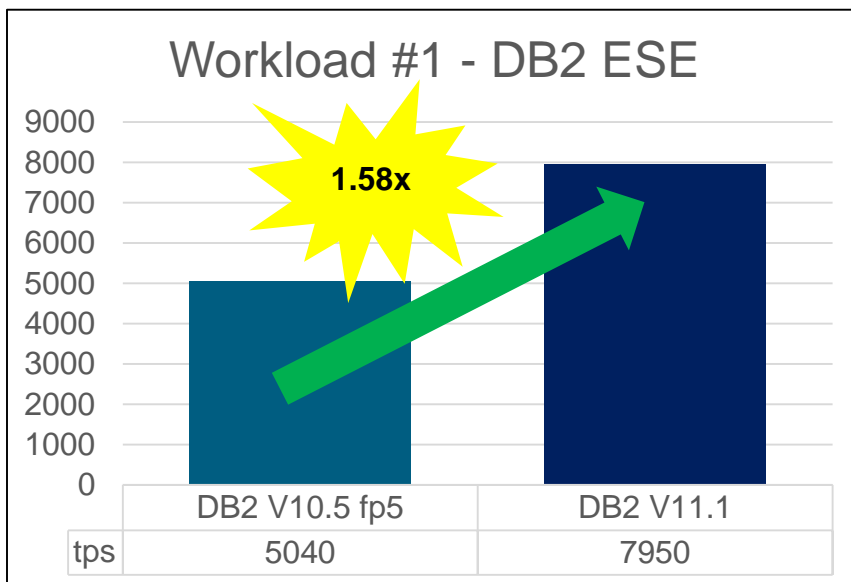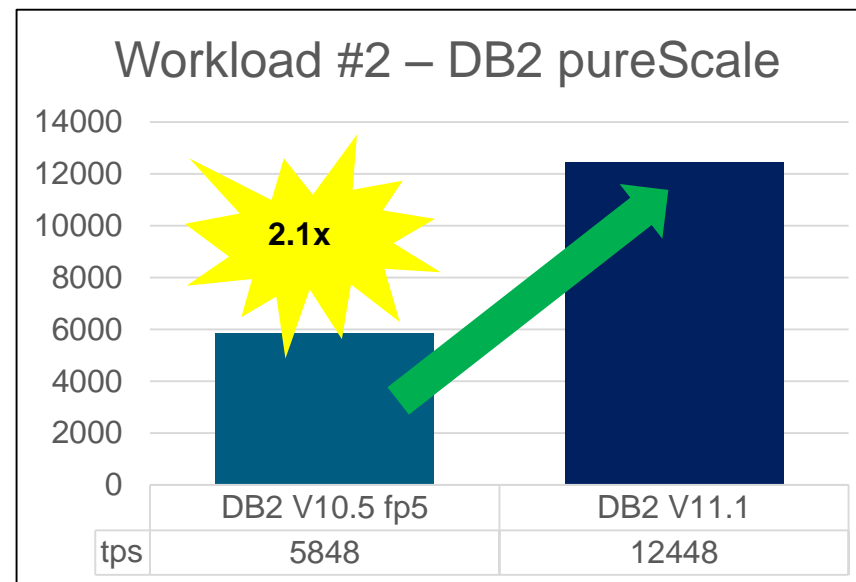
© 2016 IBM Corporation

高并发负载性能改进

- **DB2 V11.1实现了流线型缓冲池Latch管理协议**
  - 减少大型多线程系统上的冲突
  - 对联机交易类负载特别有帮助

	Workload #1 - DB2 ESE	
	DB2 V10.5 fp5	DB2 V11.1
tps	5040	7950

1.58x

	Workload #2 – DB2 pureScale	
	DB2 V10.5 fp5	DB2 V11.1
tps	5848	12448

2.1x

- • Workload 1 based on an industry benchmark standard
- • POWER7 32c, 512 GB

- • Workload 2 implements a warehouse-based transactional order system
- • 4 members, 2 CFs with 16c, 256 GB

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

流线型升级

- **可以从 DB2 V9.7, V10.1 和 V10.5 (跨3个版本)升级到DB2 V11.1**
- **支持跨数据版本执行roll-forward处理**
    - **单分区** DB2 企业版或DB2 pu reScale
    - 至少需要从DB2 V10.5 Fix Pack 7或以上补丁升级
- **升级前后不再需要执行数据库离线备份**
    - 数据库恢复支持跨数据库版本执行roll-forward处理

- **HADR 环境现在在主机升级后无需重新初始化备库**
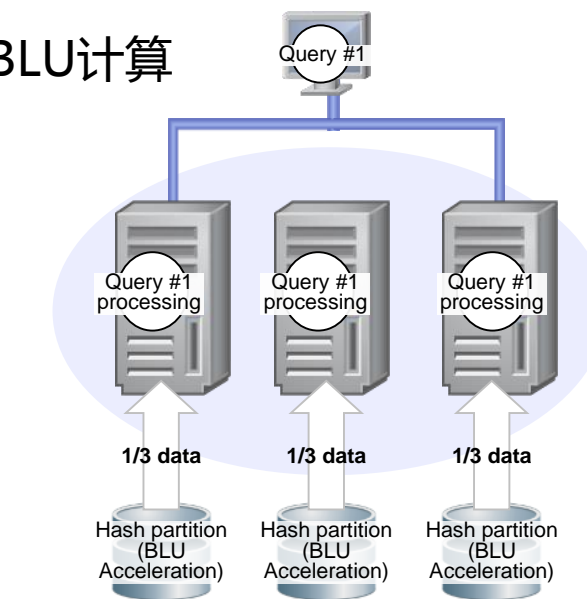   - 限**单分区**DB2企业版用户从DB2 V10.5 Fix Pack 7或以上补丁升级

BLU 加速: MPP 水平扩展

- **技术**
  - 无处不在的 SMP 和 MPP 查询并行
  - 分区内和分区间同时并行，内存优化、列式存储，SIMD支持，BLU计算

- **价值定位**
  - 改善响应时间
    - 所有服务器均参与一个查询的计算
  - 大规模扩展数据
    - 大大超过实际限制
  - 流线型BLU 使用
    - 将BLU加速功能集成到现有数据仓库

Query #1

Query #1 processing   Query #1 processing   Query #1 processing

1/3 data   1/3 data   1/3 data

Hash partition (BLU Acceleration)   Hash partition (BLU Acceleration)   Hash partition (BLU Acceleration)

DB2 10.5 BLU Capacity	DB2 V11.1 BLU Capacity
10s of TB	1000s of TB
100s of Cores	1000s of Cores

BLU MPP 线性扩展证明

- **IBM Power E850 上运行DB2 V11.1**

Scaling Hardware at constant Data Volume

	3 Node (10TB) - 6 MLNS	6 Node (10TB) - 12 MLNS
QpH	111	213

(1.92x)

Scaling Hardware along with Data Volume

	3 Node (10TB) - 6 MLNS	6 Node (20TB) - 12 MLNS
QpH	113	109

(QpH Held up!)
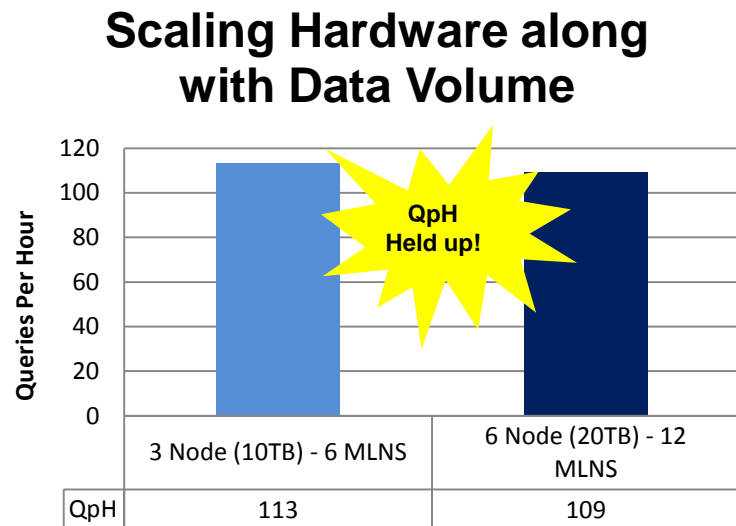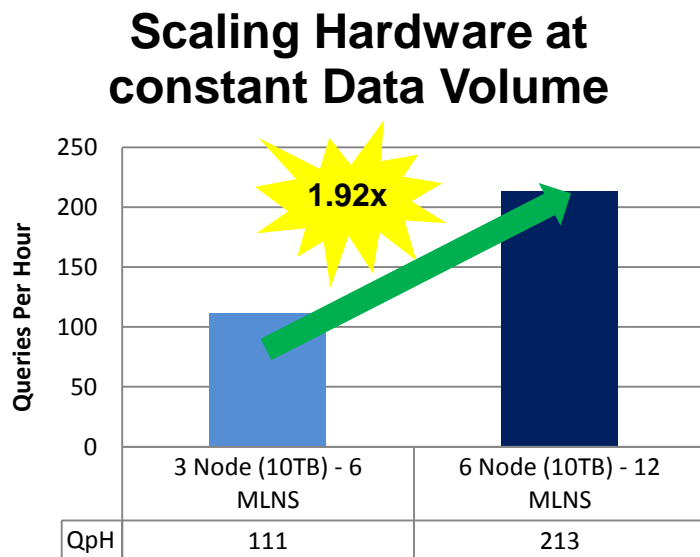
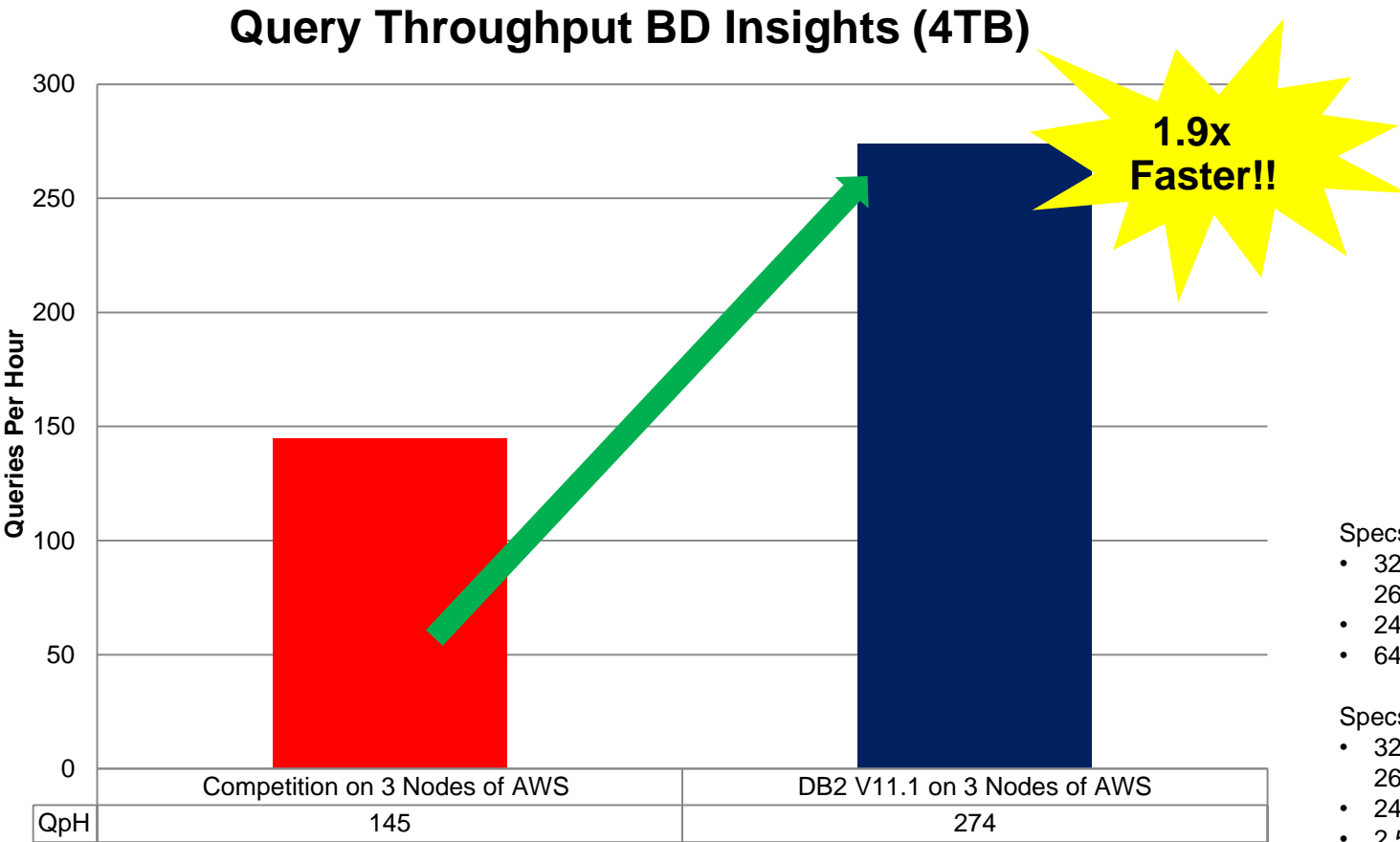- **用两种不同方式衡量扩展性**
  - 硬件配置加倍但是数据库大小不变
  - 硬件配置和数据库大下同时加倍
  - 两项测试均使用 BD Insights（TPC-DS）深度分析型负载

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

杰出的BLU MPP 性能

- **DB2 Version 11.1比友商快1.9倍**

**Query Throughput BD Insights (4TB)**

1.9x
Faster!!

Queries Per Hour

Specs of each BLU MPP node in AWS EC2:
- 32 cpu threads of Ivy Bridge (Intel Xeon E5-2670 v2) virtual cores
- 244 GiB of RAM
- 640 GB of SSD

Specs of each node of the Competition:
- 32 cpu threads of Ivy Bridge (Intel Xeon E5-2670 v2) virtual cores
- 244 GiB of RAM
- 2.56 TB of SSD

	Competition on 3 Nodes of AWS	DB2 V11.1 on 3 Nodes of AWS
QpH	145	274

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

BLU 加速: ELT 和 ISV 应用大幅性能改善

BLU 声明的全局临时表 (非记日志的DGTT) 并行
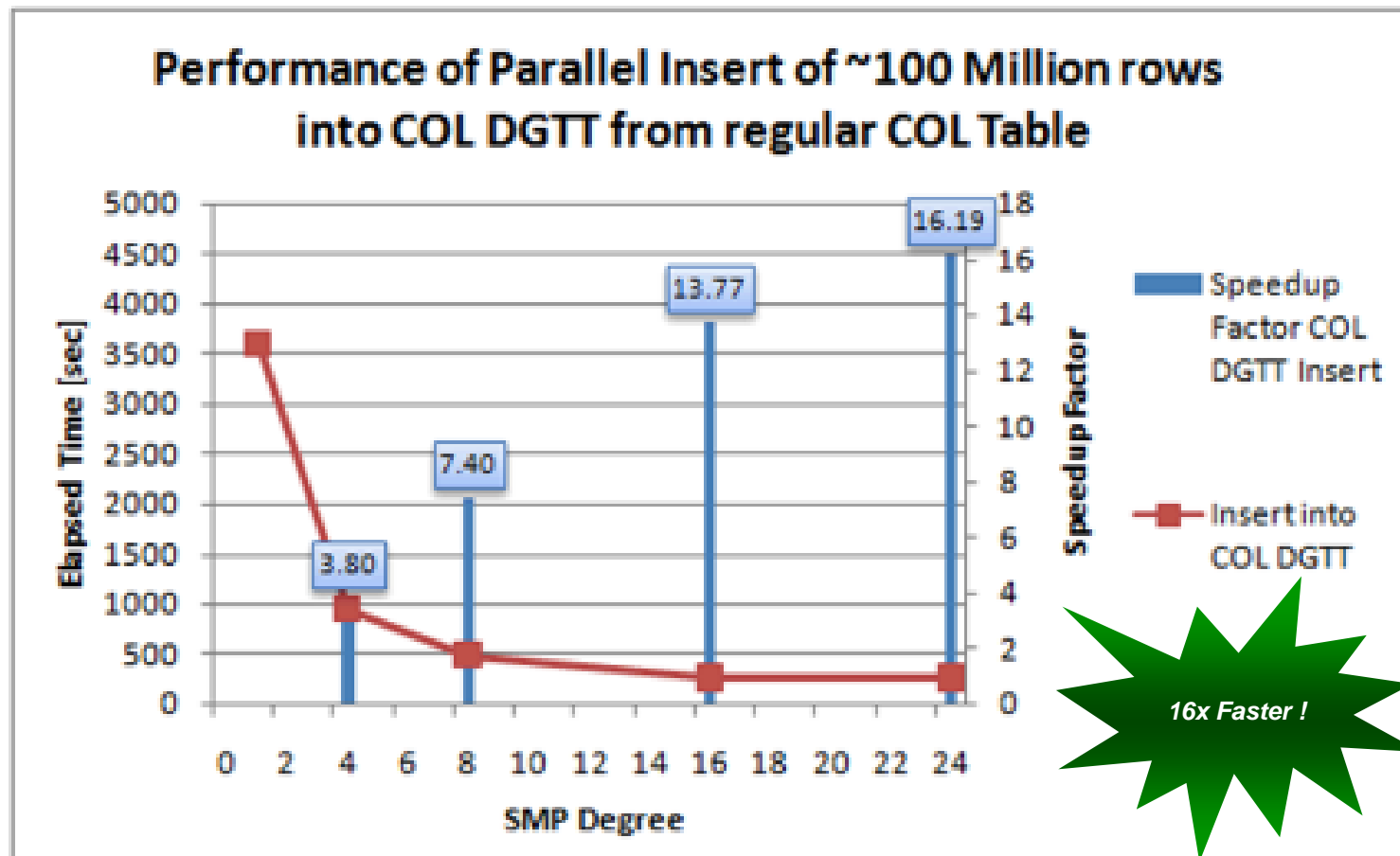
Performance of Parallel Insert of ~100 Million rows into COL DGTT from regular COL Table

16x Faster !

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

针对列式存储表优化的SQL 的支持

- **SQL OLAP improvements for deeper in-database analytics with column-organized tables**
- **Additional Oracle Compatibility Support**
  - Wide rows
  - Logical character support (CODEUNITS32)
- **DGTT support (except not logged on rollback preserve rows)**
  - Parallel insert into not-logged DGTT from BLU source
- **IDENTITY and EXPRESSION generated columns**
- **European Language support (Codepage 819)**
- **NOT LOGGED INITIALLY support**
- **Row and Column Access Control (RCAC)**
- **ROWID Support**
- **Nested Loop Join Support**

DB2 : a polyglot database

BLU性能的提高

- **BLU Acceleration includes a number of performance improvements in the columnar engine**
  - Faster SQL MERGE processing
  - Nested Loop Join Support
  - Synopsis Table Maintenance
  - Industry Leading Parallel Sort
  - Improved SORTHEAP utilization
  - Query Rewrite Improvements
  - Push-down of a number of SQL & OLAP functions into the BLU engine
  - Optimized SQL support for BLU

业界领先的并行排序技术

- **Leverages the latest sort innovations from IBM TJ Watson Research and DB2 Development**
  - Enhancements can increase BLU Acceleration performance by as much as 13.9X

- **BLU Sort+OLAP on SMP Environment**
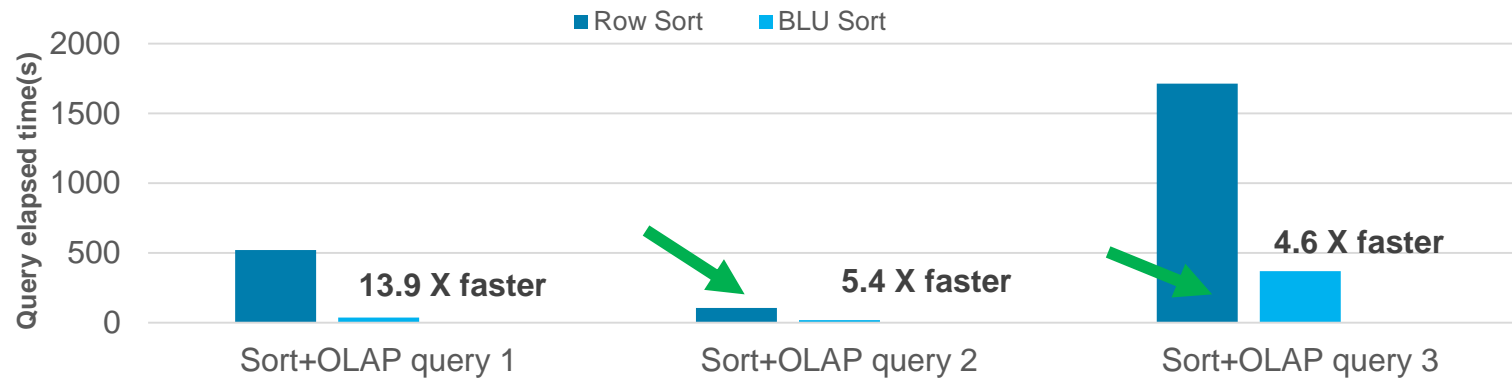
- **Configuration Details**
  - On 4-socket Intel Xeon platform with 72 Cores and 742G RAM
  - 1 TB TPC-DS database
  - Query scenarios involving multiple sort and OLAP operations

Synopsis表维护的改进

▪ **Buffer tuples in memory and write them to synopsis table every 1024 rows**

- Up to 46% improvement in insert performance!
- Up to 87% reduction in synopsis table size!
- Up to 60% reduction in elapsed time with queries with range predicates!

OLTP
+ OLAP
= HTAP
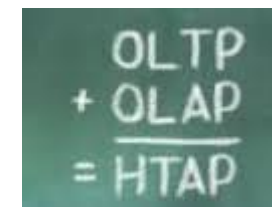
Elapsed Time of INSERT With and Without Synopsis Optimization
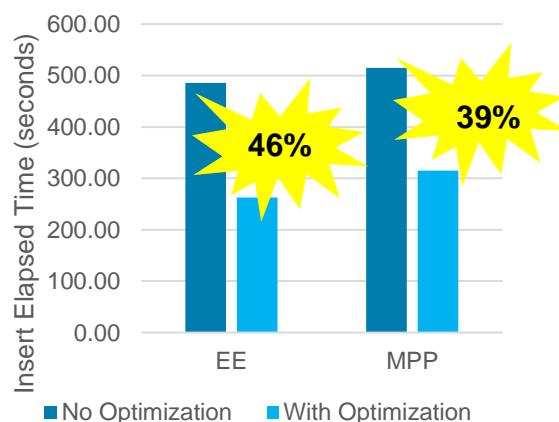
Size of Synopsis Table With and Without Synopsis Optimization
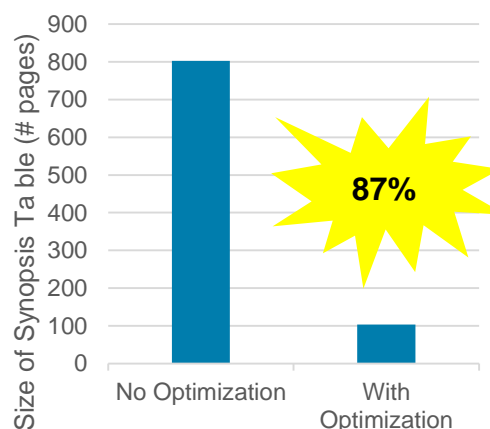
Query Elapsed Time With and Without Synopsis Optimization
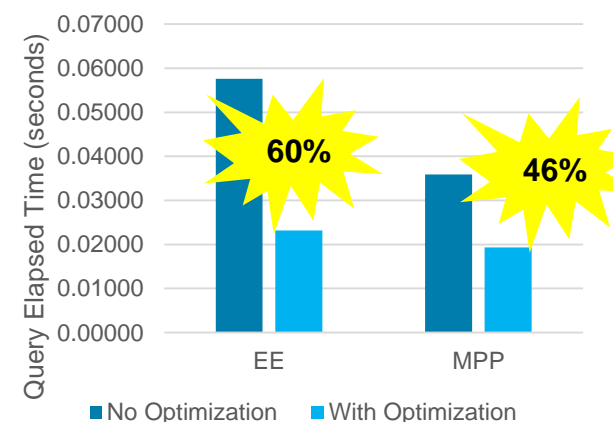
- IBM p760 / POWER7+ 32 cores / 1TB RAM
- Table with 50 columns.

- Single user
- Insert test 100K rows with commit count = 1
- Select query with 10 pairs of range predicates

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.
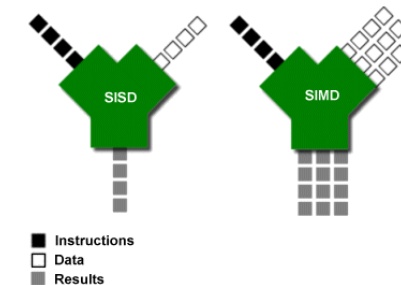
新增的BLU 性能提高

- **BLU Automatic Dictionary Creation on uncommitted data**
  - Earlier dictionary availability for Ingest/Insert/Import
    - for large insert jobs such as INSERT with *subselect*
  - Higher PCTENCODED values in SYSCAT.COLUMNS
  - Improved query performance following insert
- **BLU SIMD enhancements**
  - Intel AVX2, SSE2, SSSE3, SSE4
  - z13, Power8, and PPC LE
  - More functions and bit operations, SIMD comparisons
- **Improved BLU performance in some cases for queries with OLAP and ORDER BY through sort elimination**
- **DECIMAL data type improvements (not specific to BLU tables)**
  - Optimized conversions between row and column organized formats
    - Intel, PPCLE, z Systems (previously optimized for AIX)
  - SUM optimizations for Linux on z Systems

■ Instructions
□ Data
■ Results

BLU支持新的数据类型

▪ **BINARY and VARBINARY**

- – Binary data to be stored and manipulated without overhead of BLOB type
  - A binary string is a sequence of bytes that are used to store pictures, sound, etc
- – BINARY and VARBINARY data types are compatible with each other and are compatible with the BLOB data type
  - They are not compatible with character string data types, except those character strings that are defined as FOR BIT DATA
- – Support for BINARY and VARBINARY data types enhances compatibility with other relational database management systems
  - BINARY can contain 1-254 bytes
  - VARBINARY can be up to 32672 bytes

▪ **Column support for BOOLEAN data type – True, False, Unknown (NULL)**

- – Restrictions include
  - Replication (SQL / Q Replication, CDC)
  - Embedded SQL support – specifically host variable support
  - ALTER NICKNAME – column type change to/from BOOLEAN
  - ANALYZE_TABLE expression

工作负载管理器（WLM）增强

- **WLM 调度器: 数据库级别的CPU控制**
  - 数据库级CPU时间片和限制机制实现了多数据库环境下CPU分配优先级控制
    - wlm_cpu_shares, wlm_cpu_share_mode, 及wlm_cpu_limit 数据库参数
    - 以前只能在一个数据库内的服务类（service superclass）和服务子类（service subclass）级别进行控制

数据库X: 高优先级
(700 软CPU 时间片限制)

数据库Y: 低优先级
(300 硬CPU时间片限制)

Service Class A	Service Class B	Service Class C
CPU Share 500	CPU Share 200	CPU Share 300
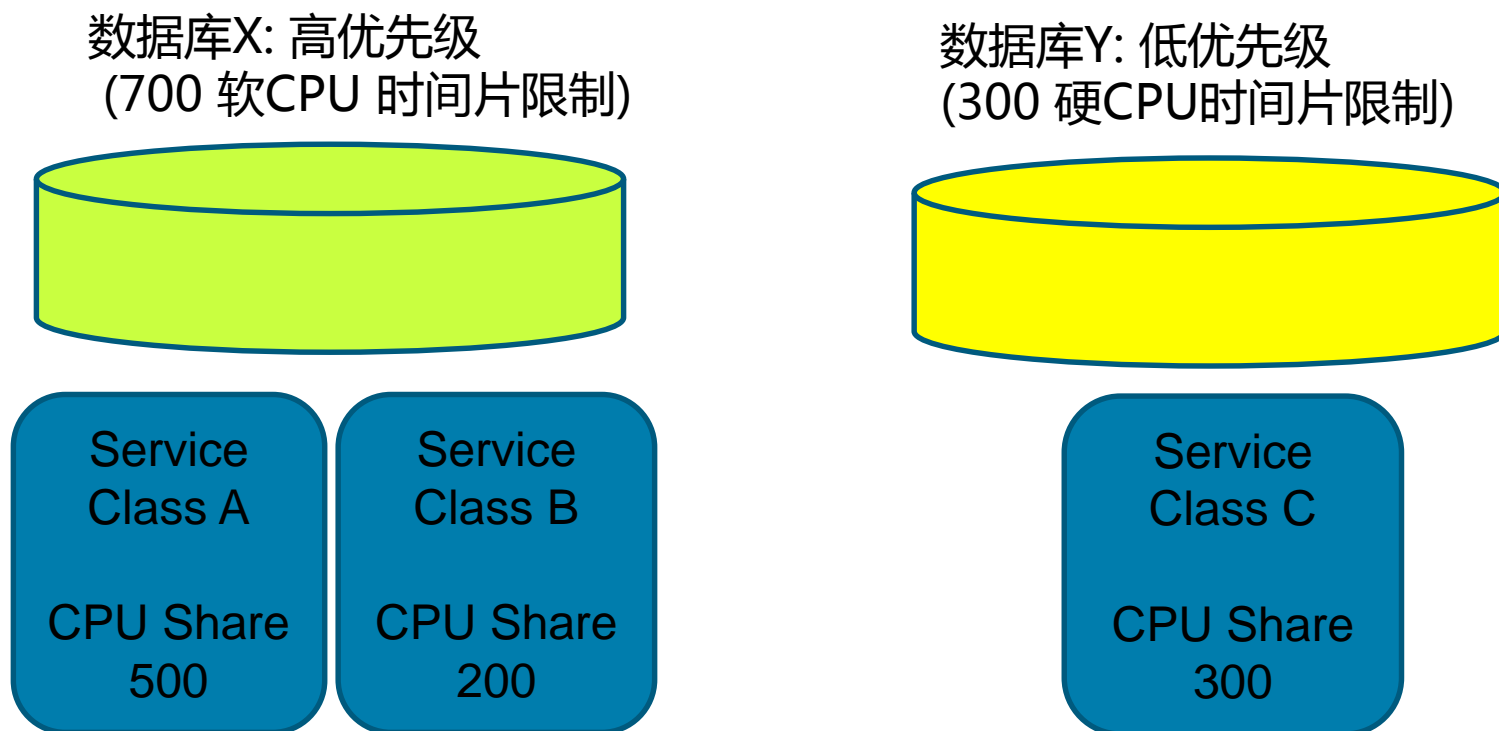
数据联邦增强和简化

▪ **数据联邦的功能扩展到更多的数据源**
  - 更多的数据源驱动在DB2安装程序中直接集成，如Hive、Netezza等
  - 扩展基于SQL和NoSQL的数据存储
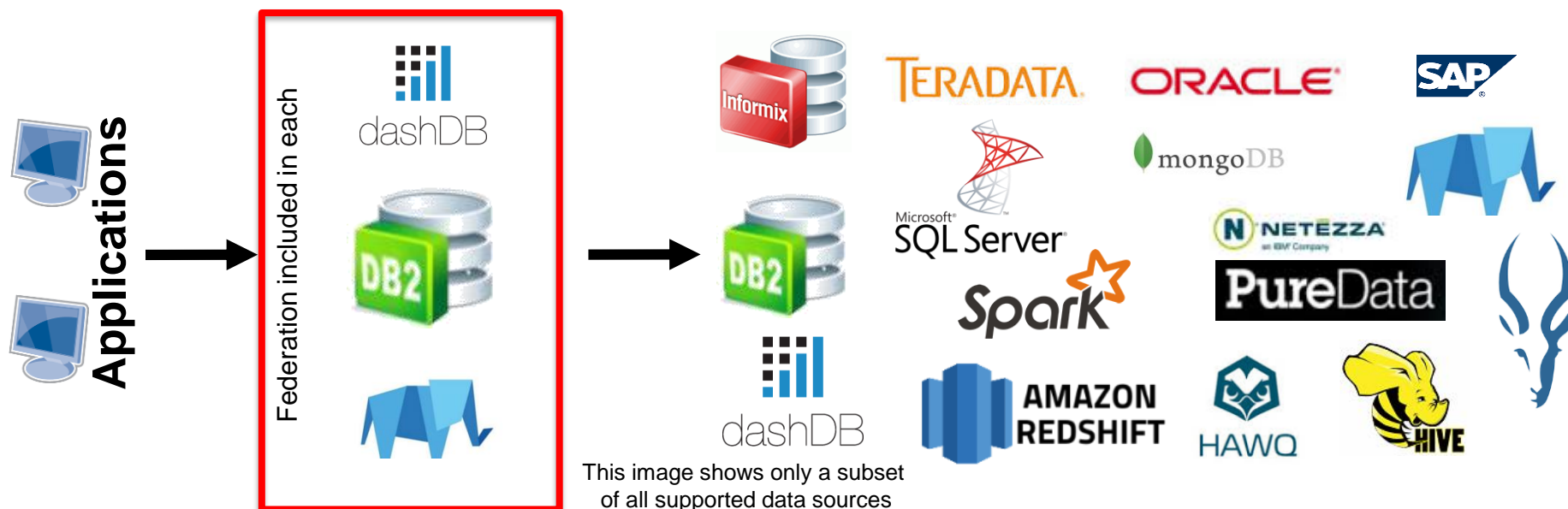 - 扩展传统平台和云平台数据源

▪ **简化配置**
  - CREATE WRAPPER 语句和in CREATE SERVER 的WRAPPER 子句现在是可选项
    • 如果没有指定wrapper字句，缺省wrapper将被自动创建
    • 没有编目远程节点和数据库时，可以在创建SERVER选项中指定服务器、端口和服务名

Applications

Federation included in each

This image shows only a subset of all supported data sources

对Netezza Performance Server兼容性的支持

▪ **Use the SQL_COMPAT global variable to activate the following optional NPS compatibility features**

– SET SQL_COMPAT='NPS'

▪ **Compatibility Features**

– Double-dot notation

• You can use double-dot notation to specify a database object

– TRANSLATE parameter syntax

• TRANSLATE (*char-string-exp*, *from-string-exp* ,*to-string-exp*)

– NPS Operator Symbols

• The operators ^ and ** are both interpreted as the exponential operator, and the operator # is interpreted as bitwise XOR

– Grouping by SELECT clause columns

• You can specify the ordinal position or exposed name of a SELECT clause column when grouping the results of a query

– Routines written in NZPLSQL

• The NZPLSQL language can be used in addition to the SQL PL language.

DB2 11.1.1.1 取消了一些限制

- **32K system temp tablespace not required for extended row sizes**

- **Alter VARCHAR/VARGRAPH length in column-organized tables**
  - Size increase only, cannot modify string units (e.g. CODEUNIT32)

- **Log file size limit increased to 64GB (from 4GB)**
  - New maximum of 64GB = 16,777,152 4k pages

- **db2convert command now supports tables in MPP and with generated columns**

- **Allow system maintained MQT in MPP that references a nickname**

- **Enable WITH clause to be used in SELECT INTO statements for Static SQL and SQL Stored Procedures or Functions**
  - WITH CTE_1(c1, c2) as (select c1, c2 from t1) SELECT sum(c2) into :outvar FROM CTE_1 WHERE c1 between :lowval and :highval

技术尝鲜: 在线崩溃恢复 (异步UNDO)

- **崩溃恢复过程中（或是HADR强制接管时），不允许建立数据库连接直到所有恢复完成**
  - REDO 阶段重做所有交易直到日志尾
  - UNDO 阶段回滚所有没有提交的交易

- **做为DB2 11.1.1.1的一项技术预览， 非生产系统 环境可以配置异步UNDO，即REDO阶段完成后，在UNDO阶段正在执行时，就可以允许建立数据库连接并执行交易。 通过简单的注册变量实现如下：**
  - **db2set DB2_ASYNC_UNDO=yes**

- **参见技术文档：**
- http://g01zciwas018.ahe.pok.ibm.com/support/dcf/preview.wss?host=g01zcidbs003.ahe.pok.ibm.com&db=support/swg/dmgtech.nsf&unid=B2A7F01D9493D64F8525806B000CA708&taxOC=SSCPMKY&MD=2016/11/21%2023:11:26&sid=

技术尝鲜: 在线崩溃恢复 (异步UNDO)

pS
No Change

DB up

prior to
11.1.1.1

DB down

DB up

11.1.1.1

DB down **DB up (*)**

DB up

← **REDO** → **UNDO**

Total Transactions Per Second

45000
40000
35000
30000
25000
20000
15000
10000
5000

pureScale

New

Connections allowed to DB after REDO,
just as asynchronous UNDO begins

Current

(*) Some data may be locked while UNDO proceeds concurrently

Time →

© 2016 IBM Corporation
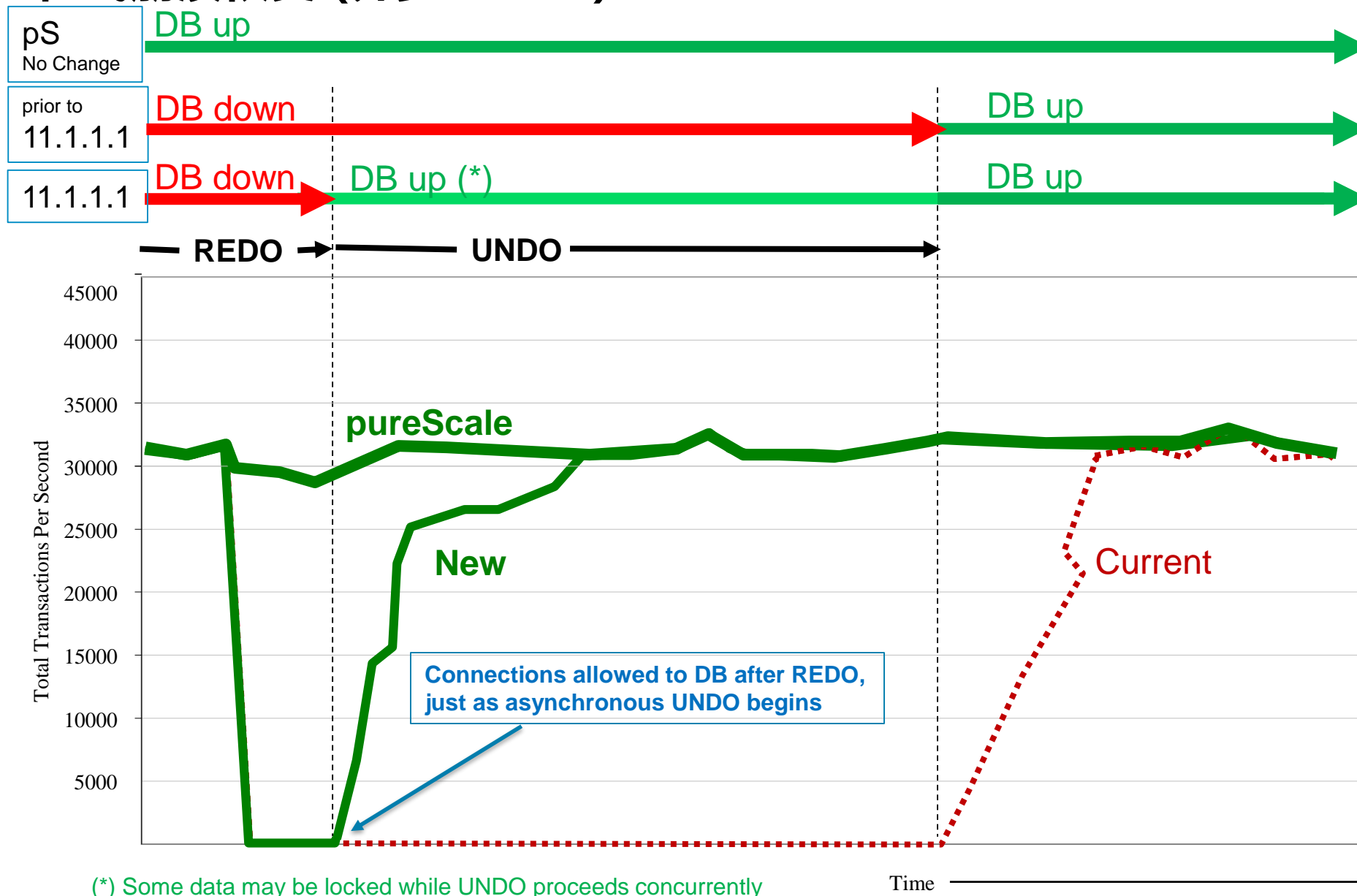
DB2: 混合数据库的策略

Designed to Bridge You to the Cloud

New Consolidated DB2 Editions
- Reduced Complexity higher value Packaging

New DB2 Offering for the Digital Market
- Direct Advanced / Direct Standard Editions
- Subscription Market based Pricing
- Simplified Metric for any cloud deployment
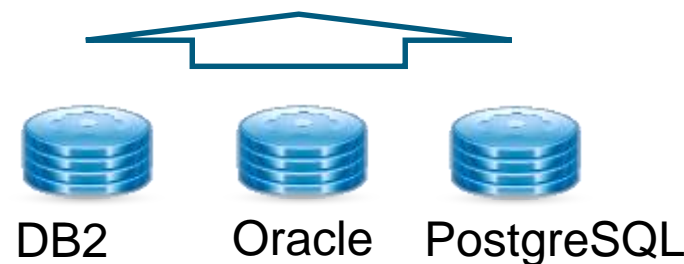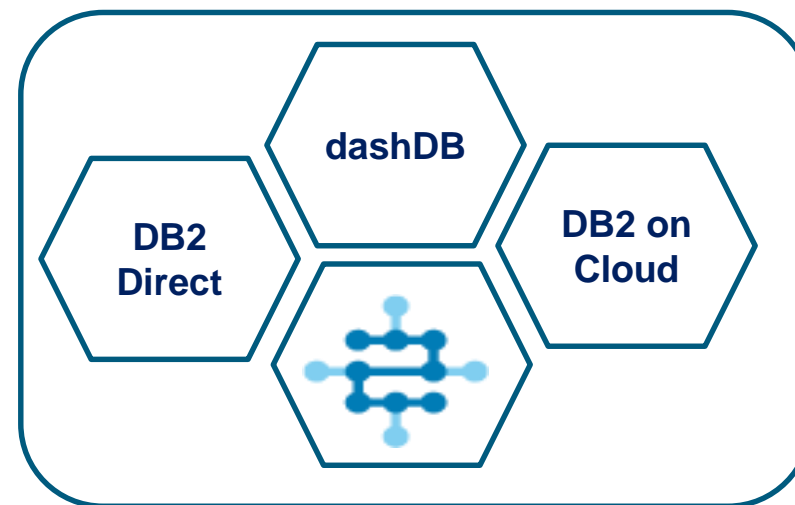
DB2 on Cloud
- Rapid Deployment powered by DB2 V11
- Coming Soon Bridge to Cloud Self Service offering

DB2 for Big Data
- Interchangeable DB2 and BigInsights package

dashDB Compatibility
- Interoperate your DB2 warehouse with dashDB as you migrate to the cloud

dashDB

DB2 Direct

DB2 on Cloud

DB2   Oracle   PostgreSQL

Bridge from Oracle/DB2/PostgreSQL
With competitive hybrid and pure Cloud Solutions

通用分析引擎

Managed Public Cloud Service

dashDB

Software-defined

dashDB Local

Appliance

dashDB appliance

Custom Deployable Software

DB2
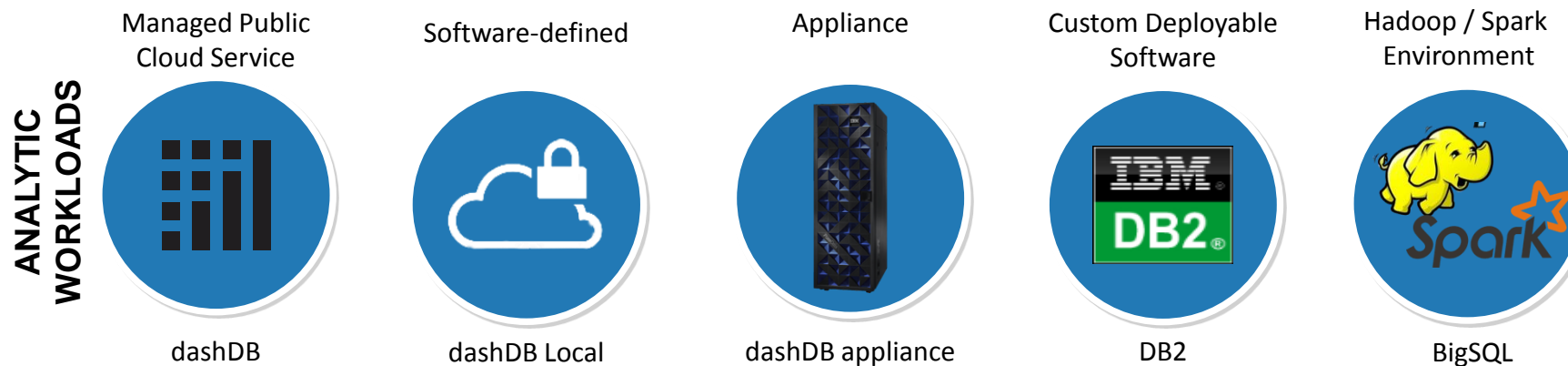
Hadoop / Spark Environment

BigSQL

A **Common Analytics SQL engine**
enabling true hybrid data warehousing solutions with portable analytics
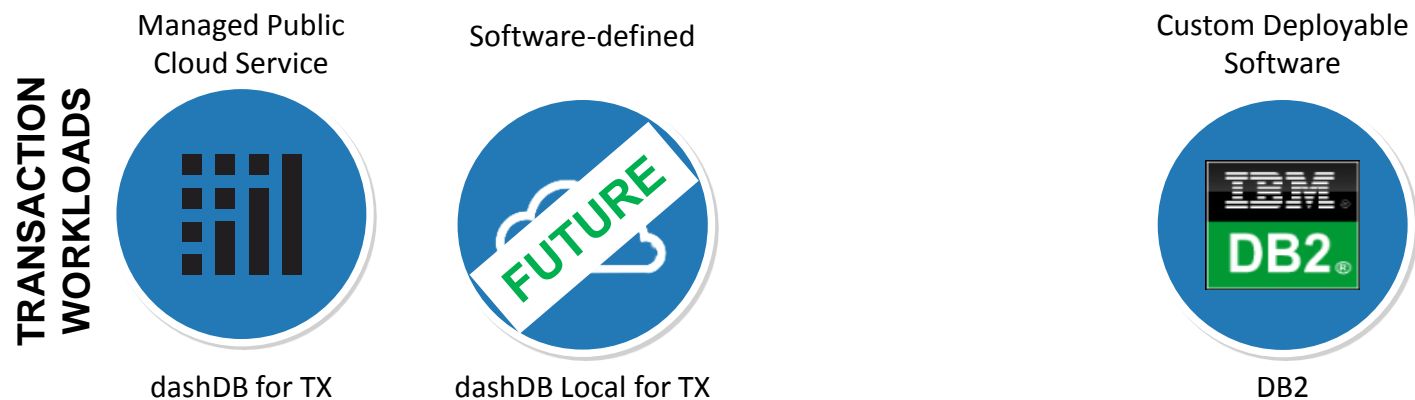
✓ Data Virtualization

✓ Load and Go deployment

✓ Built-in analytics

✓ SQL and NOSQL Capabilities

✓ Cloudant Integration

✓ Consistent Management

✓ Full MPP scalability

✓ Application portability

✓ Hybrid by design

✓ In-memory performance

✓ Advanced Security

✓ Advanced Workload Management

✓ DB2 Applications

✓ Oracle Applications

✓ Netezza Applications

通用SQL 引擎
涵盖更广泛的负载, 不仅仅是分析

ANALYTIC WORKLOADS

Managed Public Cloud Service	Software-defined	Appliance	Custom Deployable Software	Hadoop / Spark Environment
dashDB	dashDB Local	dashDB appliance	DB2	BigSQL

A Common SQL Engine
enabling true application and ecosystem portability for all workloads

TRANSACTION WORKLOADS

Managed Public Cloud Service	Software-defined	Custom Deployable Software
dashDB for TX	dashDB Local for TX	DB2

服务于更多的客户

- **Existing DB2 Footprints**
  - Explain the Common Analytic/SQL Engine
  - Take them through the value of DB2 Version 11
  - Understand their Hybrid cloud and Hybrid warehouse strategy
  - Position the portfolio for their various needs

- **Existing DB2 Accounts with Oracle**
  - Explain the Common Analytic/SQL Engine
  - Target a use-case (SAP, Analytics, Continuous Availability, etc)
  - Explain Oracle Application Compatibility Layer
  - Build a TCO story for transition from Oracle to DB2

- **New Oracle Accounts**
  - Explain the Common Analytic/SQL Engine
  - Target a use-case (SAP, Analytics, Continuous Availability, etc)
  - Explain Oracle Application Compatibility Layer
  - Build a TCO story for transition from Oracle to DB2