Best Practices for Cross Platform Data Transformation & Delivery

Ed Lynch
Executive I/T Specialist
Americas System z Information Integration Technical Sales
edlynch@us.ibm.com

Information Management software

Previous Information Integration Webcasts

- Turn Your Mainframe into an Information Integration Platform
  - Karen Durward, January 15, 2008
    - The changing face of Information Integration
    - IBM Information Server for System z
      - Understand
      - Cleanse
      - Transform
      - Deliver
    - Architecture of Information Server on System z
- Redefining What's Possible on System z with Data Warehousing and Business Intelligence
  - Beth Hamel & Mike Biere, April 29, 2008
    - Why Warehousing on System z ?
    - System z10™
    - DB2 for z/OS Enhancements in V8 and V9
    - Information Server for System z
    - Technology Preview for Cognos for System z

Challenge

- Many recognize the need for Information Integration and/or the need to Improve on the Methods utilized
- There are many methods available for addressing your requirements
  - Each method has its own advantages and disadvantages
- What are the Signs that a Mainframe Solution May Be Better
- What are Key Differentiating Benefits of a Mainframe Solution

Agenda

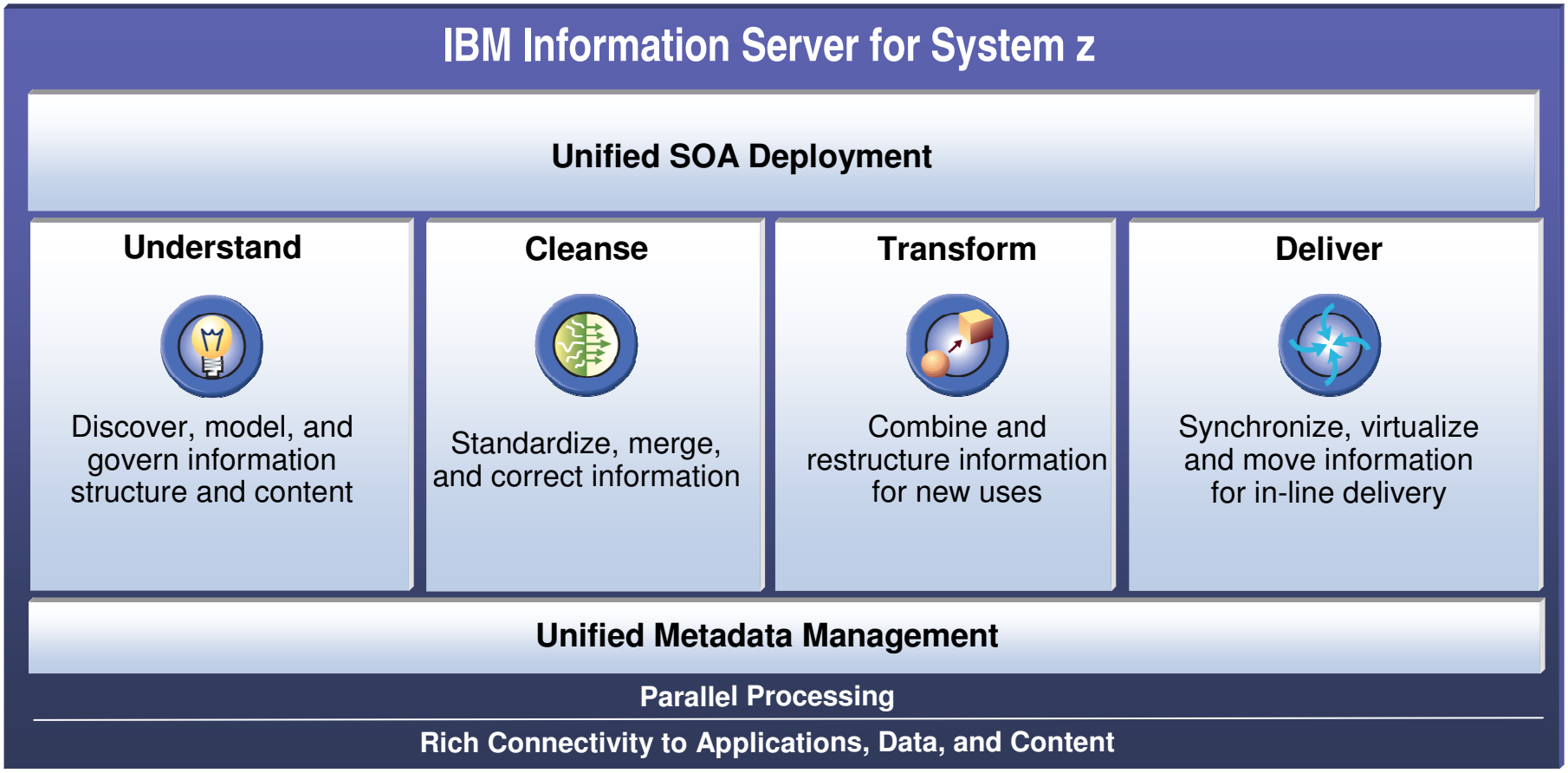
- Information Integration Product Portfolio
- System z: Specialty Processors & Linux on System z
- Architecting a Solution for Different Requirements
  - Case Study Analysis
    - Reporting
    - Event Driven Business
    - Legacy Data Access
    - ODS for Legacy Data
    - Data Warehouse
- Recap of Key Differentiators

Agenda

- **Information Integration Product Portfolio**
- System z: Specialty Processors & Linux on System z
- Architecting a Solution for Different Requirements
  - Case Study Analysis
    - Reporting
    - Event Driven Business
    - Legacy Data Access
    - ODS for Legacy Data
    - Data Warehouse
- Recap of Key Differentiators

IBM Information Server for System z

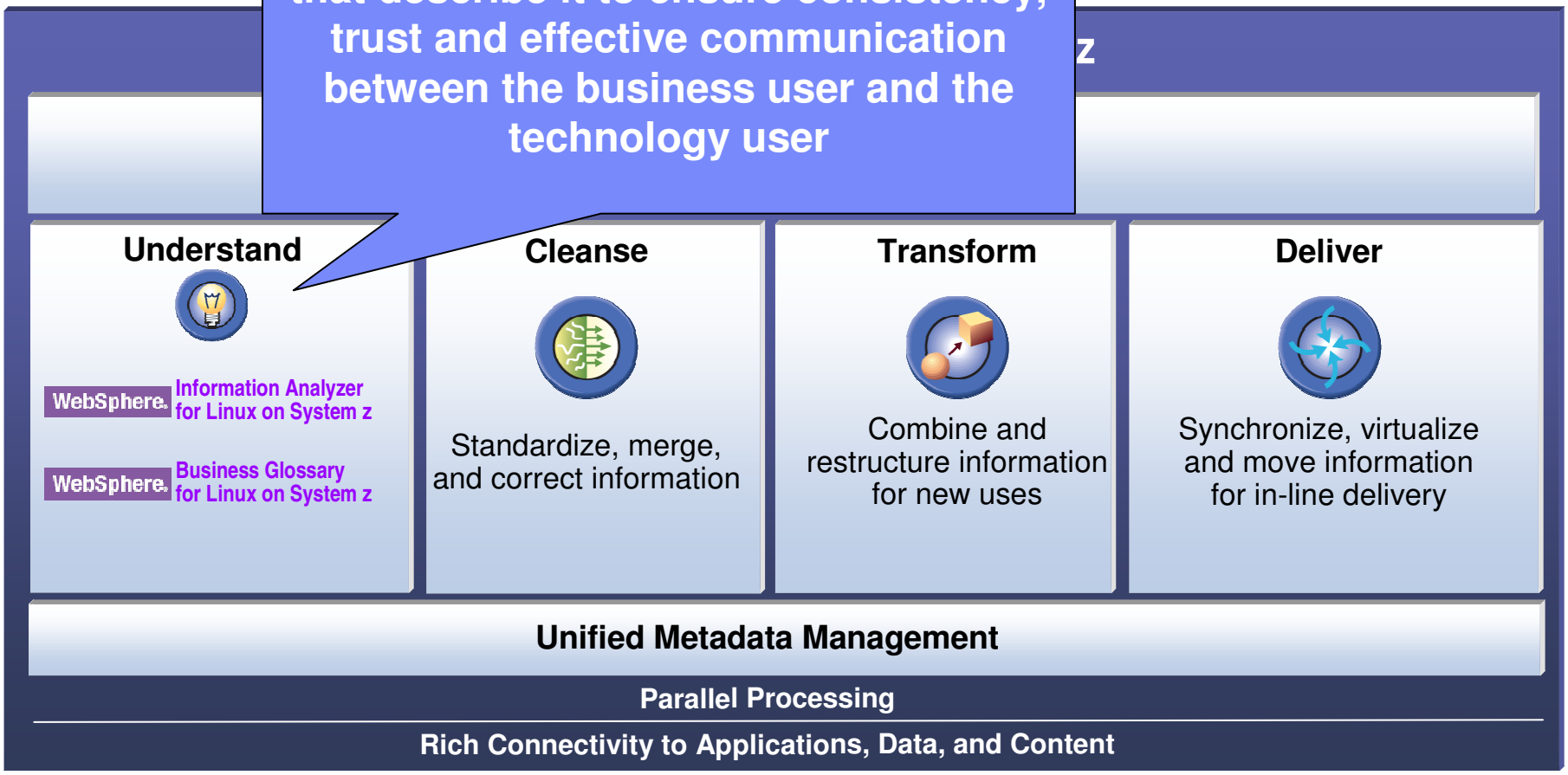
Delivering information you can trust



IBM Information Server for System z

Delivering information you can trust

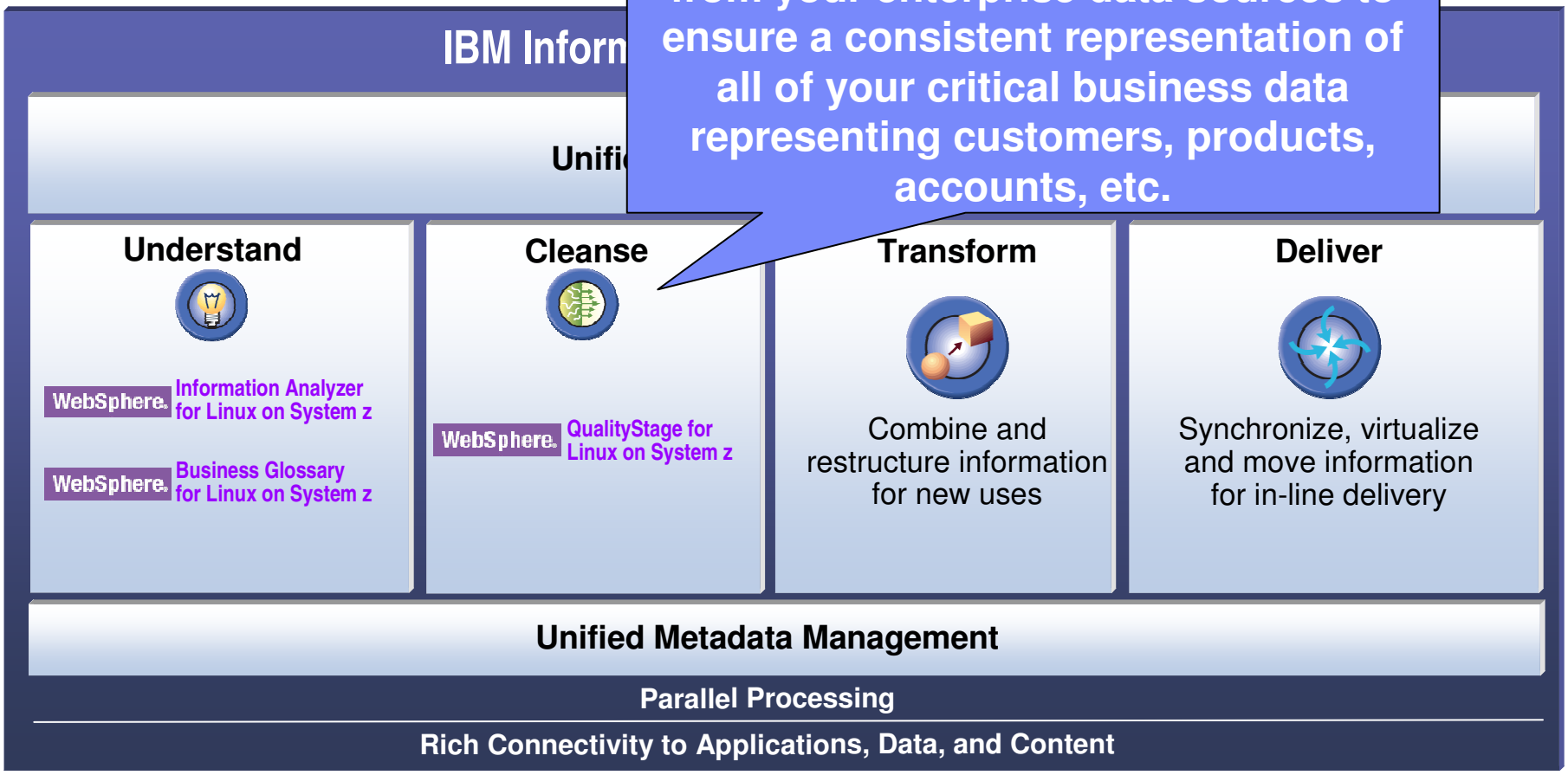
Understand your data and the terms that describe it to ensure consistency, trust and effective communication between the business user and the technology user



IBM Information Server for System z

Delivering information you can trust

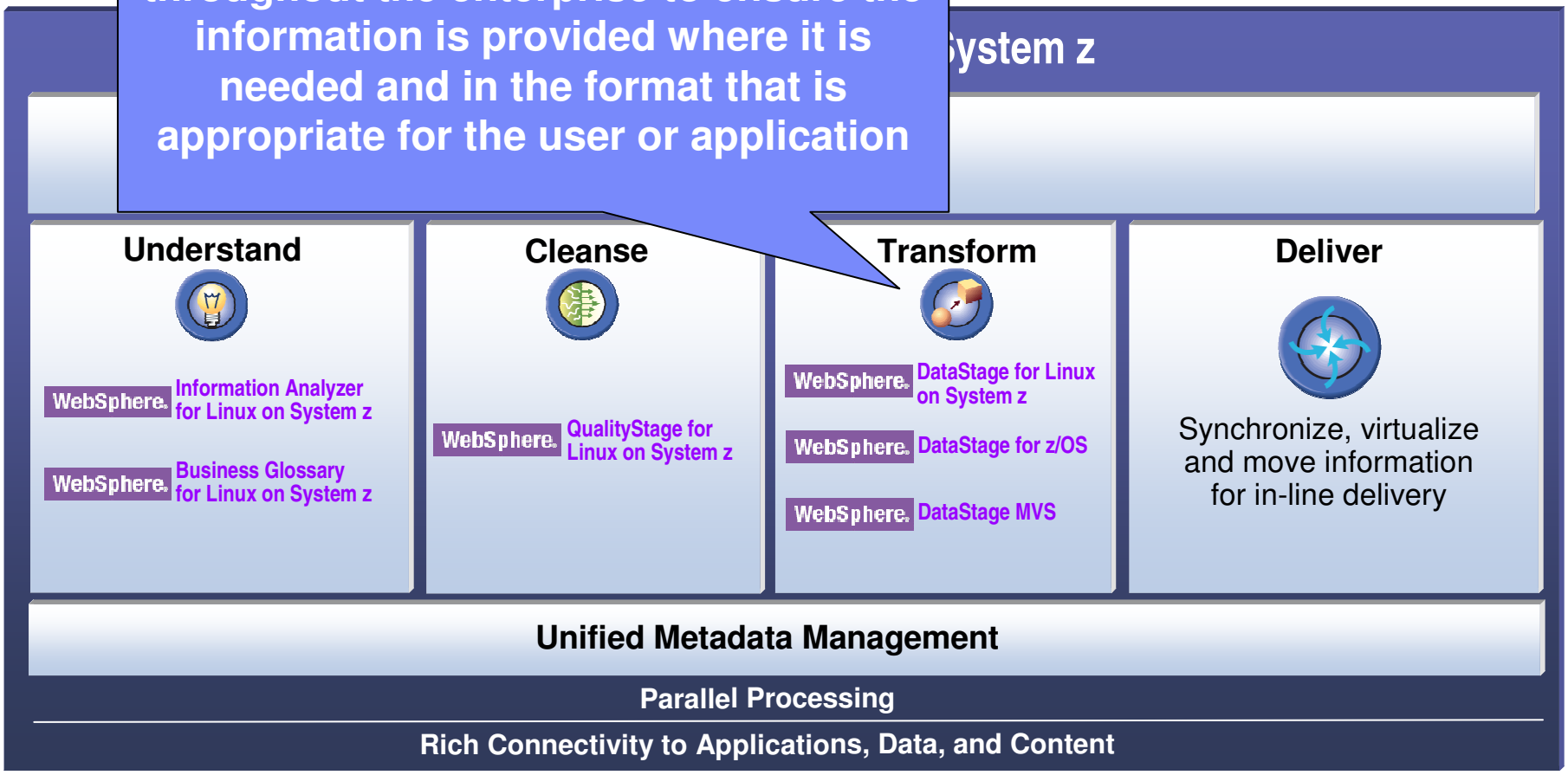
Match, merge and cleanse information from your enterprise data sources to ensure a consistent representation of all of your critical business data representing customers, products, accounts, etc.



IBM Information Server for System z

Delivering information you can trust

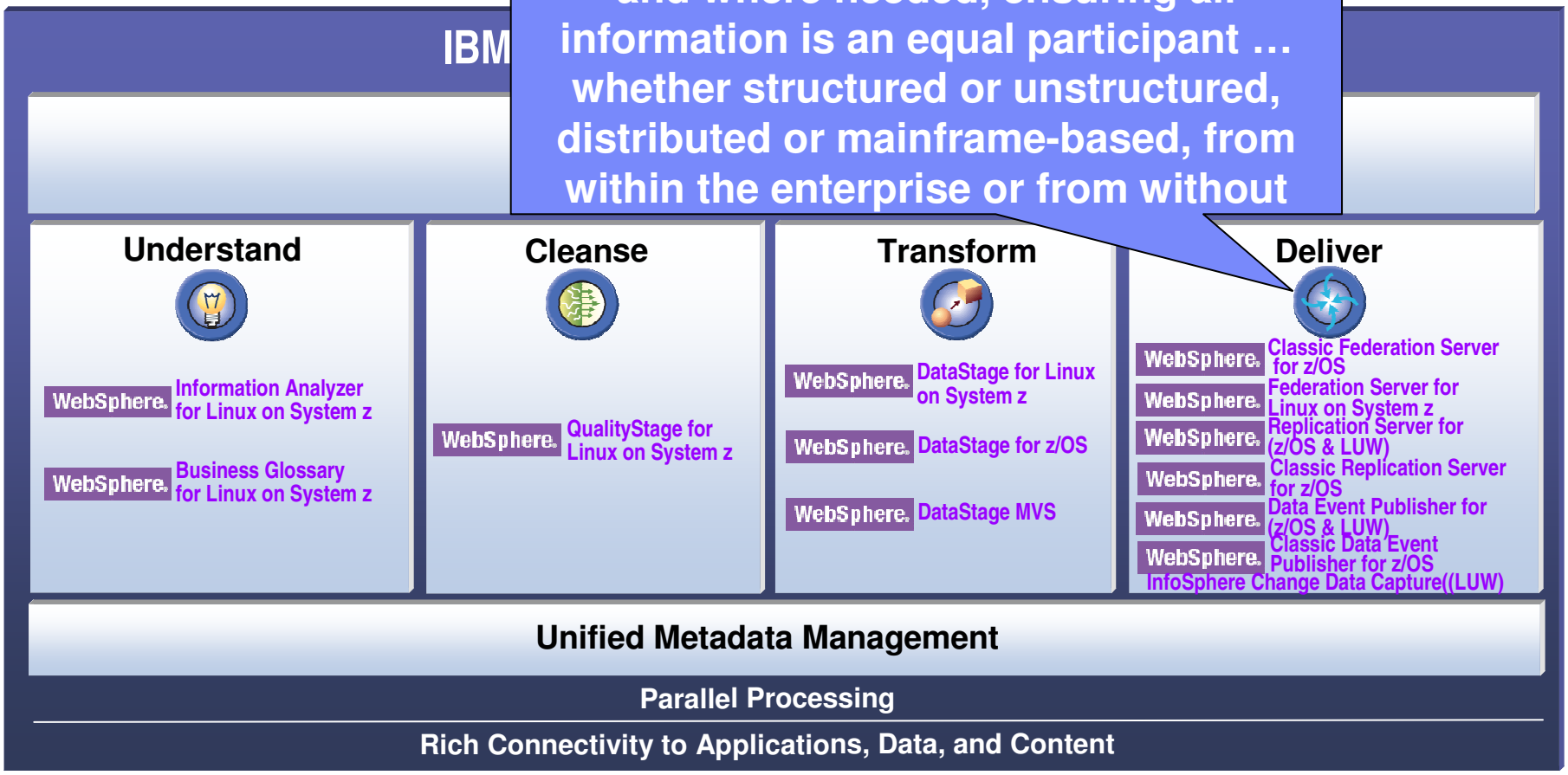
Repackage and repurpose information throughout the enterprise to ensure the information is provided where it is needed and in the format that is appropriate for the user or application



IBM Information Server for System z

Delivering information you can trust

Access and/or deliver information when and where needed, ensuring all information is an equal participant … whether structured or unstructured, distributed or mainframe-based, from within the enterprise or from without

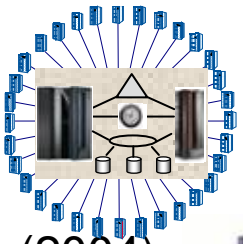


Agenda

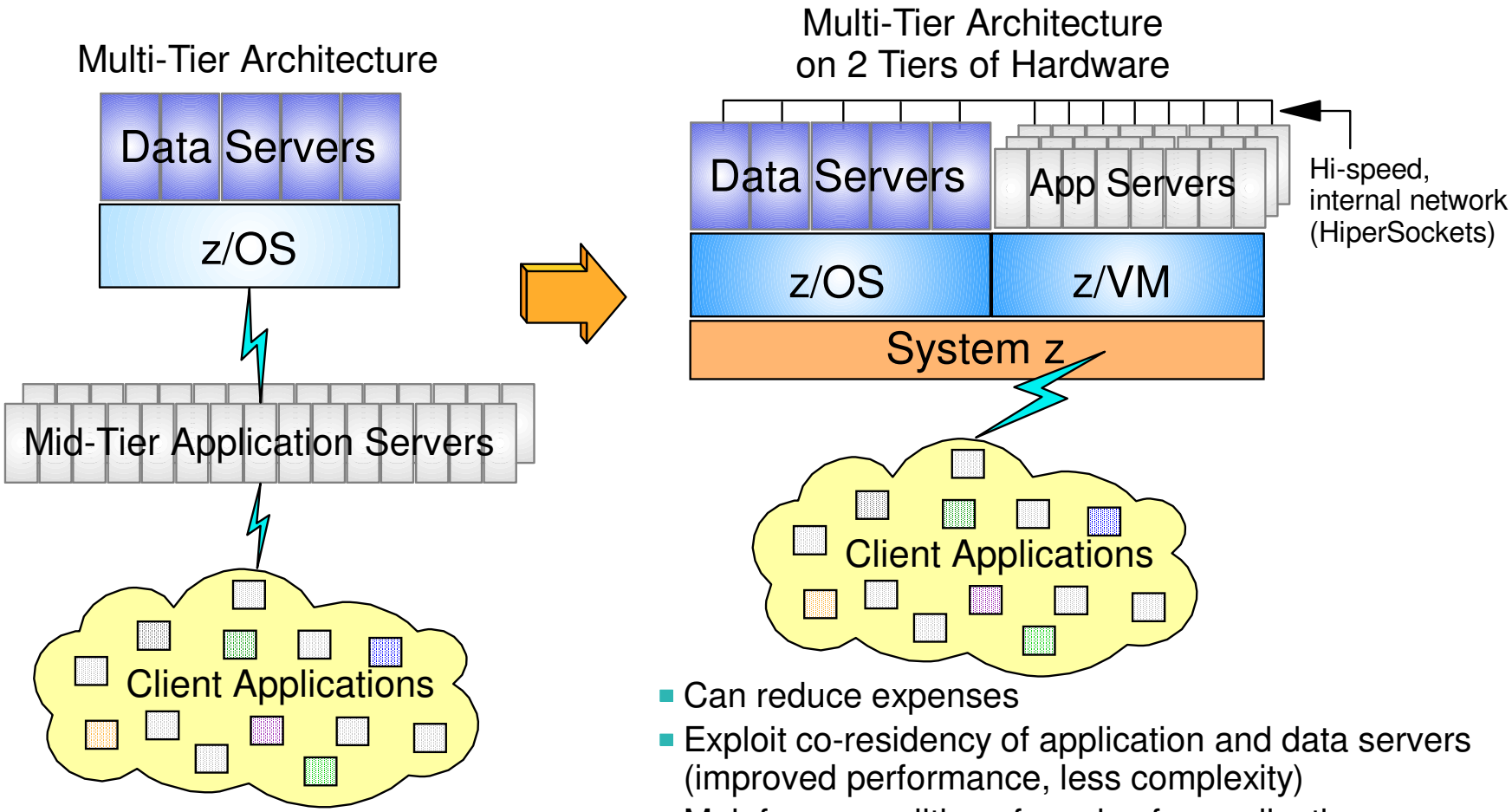
- Information Integration Product Portfolio
- **System z: Specialty Processors & Linux on System z**
- Architecting a Solution for Different Requirements
  - Case Study Analysis
    - Reporting
    - Event Driven Business
    - Legacy Data Access
    - ODS for Legacy Data
    - Data Warehouse
- Recap of Key Differentiators

System z Specialty Processors & Hipersockets

- **Specialty Processors**
  - In general, cost less than Standard Processors
  - In general, IBM does not impose additional SW fees for IBM software installed on standard processors when these processors are added.
  - Processors move with Upgrades - Capacity & Value grows with the technology
  - **ICF** – Internal Coupling Facility (1997)
  - **IFL** – Integration Facility for Linux (2001)
    - Processors dedicated to Linux only workloads
  - **zAAP** – System z Application Assist Processor (2004)
    - Execution of JAVA applications in z/OS
    - DB2z V9 – XML parsing for local requests
  - **zIIP** – System z Integrated Information Processor (2006)
    - Data Serving in z/OS environments
    - DB2 for z/OS Activities that Benefit
      - DDF (DRDA) Processing (e.g. DB2 Connect requests)
      - Portions of DB2 Utilities
      - Remotely requested SQL Stored Procedure processing (DB2z V9)
      - XML parsing for remote requests (DB2z V9)
      - Complex Parallel Queries (e.g. star schema)
- **HiperSockets** – high-speed (6 GB/sec) memory-to-memory internal network across a System z processor



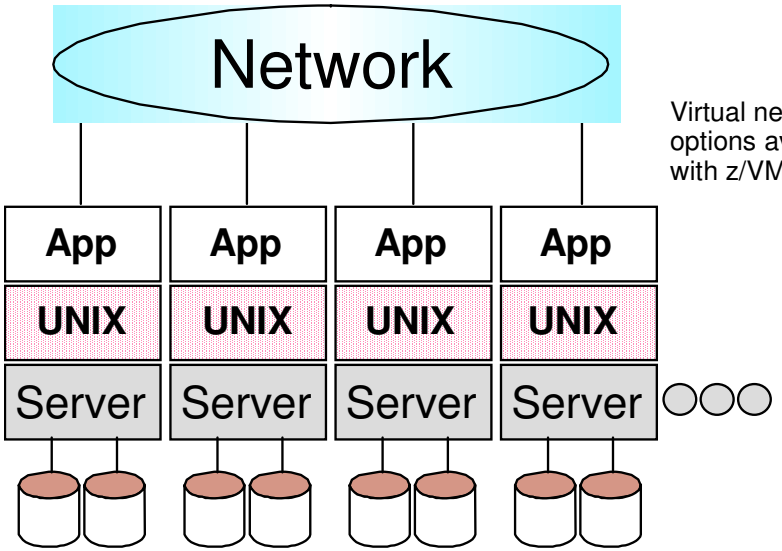
Infrastructure Simplification & Application Integration



- Can reduce expenses
- Exploit co-residency of application and data servers (improved performance, less complexity)
- Mainframe qualities of service for applications
- Little to no change to end-users

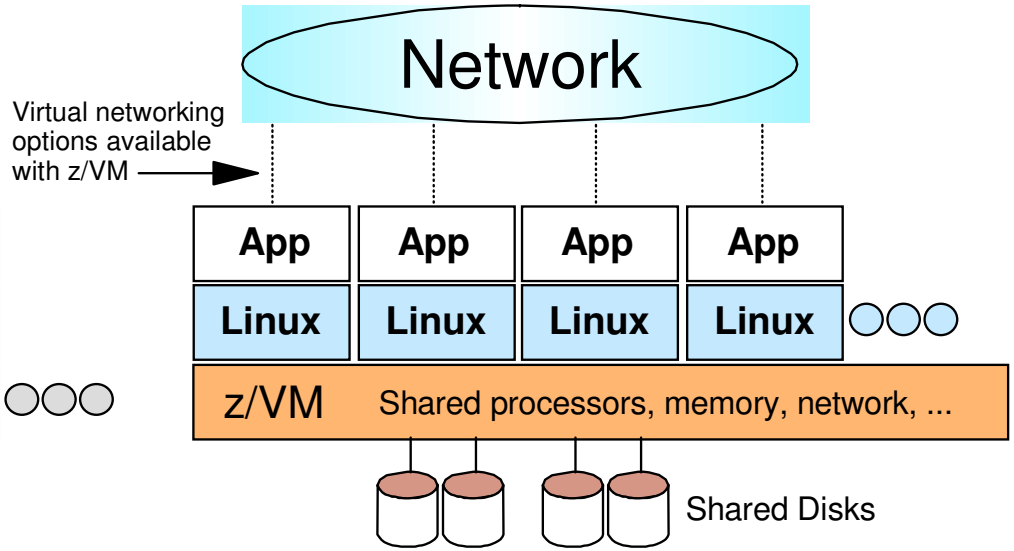
Achieving Rapid Horizontal Growth with Linux on z/VM

Typical UNIX® environment



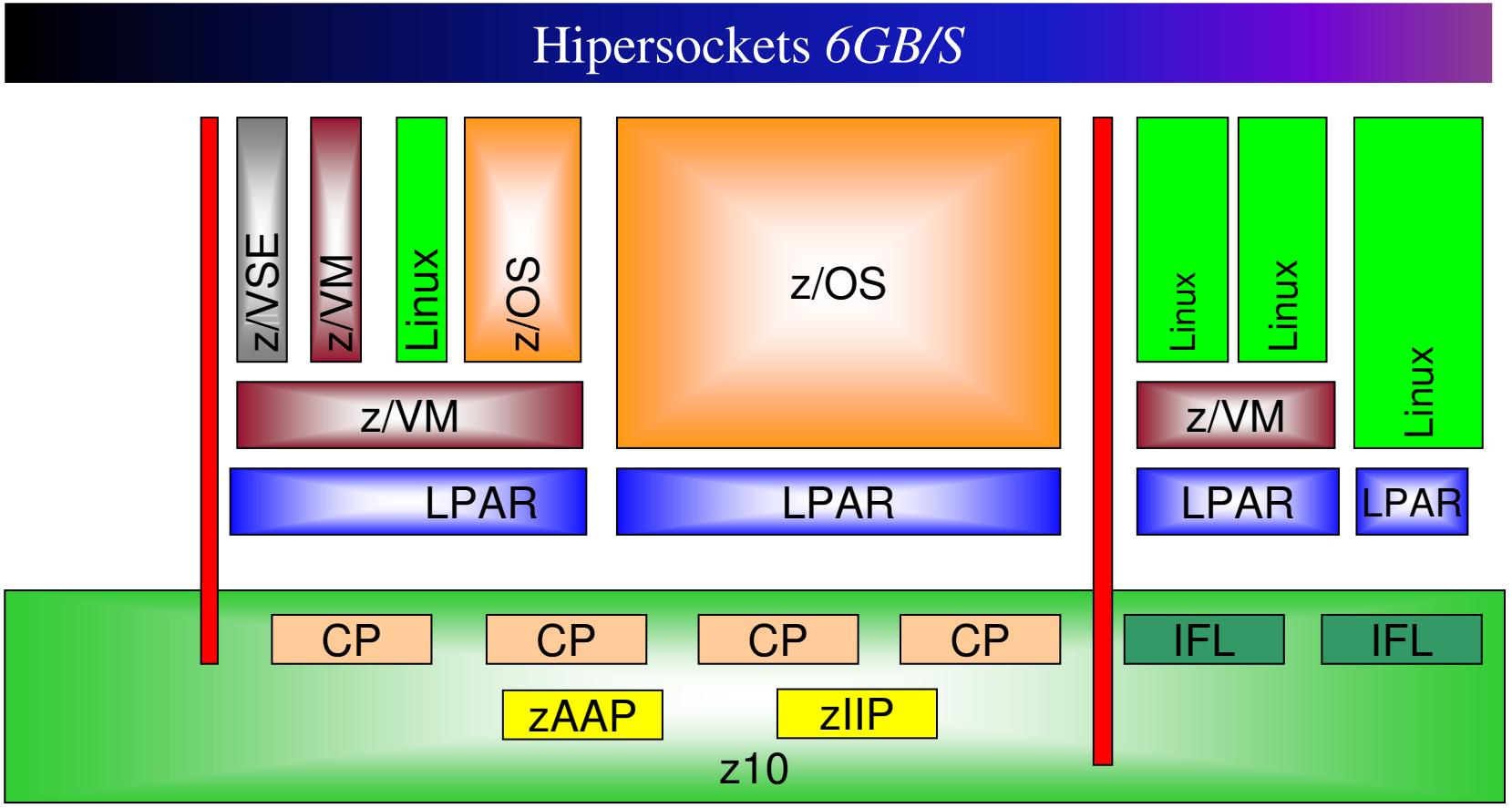
- Dedicated processors, disks, adapters, ...
  - Resources wasted when idle
- Complex system management
  - Networking and software products required for command and control
- New servers available in hours / days

Linux on z/VM

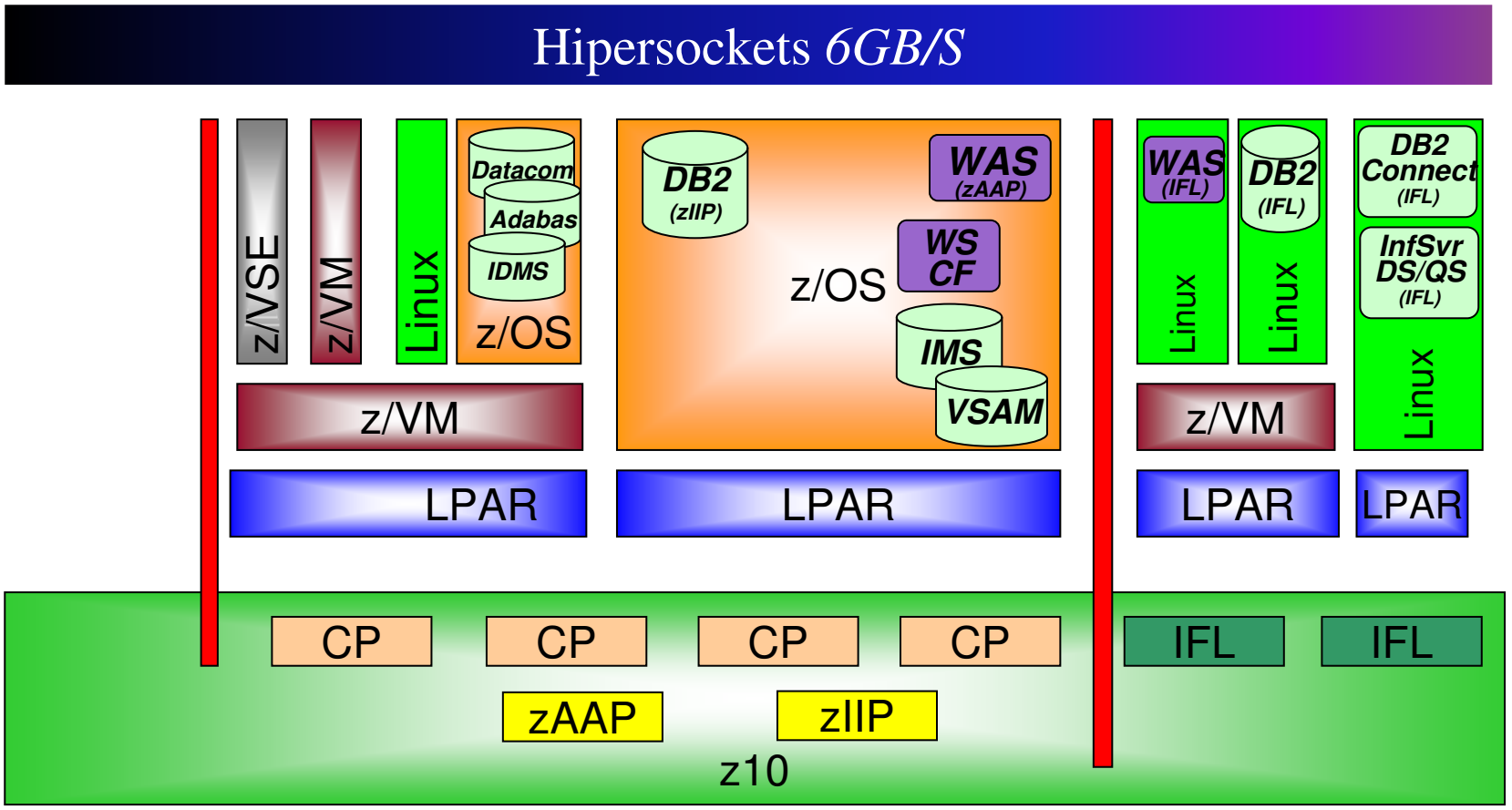


- Shared resources
  - Idle capacity given to servers that need it
- Simplified system management
  - Everything in one box
  - Automation tools included in z/VM
- New servers online in minutes / seconds
- On/Off Capacity on Demand available

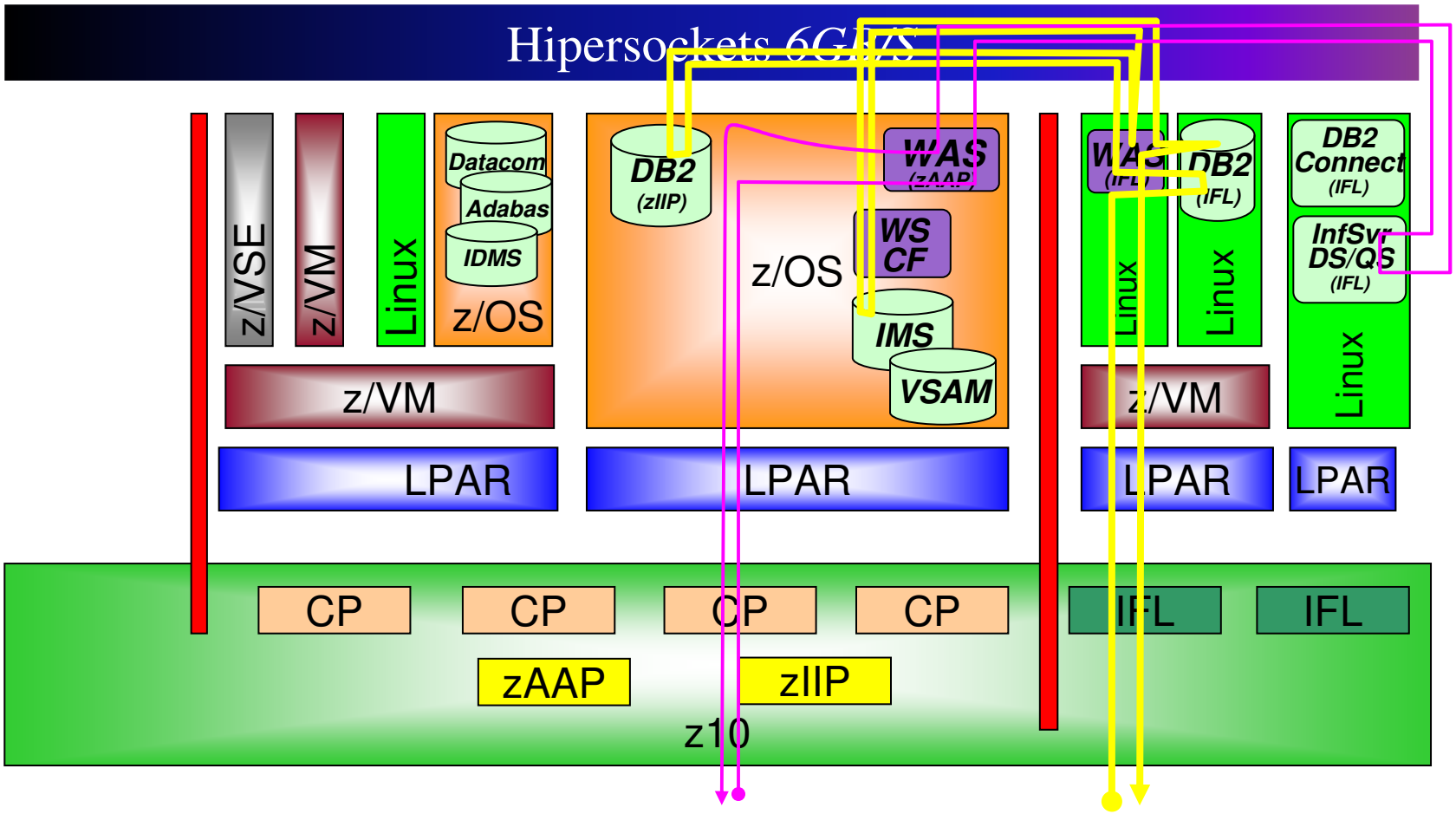
Possible Mainframe Configuration



Linking the Software to the Hardware Benefits



Creating Solutions with Multiple HW Exploitation Points



Agenda

- Information Integration Product Portfolio
- System z: Specialty Processors & Linux on System z
- **Architecting a Solution for Different Requirements**
  - **Case Study Analysis**
    - **Reporting**
    - **Event Driven Business**
    - **Legacy Data Access**
    - **ODS for Legacy Data**
    - **Data Warehouse**
- Recap of Key Differentiators

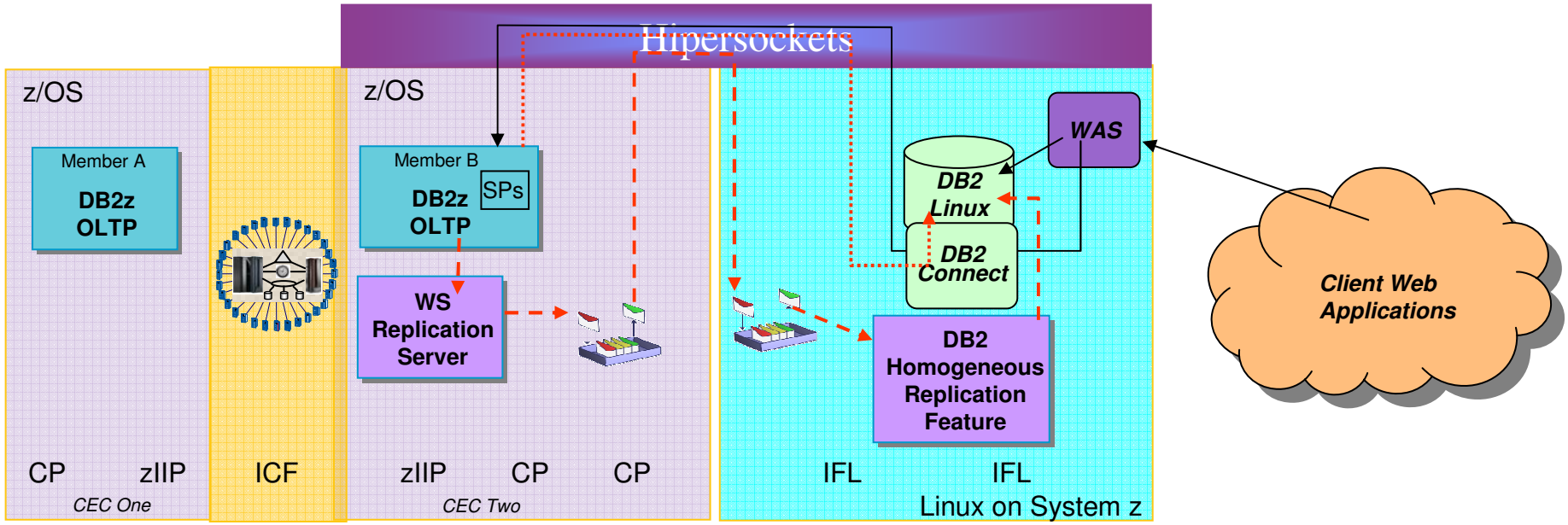
Case Study 1: Reporting - Requirements

- **Today:**
  - Unload, FTP, Load of data from production DB2 for z/OS to DB2 on LUW is performed twice daily.
- **Requirements:**
  - Reporting database (ODS) requires additional indices over production
  - Latency requirement has changed from 12 hours to 10 seconds
  - No impact to production applications
  - Copy contains most columns from tables
  - Row filtering requirement is limited, based on direct table contents
  - On a daily bases, approximately
    - 0.5% new data is inserted
    - 3.0% of data is updated
  - Usage:
    - Support Web Inquiries on Order Status
    - Support Supplemental Update Transactions (10% of users)
      - E.g. remove item, update quantity

Case Study 1: Reporting - Options

	Log Change Data Capture	Application Change Data Capture	Extract, Transform & Load
CPU	Incurred for each change plus a general overhead. Spread out over time	Incurred for each change plus a general overhead. Spread out over time	Incurred for each Record in Source at both source & target Spikes during execution.
DASD	Space to persist changes only	Space to persist changes only	Size of Source
Target Currency	Variable (continuous to long periods)	Variable (continuous to long periods)	Frequency of Process Execution
Target Availability	"Always" available	"Always" available	Outage during target load
Source Transaction Impact	Minor increase in log data generated. No impact to elapsed time (asynch)	Additional processing for each change (trigger or appl code). Direct impact to elapsed time (synch)	No impact
Transformation Capabilities	Filter, summarize	Filter, summarize	Filter, summarize

Case Study 1: Reporting - Recommended Solution



- Hardware Usage
  - IFL Usage
    - WAS, DB2 Linux, DB2 Connect, Replication Feature, MQ
  - zIIP
    - DB2z: DDF & SPs
  - Hipersockets: Linux to z/OS communications
- Variations
  - Linux on System z components placed on another physical system
  - ODS (DB2 Linux): moved to z/OS for DB2z & Sysplex exploitation.
  - WAS: moved to z/OS (zAAP & Sysplex exploitation).

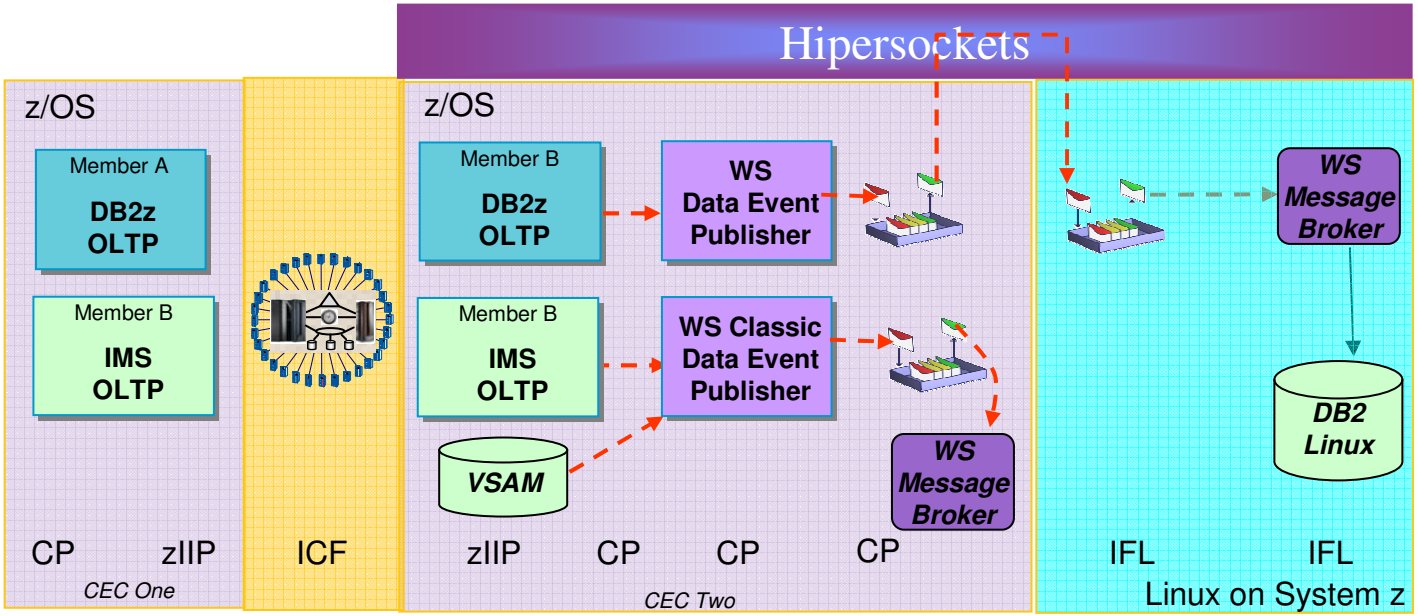
Case Study 2: Event Driven Business - Requirements

- **Today:**
  - Production DB2, IMS, and CICS/VSAM applications include code segments that invoke secondary applications that based upon the transaction and data content drive additional business processes that include a mailing of targeted marketing materials and a customer service representative placing a call.
- **Requirements:**
  - Marketing has need to vary types of marketing materials sent based upon changes in marketing campaigns.
  - Customer service calls are not always necessary
  - Application code changes must stop impacting production applications
    - Reduce Elapsed time
    - Reduce CPU time
    - Isolation of Application code failures (bugs) from impacting production (revenue) transactions
  - Improve timeliness of changes to support new marketing campaigns
  - Improve timeliness to remove code for discontinued marketing campaigns

Case Study 2: Event Driven Business - Options

	Log Change Data Capture	Database Triggers & Secondary Application	Application Maintained Timestamps & Batch Process	Batch Compare	Application Code Segments
Supported for Source DBMSs	DB2z: Yes IMS: Yes CICS/VSAM: Yes	DB2z: Yes IMS: No CICS/VSAM: No	DB2z: Yes IMS: Yes CICS/VSAM: Yes	DB2z: Yes IMS: Yes CICS/VSAM: Yes	DB2z: Yes IMS: Yes CICS/VSAM: Yes
Overall CPU	Incurred for changes only plus a general overhead. Spread out over time (Minimal)	Incurred for changes only plus a general overhead. Trigger cost spread out over time. Secondary application may incur execution spikes. (Minimal)	Incurred for changes only. Batch process incurred for each record.. Secondary application may incur execution spikes. (Medium to High)	Incurred for each Record in Source & prior copy. Spikes during execution. (Highest)	Incurred for changes only. Spread out over time (Minimal)
Source Trans. Elapsed Time Impact	None	Trigger Process - Yes Secondary application - can be isolated.	Timestamp update – nominal Batch process - none	None	All processes are inline. (Highest impact).
Isolation from Prod. Appls	High Isolation	Mostly Isolated	Mostly Isolated	High Isolation	No Isolation
Ease of Maintenance	Easiest	Easy	Difficult to assure timestamps are maintained	Most difficult	Difficult

Case Study 2: Event Driven Business - Recommended Solution



- Hardware Usage
  - IFL Usage
    - WS Message Broker, DB2 Linux, MQ
  - zIIP
    - DB2z
  - zAAP
    - Java Applications
  - Hipersockets: z/OS to Linux communications
- Variations
  - Linux on System z components placed on another physical system or into z/OS for zAAP & Sysplex exploitation.
  - Message Format from Event Publishers in either delimited or XML format

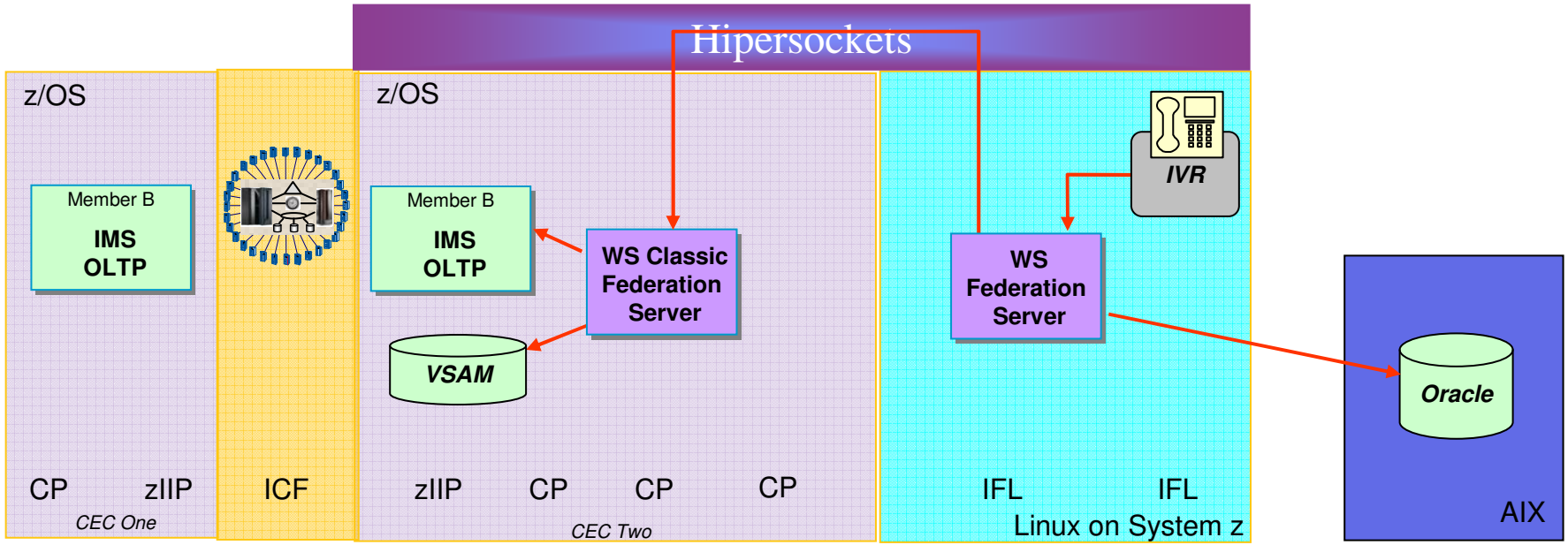
Case Study 3: Legacy Data Access - Requirements

- **Today:**
  - Highly optimized production IMS and CICS/VSAM applications drive critical revenue producing business processes. Data is required to support telephone order status inquiries. Today, skilled customer call center support representatives utilize 3270 interfaces to execute these transactions.
- **Requirements:**
  - Telephony services using Interactive Voice Response (IVR) systems will supplement call center personnel for certain transactions.
  - IVR applications require *inquiry* and *update* capabilities to IMS, VSAM, & Oracle databases utilizing two-phase commit processing.
  - Neither production applications nor call center personnel can be impacted by changes.
    - Interface
    - Elapsed Time
  - Telephony services require data in IMS, VSAM, and Oracle on AIX databases.

Case Study 3: Legacy Data Access - Options

	Federated Data Access		User Written Distributed Application	Copy of Data using Update Anywhere Replication
	Federation Server in Linux on System z	Federation Server Off-platform		
Support for DBMSs	IMS: Yes CICS/VSAM: Yes Oracle: Yes	IMS: Yes CICS/VSAM: Yes Oracle: Yes	IMS: Yes CICS/VSAM: Yes Oracle: Yes	IMS: No CICS/VSAM: No Oracle: Yes
Overall CPU	Incurred for IVR activities only. (Minimal)	Incurred for IVR activities only. (Minimal)	Incurred for IVR activities only. (Minimal)	N/A – not a viable approach
Distributed 2-Phase Commit	Yes	Yes	Requires custom application code to be developed.	N/A
Isolation from Prod. Appls	Yes	Yes	Yes	N/A
Ease of Maintenance	Easy	Easy	Harder	N/A
Duration of Legacy Locks	Short	Long	Longest	N/A

Case Study 3: Legacy Data Access - Recommended Solution



- Hardware Usage
  - IFL Usage
    - WS Federation Server, IVR
  - Hipersockets: z/OS to Linux communications
- Variations
  - Move Oracle to run under Linux on System z
  - Linux on System z components placed on another physical system

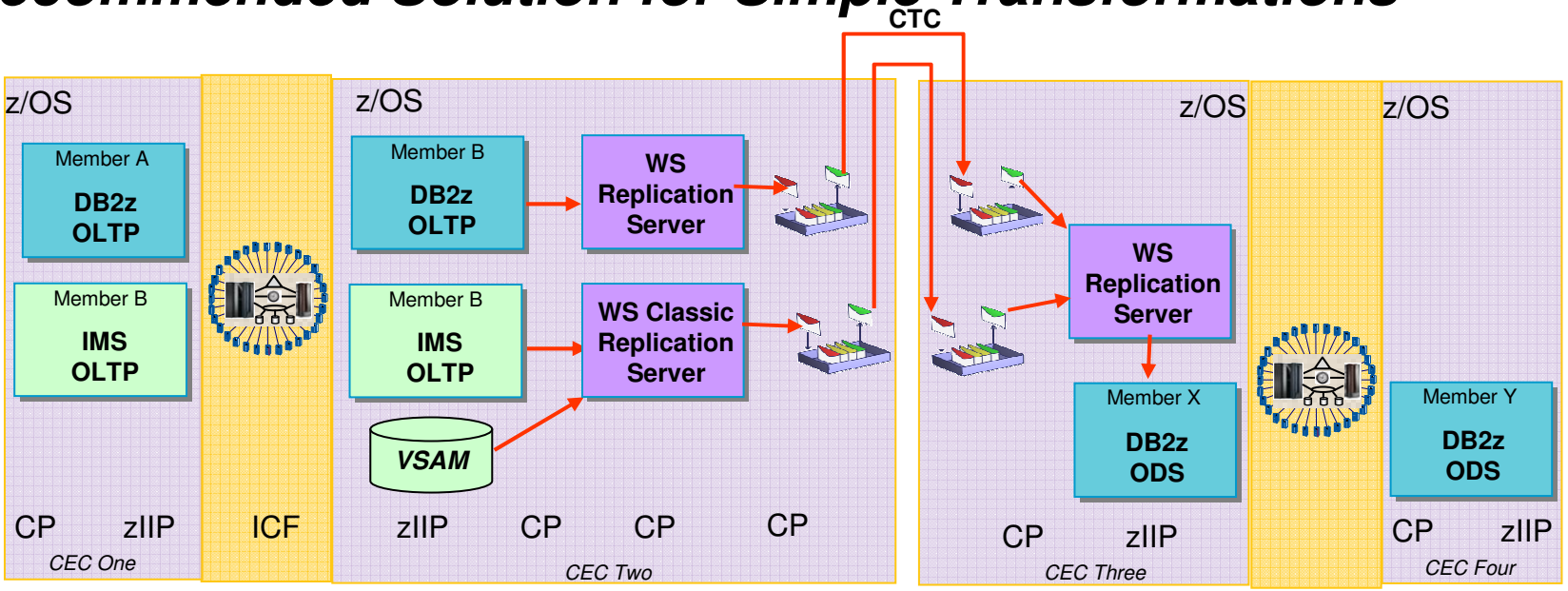
Case Study 4: ODS for Legacy Data - Requirements

- **Today:**
  - ETL of data from production DB2, IMS, and VSAM into DB2 for z/OS via custom applications.
- **Requirements:**
  - Reporting database (ODS) to support adhoc queries of DB2, IMS, & VSAM data.
  - Support customer inquiries on order status.
  - Support “web shopping” by web users of parts catalog data.
  - Latency requirement for ODS is 10 seconds
  - No impact to production applications
  - Column Filtering of data from source into ODS
  - Row filtering requirement is limited, based on direct table contents
  - On a daily bases, approximately
    - 0.1% new data is inserted into source
    - 6.0% of data is updated in source

Case Study 4: ODS for Legacy Data- Options

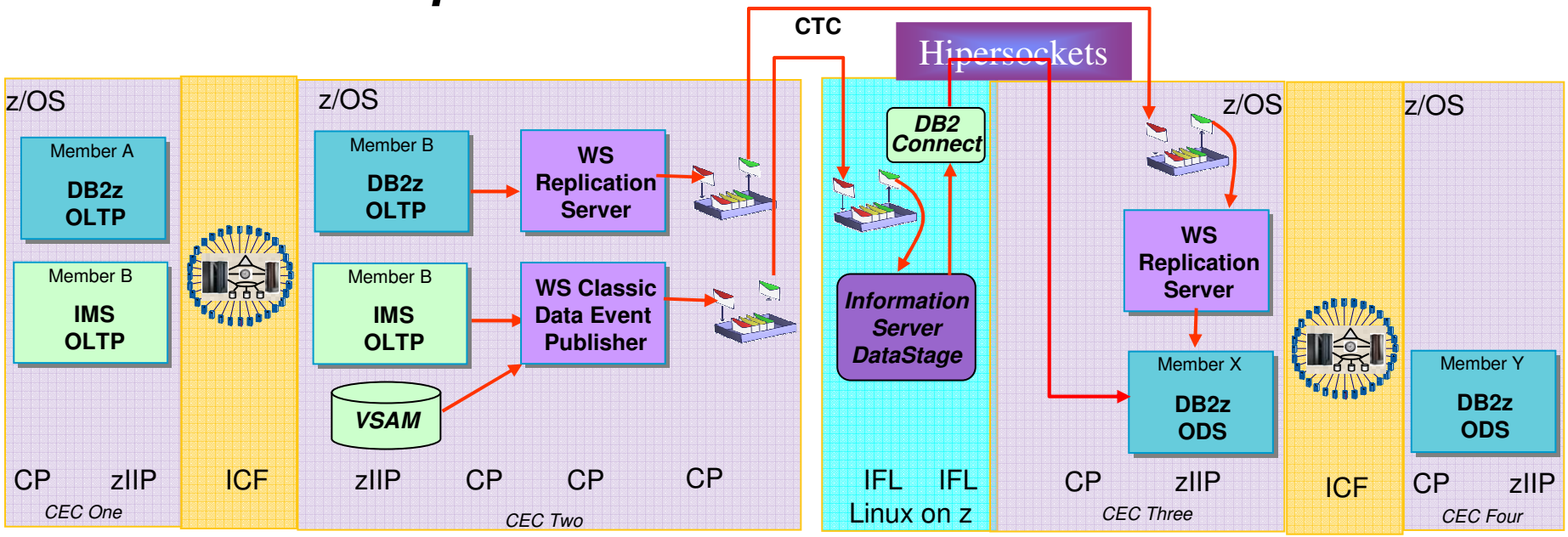
	Log Change Data Capture	Application Change Data Capture	Extract, Transform & Load
DBMS Support	IMS: Yes CICS/VSAM: Yes DB2 for z/OS: Yes	IMS: Yes CICS/VSAM: Yes DB2 for z/OS: Yes	IMS: Yes CICS/VSAM: Yes DB2 for z/OS: Yes
CPU	Incurred for each change plus a general overhead. Spread out over time	Incurred for each change plus a general overhead. Spread out over time	Incurred for each Record in Source at both source & target Spikes during execution.
DASD	Space to persist changes only	Space to persist changes only	Size of Source
Target Currency	Variable (continuous to long periods)	Variable (continuous to long periods)	Frequency of Process Execution
Target Availability	"Always" available	"Always" available	Outage during target load
Source Transaction Impact	Minor increase in log data generated. No impact to elapsed time (asynch)	Additional processing for each change via application code. Direct impact to elapsed time (synch)	No impact
Transformation Capabilities	Filter & simple transformations.	Handles complex transformations.	Handles complex transformations

Case Study 4: ODS for Legacy Data – Recommended Solution for Simple Transformations



- Hardware Usage
  - zIIP: DB2z
- Variations
  - Components on ODS CEC placed in Linux on System z or on another physical system
    - Would utilize IFL & HiperSockets

Case Study 4: ODS for Legacy Data – Recommended Solution for Complex Transformations from Classic Sources



- Hardware Usage
  - zIIP: DB2z & DDF
  - IFL: DataStage & DB2 Connect
  - HiperSockets on Target CEC
- Variations
  - Components on ODS CEC placed in Linux on System z or on another physical system
    - Would further utilize IFL

Case Study 5: Data Warehouse - Requirements

- **Requirements:**
  - DB2 for z/OS Data Warehouse
  - Data sources: DB2 for z/OS, IMS, VSAM, Oracle, & SQL Server
  - Data Cleansing of address and name information
  - Complex Transformations into DWH structures
  - Latency requirement for DWH is 5 minutes
  - No impact to production applications
  - On a daily bases, approximately 2% of data changes

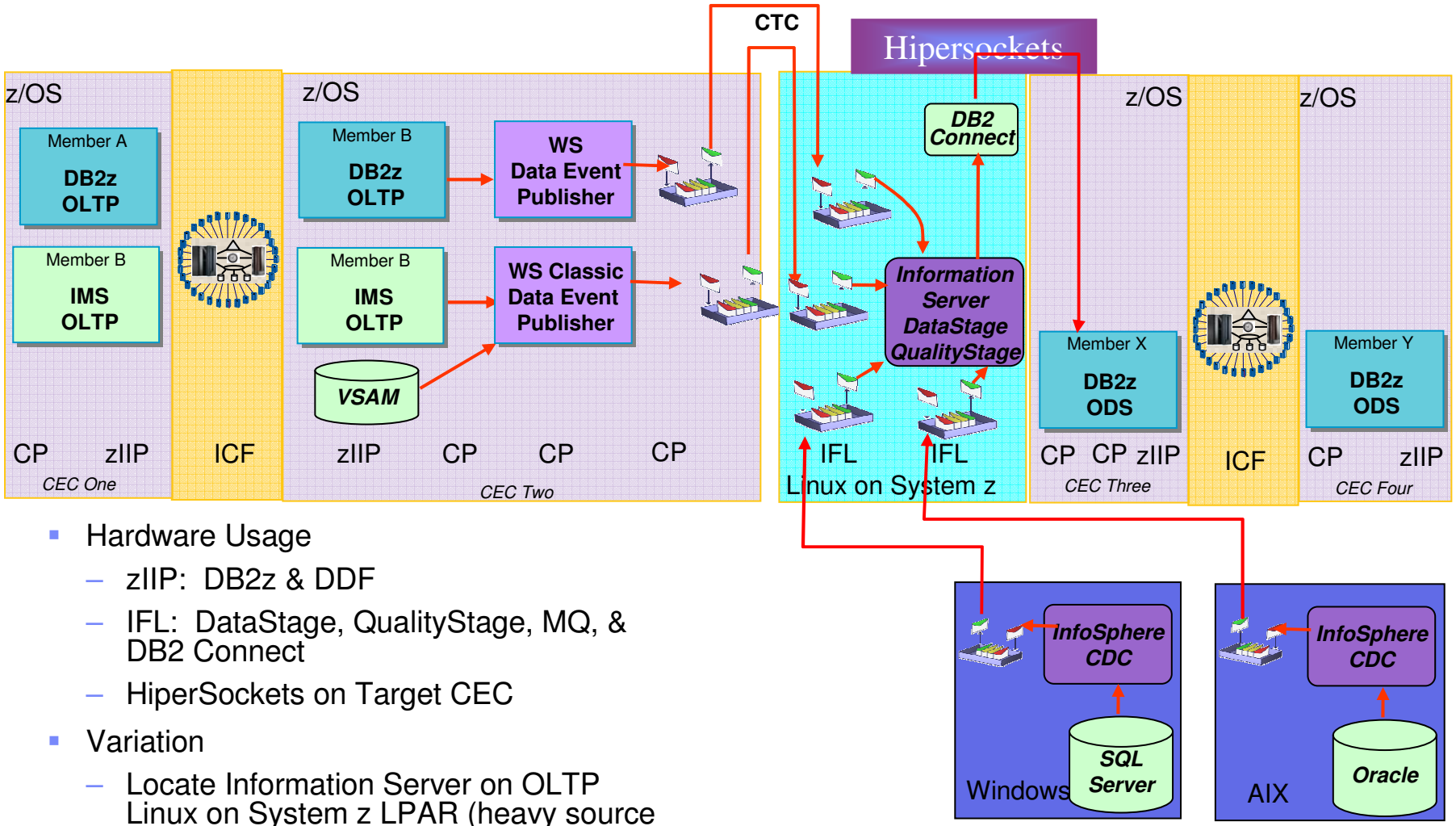
Case Study 5: Data Warehouse - Options

	Batch Extract, Compare, Transform, Load	Database Triggers & Secondary Application	Application Maintained Timestamps & Batch Process	Log Change Data Capture/ Replication	Extract, Transform, & Load Tools
Supported for Source DBMSs	DB2z: Yes IMS: Yes CICS/VSAM: Yes Oracle: Yes Sybase: Yes	DB2z: Yes IMS: No CICS/VSAM: No Oracle: Yes Sybase: Yes	DB2z: Yes IMS: Yes CICS/VSAM: Yes Oracle: Yes Sybase: Yes	DB2z: Yes IMS: Yes CICS/VSAM: Yes Oracle: Yes Sybase: Yes	DB2z: Yes IMS: Yes CICS/VSAM: Yes Oracle: Yes Sybase: Yes
Overall CPU	CPU incurred for each record and to compare. Serialized processing or manual control to parallelize. CPU spikes. (Highest)	N/A – not viable	CPU incurred for each record.. Secondary application may incur execution spikes. (High)	Incurred for changes only plus a general overhead. Does not include transformation, & load/insert. (Minimal)	CPU incurred for each record. Highly parallelized processing. CPU spikes. (High)
Latency / Throughput	Highest / Low	N/A – not viable	High / Low	Low / High	High / High
Target Availability	High	N/A – not viable	High	High	High
Source Transaction Impact	None	N/A – not viable	Nominal	None	None
Transformation Capabilities	High	N/A – not viable	High	Low	High
Isolation from Prod. Appls	High Isolation	N/A – not viable	Mostly Isolated	High Isolation	High Isolation
Ease of Maintenance	Difficult	N/A – not viable	Difficult	Easy	Easy

Case Study 5: Data Warehouse – Best Option

	Log Change Data Capture feeding ETL Tool	Log Change Data Capture / Replication	Extract, Transform, & Load Tools
Supported for Source DBMSs	DB2z: Yes IMS: Yes CICS/VSAM: Yes Oracle: Yes Sybase: Yes	DB2z: Yes IMS: Yes CICS/VSAM: Yes Oracle: Yes Sybase: Yes	DB2z: Yes IMS: Yes CICS/VSAM: Yes Oracle: Yes Sybase: Yes
Overall CPU	Incurred for changes only plus a general overhead. Transformation can be highly parallelized. Eliminates CPU spikes. (Best)	Incurred for changes only plus a general overhead. Does not include transformation, & load/insert. (Minimal)	CPU incurred for each record. Highly parallelized processing. CPU spikes. (High)
Latency / Throughput	Low / High	Low / High	High / High
Target Availability	High	High	High
Source Transaction Impact	None	None	None
Transformation Capabilities	High	Low	High
Isolation from Prod. Appls	High Isolation	High Isolation	High Isolation
Ease of Maintenance	Easy	Easy	Easy

Case Study 5: Data Warehouse - Recommended Solution



- Hardware Usage
  - zIIP: DB2z & DDF
  - IFL: DataStage, QualityStage, MQ, & DB2 Connect
  - HiperSockets on Target CEC
- Variation
  - Locate Information Server on OLTP Linux on System z LPAR (heavy source data needs for transformation).

Agenda

- Information Integration Product Portfolio
- System z: Specialty Processors & Linux on System z
- Architecting a Solution for Different Requirements
  - Case Study Analysis
    - Reporting
    - Event Driven Business
    - Legacy Data Access
    - ODS for Legacy Data
    - Data Warehouse
- **Recap of Key Differentiators**

What Makes a Good Fit

- **Leverage Classic Strengths of System z**
  - High availability
  - High I/O bandwidth capabilities
  - Flexibility to run disparate workloads concurrently
  - Requirement for excellent disaster recovery capabilities
  - Security, including Encryption & User Access
  - Scalability
- **Shortening End-to-end Path Length for Applications**
  - Source and/or Target located on System z
  - Collocation of applications
  - Consolidation of applications from distributed servers
  - Reduction in network traffic
  - Simplification of support model
- **Consolidation Effect**
  - Power requirements
  - Software costs
  - People Costs
  - Real Estate
  - Workloads requiring EXTREME Flexibility

Summary

- Information Integration Product Portfolio
- System z: Specialty Processors & Linux on System z
- Architecting a Solution for Different Requirements
  - Case Study Analysis
    - Reporting
    - Event Driven Business
    - Legacy Data Access
    - ODS for Legacy Data
    - Data Warehouse
- Recap of Key Differentiators

Thank You for Joining Us today!

Go to **www.ibm.com/software/systemz** to:

- Replay this teleconference
- Replay previously broadcast teleconferences
- Register for upcoming events