Extending Your Mainframe for More Business Value

Deliver Business Insight with a Data Warehouse on System z

Get More Business Results Out of Your Data

Our branch offices have separate databases. Each branch is analyzing customers and sales on their own.

Looking at data in isolation can miss larger trends and opportunities

Service Oriented Finance Marketing

IBM
Service Oriented Finance Needs a Data Warehouse to Make Optimal Business Decisions

- Each branch is responsible for its own marketing campaign
- Corporate marketing gets reports from each of the branches based on local results
- Corporate marketing needs to spot trends to know what campaigns are most effective region-wide
- A corporate data warehouse would give marketing the data to easily do comparisons between the branches and promote best practices

Mainframe Extension Solution – Deliver Business Insight with a Data Warehouse

- Management Tools
- ETL Processing Connections
- Analysis Tools Large Data Base
- Source Data

Data Warehouse

SQL

Generate Reports

Extract Transform Load

Collect Historical Data
Data Warehouse Capabilities on System z

- Large Capacity Data Base
  - DB2 for z/OS
  - Parallel queries, Materialized Query Table, Star Join Enhancements

- Connect, Extract, Transform and Load
  - IBM Information Server For System z
  - Runs on zLinux (DataStage also runs on z/OS)

- Analysis Tools
  - DataQuant and Cognos
    - Solution for reporting, analysis dashboards and scorecards
    - Other tools (QMF, Alphablox, Hyperion, Business Objects, SAS, IBI)
  - InfoSphere Master Data Management Server for Linux on System z
    - More effectively manage high-value operational data

- Performance Monitoring
  - IBM Tivoli Omegamon XE for DB2 Performance Expert on z/OS

- Security and Compliance
  - DB2 Data Archive Expert, DB2 Test Database Generator, DB2 Audit Manager Expert, IBM Encryption for DB2 and IMS Databases

- Dedicated Query Hardware
  - zIIPs for parallel queries & remote access
  - Superior I/O bandwidth, multiple I/O paths


Compare Scalability Achievements:
Winter Corporation’s “2005 Top Ten” Awards

- “The highest performing transaction processing system in the 2005 program, a [DB2] z/OS implementation, executed over one billion SQL statements in an hour. The average for operational systems was 35 million SQL statements or database operations per hour”

  The study lists the largest known peak workload on Oracle RAC to be 8.6 million SQL statements per hour

- “The largest transaction processing [database] in the program, 23 TB, was hosted on [DB2] z/OS, as in the last program”

  The study lists the largest transaction processing database on Oracle RAC to be 9.6 TB

http://www.wintercorp.com/WhitePapers/WI_TopTenWP.pdf
IBM Information Server for System z Can Load Your Data Warehouse

Load Your Data Warehouse with DataStage

Source Data

Branch

Extract

Transform

Load

Data Warehouse
Data Stage Transforms Data on the Fly

Different field names
Different field order
Add Branch Identifier
Different currency format

Transform

<table>
<thead>
<tr>
<th>PROD ID</th>
<th>CUST ID</th>
<th>BRANCH ID</th>
<th>QTY</th>
<th>AMT</th>
<th>SALEDATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>000 101</td>
<td>100</td>
<td>01</td>
<td>01</td>
<td>10,000.00</td>
<td>2007-02-28</td>
</tr>
<tr>
<td>000 121</td>
<td>100</td>
<td>01</td>
<td>03</td>
<td>500.50</td>
<td>2007-02-28</td>
</tr>
<tr>
<td>000 101</td>
<td>101</td>
<td>01</td>
<td>01</td>
<td>20,000.00</td>
<td>2007-03-01</td>
</tr>
</tbody>
</table>

Branch Data

<table>
<thead>
<tr>
<th>PRODUCT</th>
<th>QTY</th>
<th>CUSTNO</th>
<th>AMOUNT</th>
<th>DATE</th>
</tr>
</thead>
</table>

Data Warehouse

DEMO: Extract, Transform, Load

Use DataStage to load sales and customer data from a branch into your data warehouse
- ODBC (Input) – Branch Sales info from SQL Server
- DB2 (Update) – Corporate Sales Warehouse

Show how built-in stages make it easy to handle transformation and aggregations
BlueCross BlueShield of Tennessee

**Challenge**
- In order to compete effectively for new business in the complex healthcare market, BCBST needed to differentiate itself from competitors with targeted offerings.
- They needed a single view of information across their multiple LOBs with business intelligence capabilities.

**Solution**
- BCBST is using IBM Information Server with IBM DB2 to profile, transform, and load data to their enterprise data warehouse.
- The solution also provides intelligent search capabilities for unstructured data using IBM OmniFind and IBM DB2 Content Manager.

**Business Benefits**
- Improved customer and provider satisfaction by enabling them to run their own analytics and better manage their healthcare costs.

IBM Information Server for System z Has Connectivity and Scalability Suitable for the Large Enterprise

IBM Information Server for System z

<table>
<thead>
<tr>
<th>Understand</th>
<th>Cleanse</th>
<th>Transform</th>
<th>Deliver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Analyzer on Linux for System z</td>
<td>QualityStage on Linux for System z</td>
<td>DataStage on Linux for System z</td>
<td>Classic Fed. Server</td>
</tr>
<tr>
<td>Business Glossary on Linux for System z</td>
<td>Business Glossary on Linux for System z</td>
<td>DataStage for z/OS</td>
<td>Classic Replication</td>
</tr>
<tr>
<td>Rational Data Architect</td>
<td>Rational Data Architect</td>
<td>DataStage MVS</td>
<td>Replication Server</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Change Data Capture</td>
</tr>
</tbody>
</table>

**Platform Services**
- Parallel Processing Services
- Connectivity Services
- Metadata Services
- Administration Services
- Deployment Services
Data From 3rd Party Systems Can Be Integrated Using Connectivity Services

<table>
<thead>
<tr>
<th>Platform Services</th>
<th>Deployment Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connectivity Services</td>
<td>Information Services Director</td>
</tr>
<tr>
<td>Metadata Services</td>
<td></td>
</tr>
<tr>
<td>Administration Services</td>
<td></td>
</tr>
<tr>
<td>Parallel Processing Services</td>
<td></td>
</tr>
</tbody>
</table>

Enterprise Applications

IBM Information Server Connects to Almost All Sources of Data

<table>
<thead>
<tr>
<th>RDBMS</th>
<th>General Access</th>
<th>Standards and Real Time</th>
<th>Legacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB2 (on Z, I, P or X series)</td>
<td>Sequential File</td>
<td>WebSphere MQ</td>
<td>Alibase/SQL</td>
</tr>
<tr>
<td>Oracle</td>
<td>Complex Flat File</td>
<td>Java Messaging Services (JMS)</td>
<td>C-ISAM</td>
</tr>
<tr>
<td>Informix (IDS and XPS)</td>
<td>File / Data Sets</td>
<td>Java</td>
<td>D-ISAM</td>
</tr>
<tr>
<td>Ingres</td>
<td>Named Pipe</td>
<td>XML and XSL-T</td>
<td>Datacom/DB</td>
</tr>
<tr>
<td>MySQL</td>
<td>FTP</td>
<td>EJB</td>
<td>IMS Mumpa</td>
</tr>
<tr>
<td>Progress</td>
<td>Compressed / Encoded Data</td>
<td>Web Services (SOAP)</td>
<td>Enscribe</td>
</tr>
<tr>
<td>RedBk</td>
<td>External Command Call</td>
<td>Enterprise Java Beans (EJB)</td>
<td>Ebase</td>
</tr>
<tr>
<td>SGL/DS</td>
<td>Parallel wrapper 3rd party apps</td>
<td>EDI</td>
<td>FOCUS</td>
</tr>
<tr>
<td>SQL Server</td>
<td>EMC InfoMover</td>
<td>FIX</td>
<td>IDMS/SQL</td>
</tr>
<tr>
<td>Sybase (ASE and IQ)</td>
<td>Web logs</td>
<td>SWIFT</td>
<td>ImageSQL</td>
</tr>
<tr>
<td>Teradata</td>
<td>Unstructured: email, docs, etc.</td>
<td>HIPAA</td>
<td>Informan</td>
</tr>
<tr>
<td>Universe</td>
<td>Content Management Systems</td>
<td></td>
<td>KSAM</td>
</tr>
<tr>
<td>UniData</td>
<td>Life Sciences</td>
<td></td>
<td>M204</td>
</tr>
<tr>
<td>NonStopSQL</td>
<td></td>
<td></td>
<td>MS Analysis</td>
</tr>
<tr>
<td>And more...</td>
<td></td>
<td></td>
<td>Nomad</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Enterprise Applications</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>JDE/PeopleSoft EnterpriseOne</td>
<td>CDC / Replication</td>
<td>DB2 (on Z, I, P, X series)</td>
<td></td>
</tr>
<tr>
<td>Oracle Applications</td>
<td></td>
<td>Oracle</td>
<td></td>
</tr>
<tr>
<td>PeopleSoft Enterprise</td>
<td></td>
<td>SQL Server</td>
<td></td>
</tr>
<tr>
<td>SAS</td>
<td></td>
<td>Sybase</td>
<td></td>
</tr>
<tr>
<td>SAP R/3 and BI</td>
<td></td>
<td>Informix</td>
<td></td>
</tr>
<tr>
<td>SAP XI</td>
<td></td>
<td>IMS</td>
<td></td>
</tr>
<tr>
<td>Siebel</td>
<td></td>
<td>VSAM</td>
<td></td>
</tr>
<tr>
<td>JDA</td>
<td></td>
<td>ADABAS</td>
<td></td>
</tr>
<tr>
<td>Anaba</td>
<td></td>
<td>IDMS</td>
<td></td>
</tr>
<tr>
<td>Manugistics</td>
<td></td>
<td>NonStopSQL</td>
<td></td>
</tr>
<tr>
<td>i2</td>
<td></td>
<td>Escribe</td>
<td></td>
</tr>
<tr>
<td>And more...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IBM Information Server Connects to Almost All Sources of Data
Remove Erroneous Data Before It Gets Into the Warehouse

DataStage can Extract, Transform, and Load data into your data warehouse. But we need to make sure the data is clean.

IBM Information Server for System z Can Load Your Data Warehouse

<table>
<thead>
<tr>
<th>IBM Information Server for System z</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Understand</strong></td>
</tr>
<tr>
<td>Information Analyzer on Linux for System z</td>
</tr>
<tr>
<td>Business Glossary on Linux for System z</td>
</tr>
<tr>
<td>Rational Data Architect</td>
</tr>
<tr>
<td><strong>Cleanse</strong></td>
</tr>
<tr>
<td>QualityStage on Linux for System z</td>
</tr>
<tr>
<td><strong>Transform</strong></td>
</tr>
<tr>
<td>DataStage on Linux on for System z</td>
</tr>
<tr>
<td>DataStage for z/OS</td>
</tr>
<tr>
<td>DataStage MVS</td>
</tr>
<tr>
<td><strong>Deliver</strong></td>
</tr>
<tr>
<td>Classic Fed. Server</td>
</tr>
<tr>
<td>Classic Replication</td>
</tr>
<tr>
<td>Replication Server</td>
</tr>
<tr>
<td>Change Data Capture</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Platform Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallel Processing Services</td>
</tr>
<tr>
<td>Connectivity Services</td>
</tr>
<tr>
<td>Metadata Services</td>
</tr>
<tr>
<td>Administration Services</td>
</tr>
<tr>
<td>Deployment Services</td>
</tr>
<tr>
<td>Metadata Server</td>
</tr>
<tr>
<td>Information Services Director</td>
</tr>
</tbody>
</table>
Why Should I Care About Cleansing Information?

- **Lack of data standards**
  - Different formats and structures across different systems

- **Data surprises in individual fields**
  - Data misplaced in the database

- **Information buried in free-form fields**

- **Data myopia**
  - Lack of consistent identifiers inhibit a single view

- **The redundancy nightmare**
  - Duplicate records with a lack of standards

### QualityStage™ Fixes Data Quality and Consistency

- Ensures clean, standardized information
  - Eliminates duplications
  - Matches against reference data

- Supports global postal address verification

- Provides visual tools for designing quality rules and matching logic

- Integrated with DataStage (one platform, one user interface)

- Data quality processes can be deployed within extract, transform, and load (ETL) jobs logic or separately as shared data quality services (SOA)
Two Methods to Decide a Match

Are these two records a match?

WILLIAM J HOLDEN 128 MAIN ST 02111 12/8/62
WILLIAM JOHN HOLDEN 128 MAINE AVE 02110 12/8/62

B B A A B D B A = BBAABDBA

Deterministic Decisions Tables:
- Fields are compared
- Letter grade assigned
- Combined letter grades are compared to a vendor delivered file
- Result: Match; Fail; Suspect

Probabilistic Record Linkage:
- Fields are evaluated for degree-of-match
- Weight assigned: represents the “information content” by value
- Weights are summed to derive a total score
- Result: Statistical probability of a match

Add QualityStage to Cleanse the Data

Discard Duplicates
Match/Compare against Reference Data

Extract Transform Load

Source Data zLinux z/OS

Data Warehouse
IBM Information Server Exposes Jobs as SOA Services

Information Services Director included in IBM Information Server enables DataStage and QualityStage jobs to be invoked as SOA services

DataStage and QualityStage Utilize Parallel Processing Services for Extreme Scalability

- Provides automatic dynamic data partitioning, repartitioning, and pipelining for optimal parallel performance
- Design integration processes without concern about underlying hardware architecture or number of processors
  - Resources defined in a separate configuration file
  - Allows easy expansion to new hardware
- Benefits from processing capacity, I/O capacity, and Hipersockets on System z
IBM Beats the Competition in Data Warehouse Solutions

**IBM**
- Integrated data services platform
- Extensive connectivity
- Enterprise scalability
- Easily expose jobs as services

**Oracle**
- Two separate un-integrated products
- Warehouse Builder can only load Oracle databases
- Oracle Data Integrator (used to be Sunopsis) has no data quality capabilities
- Coding required to expose jobs as services

**Microsoft**
- Data quality limited to “Fuzzy Search” and MS SQL only
- Limited connectivity and limited support for non-Microsoft
- Lacks enterprise scalability
- Coding required to expose jobs as services

Service Oriented Finance Wants a *Dynamic* Data Warehouse

Our data is updated frequently, We need the data in our data warehouse to be more current

Change Data Capture together with Information Server can “trickle feed” data into your data warehouse as it changes
IBM Information Server for System z Can Provide Near Real-Time Data Movement

IBM Information Server for System z

- **Understand**
  - Information Analyzer on Linux for System z
  - Business Glossary on Linux for System z
  - Rational Data Architect

- **Cleanse**
  - QualityStage on Linux for System z

- **Transform**
  - DataStage on Linux for System z
  - DataStage for z/OS
  - DataStage MVS

- **Deliver**
  - Classic Fed. Server
  - Classic Replication Server
  - Replication Server
  - Change Data Capture for System z

Platform Services

- **Parallel Processing Services**
- **Connectivity Services**
- **Metadata Services**
- **Administration Services**
- **Deployment Services**
  - Information Services Director

Change Data Capture for System z

- **Data Changes**
  - IBM Information Server Change Data Capture for System z

  - **Direct connect**

  - IBM Information Server QualityStage DataStage

  - **Load**

  - Data Warehouse

- **Event is triggered automatically by a change in the data**
  - Monitors DB2 log file for data changes
  - Sends message with data changes

- **Message can initiate a business process**
  - Example: After $10,000 in sales, a service request is sent to WebSphere Information Services Director, invoking a DataStage job to load sales data for analysis

- **Automates the process of loading the data warehouse**
  - Improves the currency of data in the data warehouse
  - Alternative to statically scheduled updates - which can lead to stale data and lost opportunities
### Log-Based Change Data Capture

- **Using existing native database recovery logs to capture table and row level activity in a relational database**
  - Most databases have a native log format that can be accessed
    - DB2 and DB2 z/OS = DB2 Log, DB2(5)=OS/400 Journal
    - Oracle = Re-do Log, SQL Server = Transaction Log

- **Many advantages to log–based Change Data Capture**
  - No changes to existing applications or schemas required
  - Little performance impact to source application and system
    - 0.05% system resources required to process over 300 GB of data
  - Sends only the changes – efficient use of bandwidth
  - Scalable
  - Alternatives have many drawbacks
    - SQL Select, File Comparison, Database Triggers, Modifying Source Application

- **Wide range support**
  - Information Server CDC for - DB2 (all platforms), Oracle, SQL Server, Sybase, and more
  - Data Event Publisher for non relational sources

### Where Should We Deploy This Solution?

- Capability is important, but cost is a big concern for us
- Deploying this solution on your System z will cost less than Oracle on distributed

---

**CIO**

**IBM**
Mainframe Extension Solution – Deliver Business Insight with a Data Warehouse

- Browser
- DataQuant for WebSphere QMF
- DB2 (data and repository)
- WebSphere Application Server z/OS
- IBM Information Server for System z DataStage QualityStage z/Linux

System z Hypervisor

GP

zIIP

03 - Deliver Business Insights with System z v5.1

Storage Costs:
DB2 Provides More Storage Savings than Oracle

- DB2 for z/OS lowers TCO by reducing storage needed
  - TPC-H Benchmark: DB2 compression of 59% vs 29% for Oracle RAC

- Storage savings with DB2 vs. Oracle for a 10 TB data base

<table>
<thead>
<tr>
<th>Storage System</th>
<th>Oracle</th>
<th>DB2 for z/OS*</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP Enterprise Virtual Array 8100 Storage</td>
<td>IBM System Storage DS6800</td>
<td></td>
</tr>
<tr>
<td>Overall database compression ratio</td>
<td>29%</td>
<td>59%</td>
</tr>
<tr>
<td>Using TPC-H benchmark results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>For 10 TB uncompressed data storage needed</td>
<td>7.5 TB of HP Storage</td>
<td>4.2 TB of IBM Storage</td>
</tr>
<tr>
<td>Cost of storage (3 year TCA)</td>
<td>$319,270 + $15,113 x 3</td>
<td>$234,101 + $13,164 x 2**</td>
</tr>
<tr>
<td>= $364,609</td>
<td>= $260,429</td>
<td></td>
</tr>
</tbody>
</table>

With compression, storage for DB2 costs 29% less than for Oracle

*DB2 for z/OS achieves similar compression ratios to those of DB2 for LUW
**IBM storage maintenance fee for the first year is included in the warranty

03 - Deliver Business Insights with System z v5.1
16

US Retailer Improves Response Time by Co-locating Data Warehouse and Operational Data

- A major US retailer moved their 5.5TB data warehouse from distributed servers to System z
  - Operational data bases were already located on System z

- On average they reduced query processing times by 80% due to better query parallelism in DB2 for z/OS
  - (17 minutes to 3 minutes)

- They saved CPU cycles to load the data warehouse
  - Avoided network processing

The Value of Business Insight

- We learned the best promotion to maximize our business profits
- And I saved money by deploying our data warehouse solution on System z

Marketing VP  
CIO
Accelerate Development with Pre-Built Data Models for Industry Data Warehouses

- Quick start your data warehouse design with pre-built IBM Industry Data Models

- Use Enterprise Model Extender to customize the data model
  - Eclipse plug-in for Rational Data Architect

- Models capture the best practices of over 400 IBM customers
  - Banking, finance, health, insurance, retail and Telco