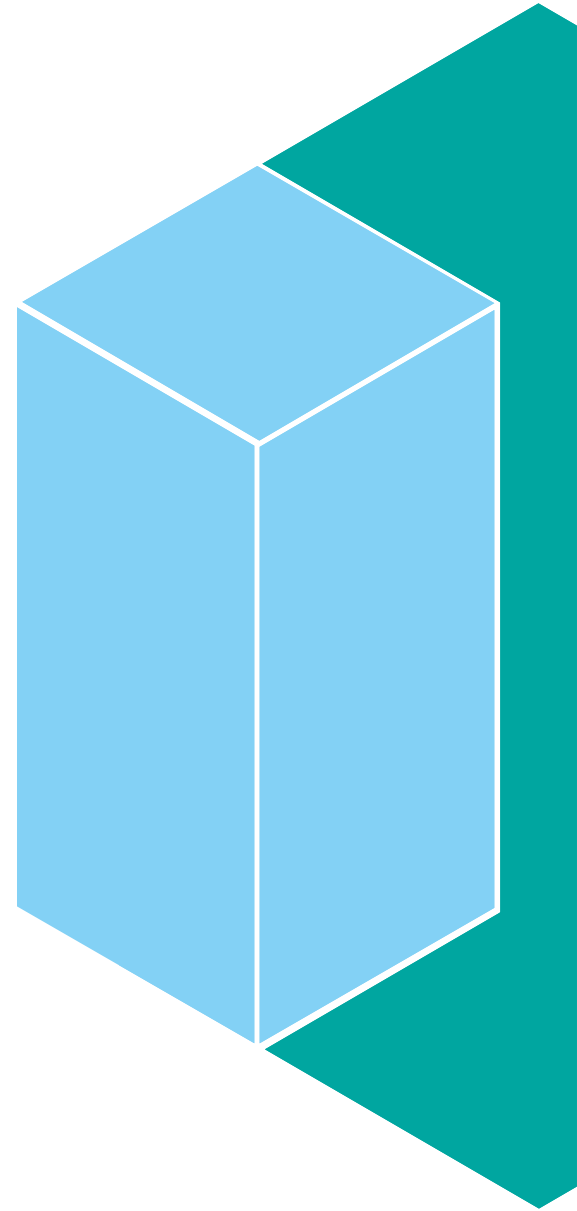


Unleash the analytic power of your IMS data using IBM DB2 Analytics Accelerator for z/OS

Timm Zimmermann, IMS Tools SME



Sharpen your competitive edge
2016 IMS Technical Symposium
March 7 – 10, 2016
Wiesbaden, Germany

www.ims-symposium.com

DB2 ANALYTICS ACCELERATOR 101

Hybrid transaction/Analytical processing



The hybrid computing platform on z Systems

Supports transaction processing and analytics workloads concurrently, efficiently and cost-effectively

Delivers industry leading performance for mixed workloads

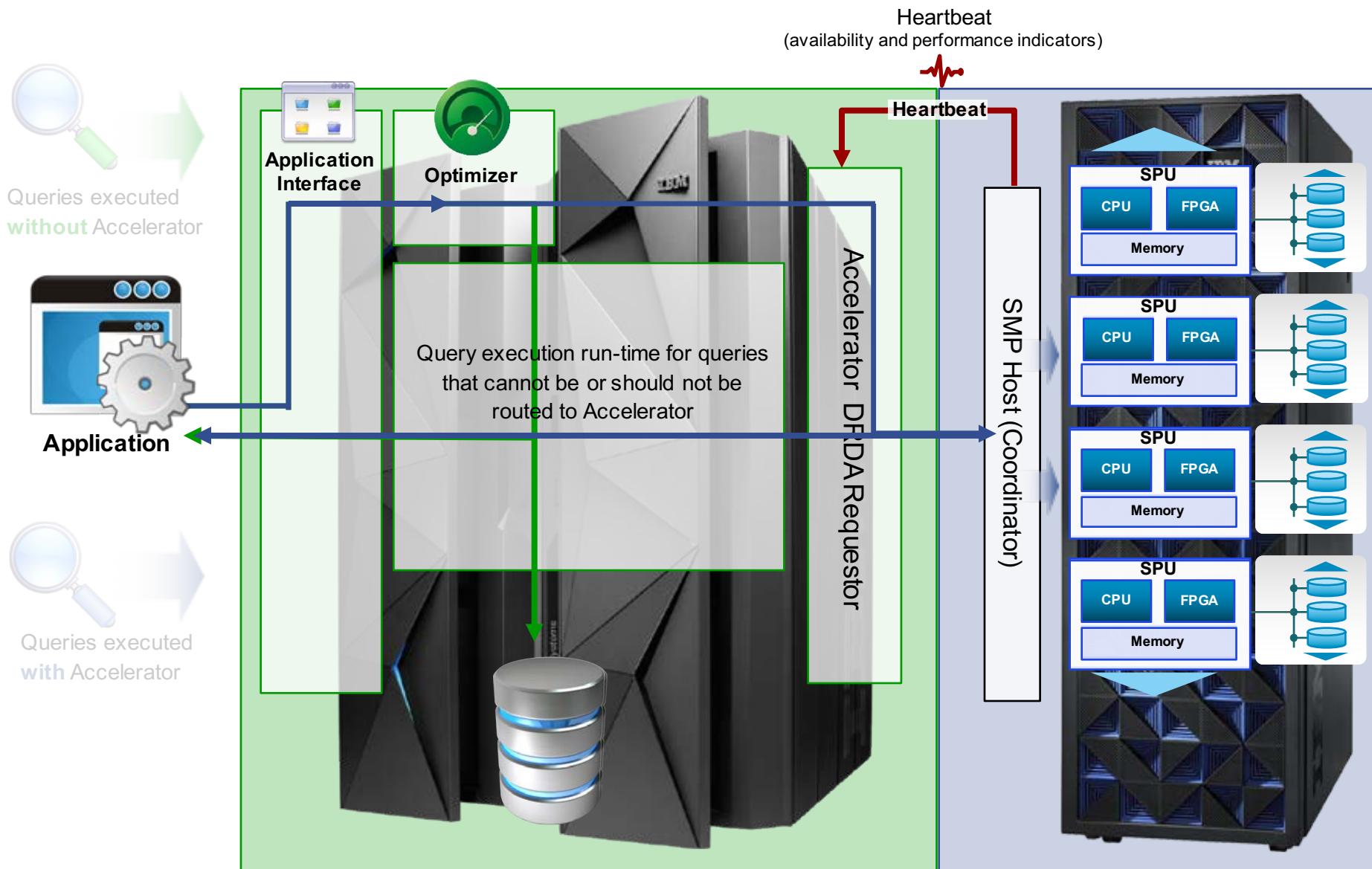
The unique heterogeneous scale-out platform in the industry

Superior availability, reliability and security

DB2 Analytics Accelerator and DB2 for z/OS

A self-managing, hybrid workload-optimized database management system that runs every query workload in the most efficient way, so that each query is executed in its optimal environment for greatest performance and cost efficiency

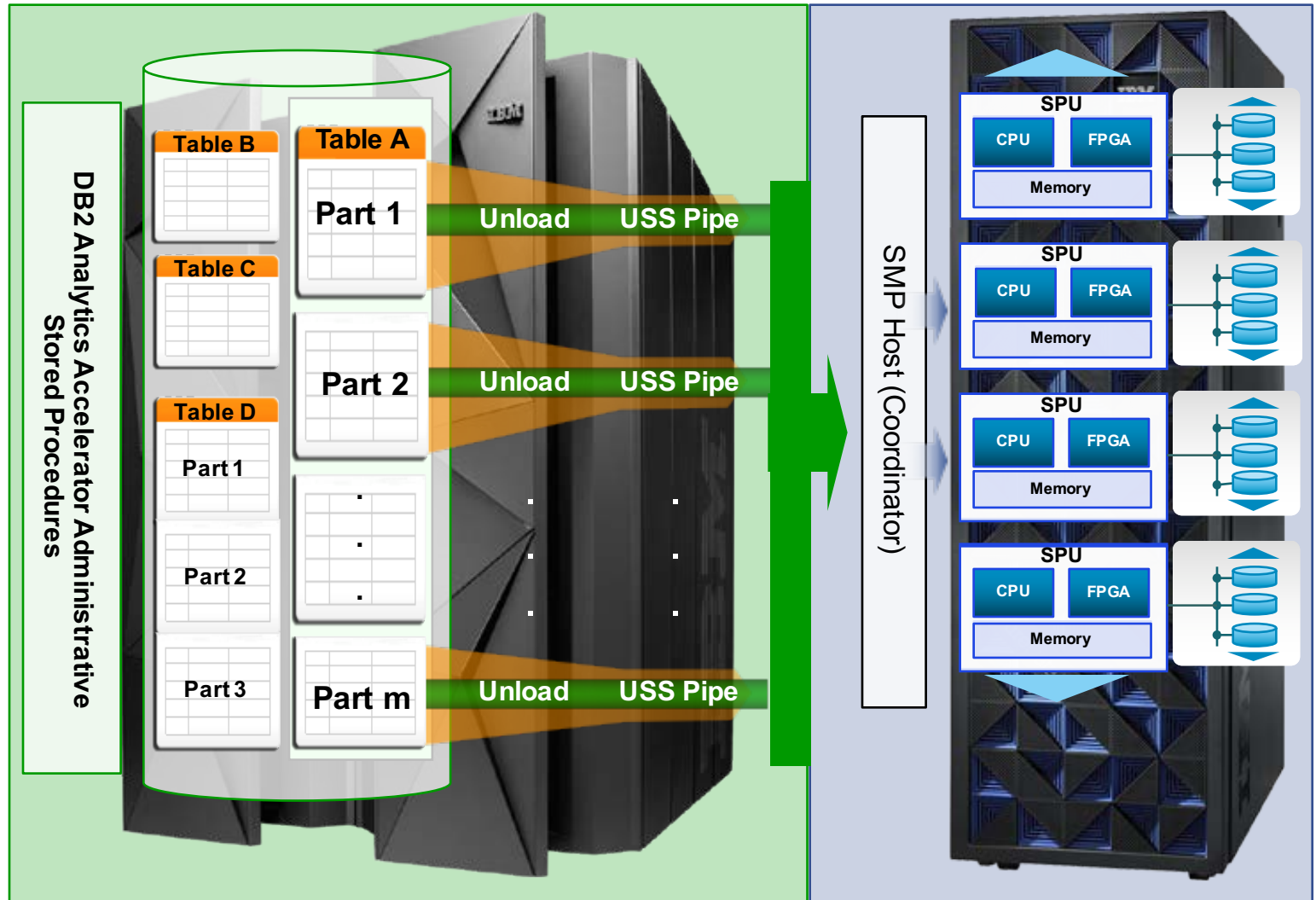
Query execution process flow



Loading the Accelerator



DB2 Analytics
Accelerator
Studio



Introducing Accelerator-only table type in DB2 for z/OS

Creation (DDL) and access remains through DB2 for z/OS in all cases

Non-accelerator DB2 table

- Data in DB2 only

Accelerator-shadow table

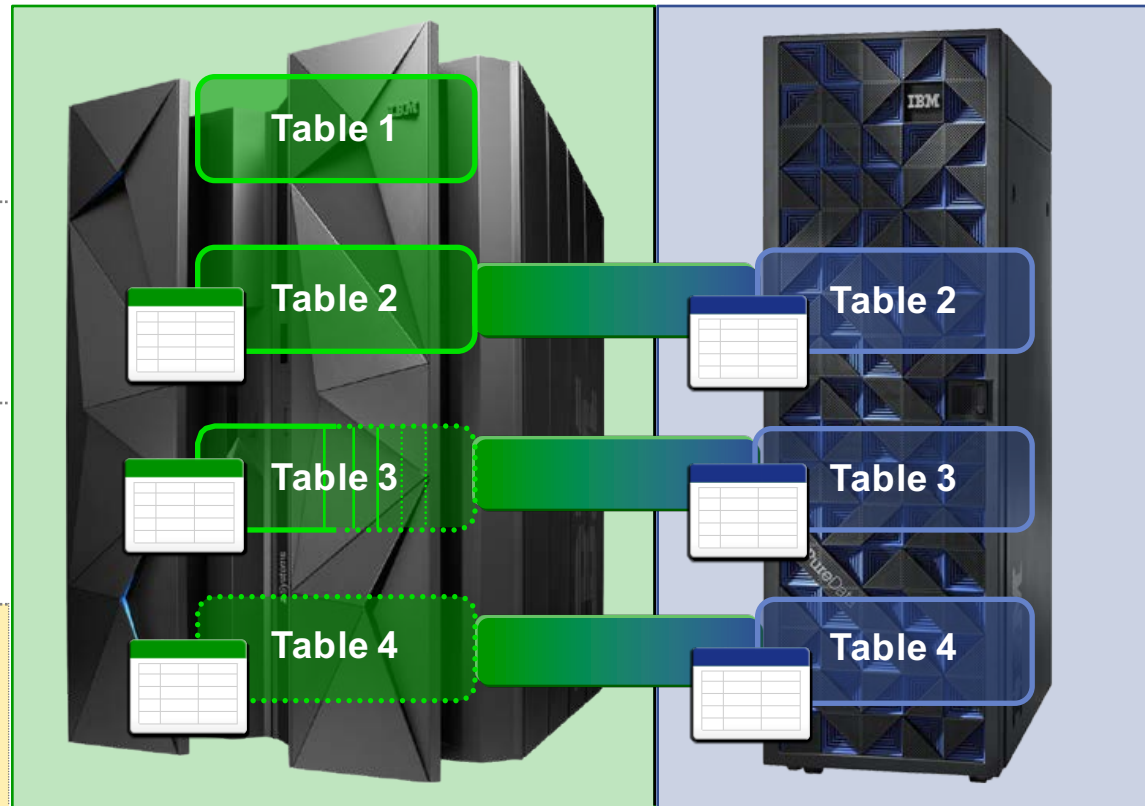
- Data in DB2 and the Accelerator

Accelerator-archived table / partition

- Empty read-only partition in DB2
- Partition data is in Accelerator only

Accelerator-only table (AOT)

- “Proxy table” in DB2
- Data is in Accelerator only



New with DB2 Accelerator V4.1 PTF5

DB2 11 support requires the following DB2 APARs: PI35817, PI35818, PI35819, PI35820, PI35821

DB2 Analytics Accelerator – Usage scenarios

How organizations leverage the Accelerator today

1 Rapid acceleration of Existing Business Critical Queries

If the data is analyzed on the mainframe

- *Performance improvements* and *cost reduction* while retaining z Systems security and reliability

2 Reduce IT sprawl for analytics

If the data is offloaded to a distributed data warehouse or data mart

- *Simplify* and *consolidate* complex infrastructures, *low latency*, *reliability*, *security* and *TCO*

3 Derive business insight from z/OS transaction systems

If the data is not being analyzed yet

- One integrated, *hybrid platform*, optimized to run *mixed workload*
- *Simplicity* and *time to value*

4 Improve access to historical data and lower storage costs

If the analysis is based on a lot of historical data

- *Performance improvements* and *cost reduction*



DB2 Analytics Accelerator Version 5.1

Adding new dimensions in functionality to expand use cases

1 Rapid acceleration of existing Business Critical Queries

Adding application support for temporary objects (QMF, Multi-step Reporting, IBM Campaign, etc.)

- Insight into now to *maximize business opportunities* in today's dynamic environment

2 Reduce IT sprawl for analytics

In-database transformation to support Data Stage Balanced Optimization and the consolidation of ETL/ELT processing in DB2 for z/OS

- *Business agility* through *simplified architecture*

3 Derive business insight from z/OS transaction systems

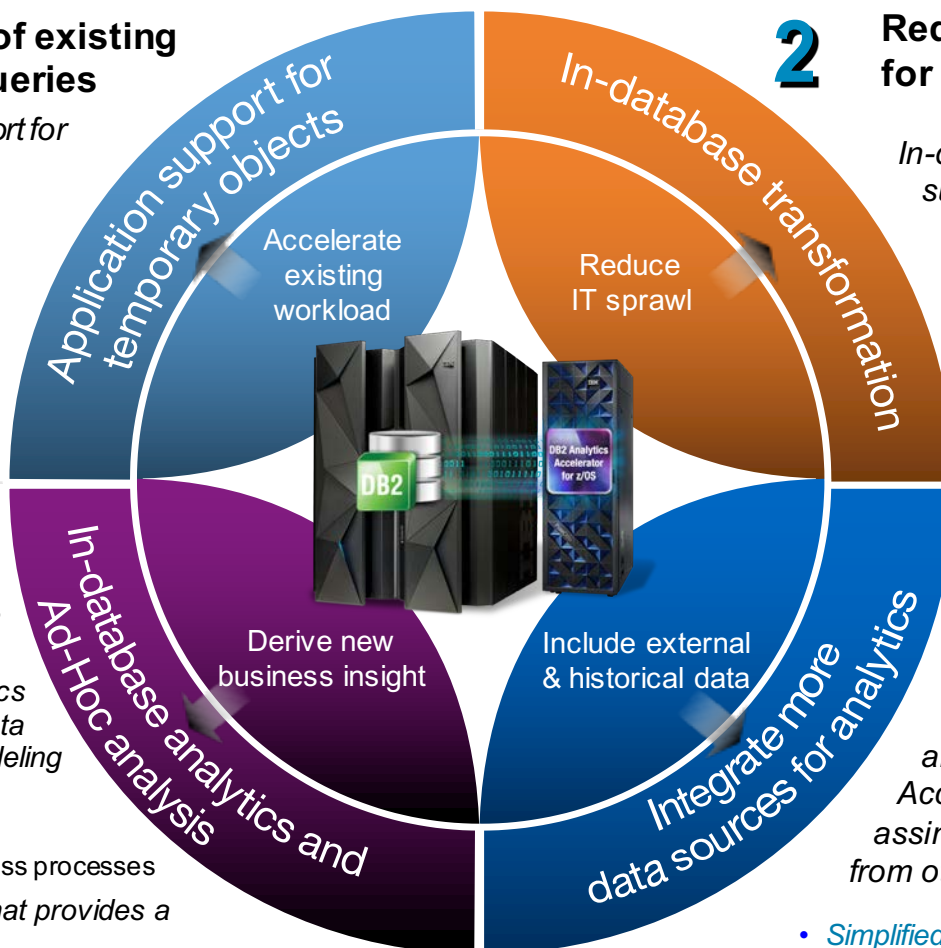
In-database analytics to accelerate predictive analytics applications; SPSS/INZA data mining and in-database modeling can be processed within the Accelerator

- *Real-time, actionable* business processes
- Individual ad-hoc analysis that provides a Data Scientist Work Area*
- Environment to efficiently, continuously test and *improve analytic results* to drive *better customer understanding*

4 Improve access to historical data and lower storage costs

Integrate more data sources for analytics, using DB2 Analytics Accelerator Loader for z/OS to assimilate with IMS data or data from other sources

- *Simplified access to information* – when you need it



Deliver Right-Time Analytics to drive better business outcomes

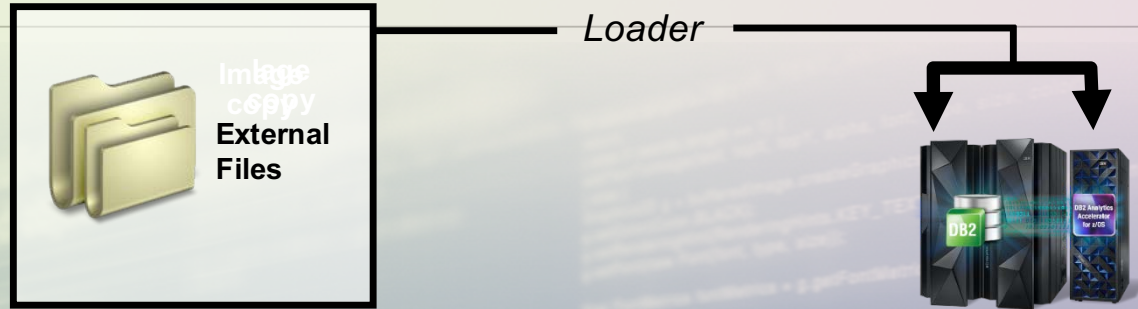
DB2 ANALYTICS ACCELERATOR LOADER FOR Z/OS

- **Announced the IBM DB2 Analytics Accelerator Loader for z/OS**
 - High speed, zIIP enabled accelerator loading with additional functions

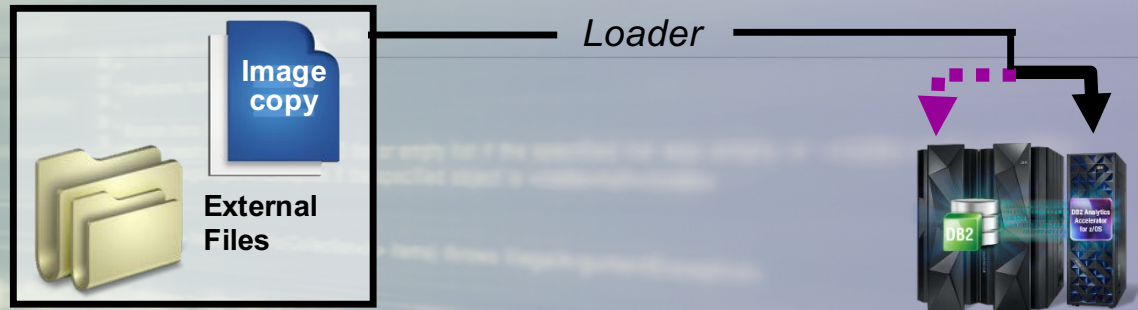
- **Addressed challenges in loading the Accelerator**
 - Elongated two-step process
 - Fast load from data in a file to accelerator only
 - Related set of tables put into read-only state
 - Inability to load to desired point-in-time
 - Plus
 - zIIP-eligible data preparation to load data faster
 - Load stand-alone backups directly to accelerator

DB2 Analytics Accelerator Loader for z/OS*

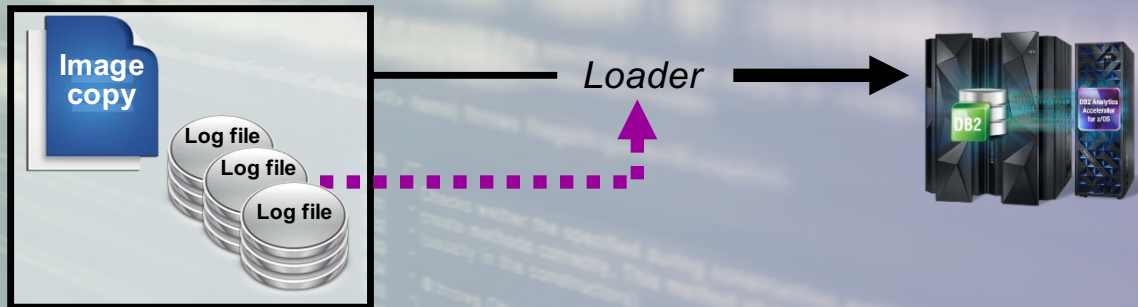
- Load data from image copy or other sources into DB2 and the Accelerator in parallel
EXTERNAL LOAD (IDAA_DUAL)



- Load data into the Accelerator only (meta data required in DB2)
EXTERNAL LOAD (IDAA_ONLY)



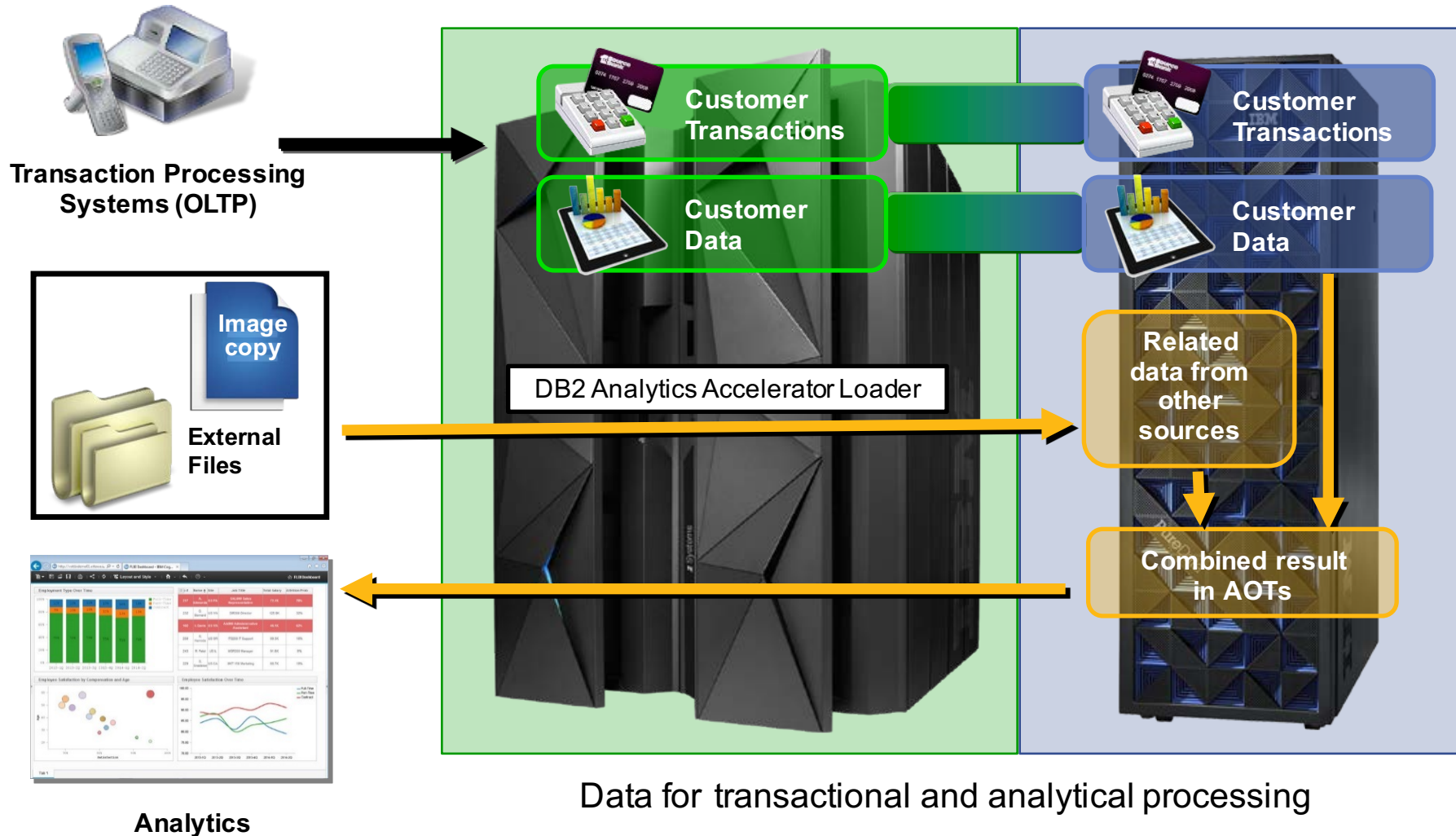
- Direct loading into Accelerator, point-in-time loading (i.e. roll-forward to point-in-time from log files)
CONSISTENT LOAD



*Separately licensed option

Integrate more data sources for analytics

Load external data to the accelerator and save combined analytic results in accelerator-only tables



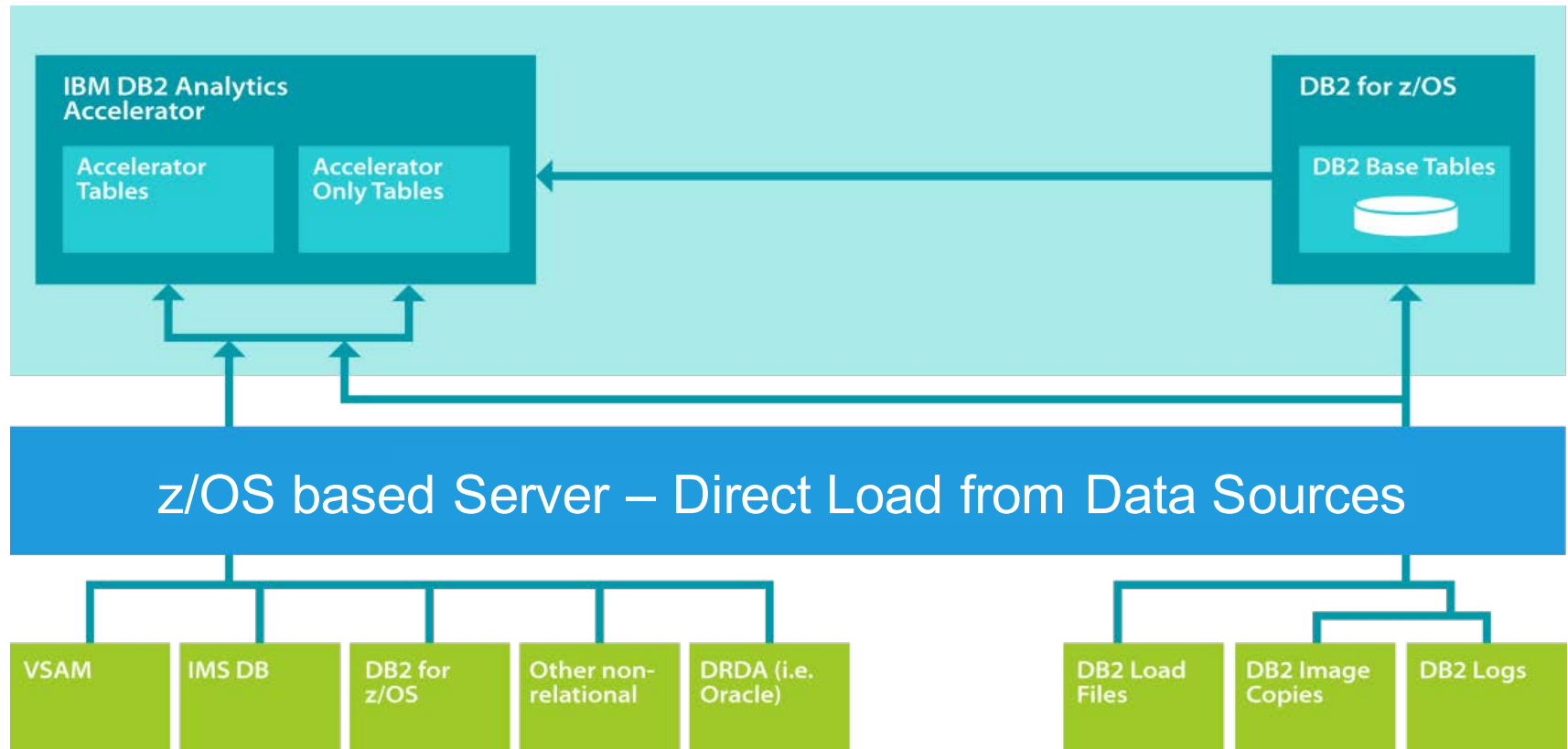
Existing Accelerator Loader Customers

▪ Identified challenges in loading non-DB2 data

- Complexity in converting non-relation data to relational
- Manual / ETL process
- Slow, due to intermediate write to disk



IBM DB2 Analytics Accelerator Loader for z/OS v2.1



Loading non-DB2 Data Sources into Accelerator

Product Comparison

- **IBM DB2 Analytics Accelerator Loader – The user must:**
 - Extract data from source (IMS, VSAM, Oracle, SMF, etc.)
 - Convert extracted data to DB2 external load file format
 - DataStage or other tooling can be used
 - Create a DB2 table that matches format of extracted data
 - Add newly created table to the Accelerator
 - Construct a DB2 Load utility field specification that describes the input data
 - Run Accelerator Loader batch job to load data to accelerator

- **Accelerator Loader V2.1– Automates entire process:**
 - User builds a select statement from data source(s) (IMS, VSAM, SMF, Oracle, ...)
 - Automatically creates the DB2 table
 - Automatically adds table to Accelerator
 - Automatically extracts specified source data
 - Automatically converts data to necessary DB2 format (in memory)
 - Automatically loads data to Accelerator
 - Automatically enables table for acceleration



Single batch job!

Other Accelerator Loader v2.1 Improvements

▪ **LOAD RESUME support**

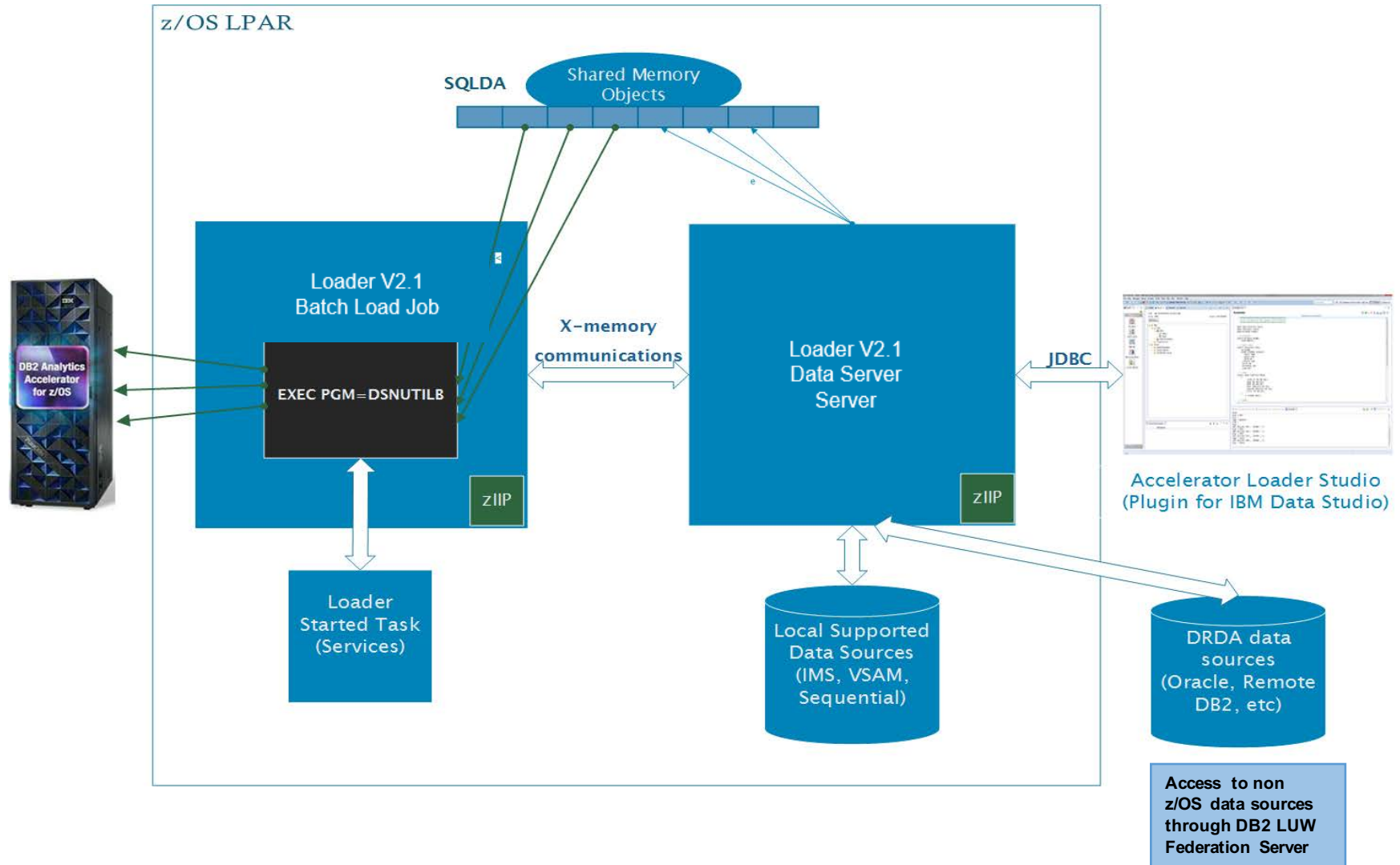
- **Avoid the need for full LOAD replace**
- New function to append data from a file to existing data in a table
 - Replaces the need for full data replace
- LOAD RESUME can provide significant CPU / elapsed time savings vs. REPLACE
- LOAD RESUME to both DB2 table and Accelerator table in parallel
- LOAD RESUME to Accelerator only and to Accelerator-only tables

▪ **Accelerator-only tables (AOTs)* – new in Accelerator V4.1 PTF5**

- New table type in DB2 for z/OS
 - Data resides only in Accelerator version of table, not in front-end DB2
- All queries targeting these tables are routed to Accelerator
 - All SELECT statements automatically routed to Accelerator
 - INSERT/UPDATE/DELETE also supported
 - No logging or point in time recovery
- **Without tooling, data must be inserted to load**
 - **Insert directly to Accelerator-only table**
 - **If data sourced from outside DB2, multi-step process to load**

*enabled via PTFs (2016)

DB2 Analytics Accelerator Loader V2.1 Architecture



IBM DB2 Analytics Accelerator Loader V2.1

- **Provides a complete solution to load DB2 and non DB2 data to the Analytics Accelerator**
- **Supplies a relational view of non-relational data sources**
 - VSAM, IMS, Sequential, Adabas (Beta), SMF, etc.
 - Allows users to code familiar SQL against data sources for purpose of loading the Accelerator
- **Consistent access to the data no matter the source**
- **Provides end-to-end support for data extraction, conversion, and loading to accelerator in one step**
 - No data is landed in any intermediate files
 - Automates process of creating DB2 and Accelerator tables
- **Leverages zIIP for**
 - Data Extraction
 - Data Conversion
 - Data Loading to the Accelerator

Accelerator Loader V2.1 Server Details

- **Accelerator Loader Server**

- Runs as z/OS address space
- Acts as a relational database server
- Provides mapping facility for non-relational data sources
- Manages Virtual Table/View metadata
- Manages all access to supported data sources
- All data access controlled by z/OS security (pass-through)

- **Accelerator Loader V2.1 will require two started tasks**

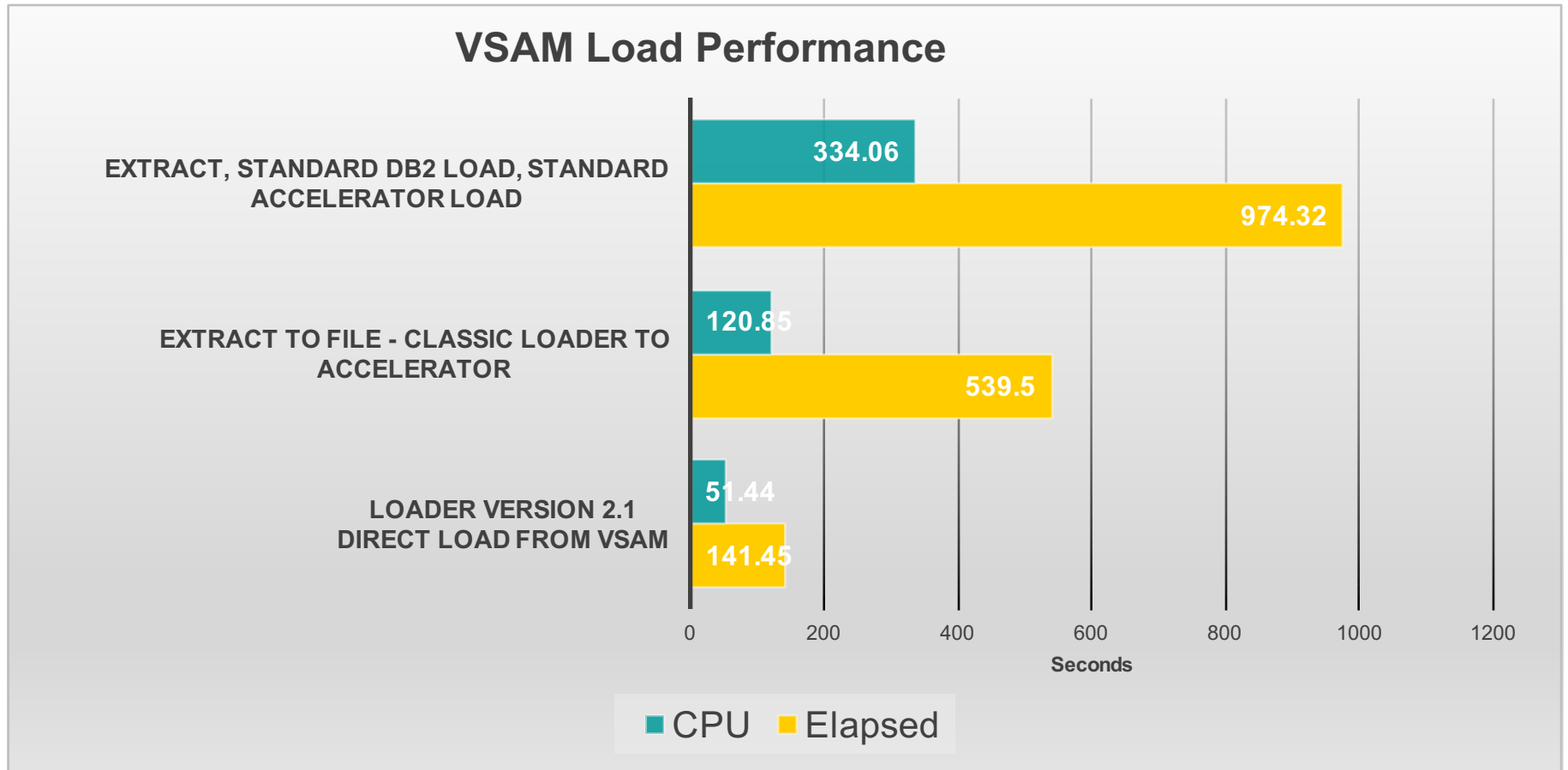
- Data Server
- Existing Accelerator Loader Started Task

- **New Data Server must be running on same LPAR that is connected to Accelerator**

- **Remote access provided to relational data sources**

- Remote DB2 (DRDA)
- DB2 LUW AESE
 - Provides access to Oracle (and other data sources)

Lab Test Results Loading VSAM Data

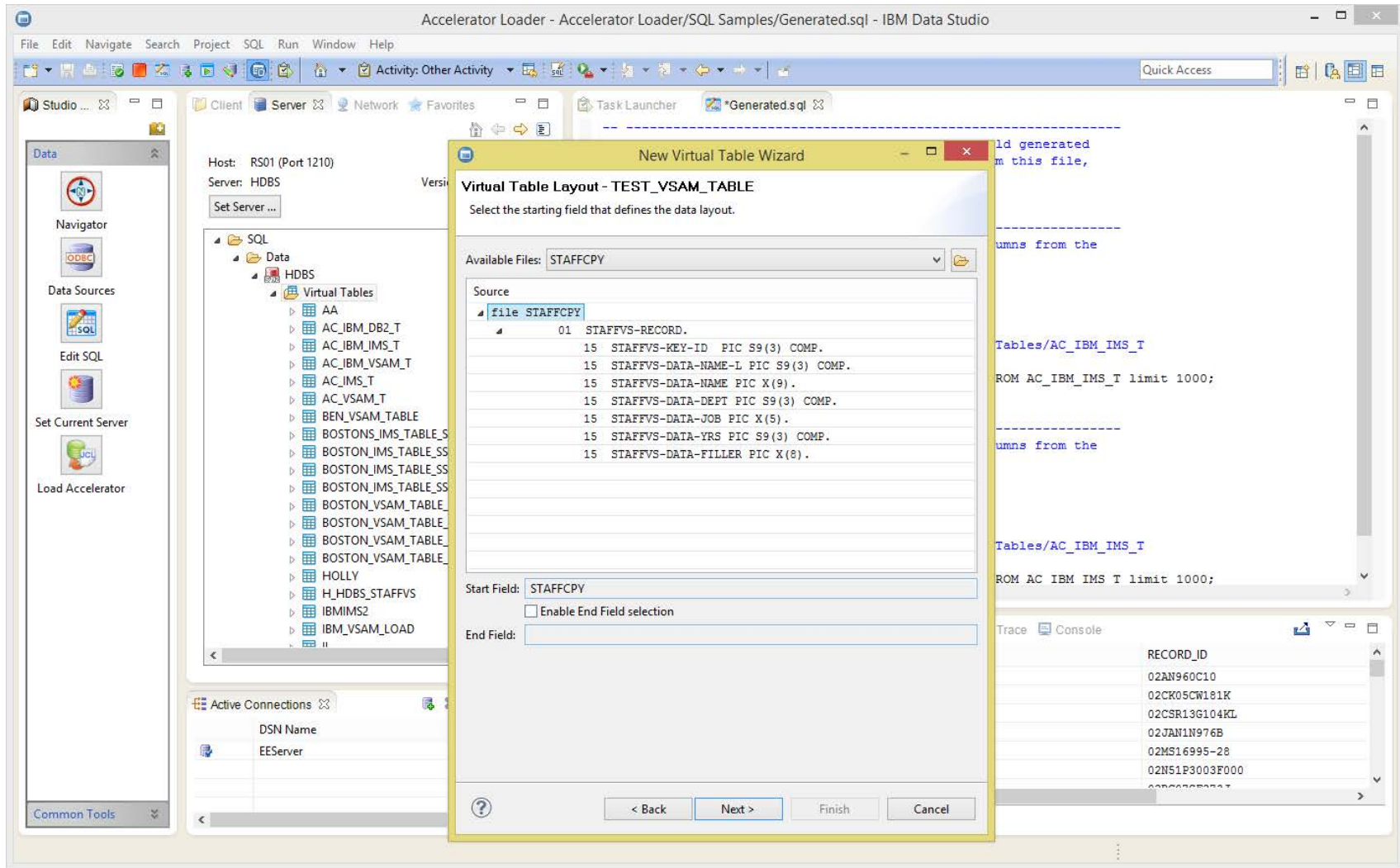


25 GB of Data - 250 Million Rows

Accelerator Loader V2.1 Direct Data Load Details

- **Only Accelerator version of table will be loaded**
- **Source data is not locked during load**
 - User can control (at the source) if necessary for data consistency
 - Cloning Tool an option for IMS data
- **Performance can be tuned**
 - Read Parallelism: VSAM and Sequential now
 - Write Parallelism: Number of pipes to accelerator
- **Data Validity Checking – Discard Processing**
 - Data type checking performed during load
 - Only messages issues on invalid data (no discard file support)
 - No duplicate key checking
 - Can be done via SQL post load
 - Roadmap for improvement

Create Virtual Table (if non-relational)



Test select on small result set to verify it returns desired data

Accelerator Loader - Accelerator Loader/SQL Samples/Generated.sql - IBM Data Studio

File Edit Navigate Search Project SQL Run Window Help

Activity: Other Activity Quick Access

Stu... Client Server Network Favori...

Data

Host: RS01 (Port 1210)
Server: HDBS Version: 02.01.00.0000
Set Server ...

Navigator

Data Sources

Edit SQL

Set Current Server

Load Accelerator

OPERLOG_MDB
OPERLOG_MDB_MDB_CONTROL_C
OPERLOG_MDB_MDB_TEXT_OBJEC
OPERLOG_SYSLOG
PATRIC
POOLE
REMOTE_ORACLE_TABLE
SAGHI
SEQUENTIAL_STAFF
STAFF_VIRTUAL_TABLE
SYSLOG
TESTMAP
TEST_VSAM_TABLE
TIM
TRACE_DESCRIPTION
TRACE_TYPES
TSTTZ_IMS_STUFF
V
VSAM_KSDS_STAFF
VV
WAYNE
WW
Virtual Views
Other Subsystems

Active Connections

DSN Name
EEServer

Task Launcher *Generated.sql TEST_VSAM_TABLE

```
-- This statement will return all rows and all columns from the  
-- following table:  
-- Name : TEST_VSAM_TABLE  
-- Catalog : null  
-- Schema : HLVSQ  
-- Remarks : VSAM - HLVSQ.STAFF.VSAM  
-- Tree Location: RS01/1210/SQL/Data/HDBS/Virtual Tables/TEST_VSAM_TABLE  
-- The sql statement:  
SELECT STAFFVS_KEY_ID, STAFFVS_DATA_NAME_L, STAFFVS_DATA_NAME, STAFFVS_DATA_DEPT,  
STAFFVS_DATA_JOB, STAFFVS_DATA_YRS, STAFFVS_DATA_FILLER  
FROM TEST_VSAM_TABLE limit 1000;
```

TEST_VSAM_TABLE (Table Node) SQL Results View Server Trace Console

| | STAFFVS_KEY_ID | STAFFVS_DATA_NAME_L | STAFFVS_DATA_NAME | STAFFVS_DATA_DEPT | STAFFVS_DATA_JOB | STAFFVS_DATA_YRS |
|----|----------------|---------------------|-------------------|-------------------|------------------|------------------|
| 0 | 10 | 7 | SANDERS | 20 | MGR | 7 |
| 1 | 20 | 6 | PERNAL | 20 | SALES | 8 |
| 2 | 30 | 8 | MARENGHI | 38 | MGR | 5 |
| 3 | 40 | 7 | O'BRIEN | 38 | SALES | 6 |
| 4 | 50 | 5 | HANES | 15 | MGR | 10 |
| 5 | 60 | 7 | QUIGLEY | 38 | SALES | 0 |
| 6 | 70 | 7 | ROTHMAN | 15 | SALES | 7 |
| 7 | 80 | 5 | JAMES | 20 | CLERK | 0 |
| 8 | 90 | 7 | KOONITZ | 42 | SALES | 6 |
| 9 | 100 | 5 | PLOTZ | 42 | MGR | 7 |
| 10 | 110 | 4 | NGAN | 15 | CLERK | 5 |
| 11 | 120 | 8 | NAUGHTON | 38 | CLERK | 0 |
| 12 | 130 | 9 | YAMAGUCHI | 42 | CLERK | 6 |
| 13 | 140 | 5 | FRAYE | 51 | MGR | 6 |
| 14 | 150 | 8 | WILLIAMS | 51 | SALES | 6 |

35 rows SQL Messages

Build, submit JCL and view results

The screenshot displays the IBM Data Studio Accelerator Loader interface. The main window shows the JCL code for loading data into the TEST1.ACCEL_STAFF table. The code includes a DROP TABLE statement, a COMMIT, a CREATE TABLE statement, and a LOAD DATA statement. The database structure is visible in the left pane, showing the TEST1 database and the ACCEL_STAFF table. The bottom pane shows the JCL Library and Host information.

Host: RS01 (Port 1210)
Server: HDBS
Version: 02.01.00.0000

Set Server ...

Task Launcher: *Generated.sql, TEST_VSAM_TABLE, *ABC.jcl

```
END
/*
//SYSIN DD *
-- DROP TABLE "TEST1"."ACCEL_STAFF";
-- COMMIT;

CREATE TABLE "TEST1"."ACCEL_STAFF"
(
  "STAFFVS_KEY_ID" SMALLINT NOT NULL,
  "STAFFVS_DATA_NAME_L" SMALLINT NOT NULL,
  "STAFFVS_DATA_NAME" VARCHAR(9) NOT NULL,
  "STAFFVS_DATA_DEPT" SMALLINT NOT NULL,
  "STAFFVS_DATA_JOB" VARCHAR(5) NOT NULL,
  "STAFFVS_DATA_YRS" SMALLINT NOT NULL,
);
COMMIT;
/*
//LOAD EXEC PGM=DSNUTILB,
// PARM=('PA1A')
//STEPLIB DD DISP=SHR,DSN=TEST.DB2.LOAD
// DD DISP=SHR,DSN=HLR.SHLRLOAD
// DD DISP=SHR,DSN=HLV.SHLVLOAD
//SYSIN DD *
EXEC SQL DECLARE HLVCUR CURSOR FOR
SELECT * FROM TEST_VSAM_TABLE
ENDEXEC

LOAD DATA
IDAA_ONLY ON PA1AACCI
REPLACE
LOG NO NOCOPYPEND
ENFORCE NO
ACCEL_CURSOR HLVCUR
ACCEL_HLV_SSID HDBS
ACCEL_REMOVE_AND_ADD_TABLES
ACCEL_ON_SUCCESS ENABLE YES
ACCEL_LOAD TASKS 1
```

TEST_VSAM_TABLE (Table Node) | SQL Results View | Server Trace | Console | JCL View

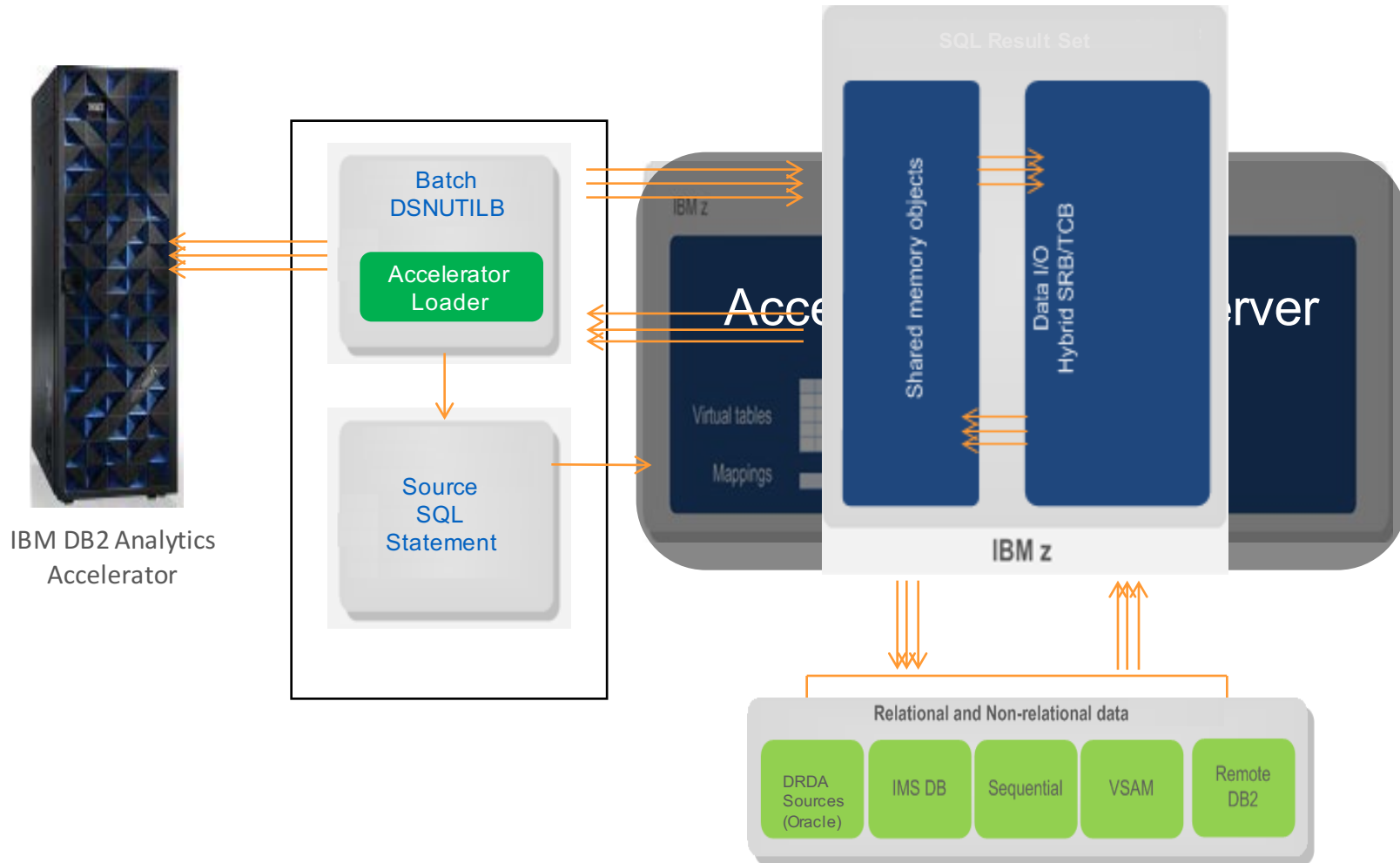
JCL Library: Host: RS01

Loading Non-Relational Data Source

- **VSAM, IMS, Sequential, Adabas**
 - Define copy book source library
 - Import copy book
 - Each field of copy book will be represented as table column
 - Choose parallelism to read data source
 - Designate read exit (required for compression)
- **Virtual table not required for relational data sources unless there is need to join data with non-relational or other relational data sources**
- **Create virtual view to:**
 - Refine data via SQL functions or casts
 - Subset data via where clause
 - Join two or more virtual tables

Analytics Accelerator Loader V2.1

In Memory Data Transformation



IMS Data Load Specifics

- **Requires IMS running on same LPAR as DB2**
- **IMS access is single threaded and higher general processor CPU**
 - Performance won't be on a par with data loading from datasets
- **There will be additional workload against IMS system**
- **Data consistency not guaranteed when loading**
 - If these are concerns, IMS Cloning Tool can be used
 - Create consistent clone, then Loader at IMS Clone
- **Validate customer has zIIP capacity for process**
- **Performance tuning**
 - Set Loader Data Server just below DB2/IMS
 - Batch job can stay at default or if needed, set just below Data Server

Potentially supported data sources roadmap

Databases

- PostgreSQL
- MySQL
- DB2 LUW
- Microsoft SQL Server
- Sybase
- Microsoft Access
- Derby
- H2
- HSQL
- Ingres
- Informix
- MetaMatrix
- Teradata
- Generic ANSI SQL - for any JDBC/ODBC source

Web Services and NoSQL

- **Web Services**
 - SOAP
 - REST
 - Arbitrary HTTP(S)
- **OData**
- **Big Data / No SQL / Search Engines / JCR and Other Sources**
 - Amazon SimpleDB
 - Apache Accumulo
 - Apache Cassandra DB
 - Apache SOLR
 - Greenplum
 - Hive / Hadoop
 - ModeShape JCR Repository
 - Mongo DB
 - Mondrian OLAP
 - Netezza data warehouse appliance

Enterprise

- **Enterprise Systems**
 - Salesforce
 - SAP Netweaver Gateway
 - Applications running on IBM i , UNIX, Windows
- **Object Sources**
 - JDG / Infinispan
 - Intersystems Cache Object Database
 - JPA sources
- **LDAP**
 - RedHat Directory Server
 - Active Directory
- **Files**
 - Delimited / Fixed width
 - XML
- **Spreadsheets**
 - Excel
 - Google Spreadsheet

Questions ?

