# How and why to archive & purge application data?

Bharath Nunepalli, Sr. Db2 DBA

HCA

10/2/19, 2:20 PM

# Agenda

About HCA Inc.

Our ERP Environment

Why we needed a data archive strategy

What is Optim archive?

How did we achieve data archive and purge?

Limitations in using Optim archive tool

Q&A

# HCA Inc. – Some facts about us:

- HCA is named one of the world's most ethical companies for nine years in a row
- 184 hospitals and approximately 2,000 sites of care, including surgery centers, freestanding ERs, urgent care centers, and physician clinics in 21 states and the United Kingdom.
- Ranked 63rd in Fortune 500
- 249,000 employees

    38,000 active physicians

    90,000 nurses

    5,300 IT employees
- 28 million patient encounters per year

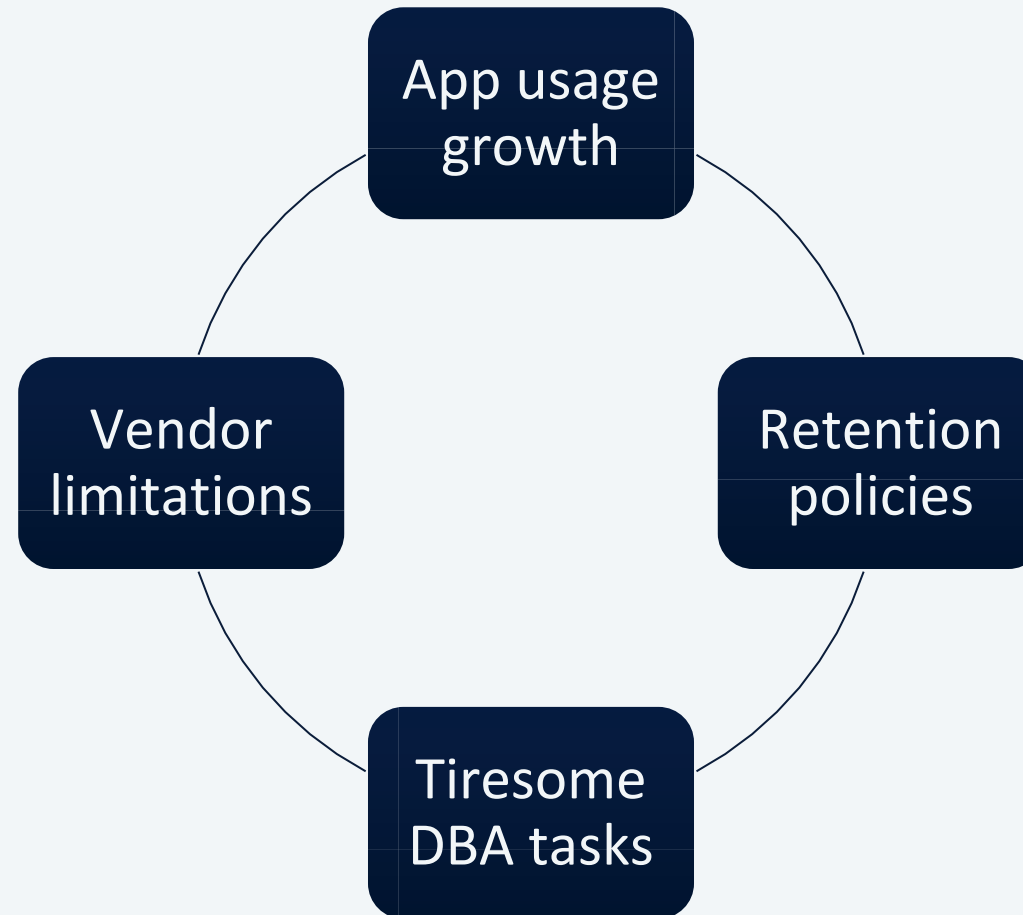    8.6 million emergency visits per year

**Enterprise Resource Planning (ERP) Environment:**

- 120+ databases and different swim lanes supporting ERP development and maintenance

- 1000+ Tablespaces, 2,800 Tables & 7,500 IX per DB

- Largest table has 1.5+ billion rows and 7 Indexes
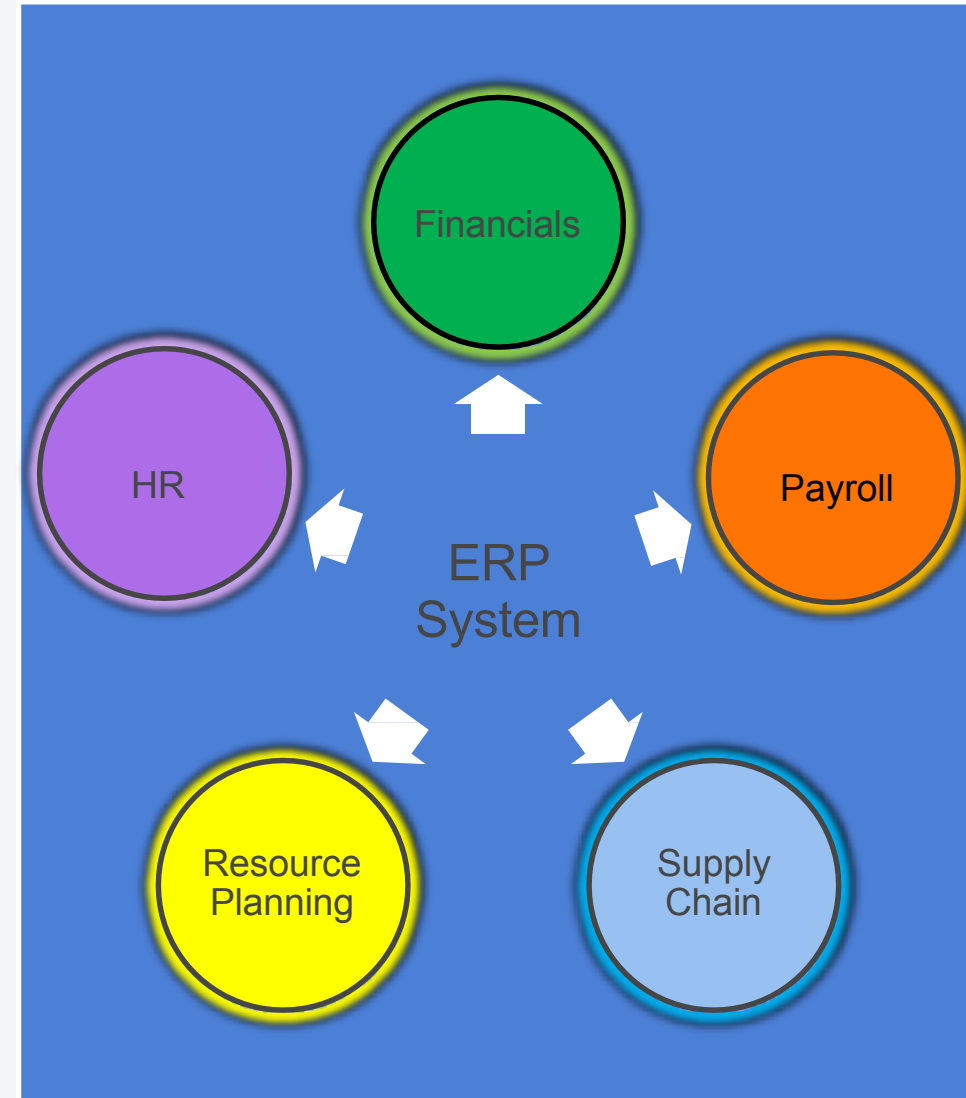
# Why we needed a data archive strategy

# What is Optim archive?

IBM InfoSphere Optim Data Growth Solution for z/OS provides everything you need to create and manage archives of relationally intact data from databases with any number of tables and relationships. Using the archiving features in Optim Data Growth Solution for z/OS, you can:

- Isolate historical data from current activity and safely remove it to a secure archive.
- Access archived historical data easily, using familiar tools and interfaces.
- Restore archived data to its original business context when it requires additional processing.
- Build repetitive process which can be executed whenever needed.
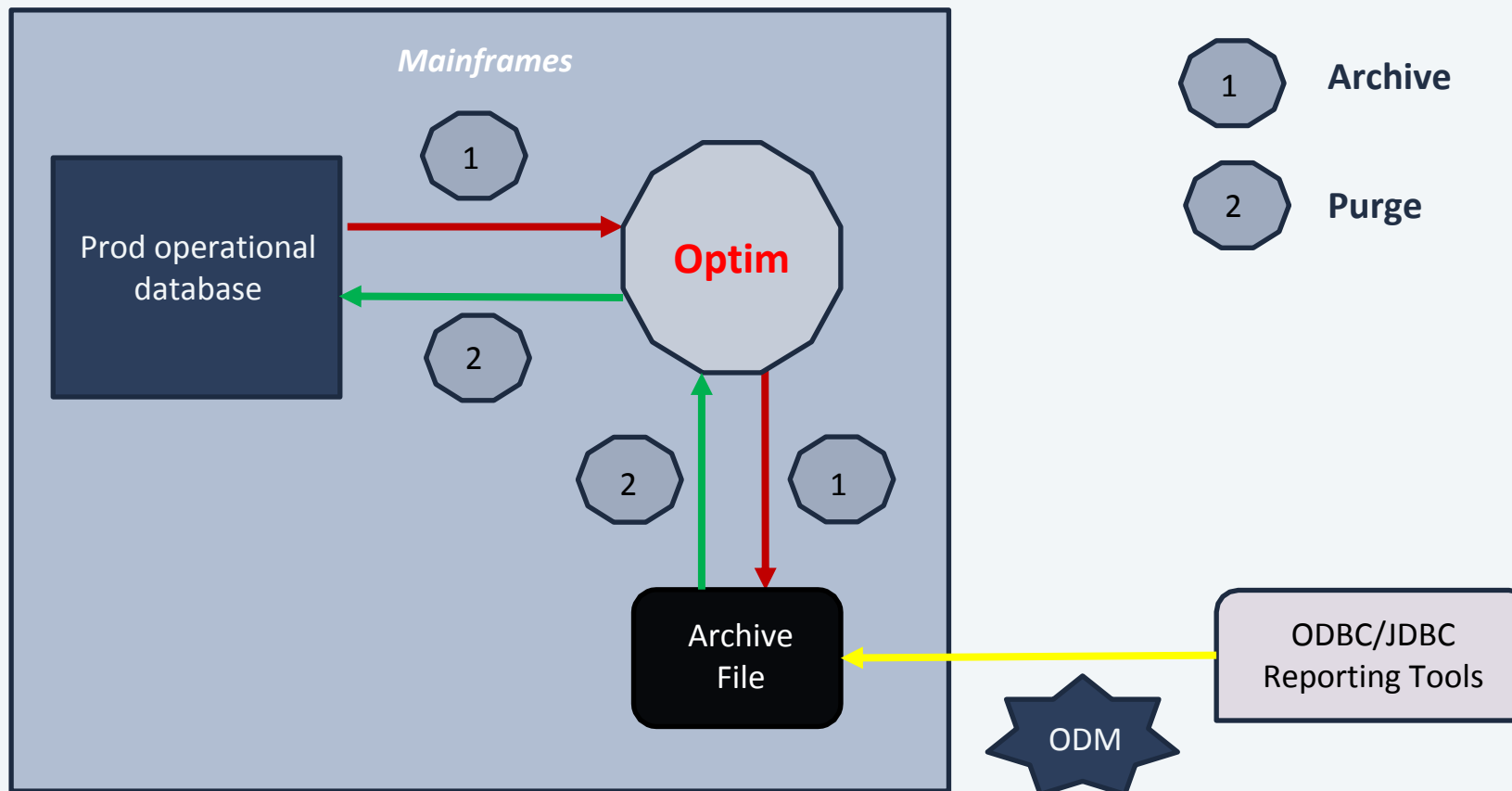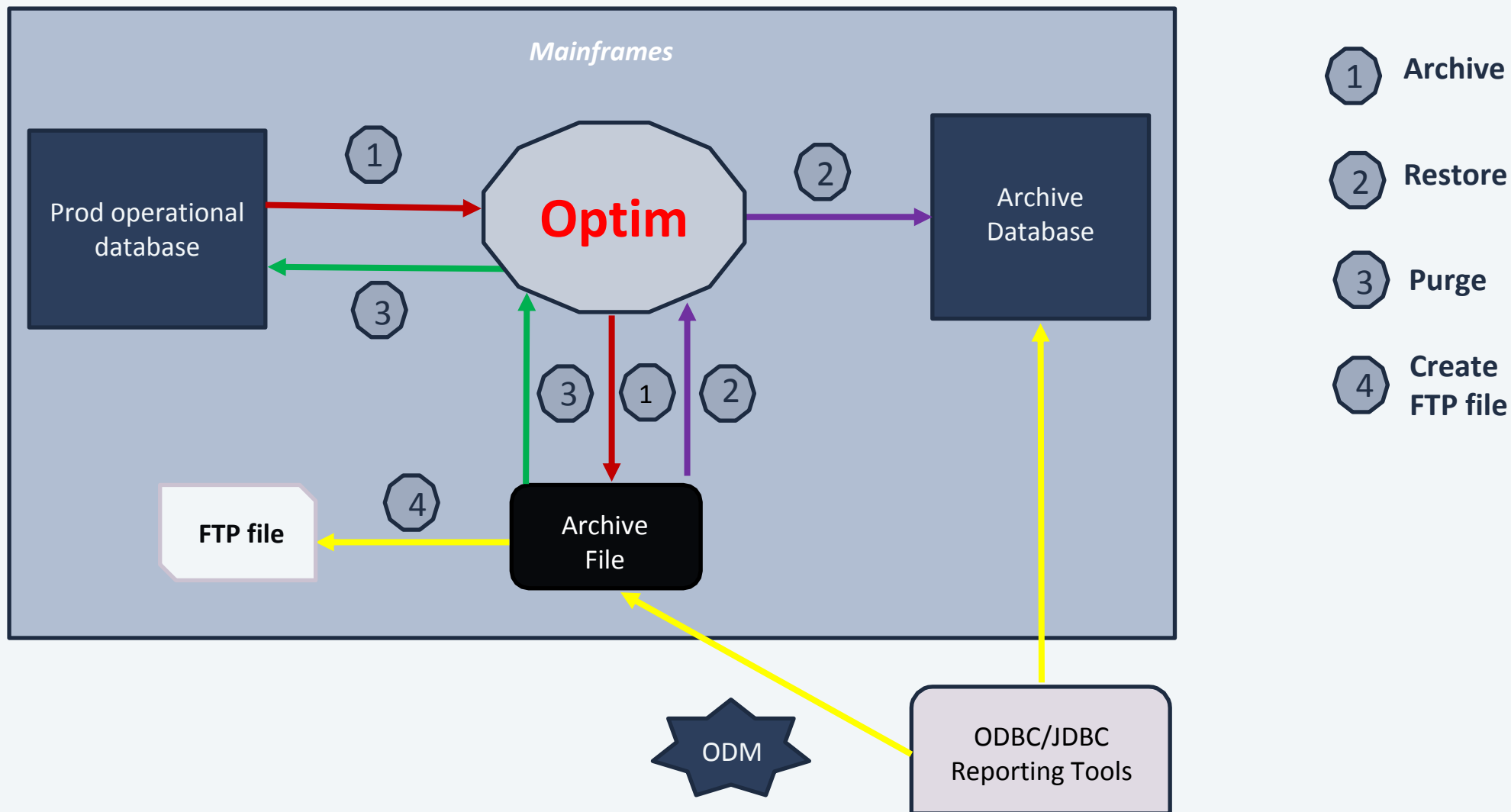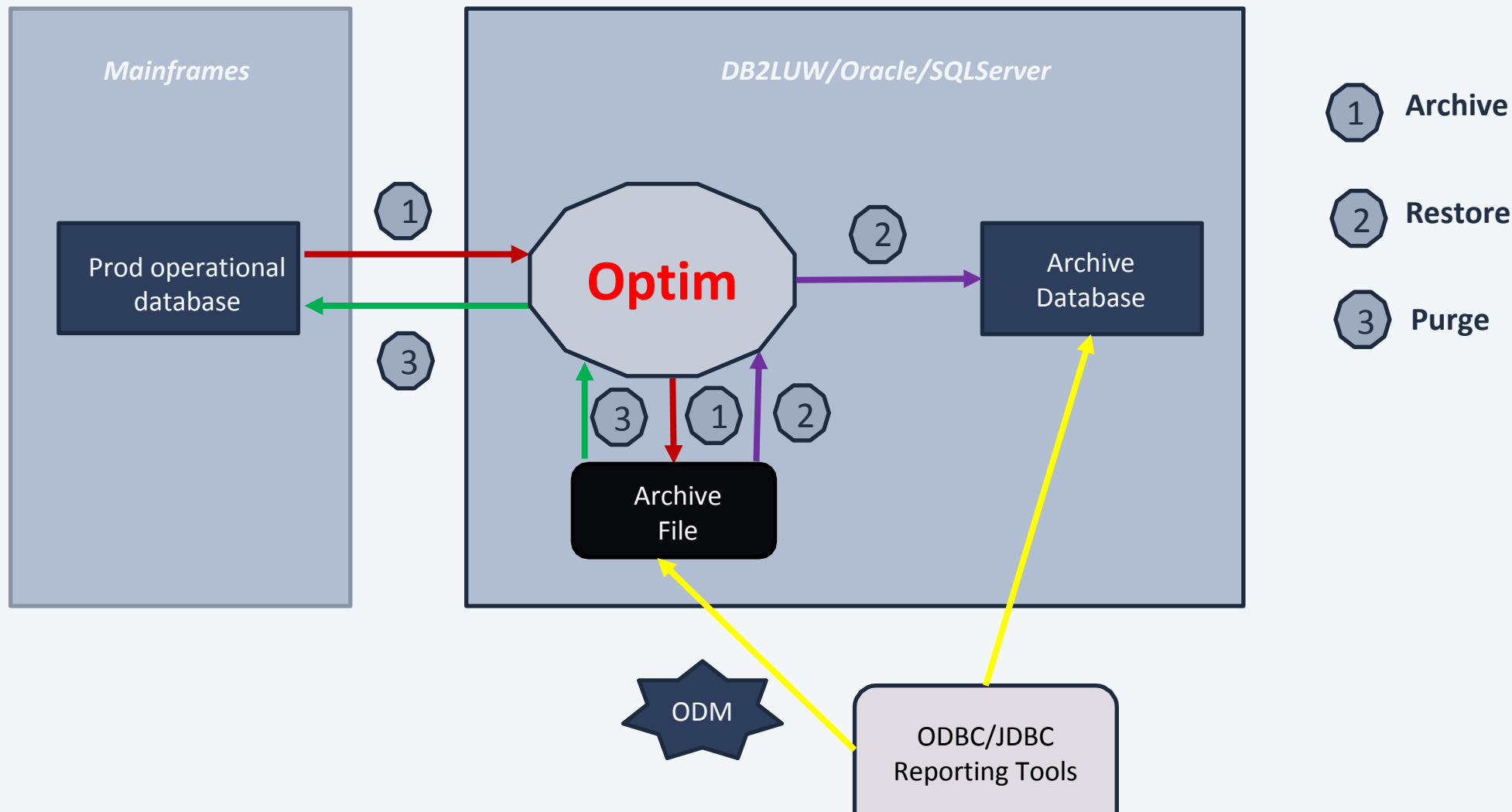
# How did we achieve data archive and purge?

Choosing the suitable archive
path

Build and execute JCLs

Creating Access Definitions (AD)
and relationships

# 1. Choosing the suitable archive path
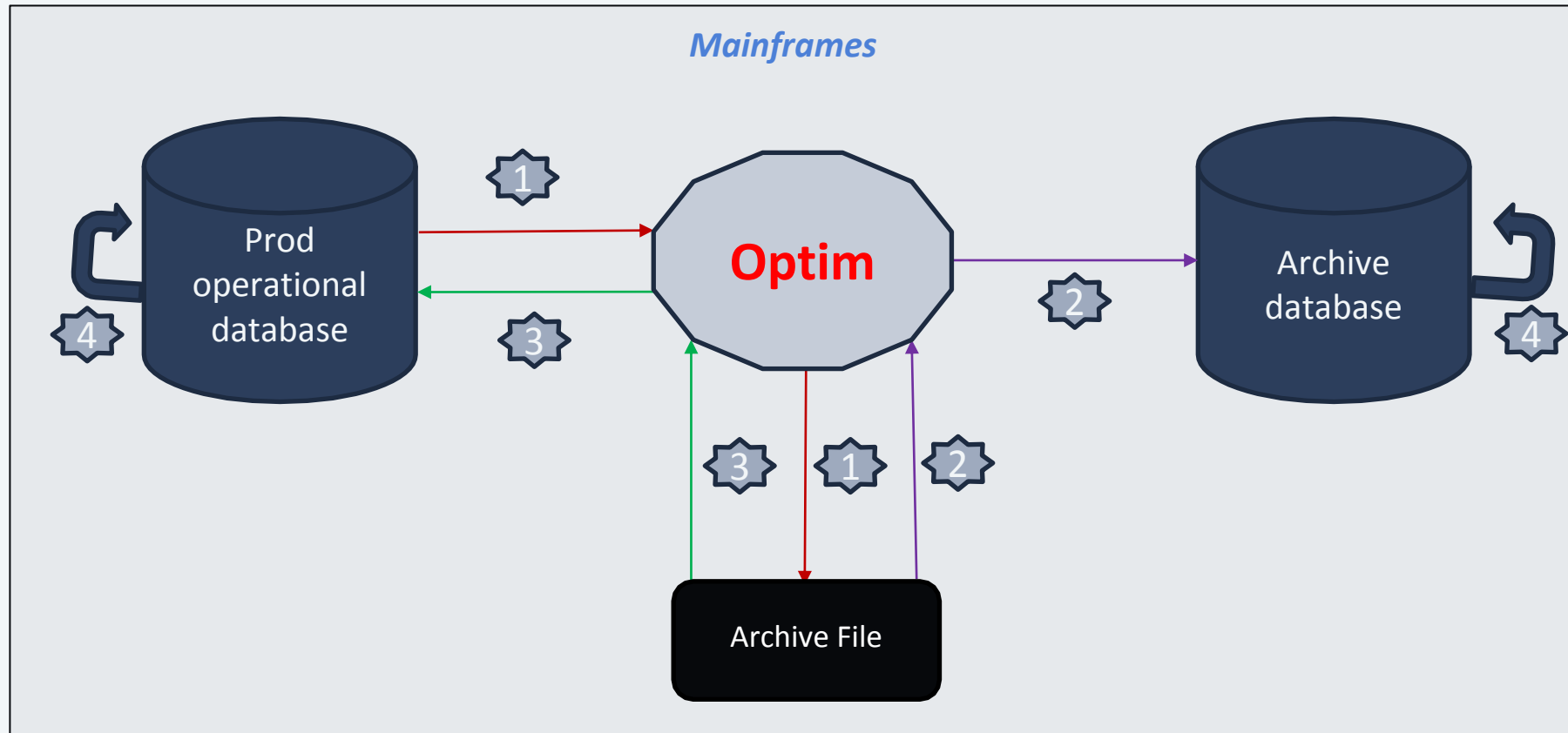
# 2. Creating Access Definitions (AD) and relationships

An Access Definition describes the data to be extracted from the source database.

The components of an Access Definition include the following:

- A list of tables from which the data is extracted.

- Selection criteria (WHERE clause in SQL query).

- The list of relationships to be traversed.

```
FOPRM000 ------------------------ IBM's InfoSphere Optim  ------------------------
OPTION   ===> 5
                                                              More:            +
0   OPTIONS          - Site and User Options      SQLID   ===> XXX1234
1   BROWSE TABLE     - Browse a DB2 Table         SUBSYS  ===> SSID
2   EDIT TABLE       - Edit a DB2 Table
3   BROWSE USING AD  - Browse DB2 Tables Using Access Definition
4   EDIT USING AD    - Edit DB2 Tables Using Access Definition
5   ADS              - Create or Modify Access Definitions
6   DEFINITIONS      - Maintain InfoSphere Optim Definitions (Keys, Maps, ...)
7   MIGRATION        - Data Migration - Extract, Insert, Update, ...
8   COMPARE          - Compare Two Sets of Data
9   ARCHIVE          - Archive and Restore Data


T   TUTORIAL         - Information About IBM's InfoSphere Optim
C   CHANGES          - Changes from Prior Release(s)
X   EXIT             - Terminate Product Use



5655-QCA (C) Copyright IBM Corporation 1989, 2014.
All rights reserved.  Licensed materials - property of IBM.
US Government Users Restricted Rights - Use, duplication or disclosure
restricted by GSA ADP schedule contract with IBM Corp.
```
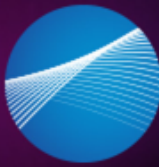
```
FOPDF001 ----------------- Choose an Access Definition ----------------

Command ===> _


                                           SQLID   ===> XXX1234
                                           SUBSYS ===> SSID
Access Definition:
   Group ===> GRPNAME
   User  ===> OPTIM1
   Name  ===> HR



Use "_" for DB2 LIKE character ===> NO    (Y-Yes, N-No)




To limit selection list to Access Definitions with certain start tables, enter
the start table name below.  A wild card is allowed at the end of each part.


   Start Table Creator ID ===>                                          >
   Start Table Name       ===>                                          >
```

```
FOPF0000   Tables/Views for AD:      .OPTIM1.HR1 ---------------------------
Command ===>                                        Scroll ===> CSR


  Primary : COL,SEL,SQL,REL,POINT,GROUP,GET TABLES RELATED,INDENT,ARC,LIST SUBS
     Line : COL,SEL,SQL,ALL,GR(A),GP(A),GC(A),DR(A),OR(A),DP(A),OP(A),
            DC(A),OC(A),EXP,ARC,ACT,STA

                                                    Table 1 of 2 <<MORE
Default Creator ID ===> CREATORID                                        >>
Start Table        ===> EMPLOYEE                                         >>
Start Table Options    : None

                                            R  D  --Extract Parms---
Cmd    Status      (CreatorID.)Table/View Name  F  A  EveryNth Row Limit   Type
---  ----------  --------------------------->> -  -  -------- ---------- ------
*** *********************************** TOP *******************************
             EMPLOYEE                             N  _____  _____      TABLE
SQL          HRHISTORY                         N  Y  _____  _____      TABLE
*** ************************** BOTTOM ************************************
```
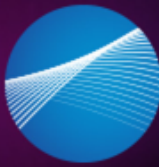
```
FOPRM000 ------------------------ IBM's InfoSphere Optim -------------------------

OPTION  ===> 6
                                                        More:         +

0   OPTIONS         - Site and User Options      SQLID  ===> XXX1234
1   BROWSE TABLE    - Browse a DB2 Table         SUBSYS ===> SSID
2   EDIT TABLE      - Edit a DB2 Table
3   BROWSE USING AD - Browse DB2 Tables Using Access Definition
4   EDIT USING AD   - Edit DB2 Tables Using Access Definition
5   ADS             - Create or Modify Access Definitions
6   DEFINITIONS     - Maintain InfoSphere Optim Definitions (Keys, Maps, ...)
7   MIGRATION       - Data Migration - Extract, Insert, Update, ...
8   COMPARE         - Compare Two Sets of Data
9   ARCHIVE         - Archive and Restore Data


T   TUTORIAL        - Information About IBM's InfoSphere Optim
C   CHANGES         - Changes from Prior Release(s)
X   EXIT            - Terminate Product Use
```

```
FOPAM003 ----------------- Choose a Definition Option ------------

OPTION  ===> 2

                                    SQLID  ===> XXX1234

1  PRIMARY KEYS    - Maintain Primary Keys      SUBSYS ===> SSID
2  RELATIONSHIPS   - Maintain Relationships
3  COLUMN MAPS     - Maintain Column Maps
4  TABLE MAPS      - Maintain Table Maps
5  ADS             - Maintain Access Definitions
6  COLLECTIONS     - Maintain Archive Collections
A  PROCEDURES      - Maintain Column Map Procedures


E  EXPORT          - Export Optim Object Definitions
I  IMPORT          - Import Optim Object Definitions
```

```
FOPE0002 ----------------- Choose a Relationship --------------------
OPTION ===> 1                                      SCROLL ===> CSR


   1  CREATE   - Create a Relationship for Specified Parent or Child Table
   2  MODIFY   - Modify a Relationship for Specified Child Table
   3  LIST     - List All Relationships for Specified Table


Specify Table Name (Child for OPTION 2, Parent or Child for OPTIONS 1 and 3)
   Creator ID ===> CREATORID                                          >>
   Table Name ===> EMPLOYEE                                           >>


Specify Relationship Name (OPTIONS 1 and 2)
   Relationship Name ===> EMP2HRHIST                                    >>


Specify Relationship Type (OPTIONS 2 and 3)
   Relationship Type ===> BOTH              (P|O-OPT, D-DB2, B-Both)


Use "_" for DB2 LIKE character ===> NO        (Y-Yes, N-No)
```

```
+----------------------Create a New Relationship----------------------+
|                                                                     |
|                                                                     |
| Specified Table   :CREATOR.EMPLOYEE                                 |
| Table Type     ===> P            (P-Parent, C-Child)                |
|                                                                     |
|                                                                     |
|                                                                     |
| Leave blank or include wild cards for Table Selection List          |
|                                                                     |
|                                                                     |
|                                                                     |
| Other Table:                                                        |
|   Creator ID ===> CREATORID                                    >>   |
|                                                                     |
|   Table Name ===> HRHISTORY                                    >>   |
|                                                                     |
|                                                                     |
| Relationship Name ===> EMP2HRHIST                              >>   |
|                                                                     |
+---------------------------------------------------------------------+
```

```
FOPF0000 ----------------- Define Relationship ---------------------

Command ===> _                                       Scroll ===> CSR


                 Define OPTIM Relationship EMP2HRHIST
           Special Commands: LIST COLUMNS, EXPAND, GENERIC, MODEL


     Parent:      .EMPLOYEE           Child:      .HRHISTORY


     Exit Name ===>
                                                              1 OF 2

Cmd      Column Name     Data Type     Column Name     Data Type
--- -------------------->> ---------- -------------------->> ----------
*** ************************* TOP *****************************
___ COMPANY             SMALLINT   COMPANY             SMALLINT
___ EMPLOYEE            INTEGER    EMPLOYEE            INTEGER
*** ************************ BOTTOM **************************
```

```
SELECT * FROM creator.EMPLOYEE;


SELECT *
FROM creator.HRHISTORY A
     INNER JOIN
  creator.EMPLOYEE B ON
      A.COMPANY = B.COMPANY AND
      A.EMPLOYEE = B.EMPLOYEE
WHERE YEAR(A.DATE_STAMP)<= archive_year;
```

# 3. Building and executing JCLs

```
FOPRM000 -------------------- IBM's InfoSphere Optim  --------------------
OPTION   ===> 9

                                                          More:        +
0  OPTIONS         - Site and User Options      SQLID   ===> XXX1234
1  BROWSE TABLE    - Browse a DB2 Table         SUBSYS  ===> SSID
2  EDIT TABLE      - Edit a DB2 Table
3  BROWSE USING AD - Browse DB2 Tables Using Access Definition
4  EDIT USING AD   - Edit DB2 Tables Using Access Definition
5  ADS             - Create or Modify Access Definitions
6  DEFINITIONS     - Maintain InfoSphere Optim Definitions (Keys, Maps, ...)
7  MIGRATION       - Data Migration - Extract, Insert, Update, ...
8  COMPARE         - Compare Two Sets of Data
9  ARCHIVE         - Archive and Restore Data


T  TUTORIAL        - Information About IBM's InfoSphere Optim
C  CHANGES         - Changes from Prior Release(s)
X  EXIT            - Terminate Product Use
```

```
FOPAR000 ---------------- Archive and Restore ----------------
OPTION  ===>

                                        SQLID  ===>  XXX1234
                                        SUBSYS ===>  SSID

1   ARCHIVE      - Archive Data
2   RESTORE      - Restore Data
3   DELETE       - Delete Archived Data from DB2 Tables
4   LIST         - List Archive Directory
5   IMPORT       - Import Archive File and Populate Directory
6   UPDATE       - Update Archive File Indexes
7   LOAD         - Create Load Files and Perform Load
8   SUBSET       - Create Subset of Existing Archive File
9   CREATE       - Create Tables and Related Object Definitions
A   CONVERT      - Convert Archive File using Table and Column Maps


R   RETRY/RESTART  - Retry/Restart Delete After Archive or Restore process
B   BROWSE         - Browse Content of Archive, Extract, or Control File
L   LOG            - Browse Archive Log
```

```
FOPF0001 ----------- Specify ARCHIVE Parameters and Execute ----------------
Command ===> _


Current AD Name            :        .OPTIM1.HR
Archive File DSN        ===>  'DSNDBA.OPTIM.ABCDEF.HR.YR2012.ARCH'
Control File DSN        ===>  'DSNDBA.OPTIM.ABCDEF.HR.YR2012.CNTL'
Archive Group           ===>                    (Any 8 Character Designation)
Archive Description     ===>
Archive Collection      ===>                    (*=Selection List)


Create Duplicate Archive File    ===> N         (Y-Yes, N-No)
Limit Number of Archive Rows     ===>           (1-4294967295, Blank-Site Max)
Archive Data to Tape             ===> N         (Y-Yes, N-No)
Defer Delete After Archive       ===> Y         (Y-Yes, N-No)


Run Archive in Batch or Online   ===> B         (B-Batch, O-Online)
  If Batch, Review or Save JCL   ===> R         (N-No, R-Review, S-Save)


Process Report Type              ===> S         (D-Detailed, S-Summary)
```

```
FOPE0002 ------------- Delete Rows From Archive Process ------------------
OPTION ===> _                                          SCROLL ===> CSR



Specify Data Set Names for Archive and Control Files
   Archive File DSN   ===> 'DSNDBA.OPTIM.ABCDEF.HR.YR2012.ARCH'
   Control File DSN   ===> 'DSNDBA.OPTIM.LPAHCA.HR.YR2012.CNTL'


Process Options:
   Lock Tables During Delete          : N            (Y-Yes, N-No)
   Commit Frequency Rate          ===>               (1-10000, Blank/SL)
   Limit Number of Discarded Rows ===>               (1-4294967295, Blank/NL)
   Compare Row Contents           ===> Y             (Y-Yes, N-No)


   Review ACM and Key Lookup Limit ===> N            (Y-Yes, N-No)
Run Process in Batch or Online    ===> B             (B-Batch, O-Online)
   If Batch, Review or Save JCL?   ===> R            (N-No, R-Review, S-Save)
```
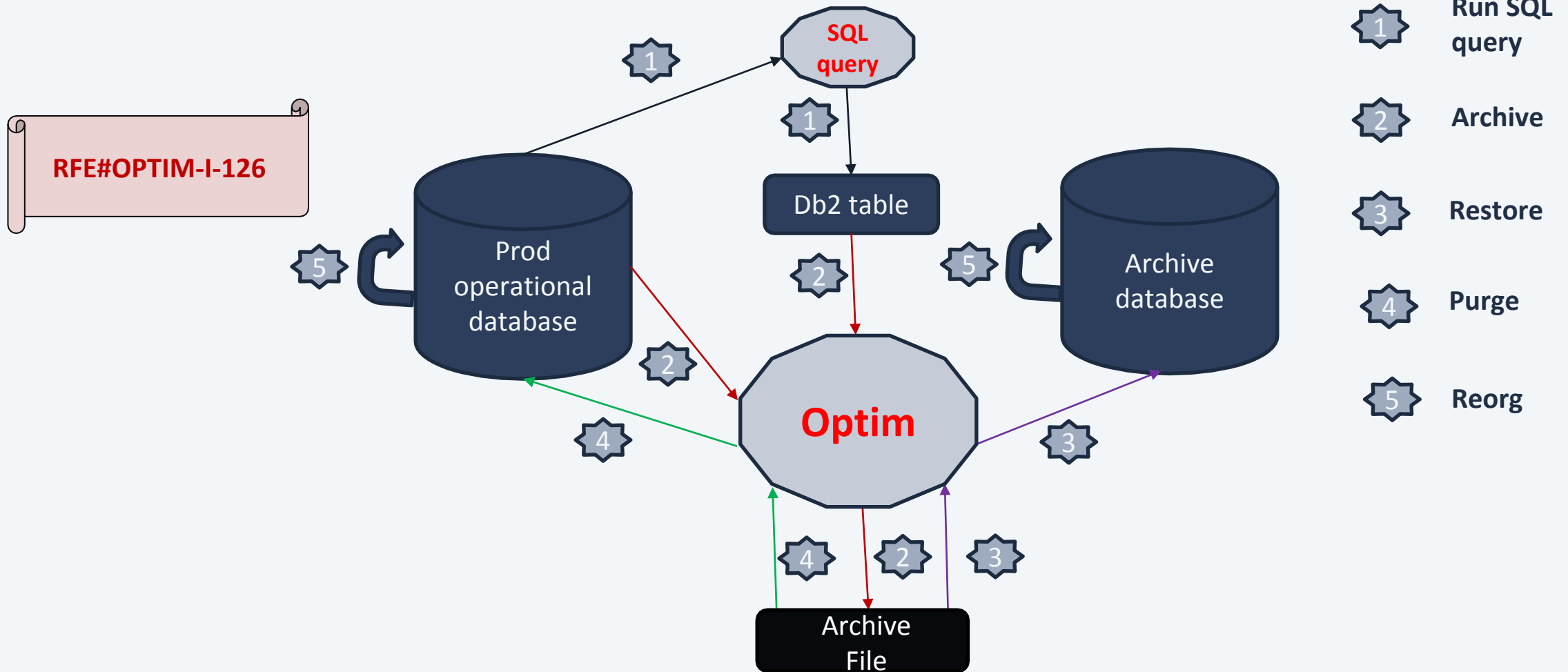
# Just some stats to wow you!!!

| Group | Archive year | Month | Rows archived | Archive Elap time | Archive CPU time | Rows purged | Purge Elap time | Purge CPU time | Commit Frequency |
|---|---|---|---|---|---|---|---|---|---|
| HR | 2006 | Jan | 1,153,820 | 3.0 mins | 1.01 mins | 707,388 | 7.3 mins | 1.14 mins | 10,000 |
| HR | 2006 | Feb | 1,464,290 | 18.4 mins | 5.19 mins | 618,741 | 5.8 mins | 0.91 mins | 10,000 |
| HR | 2006 | Mar | 1,132,256 | 6.7 mins | 2.09 mins | 686,489 | 5.4 mins | 0.96 mins | 10,000 |
| HR | 2006 | Apr | 1,253,081 | 6.1 mins | 2.02 mins | 777,012 | 5.8 mins | 1.06 mins | 10,000 |
| HR | 2006 | May | 3,088,530 | 15.3 mins | 3.51 mins | 824,106 | 9.9 mins | 1.53 mins | 10,000 |
| HR | 2006 | Jun | 1,460,621 | 7.4 mins | 2.39 mins | 921,156 | 7.8 mins | 1.50 mins | 10,000 |
| HR | 2006 | Jul | 1,253,396 | 6.9 mins | 2.31 mins | 784,265 | 7.0 mins | 1.30 mins | 10,000 |
| HR | 2006 | Aug | 1,330,779 | 6.6 mins | 2.30 mins | 832,924 | 8.2 mins | 1.45 mins | 10,000 |
| HR | 2006 | Sep | 5,040,853 | 14.7 mins | 4.02 mins | 822,098 | 10.2 mins | 1.53 mins | 10,000 |
| HR | 2006 | Oct | 1,717,699 | 6.6 mins | 2.23 mins | 1,136,446 | 10.1 mins | 1.82 mins | 10,000 |
| HR | 2006 | Nov | 1,294,788 | 6.0 mins | 2.09 mins | 796,753 | 8.8 mins | 1.37 mins | 10,000 |
| HR | 2006 | Dec | 1,600,759 | 5.8 mins | 2.17 mins | 1,031,153 | 11.1 mins | 1.78 mins | 10,000 |
| HR | 2007 | Jan | 1,393,391 | 16.0 mins | 2.76 mins | 685,948 | 6.7 mins | 1.04 mins | 10,000 |
| HR | 2007 | Feb | 1,343,581 | 15.2 mins | 2.59 mins | 687,034 | 5.3 mins | 0.94 mins | 10,000 |
| HR | 2007 | Mar | 1,488,544 | 14.2 mins | 2.65 mins | 763,036 | 5.8 mins | 1.15 mins | 10,000 |
| HR | 2007 | Apr | 1,466,716 | 13.9 mins | 2.62 mins | 724,713 | 5.4 mins | 1.07 mins | 10,000 |
| HR | 2007 | May | 4,341,217 | 18.0 mins | 4.84 mins | 746,895 | 8.7 mins | 1.53 mins | 10,000 |
| HR | 2007 | Jun | 1,621,649 | 16.7 mins | 4.04 mins | 792,587 | 7.6 mins | 1.47 mins | 10,000 |
| HR | 2007 | Jul | 1,514,917 | 9.9 mins | 2.81 mins | 769,469 | 8.9 mins | 1.27 mins | 10,000 |
| HR | 2007 | Aug | 1,581,250 | 16.1 mins | 3.91 mins | 795,094 | 8.4 mins | 1.29 mins | 10,000 |
| HR | 2007 | Sep | 7,410,698 | 20.9 mins | 5.51 mins | 762,722 | 7.5 mins | 1.47 mins | 10,000 |
| HR | 2007 | Oct | 1,635,793 | 17.3 mins | 3.83 mins | 776,331 | 7.0 mins | 1.46 mins | 10,000 |
| HR | 2007 | Nov | 1,459,097 | 9.0 mins | 2.79 mins | 705,562 | 6.6 mins | 1.35 mins | 10,000 |
| HR | 2007 | Dec | 1,981,865 | 17.2 mins | 3.94 mins | 1,018,550 | 7.1 mins | 1.78 mins | 10,000 |

# Limitations in using Optim archive tool



RFE#OPTIM-I-126

SQL query

Db2 table

Prod operational database

Optim

Archive database

Archive File

1. Run SQL query
2. Archive
3. Restore
4. Purge
5. Reorg

# Special Thanks to

*Greg Czaja (greg.czaja@unicomsi.com)*