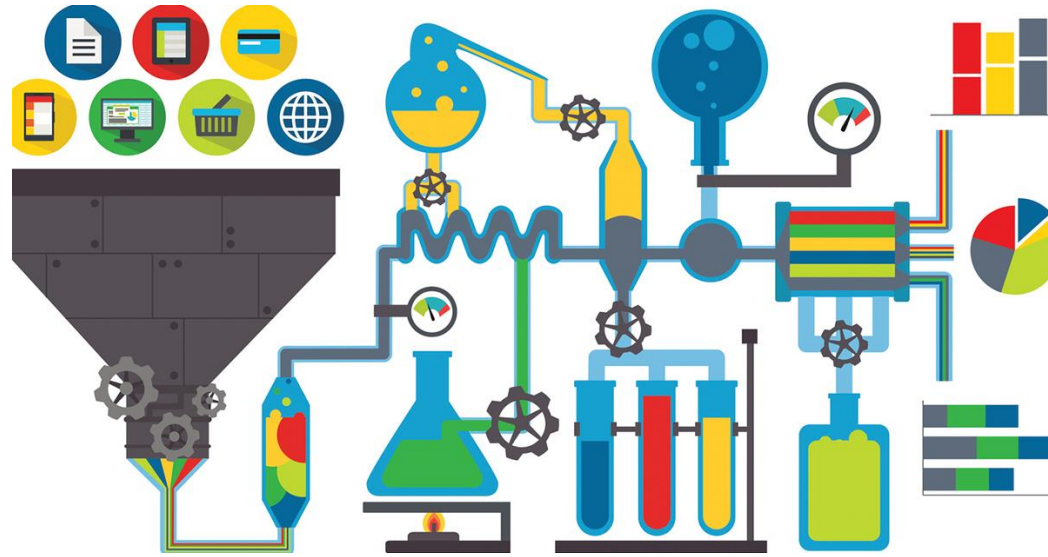
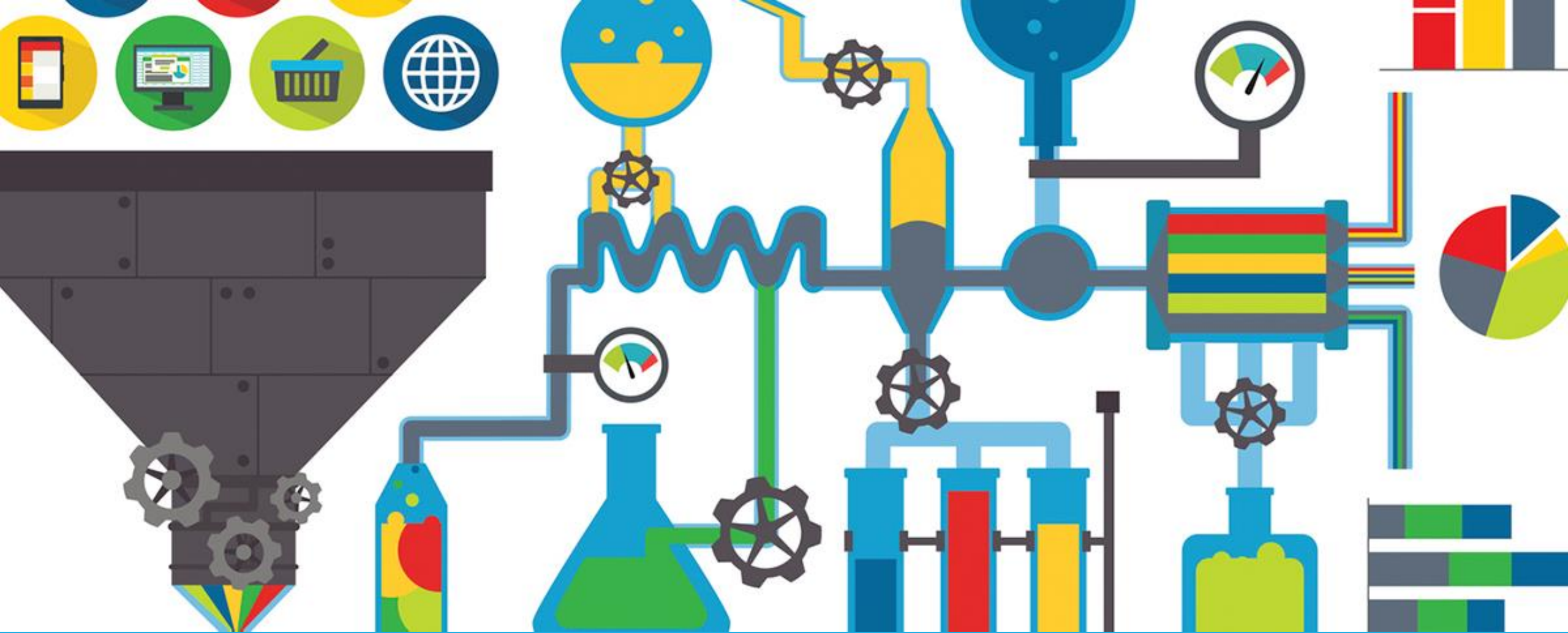




Data Warehouse and Data Analysis Trend in the ICT Industry

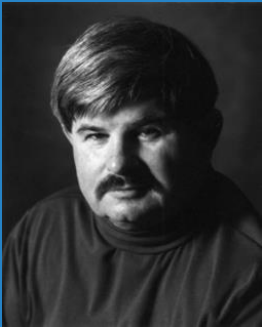
Case Study: CIF Monitoring Bank BPRKS Bandung

Daud Rusyad Nurdin
David Andriansyah

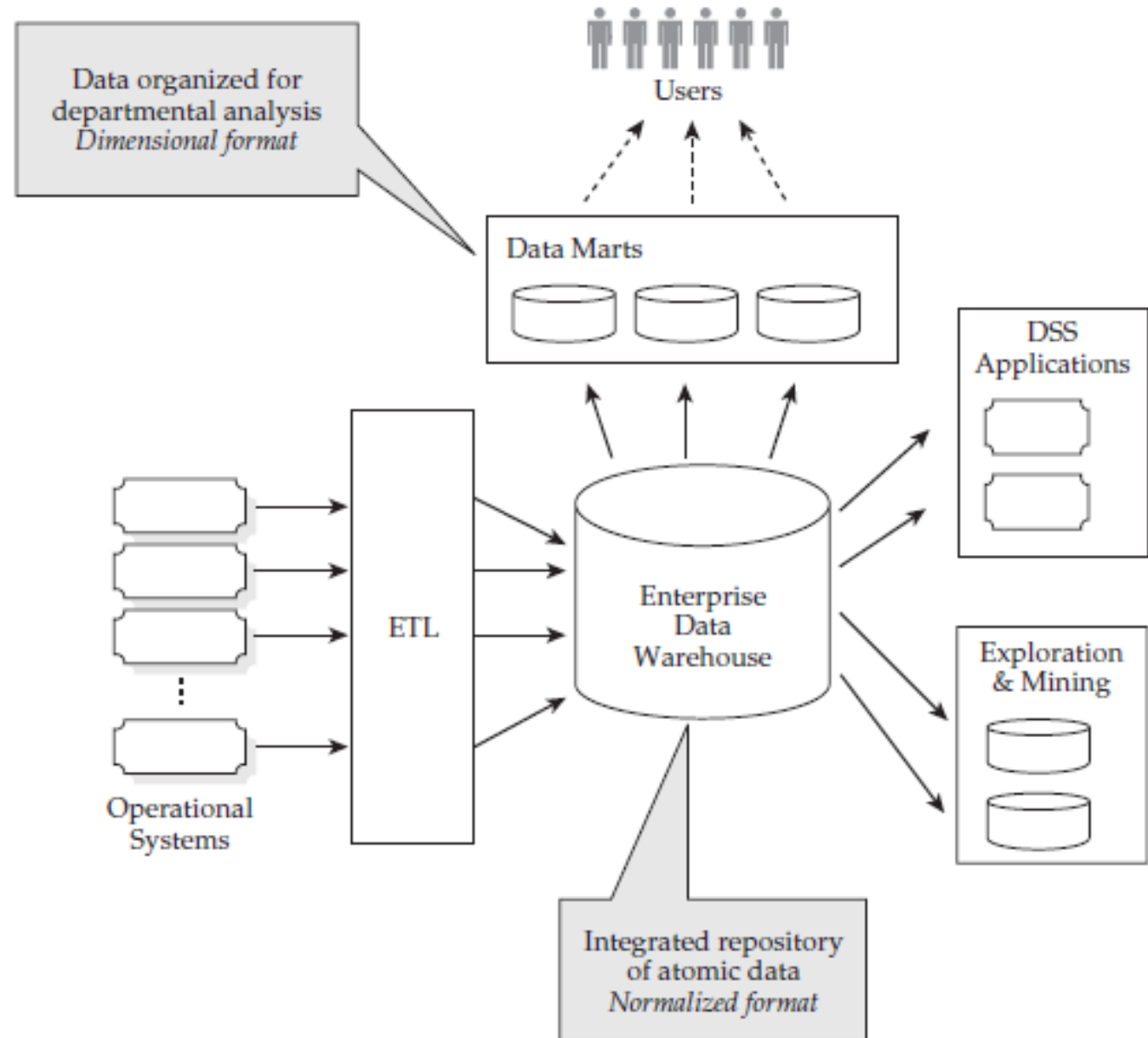


Data Warehouse Architecture

Data Warehouse Architecture Inmon Corporate Information Factory (CIF)



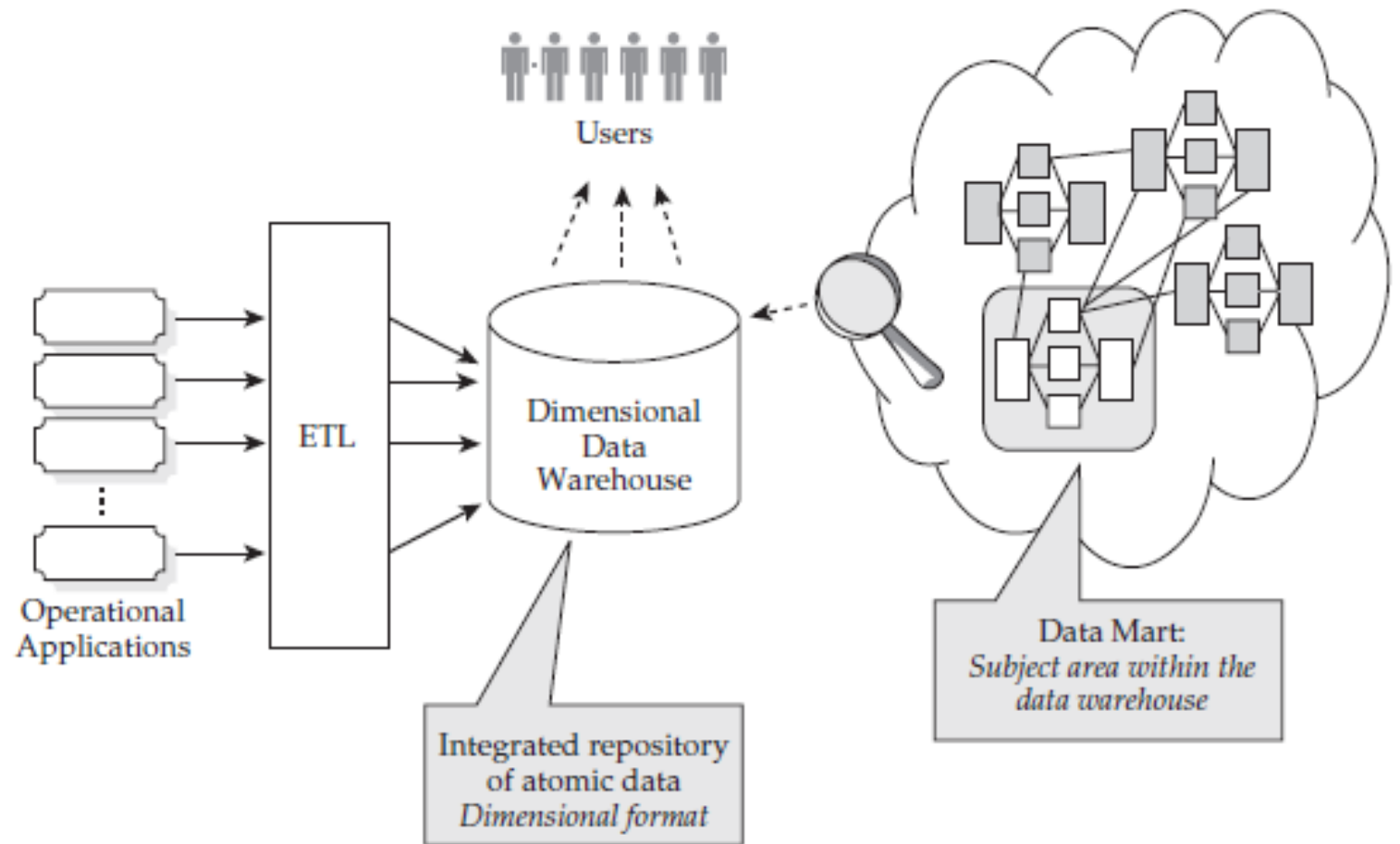
William H. Inmon



Data Warehouse Architecture Kimball – Star Schema

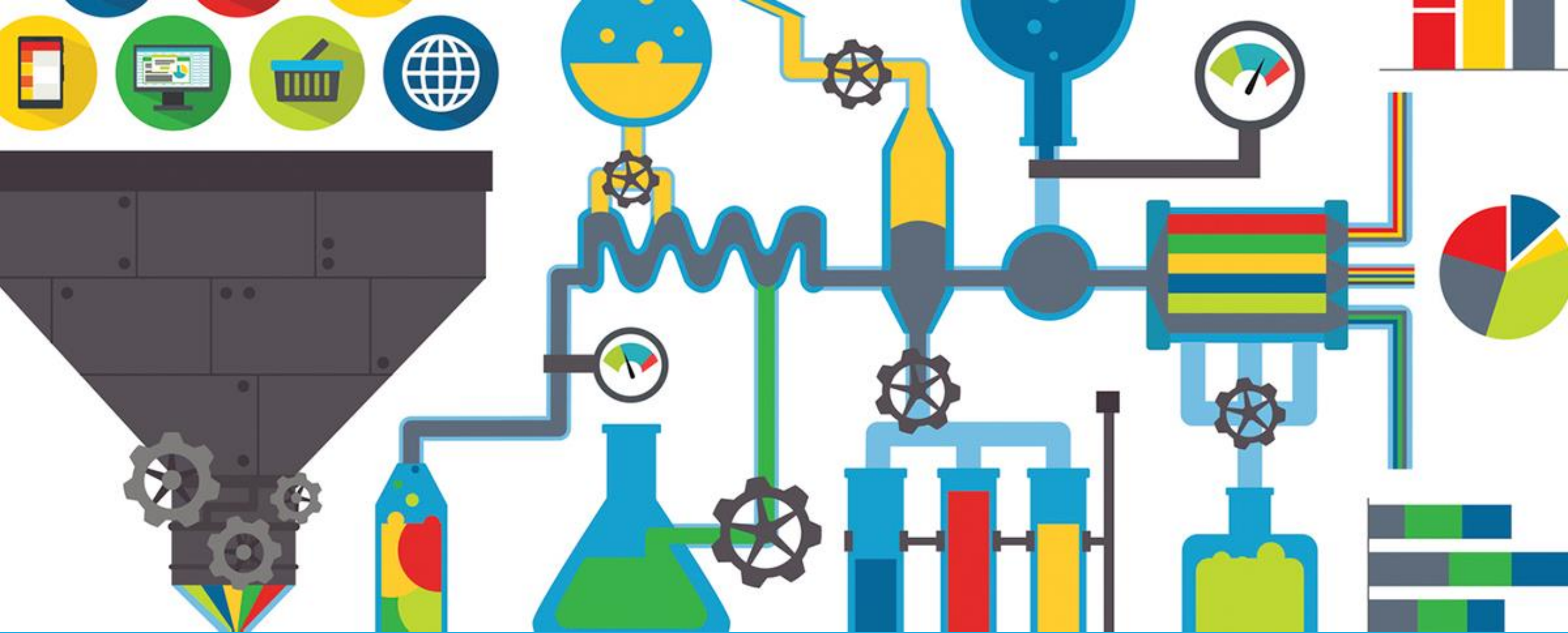


Ralph Kimball



Comparison of Data Warehouse Architectures

Architecture	Advocate	Also Known As	Description	Role of Dimensional Design
Corporate Information Factory (CIF)	William H. Inmon	- Atomic Data Warehouse - Enterprise Data Warehouse	- Enterprise data warehouse is an integrated repository of atomic data - It is not accessed directly - Data marts reorganize data for departmental use/analysis	Dimensional design used for data marts only
Multidimensional Modeling/ Star Schema	Ralph Kimball	- Enterprise Data Warehouse - Bus architecture - Architected data marts - Virtual data marts	- Dimensional data warehouse is an integrated repository of atomic data - Subject areas within the dimensional data warehouse sometimes called data marts - Data marts not required to be separate databases	All data organized dimensionally



Data Analysis

To identify anomaly data for Data Cleansing

Lost “Primary Key”

```
SELECT *  
FROM MANIFEST_ITEM_SRAH@db_admail  
WHERE TRIM(KDITEM) IS NULL
```

Grid

Data Grid DBMS Output (disabled)

KDITEM	KDSTATUS	KDMI	BRT_ITEM_SRH	WKTLOKAL	WKTSRV	KDNOPENT
1100			0	1/25/2013 3:39:52 PM	1/28/2013 4:43:44 PM	52200
1100		13000~110~13000~120~13~000024	10890	4/2/2013 4:25:41 PM	4/4/2013 1:12:45 PM	13000
1100		24300~110~24392~110~13~000001	0	4/17/2013 9:41:05 AM	4/17/2013 10:19:59 AM	24392
1100		40125A~110~40400~120~12~000002	0	12/28/2012 2:45:50 PM	12/29/2012 7:42:17 AM	40400
1100		401618~110~40400~120~13~000012	74	1/17/2013 1:10:01 PM	1/17/2013 1:57:14 PM	40400
1100		40383~110~40383~120~13~000001	0	1/7/2013 2:13:37 PM	1/8/2013 7:56:44 AM	40400
1100		40511A~110~40500~120~13~000028	8500	2/18/2013 2:32:59 PM	2/19/2013 2:11:19 PM	40500
1100		41263~110~41263~120~13~000005	200	1/11/2013 11:25:51 AM	1/11/2013 11:28:35 AM	41263
1100		41371~110~41300~140~13~000023	0	3/21/2013 12:52:32 PM	3/21/2013 3:27:06 PM	41300
1100		42100~110~42100~120~13~000022	0	4/10/2013 3:26:47 PM	4/11/2013 5:02:51 PM	42100
1100		43100~110~43100~140~13~000032	0	3/4/2013 12:37:51 PM	3/4/2013 6:34:35 PM	43100
1100		43200~110~43200~120~13~000094	0	2/25/2013 5:43:06 PM	2/25/2013 5:44:35 PM	43200
1100		43253~110~43255~140~13~000007	10	1/31/2013 9:34:11 AM	1/31/2013 12:52:28 PM	43255
1100		43266~110~43200~120~13~000002	0	2/23/2013 11:57:26 AM	2/25/2013 11:57:47 AM	43200
1100		44100~110~40400~120~13~000011	0	2/11/2013 3:06:07 PM	2/11/2013 4:01:20 PM	40400

2 Row 1 of 65 total rows STAGING1@10.33.41.176:1521/DWHPOS Modified

“Strange” Data

The screenshot shows a database query tool interface. At the top, a SQL query is entered in a text area:

```
SELECT *  
FROM MANIFEST_ITEM_SERAH@db_admail A  
WHERE A.BRT_ITEM_SRH=0  
;
```

Below the query is a toolbar with various icons and a 'Cancel' button. The main area displays a data grid with the following columns: KDITEM, KDSTATUS, KDMI, BRT_ITEM_SRH, WKTLOKAL, WKTSRV, and KDNOPENT. A red dashed box highlights the BRT_ITEM_SRH column.

KDITEM	KDSTATUS	KDMI	BRT_ITEM_SRH	WKTLOKAL	WKTSRV	KDNOPENT
	1100		0	1/25/2013 3:39:52 PM	1/28/2013 4:43:44 PM	52200
	1100	24300~110~24392~110~13~000001	0	4/17/2013 9:41:05 AM	4/17/2013 10:19:59 AM	24392
	1100	40125A~110~40400~120~12~000002	0	12/28/2012 2:45:50 PM	12/29/2012 7:42:17 AM	40400
	1100	40383~110~40383~120~13~000001	0	1/7/2013 2:13:37 PM	1/8/2013 7:56:44 AM	40400
	1100	41371~110~41300~140~13~000023	0	3/21/2013 12:52:32 PM	3/21/2013 3:27:06 PM	41300
	1100	42100~110~42100~120~13~000022	0	4/10/2013 3:26:47 PM	4/11/2013 5:02:51 PM	42100
	1100	43100~110~43100~140~13~000032	0	3/4/2013 12:37:51 PM	3/4/2013 6:34:35 PM	43100
	1100	43200~110~43200~120~13~000094	0	2/25/2013 5:43:06 PM	2/25/2013 5:44:35 PM	43200
	1100	43266~110~43200~120~13~000002	0	2/23/2013 11:57:26 AM	2/25/2013 11:57:47 AM	43200
	1100	44100~110~40400~120~13~000011	0	2/11/2013 3:06:07 PM	2/11/2013 4:01:20 PM	40400
	1100	45471~110~45400~120~13~000006	0	1/22/2013 12:06:42 PM	1/22/2013 12:08:49 PM	45400

“Strange” Data

```
SELECT *  
FROM MANIFEST_ITEM_SERAH@db_admail  
WHERE CK_STR(TRIM(KDITEM))='INVALID'  
;
```

Grid

Data Grid | DBMS Output (disabled)

Navigation icons: First, Previous, Next, Last, Find, Filter, Refresh, Cancel

KDITEM	KDSTATUS	KDMI	BRT_ITEM_SRH	WKTLOKAL	WKTSRV	KDNOPE
!@(%^@&*(^#	1200	40400~120~40400~140~13~000069	0	2/26/2013 7:52:28 PM	2/26/2013 8:29:27 PM	40400
!@(%^@(\$#%@	1200	40400~120~40400~140~13~000069	0	2/26/2013 7:52:16 PM	2/26/2013 8:29:27 PM	40400
!@((*@%(%**	1200	40400~120~40400~140~13~000069	0	2/26/2013 7:48:14 PM	2/26/2013 8:29:26 PM	40400
!@((*@%(&)@	1200	40400~120~40400~140~13~000069	0	2/26/2013 7:48:17 PM	2/26/2013 8:29:26 PM	40400
!@((*@%(*&&	1200	40400~120~40400~140~13~000069	0	2/26/2013 7:50:16 PM	2/26/2013 8:29:26 PM	40400
!@((*@%(^\$)	1200	40400~120~40400~140~13~000069	0	2/26/2013 7:48:07 PM	2/26/2013 8:29:26 PM	40400
!@((*@%(^%#	1200	40400~120~40400~140~13~000069	0	2/26/2013 7:48:08 PM	2/26/2013 8:29:26 PM	40400
!@((*@%(^&(&	1200	40400~120~40400~140~13~000069	0	2/26/2013 7:48:10 PM	2/26/2013 8:29:26 PM	40400
!@((*@%(^)*	1200	40400~120~40400~140~13~000069	0	2/26/2013 7:48:12 PM	2/26/2013 8:29:26 PM	40400
!@((*@%(^@\$	1200	40400~120~40400~140~13~000069	0	2/26/2013 7:48:11 PM	2/26/2013 8:29:26 PM	40400

Out of range Data

```
SELECT TO_CHAR(WKTLOKAL, 'RRRR-MM') YRMON, COUNT(1)
FROM MANIFEST_ITEM_SERAH@db_admail A
WHERE A.BRT_ITEM_SRH=0
GROUP BY TO_CHAR(WKTLOKAL, 'RRRR-MM')
ORDER BY 1
```

YRMON	COUNT(1)
2008-01	33
2009-03	268
2012-08	6016
2012-09	20580
2012-10	19973
2012-11	20373
2012-12	239491
2013-01	2353606
2013-02	3602324
2013-03	4446627
2013-04	5108900
2013-12	2

Untuk WKTLOKAL masih saja mengandung tanggal yang janggal, misalnya dari data di atas terdapat data '2013-12' (Desember 2013), padahal seharusnya belum ada transaksi kiriman di bulan tersebut. Atau juga data transaksi yang dicatat sebagai tahun '2008' dan '2009'.

Inconsistent Data

```
SELECT SUBSTR(A.KDMI, INSTR(A.KDMI, '~', 1, 2)+1, INSTR(A.KDMI, '~', 1, 3)-INSTR(A.KDMI, '~', 1, 2)-1) KDNOPENTX  
      , SUBSTR(A.KDMI, INSTR(A.KDMI, '~', 1, 3)+1, INSTR(A.KDMI, '~', 1, 4)-INSTR(A.KDMI, '~', 1, 3)-1) KDTRX  
      , A.*  
FROM MANIFEST_ITEM_SERAH@db_admail A  
WHERE TRIM(A.KDMI) IS NOT NULL  
AND LENGTH(A.KDNOPENT)>5  
AND SUBSTR(A.KDMI, INSTR(A.KDMI, '~', 1, 2)+1, INSTR(A.KDMI, '~', 1, 3)-INSTR(A.KDMI, '~', 1, 2)-1)<>A.KDNOPENT  
;
```

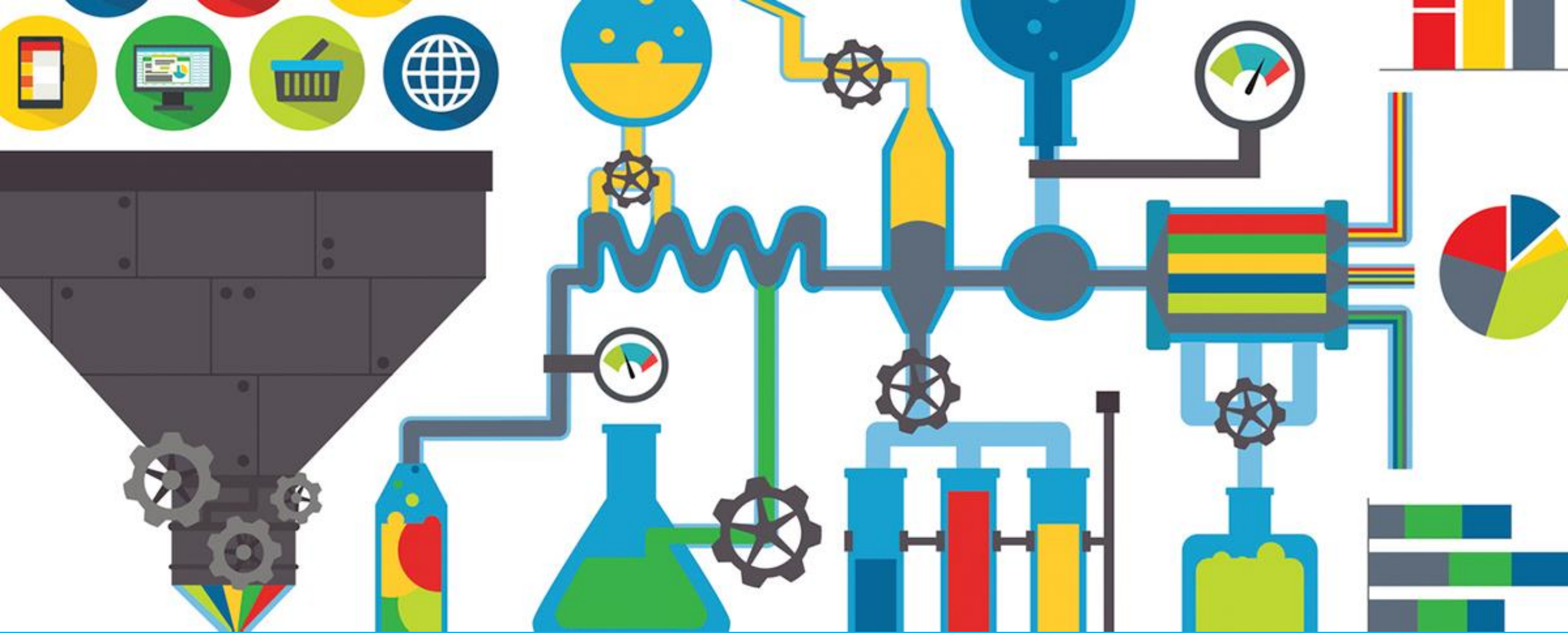
KDNOPENTX	KDTRX	KDITEM	KDSTATUS	KDMI	BRT_ITEM_SRH	WKTLOKAL	WKTSRV	KDNOPENT
75300	140	12530098349	1200	75300~110~75300~140~13~000001	0	4/1/2013 10:38:06 AM	4/1/2013 9:41:18 AM	75324A
75300	140	12674303505	1200	75300~110~75300~140~13~000001	0	4/1/2013 10:38:36 AM	4/1/2013 9:41:19 AM	75324A
40565	140	12733563454	1100	40500~110~40565~140~12~000001	35	12/29/2012 4:39:04 PM	12/29/2012 5:03:08 PM	405C101
40565	140	12733616703	1100	40500~110~40565~140~12~000001	60	12/29/2012 4:39:04 PM	12/29/2012 5:03:08 PM	405C101
40565	140	12733616716	1100	40500~110~40565~140~12~000001	60	12/29/2012 4:39:04 PM	12/29/2012 5:03:08 PM	405C101
40565	140	12733844130	1100	40500~110~40565~140~12~000001	10	12/29/2012 4:39:04 PM	12/29/2012 5:03:08 PM	405C101
40565	140	12733844234	1100	40500~110~40565~140~12~000001	40	12/29/2012 4:39:04 PM	12/29/2012 5:03:08 PM	405C101
40565	140	12733844510	1100	40500~110~40565~140~12~000001	15	12/29/2012 4:39:04 PM	12/29/2012 5:03:08 PM	405C101
40565	140	12733844523	1100	40500~110~40565~140~12~000001	15	12/29/2012 4:39:04 PM	12/29/2012 5:03:08 PM	405C101
40565	140	12733844536	1100	40500~110~40565~140~12~000001	15	12/29/2012 4:39:04 PM	12/29/2012 5:03:08 PM	405C101
40565	140	12733844549	1100	40500~110~40565~140~12~000001	15	12/29/2012 4:39:04 PM	12/29/2012 5:03:08 PM	405C101

2 Row 1 of 42 total rows STAGING1@10.33.41.176:1521/DWHPOS Modified

Date data in String Data Type

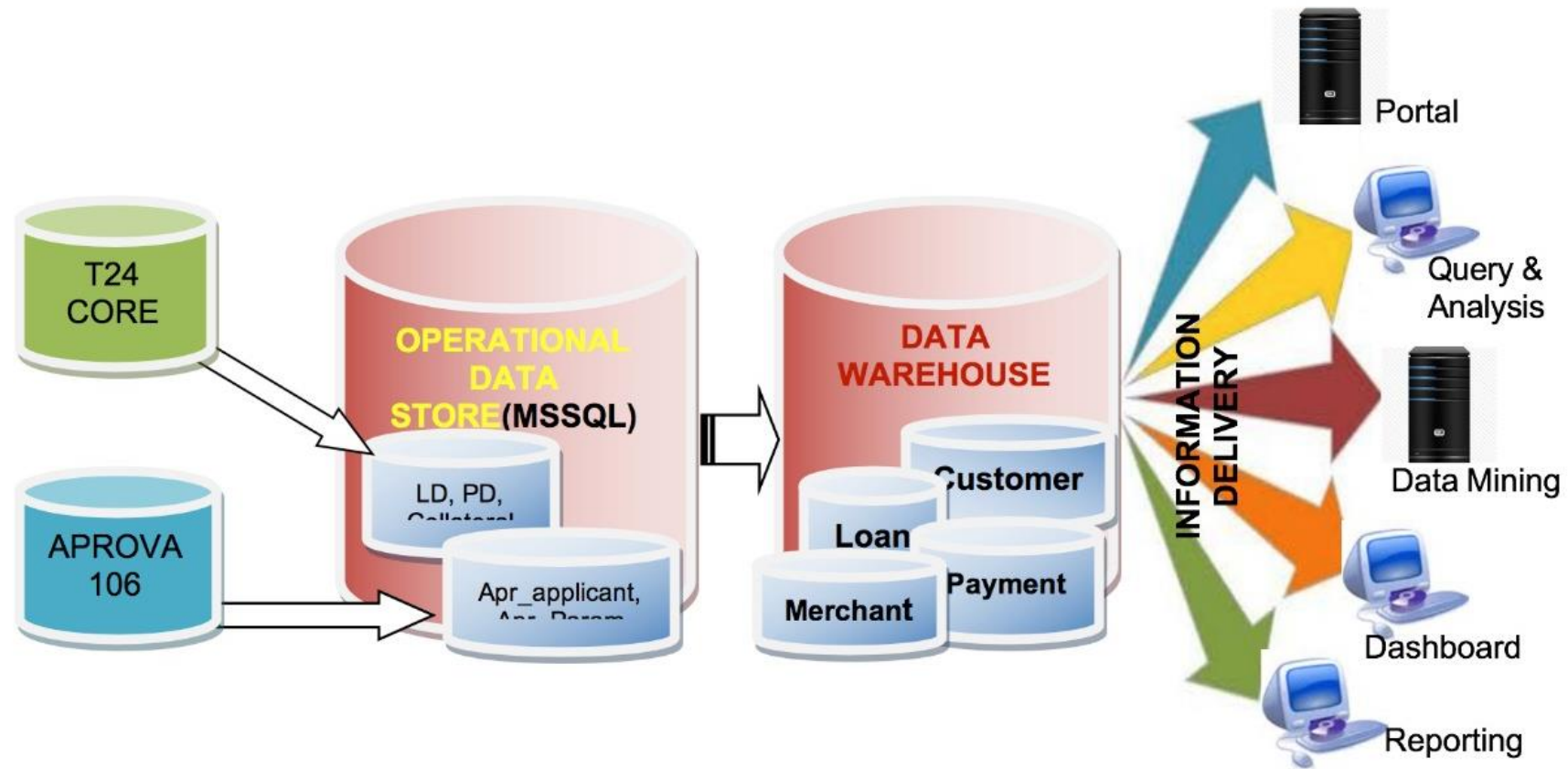
```
SELECT rs_db.dbo.REPL_NUM BY X(WKT_BACKSHEET) WKT_BACKSHEET, COUNT (1)
FROM rs_db_20140522.dbo.TBACKSHEET
GROUP BY rs_db.dbo.REPL_NUM BY X(WKT_BACKSHEET)
ORDER BY 2 DESC
;
```

	WKT_BACKSHEET	(No column name)
1	XXXX-XX-XX XX:XX:XX.XXX	12313407
2	XXXX/XX/XX XX:XX:XX	89
3	MON XX XXXX X:XXZM	4
4	XXXX-XX-XX XX.XX.XX.XXX	4
5	MON X XXXX X:XXZM	2



CIF Monitoring case study

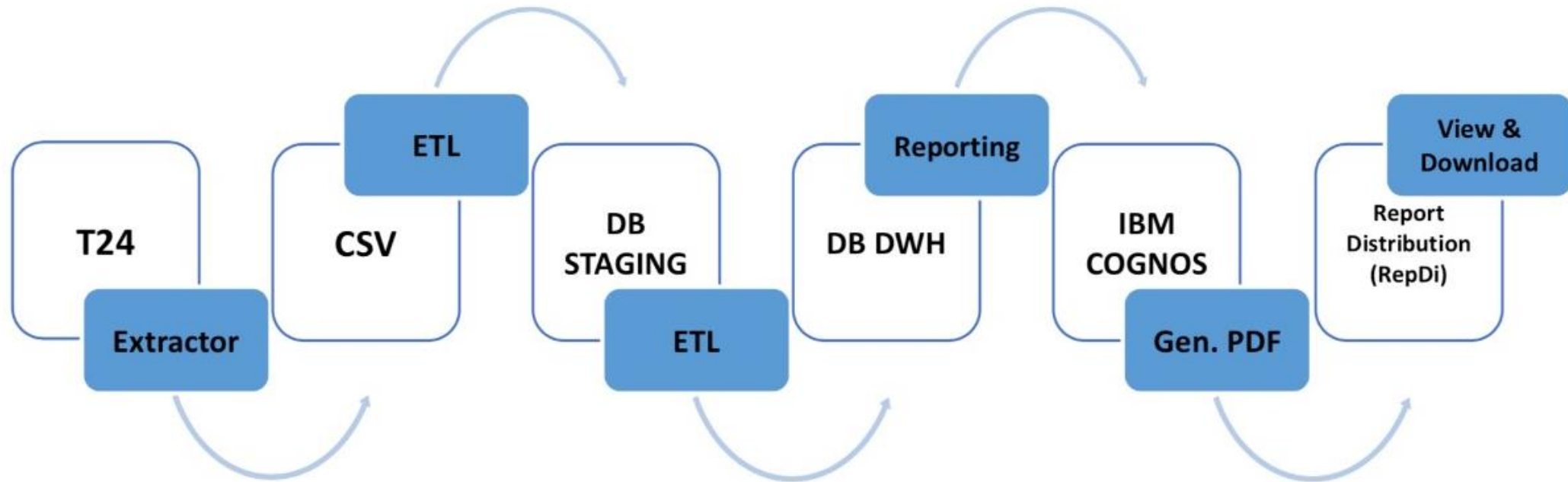
System Architecture



Software

Operating System	Linux RHEL
RDBMS	IBM DB2 BLU Acceleration
ETL tool	IBM ® InfoSphere™ Data Stage® and QualityStage v. 9.1
BI tool	IBM Cognos Analytics

MIS Report Process Flow



Steps of Development

- ✓ To understand the requirement
 - What is CIF scoring?
 - Data presentation needed by business user
- ✓ To understand data source and target table(s)
 - Source data comes from core banking system (T24) – loan and payment data
 - To design star schema (tables and their structure) of target (data warehouse)
- ✓ To do mapping from source to target
- ✓ To develop ETL process
 - ETL processes separated into two parts: (1) source to staging; (2) staging to data warehouse
 - To create jobs and sequence
 - To schedule sequence for daily run
- ✓ To develop Business Intelligence dashboard/ report
 - To develop metadata model with framework
 - To develop dashboard/ report

CIF Scoring

- Report Scoring CIF adalah report untuk memantau "kinerja" nasabah dari sisi peminjaman kredit dan pembayarannya. Semakin nasabah tertib (tidak terlambat) dalam melakukan pembayaran terhadap cicilan kredit-nya, maka penilaian (scoring) terhadap nasabah tersebut akan menjadi baik. Penilaian dilakukan per nasabah terhadap semua kredit yang berada di bawah nasabah tersebut. Untuk menentukan scoring dari kinerja pembayaran kredit per nasabah mengacu pada tabel scoring yang ada.
- Nilai-nilai parameter yang mempengaruhi nilai Score ada 2 (dua), yakni: (1) Overdue Class - Nilai ini ditentukan dari nilai overdue, selisih antara Tanggal Bayar dan Tanggal Jatuh Tempo. Pada contoh report di bawah terdapat pada kolom "Dpd"; (2) Usia relatif dari bulan Tanggal Jatuh Tempo terhadap bulan Tanggal Scoring. Nilainya akan bergerak terus seiring dengan waktu proses generasi data Scoring.

*) CIF = Customer Information System

Report/ Data Presentation

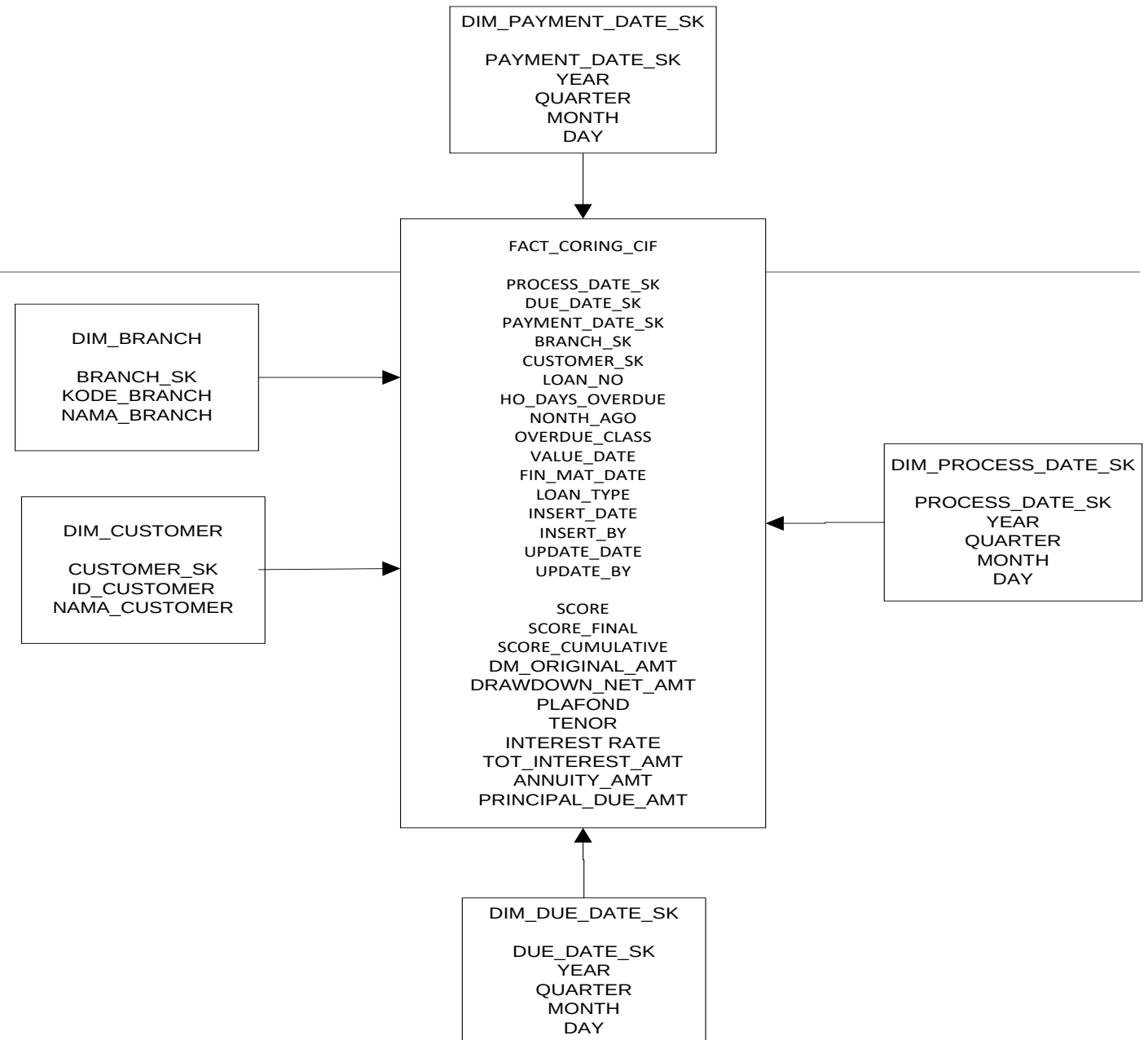
Tgl. Score 01 April 2015.

CIF 30305837

Nama

LD1314101389			Kelompok	Score	LD1415392220			Kelompok	Score	LD1502088513			Kelompok	Score	Score Akhir	Kumulatif
TglJT	TglByr	Dpd			TglJT	TglByr	Dpd			TglJT	TglByr	Dpd				
6/21/2013	6/21/2013	0	22A	19.0583											19.0583	19.0583
7/21/2013	7/21/2013	0	21A	19.4320											19.4320	38.4903
8/21/2013	8/21/2013	0	20A	19.8057											19.8057	58.2960
9/21/2013	9/21/2013	0	19A	20.1794											20.1794	78.4754
10/21/2013	10/21/2013	0	18A	20.5531											20.5531	99.0285
11/21/2013	11/21/2013	0	17A	20.9268											20.9268	119.9553
12/21/2013	12/21/2013	0	16A	21.3004											21.3004	141.2557
1/21/2014	1/21/2014	0	15A	21.6741											21.6741	162.9298
2/21/2014	2/21/2014	0	14A	22.0478											22.0478	184.9776
3/21/2014	3/24/2014	3	13A	22.4215											22.4215	207.3991
4/21/2014	4/21/2014	0	12A	22.7952											22.7952	230.1943
5/21/2014	5/21/2014	0	11A	23.1689											23.1689	253.3632
6/21/2014	6/23/2014	2	10A	23.5426											23.5426	276.9058
7/21/2014	8/2/2014	12	9A	23.9163	7/2/2014	7/2/2014	0	9A	23.9163						23.9163	300.8221
8/21/2014	8/25/2014	4	8A	24.2900	8/2/2014	8/2/2014	0	8A	24.2900						24.2900	325.1121
9/21/2014	9/27/2014	6	7A	24.6637	9/2/2014	9/2/2014	0	7A	24.6637						24.6637	349.7758
10/21/2014	10/22/2014	1	6A	25.0374	10/2/2014	10/2/2014	0	6A	25.0374						25.0374	374.8132
11/21/2014	11/24/2014	3	5A	25.4111	11/2/2014	11/5/2014	3	5A	25.4111						25.4111	400.2243
					12/2/2014	12/2/2014	0	4A	25.7848						25.7848	426.0091
					1/2/2015	1/9/2015	7	3A	26.1584						26.1584	452.1675
					2/2/2015	2/3/2015	1	2A	26.5321	2/21/2015	2/23/2015	2	2A	26.5321	26.5321	478.6996
					3/2/2015	3/2/2015	0	1A	26.9058	3/21/2015	Bim Bayar				26.9058	505.6054
										4/21/2015	x					

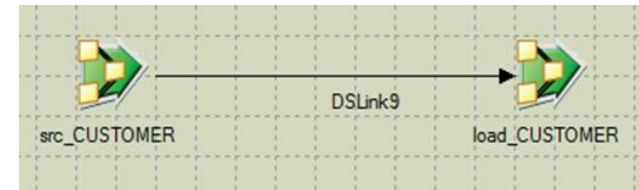
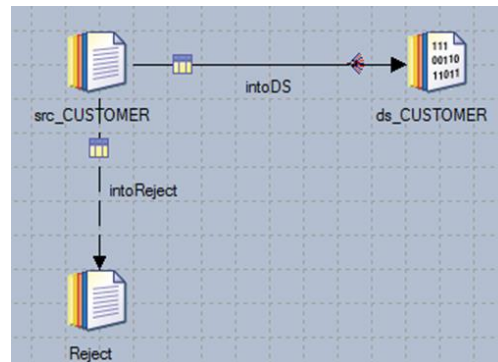
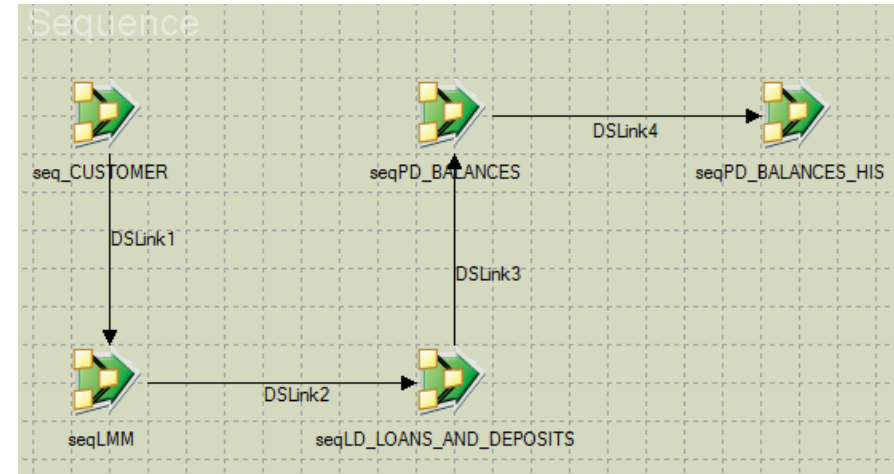
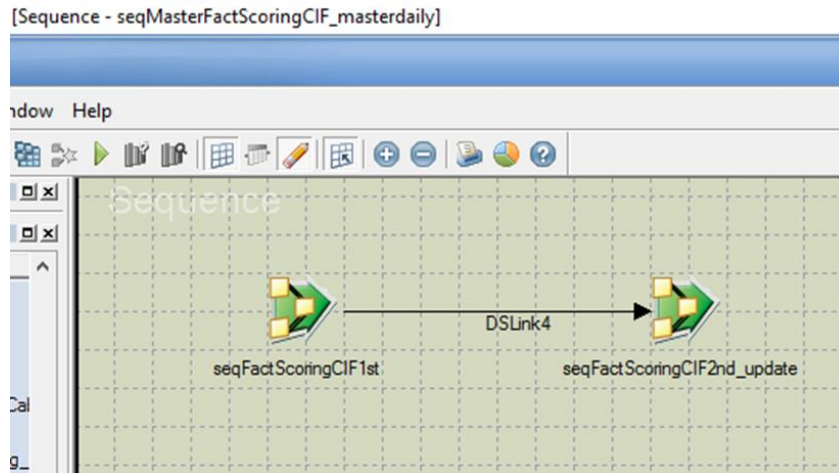
Star Schema



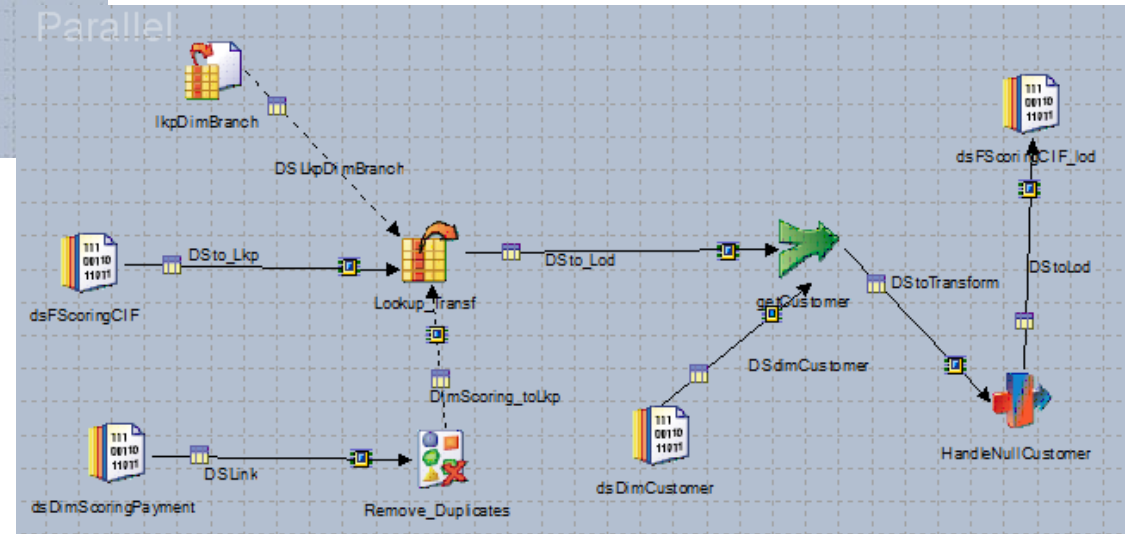
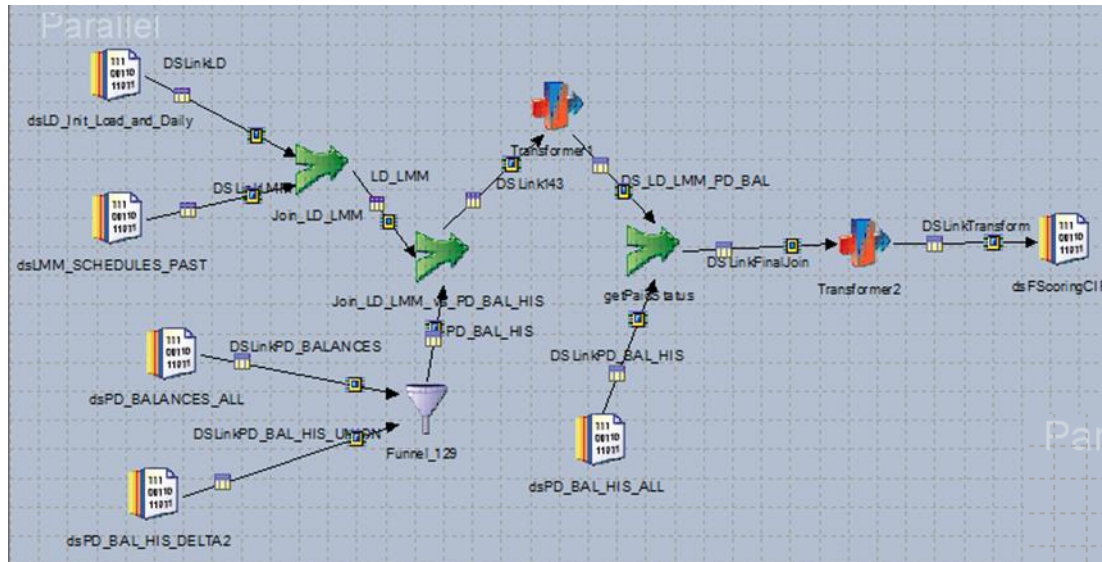
Mapping Source to Target

No	Report Field/Column	Source			Description/ Business Rule
		System	Table	Column	
1.	CIF	T24	CUSTOMER	ID	Kode unik nasabah (Customer Information File) --> Pada tabel Customer di T24 sama dengan ID Customer. LD.CUSTOMER_ID = CUSTOMER.ID
2.	Nama	T24	CUSTOMER	Name_1	Nama nasabah LD.CUSTOMER_ID = CUSTOMER.ID
3.	Nomor PK	T24	LD	ID	Nomor Perjanjian Kredit (Loan Account Number) <u>Clue:</u> SELECT CUSTOMER CIF , ID NO_PK , CO_CODE BRANCH_CODE , MAX(REPORT_DATE) FROM LD GROUP BY CUSTOMER, ID, CO_CODE ⇒ Data yang di-backup bersifat historical.

ETL Development: Sequence & Job



ETL Development: Jobs



ETL Development: Transformation (e.g.)

PAYMENT_DATE_SK

= IF IsNull(PAYMENT_STATUS) AND MAX_NO_DAYS_OVERDUE>0

THEN 99990101

ELSE AsInteger(DateToString(IF NullToValue(MAX_NO_DAYS_OVERDUE,0)=0

THEN DATE_REC

ELSE DateOffsetByDays(DATE_REC,
Max(NullToValue(MAX_NO_DAYS_OVERDUE,0),NullToValue(NO_DAYS_OVERDUE,0))
),"%yyyy%mm%dd"

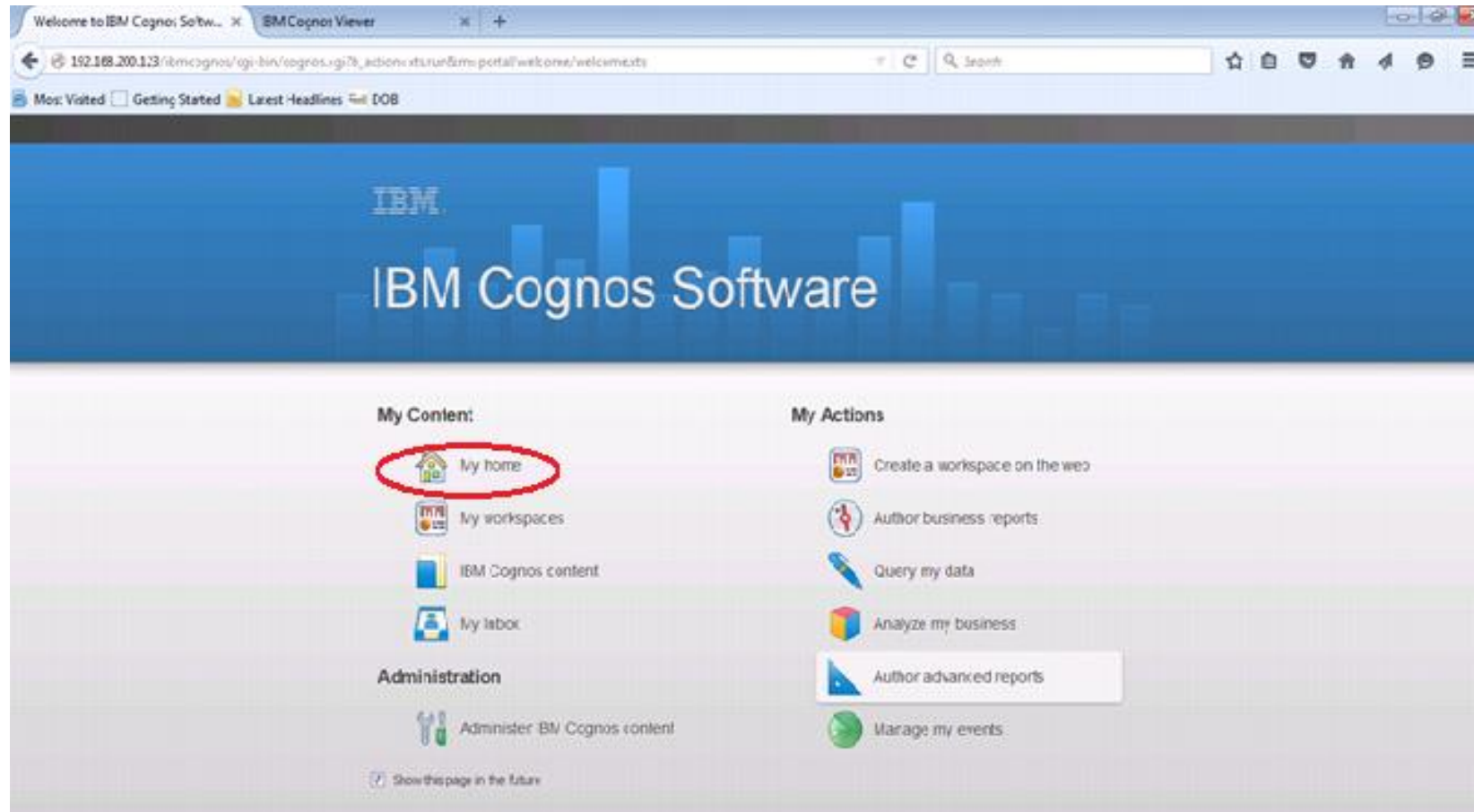
)

)

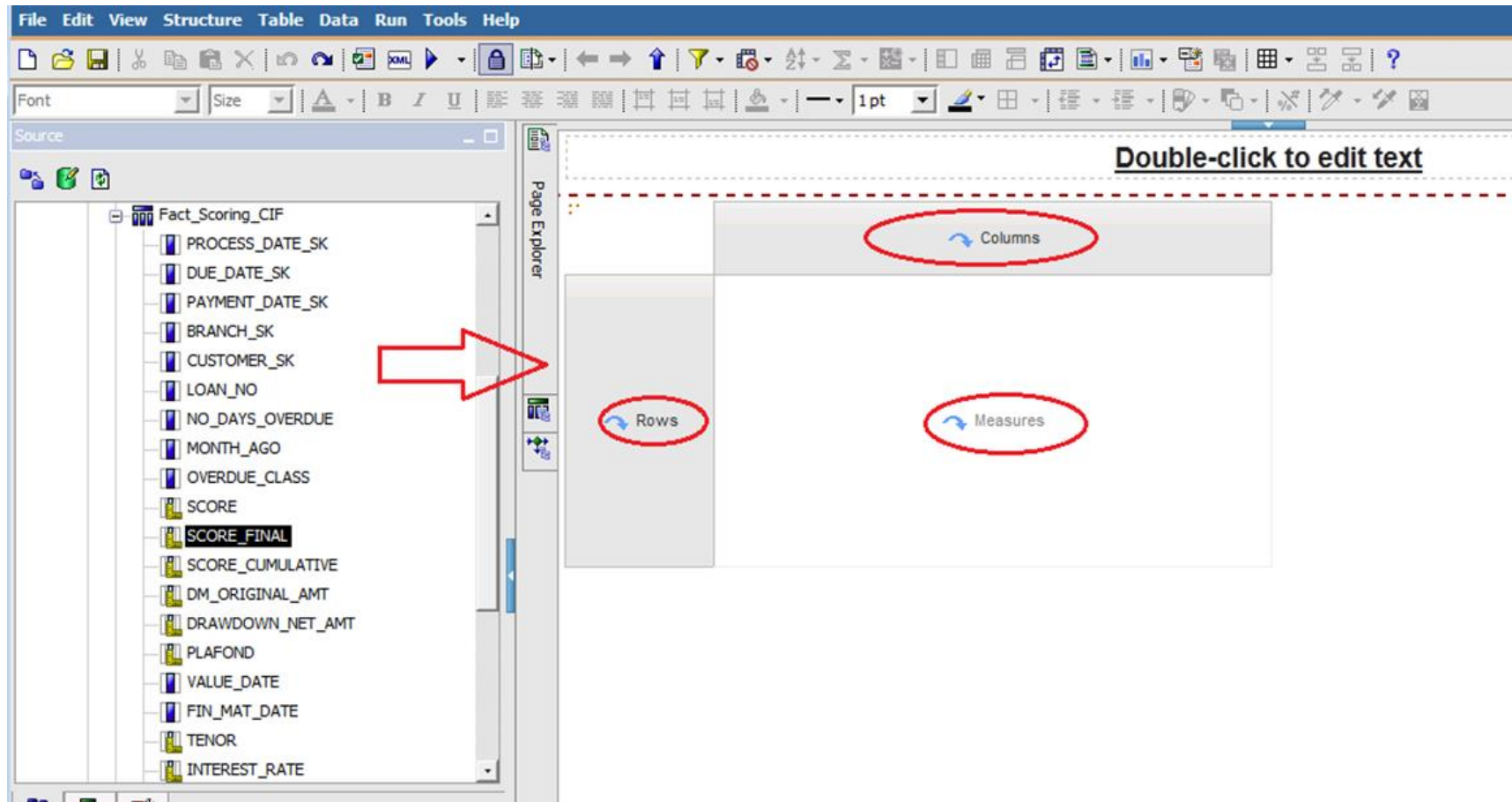
ETL Data Stage Demo Session



BI Cognos: Entering application



BI Cognos: Development Environment



BI Cognos: Report Parameter

192.168.200.123/ibmcognos/cgi-bin/cognos.cgi?b_action=cognosViewer&ui.action=run&ui.object=%2fcontent%2fpackage%40name%3c Search

Most Visited Getting Started Latest Headlines DOB

BPR
BPR KS Karyajatnika Sadaya

CREDIT SCORE BY CIF

Keywords:
Type one or more keywords separated by spaces

40722709 Search

Options

Choices:

40722709

Finish Deselect

BI Cognos - Report

Tgl. Score : 16 Oct 2015
 CIF :
 Nama :

LD1406250034						LD1432475715						LD1516640770						Score Akhir	Kumulatif
No	TglJT	TglByr	Dpd	Kelompok	Score	No	TglJT	TglByr	Dpd	Kelompok	Score	No	TglJT	TglByr	Dpd	Kelompok	Score		
1	3 Apr 2014	3 Apr 2014	0	18 A	20.5531													20.5531	20.5531
2	3 May 2014	9 May 2014	6	17 A	20.9268													20.9268	41.4799
3	3 Jun 2014	3 Jun 2014	0	16 A	21.3004													21.3004	62.7803
4	3 Jul 2014	3 Jul 2014	0	15 A	21.6741													21.6741	84.4544
5	3 Aug 2014	3 Aug 2014	0	14 A	22.0478													22.0478	106.5022
6	3 Sep 2014	2 Oct 2014	29	13 B	22.4215													22.4215	128.9237
7	3 Oct 2014	10 Nov 2014	38	12 C	-21.4537													-21.4537	107.4700
8	3 Nov 2014	20 Nov 2014	17	11 B	23.1689													23.1689	130.6389
						1	20 Dec 2014	10 Jan 2015	21	10 B	23.5426							23.5426	154.1815
						2	20 Jan 2015	23 Feb 2015	34	9 C	-22.4143							-22.4143	131.7672
						3	20 Feb 2015	24 Mar 2015	32	8 C	-22.7346							-22.7346	109.0326
						4	20 Mar 2015	6 Apr 2015	17	7 B	24.6637							24.6637	133.6963
						5	20 Apr 2015	4 May 2015	14	6 A	25.0374							25.0374	158.7337
						6	20 May 2015	4 Jun 2015	15	5 A	25.4111							25.4111	184.1448
						7	20 Jun 2015	11 Jul 2015	21	4 B	25.7848							25.7848	209.9296
						8	20 Jul 2015	30 Aug 2015	41	3 C	-24.3356	1	16 Jul 2015	27 Sep 2015	73	3 D	-121.6779	-24.3356	185.5940
						9	20 Aug 2015	-	56	2 C	-24.6558	2	16 Aug 2015	30 Sep 2015	45	2 C	-24.6558	-24.6558	160.9382
						10	20 Sep 2015	-	25	1 B	26.9058	3	16 Sep 2015	-	29	1 B	26.9058	26.9058	187.8440

BI Cognos Demo Session



Thank you