



WITS ENTERPRISE DATA WAREHOUSE AND HOUSEKEEPING

SESSION 8045
01:30PM – 02:30PM
June 28, 2016

PRESENTERS

Amer Nazir

BI Data Architect and Infrastructure Engineer

Business Intelligence Services

University of the Witwatersrand

amer.nazir@wits.ac.za

MSc Data Warehousing, BSc Honours IT
Management, MSc Computer Science, BSc
Physics and Mathematice



**UNIVERSITY OF THE
WITWATERSRAND**

OVERVIEW

1. WITS Enterprise Data Warehouse
2. Challenges - keeping BI Data in sync
3. Housekeeping
4. Audit / Discrepancy Reports
5. Housekeeping Benefits

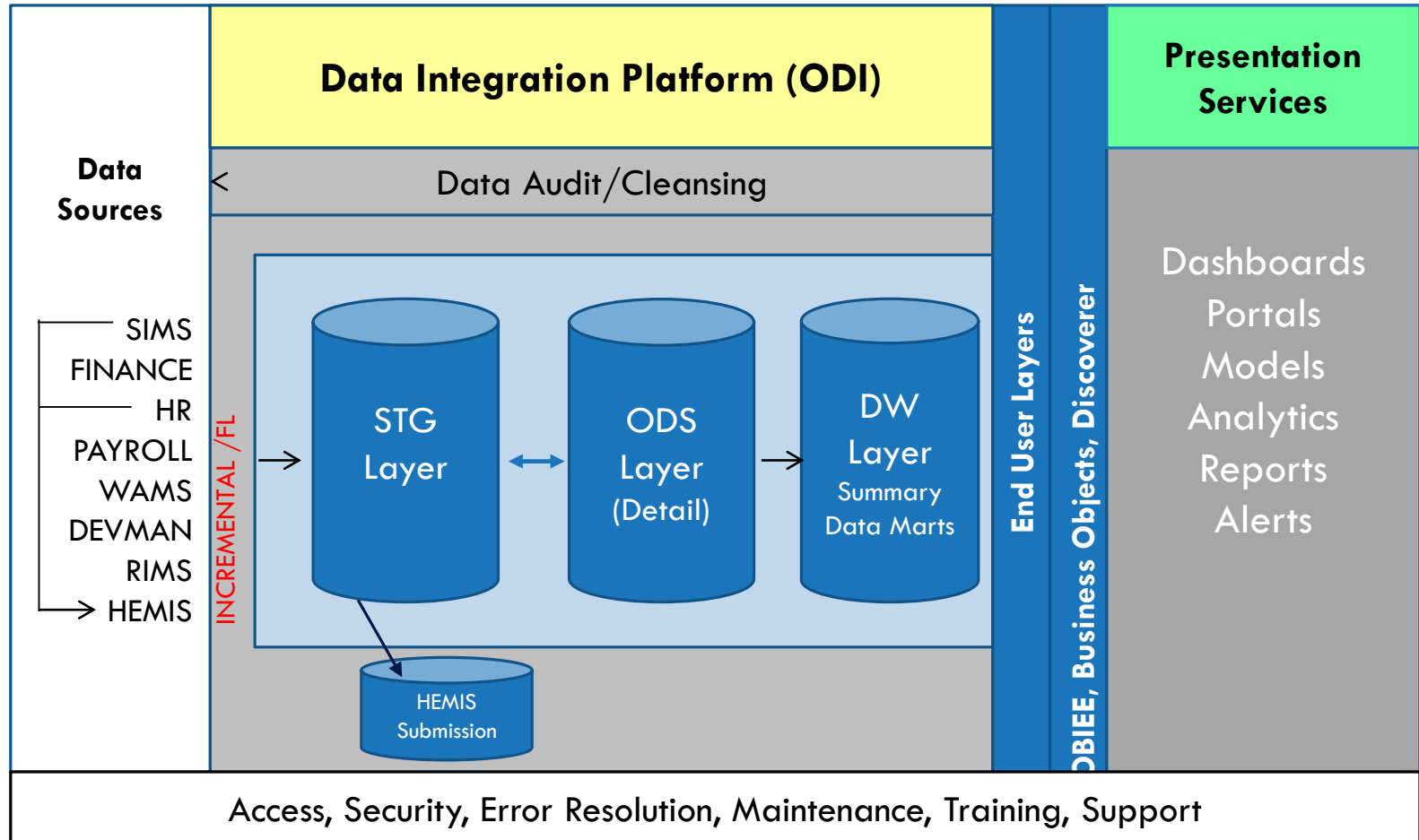
1. WITS ENTERPRISE DATA WAREHOUSE

ENTERPRISE DATA WAREHOUSE

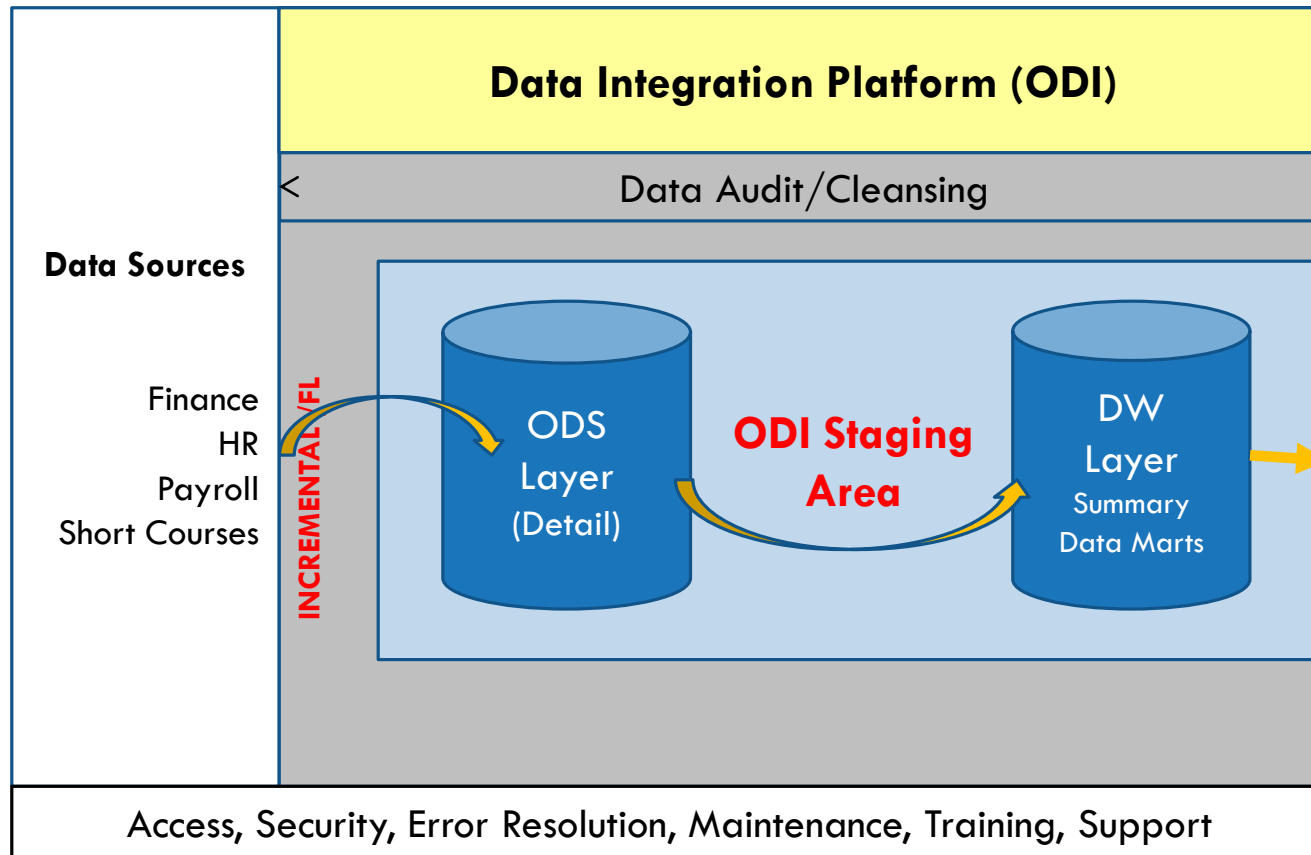
- An enterprise data warehouse is a unified database that holds all the business information an organization and makes it accessible all across the company.
- Data is one of the most valuable assets an organization has.
- Through appropriate use of accurate data, enterprises can make informed decisions that significantly impact their business success.



ARCHITECTURE

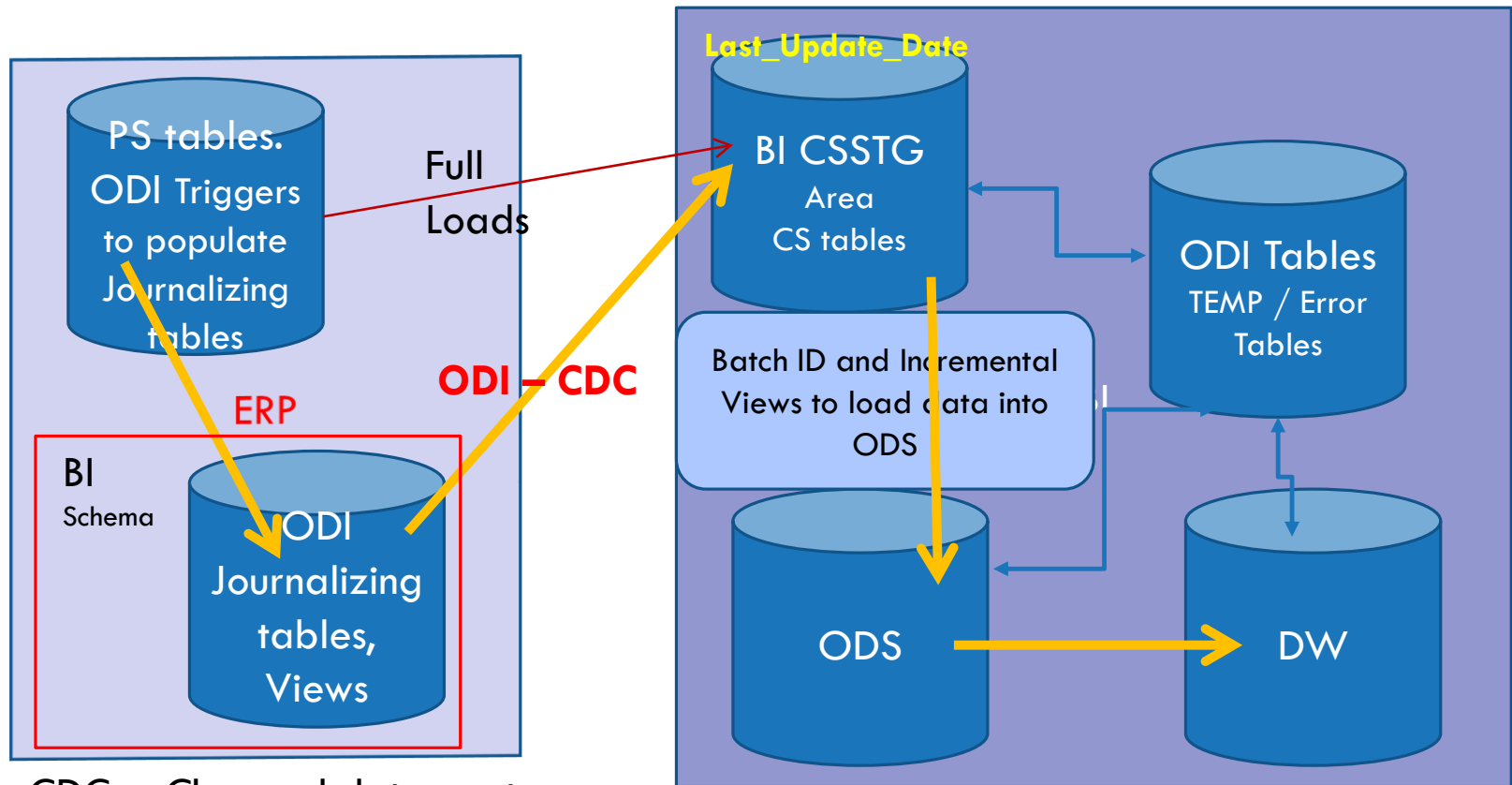


ARCHITECTURE FOR EBS EXTRACTS



Source systems keep timestamp against every record. Easy to identify incremental data.

ARCHITECTURE FOR SIMS EXTRACTS



CDC = Changed data capture

Source system uses effective date that is not equivalent to timestamp. CDC approach was used by replicating Campus Solution tables required for BI use by adding timestamp and other required columns.

2. CHALLENGES - KEEPING BI DATA IN SYNC

DATA DISCREPANCY

- Data **deleted** on source systems
- Data **modified** on primary key(s) at source side
- Data **missing** at BI side.

Source System	ACTION	REC-NO (PRIMARY KEY)	VAL1	VAL2
	NEW	1423		CDQ
	DELETED	1234	ABC	DEF
	Modified	1356	DJK	CFQ
		1357	DJJ	CFQ

BI System		REC-NO (PRIMARY KEY)	VAL1	VAL2
	EXTRA	1234	ABC	DEF
	EXTRA	1356	DJK	CFQ
	CORRECT	1357	DJJ	CFQ

New record missing at BI side can be due to incomplete data captured at source side and moved by the extracts into ODI error tables.

APPROACHES TO TACKLE DATA DISCREPANCY

- Sending errors records to business data owners for corrections.
- Truncating and reloading tables where the keys are not referenced into further tables and subsequent layers.
- Deleting subset of data – Snapshot Models
- ODI knowledge module customization and CDC, Jflag for identifying deleted and incremental records.
- Using audit/discrepancy programs to identify extra / missing records.



CUSTOMIZE KM AND CDC JFLAG

Customize KM managing overlapping dates

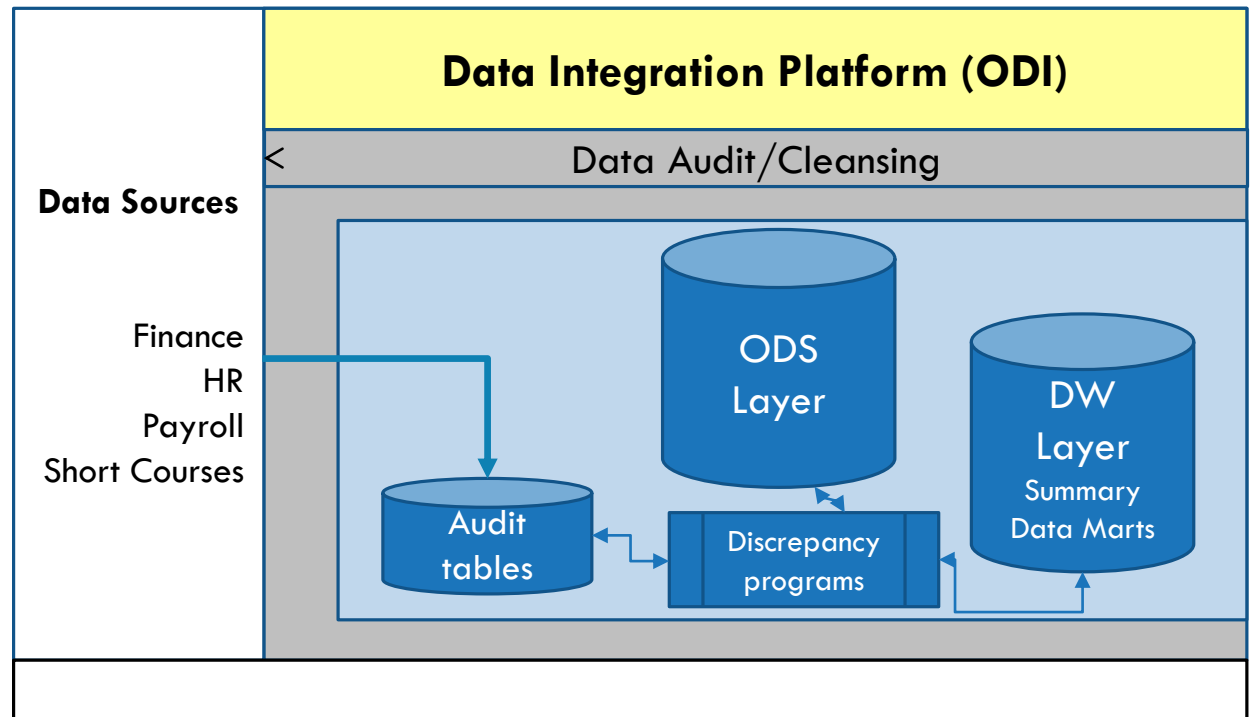
PERSON_ID	EFFECTIVE_START_DATE	EFFECTIVE_END_DATE	AUDIT_STATUS	AUDIT_DATE
14945	14/09/1997	16/04/2008	O	27/06/2016 22:32:57
14948	14/09/2015	15/09/2015	D	27/06/2016 22:32:57

CDC JFlag

JRN_SUBSCRIBE R	JRN_CONSUMED MED	JRN_FLAG G	JRN_DATE	EMPLID	ACAD_CAREER R	STDNT_CAR_N BR	EFFDT	EFFSEQ	ACAD_PLAN
SUNOPSIS	0	I	28/06/2016 08:35:18	4955R	R	0	27/06/2016	2	SFOSMATH70
SUNOPSIS	0	I	28/06/2016 08:35:18	4955R	R	0	27/06/2016	1	SFOSMATH70
SUNOPSIS	0	D	28/06/2016 08:35:18	4955R	R	0	06/06/2016	1	SFOSMATH70

CDC captures deleted records with Jflag = D and the extra records flagged into CS_STG, ODS and DW layers.

AUDIT / DISCREPANCY PROGRAMS



Audit Tables

STATUS	REC-NO (PRIMARY KEY)	VAL1
MISSING	1423	
EXTRA	1234	ABC
EXTRA	1356	DJK

3. HOUSEKEEPING

The routine tasks that must be done in order for a system to function or to function efficiently.

[Merriam Webmaster]



DATA WAREHOUSE HOUSEKEEPING

- The amount of data loaded into a data warehouse demands that a careful regime of management is imposed to ensure maximum efficiency and performance.
- It is important that regular housekeeping processes are run to
 - Regularly backup the data
 - Optimize the data storage
 - Deleting obsolete data.












OPTIMIZE THE DATA STORAGE

- Truncating / dropping temp tables releases space.
- Deleting high volume of data from a table leaves it fragmented, and this make two problems:
 - The High Water Mark (which show the max usage of a table) is still at maximum size of table and make worst performance on Full table Scan
 - Space (extents) are not freed.
- Running DBMS utilities regularly addresses fragmentation issues and enhance query performance.
- Rebuilding indexes
 - In systems that do batch DML jobs, index rebuilding "often" improves SQL performance.



DELETING OBSOLETE DATA

- End of the each load, delete programs run to remove extra records that were marked by the discrepancy and Jflag programs. D\$ tables were created to keep history of the deleted records.

+		✓	262860 - PROC_DEL_ASGN_TAX_INFO -
+		✓	262859 - PROC_HR_TRAINING_ENROLME
+		✓	262858 - PROC_HR_QUAL - 27-Jun-2016
+		✓	262857 - PROC_EMP_ANN_SAL - 27-Jun-
+		✓	262856 - PROC_DEL_LEAVE - 27-Jun-201
+		✓	262855 - PROC_DEL_PROBATION - 27-Ju
+		✓	262854 - PROC_DEL_PENSION - 27-Jun-
+		✓	262853 - PROC_DEL_EMP_PCS_DTLS - 27
+		✓	262852 - PROC_DEL_BURS_INFO - 27-Ju

4. AUDIT / DISCREPANCY REPORTS

HR DISCREPANCY REPORT

AUDIT_AREA	REC_STATUS	SOURCE_RECORD					REPORT_DATE	ODI_ERROR_RECORD
		REPORT_YEAR	COUNTS	BI_RECORD_COUNTS	EXTRA_COUNTS	MISSING_COUNTS		
ASGN_TAX_INFO	ASGN_TAX_INFO	2016	135,536	135,153	0	383	25/06/2016	0
BURS_INFO	BURS_INFO	2016	26,429	26,430	1	3	25/06/2016	12
DISABILITY	DISABILITY	2016	280	280	0	0	25/06/2016	0
EMP_ANN_SAL	EMP_ANN_SAL	2016	58,192	59,881	0	0	25/06/2016	2
EMP_PCS_DTLS	EMP_PCS_DTLS	2016	17,728	17,728	0	0	25/06/2016	0
HR_ASSIGNMENT	ACTIVE_ASSIGNMENT	2016	153,635	153,625	0	6	25/06/2016	0
HR_EMPLOYEE	ACTIVE_EMPLOYEE	2016	178,643	178,643	0	0	25/06/2016	0
HR_PENSION	HR_PENSION	2016	8,059	8,059	0	0	25/06/2016	0
HR_PROBATION	HR_PROBATION	2016	5,615	5,550	0	12	25/06/2016	0
HR_QUALIFICATION	ACTIVE_QUALIFICATION	2016	29,089	29,065	0	0	25/06/2016	0
LEAVE	LEAVE	2016	279,762	279,762	0	0	25/06/2016	0
ORG_MANAGER	ORG_MANAGER	2016	1,350	1,353	3	0	25/06/2016	0
TRAINING_ENROL	TRAINING_ENROL	2016	22,860	22,860	0	0	25/06/2016	0

PROJECTS AUDIT REPORT



Mon 27/06/2016 05:41

dwsysadmin@miu.wits.ac.za

Projects Audit Report on 27-JUN-16

To

From: dwsysadmin@miu.wits.ac.za

To: [\\$dept-MIU-DataWarehouse@wits.ac.za](mailto:$dept-MIU-DataWarehouse@wits.ac.za)

???????No Differences found in

Actual Cost, Actual Revenue, Billed Revenue, Burdened Cost, Approved Cost Budget, Approved Revenue Budget

Extra Commitment Records Deleted: 0

5. HOUSEKEEPING BENEFITS

HOUSEKEEPING BENEFITS

- Providing correct and audited data to users.
- Through housekeeping approaches optimizing data storage and removing obsolete records.
- Assisting business data owners in data validation and verification, so data can be fixed on source systems timeously.

CONCLUDING THOUGHTS

ANY QUESTIONS?

PRESENTERS

Amer Nazir

BI Data Architect and Infrastructure Engineer

Business Intelligence Services

University of the Witwatersrand

amer.nazir@wits.ac.za

**ALL SAHEUG ALLIANCE PRESENTATIONS WILL BE
AVAILABLE FOR DOWNLOAD FROM THE CONFERENCE
SITE**



THANK YOU!

