



Тестирование в Business Intelligence проектах

Дмитрий Романов. Itera Consulting
dmitriy.romanov@gmail.com

Тестирование в BI проектах. Содержание.

Часть 1. Что такое Business Intelligence

Microsoft Business Intelligence Technology Stack

- Extraction-Transformation-Load
- Analytical Services
- Reporting Services

Тестирование в BI проектах

- Задачи
- Особенности
- Сложности тестирования

Основные вызовы стоящие перед QA в BI

Часть 2. Пример построения Framework для юнит тестирования

Заключение

Тестирование в BI проектах

Часть 1. Что такое Business Intelligence

Microsoft Business Intelligence Technology Stack

- Extraction-Transformation-Load
- Analytical Services
- Reporting Services

Тестирование в BI проектах

- Задачи
- Особенности
- Сложности тестирования

Основные вызовы стоящие перед QA в BI

Часть 2. Пример построения Framework для юнит тестирования

Заключение

Что такое Business Intelligence

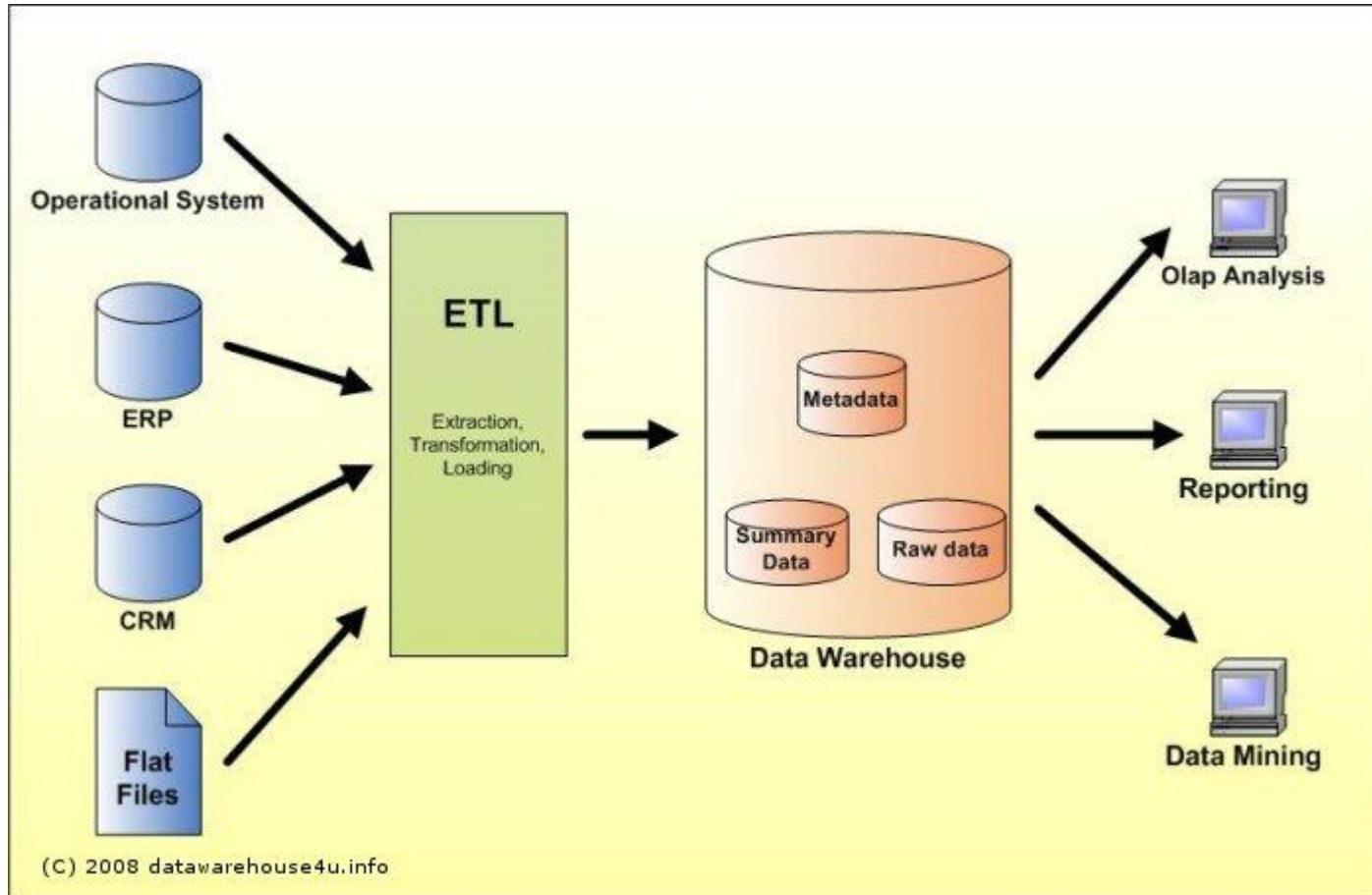
Business intelligence (BI) это широкая категория приложений и технологий для сбора, хранения, анализа и предоставления доступа к данным

Data Warehouse (DWH) – централизованный репозиторий который содержит данные с различных источников и трансформирует их в общепринятую, многомерную модель данных для эффективного анализа

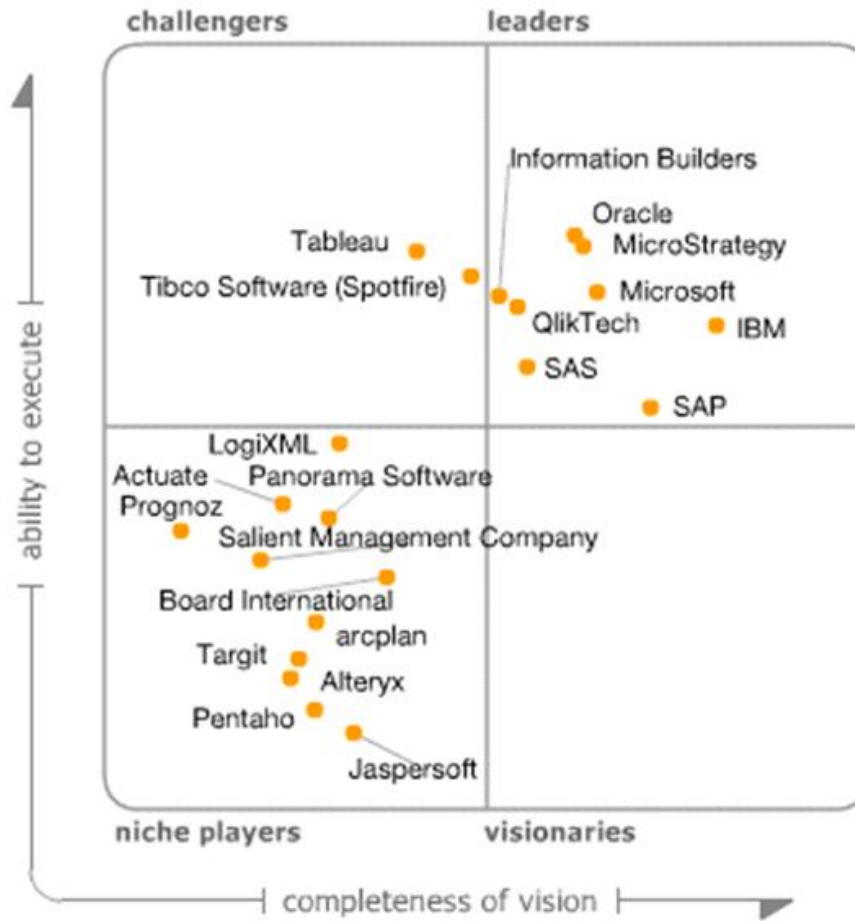
ETL (от англ. *Extract, Transform, Load*) – один из основных процессов в управлении DWH, который включает в себя:

- извлечение данных из внешних источников;
- их трансформация и очистка
- загрузка их в хранилище данных

Что такое Business Intelligence



2012 Gartner Magic Quadrant for BI Platforms



As of February 2012

Тестирование в BI проектах

Часть 1. Что такое Business Intelligence

Microsoft Business Intelligence Technology Stack

- Extraction-Transformation-Load
- Analytical Services
- Reporting Services

Тестирование в BI проектах

- Задачи
- Особенности
- Сложности тестирования

Основные вызовы стоящие перед QA в BI

Часть 2. Пример построения Framework для юнит тестирования

Заключение

Тестирование в BI проектах

Часть 1. Что такое Business Intelligence

Microsoft Business Intelligence Technology Stack

- Extraction-Transformation-Load
- Analytical Services
- Reporting Services

Тестирование в BI проектах

- Задачи
- Особенности
- Сложности тестирования

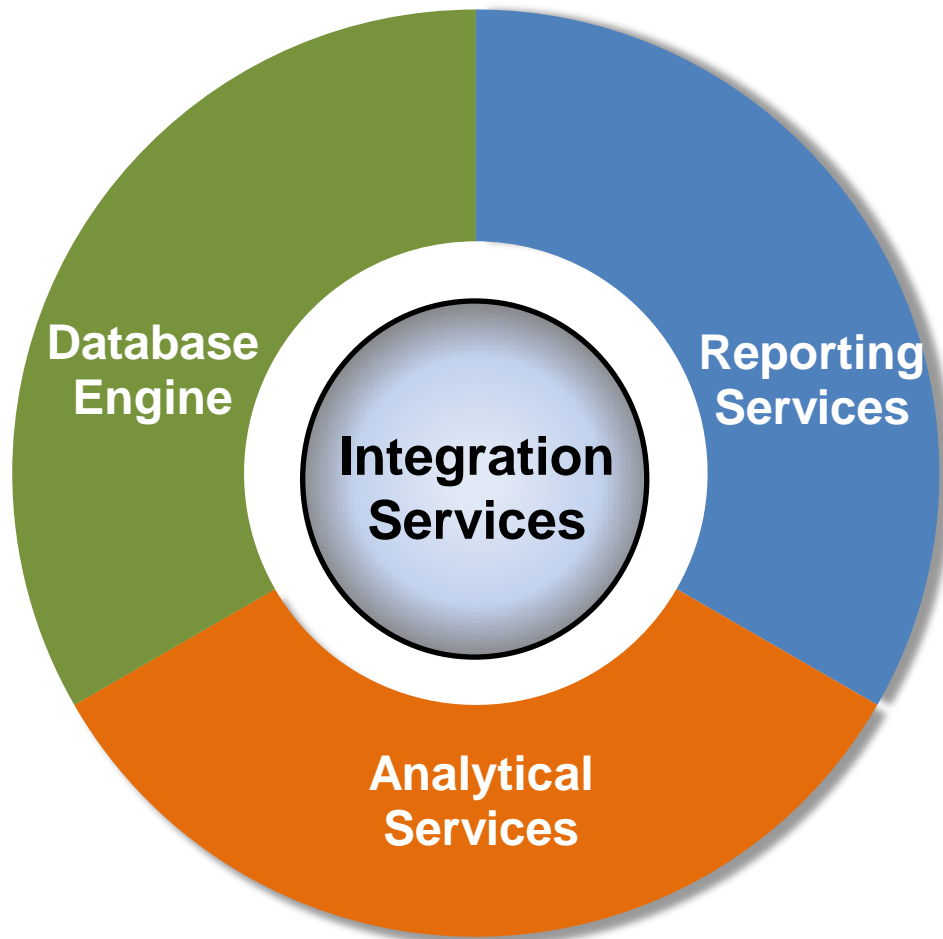
Основные вызовы стоящие перед QA в BI

Часть 2. Пример построения Framework для юнит тестирования

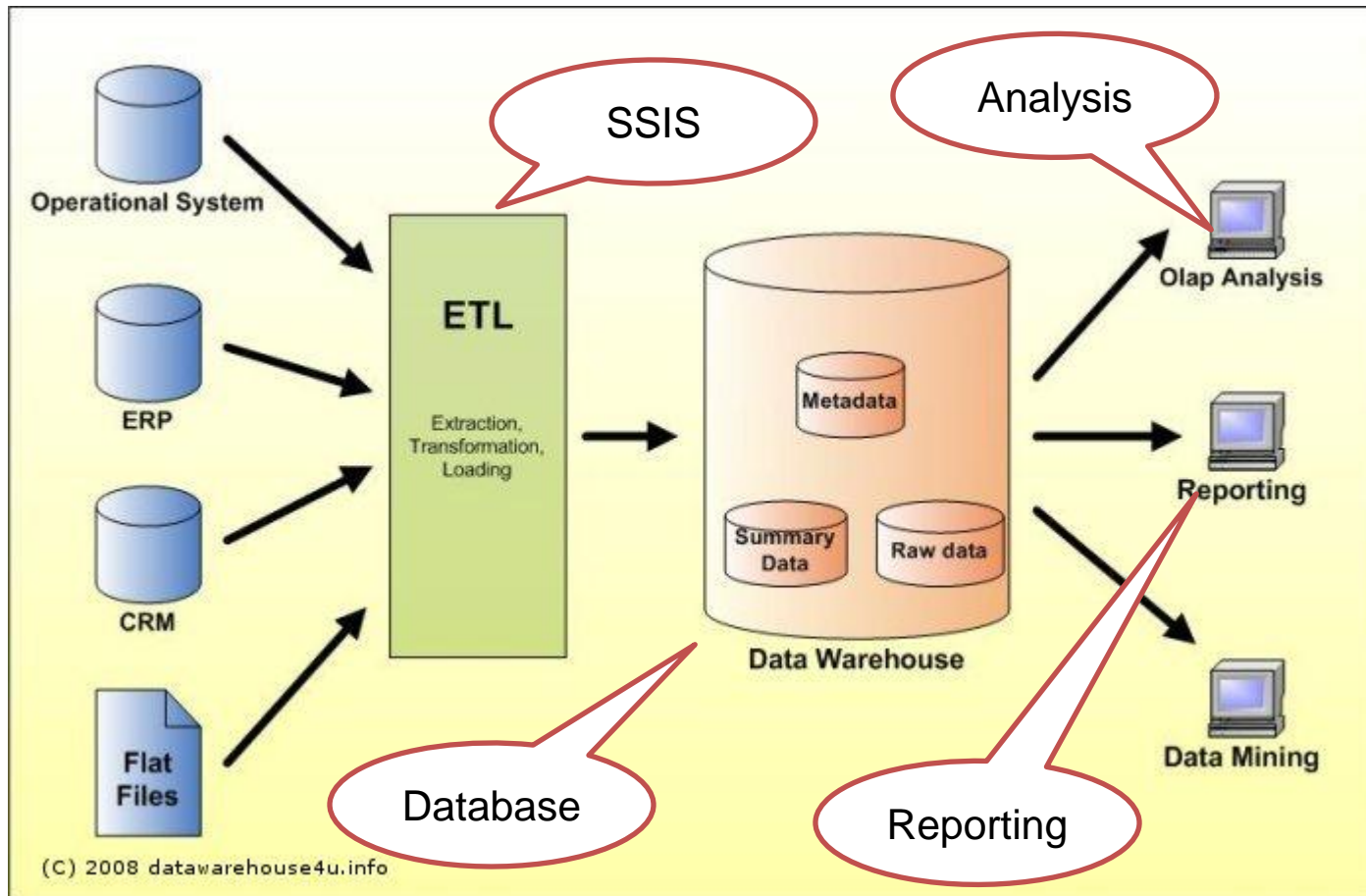
Заключение

Что такое Business Intelligence

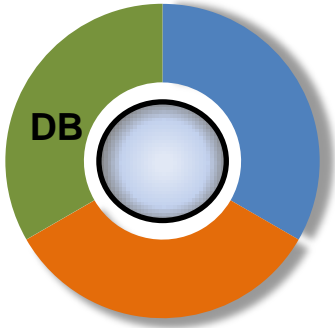
Microsoft BI



Что такое Business Intelligence



Database Services. Особенности



Как ведется разработка ?

Что тестировать ?

Какой существует тест-инструментарий ?

Development:

- Performed in T-SQL
- Results stored in Database

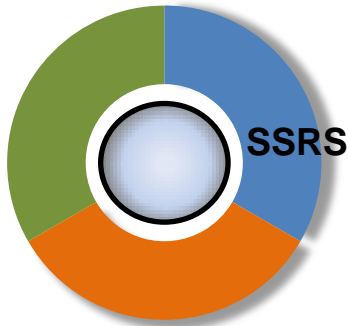
Test tools:

- DB related frameworks*
- *DBUnit, DBTestUnit, ...

What to test:

- Stored procedures
- DB constrains
- Deployment routine
- Integrity & System tests
- Performance

Reporting Services. Особенности



Как ведется разработка ?

Что тестировать ?

Какой существует тест-инструментарий ?

Development:

- Performed in VS Designer
- Displayed in browser

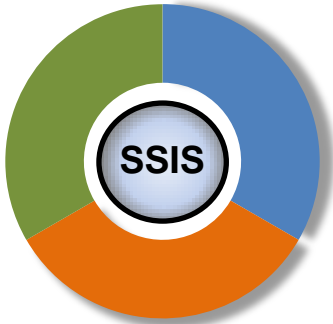
Test tools:

- Web testing tools*
- *Selenium, QTP, TestComplete,...

What to test:

- Page layout
- Performance
- Report values ?

Integration Services. Особенности



Как ведется разработка ?

Что тестировать ?

Какой существует тест-инструментарий ?

Development:

- Performed in SSIS Designer
- Results stored in database

Integration Services



Extract

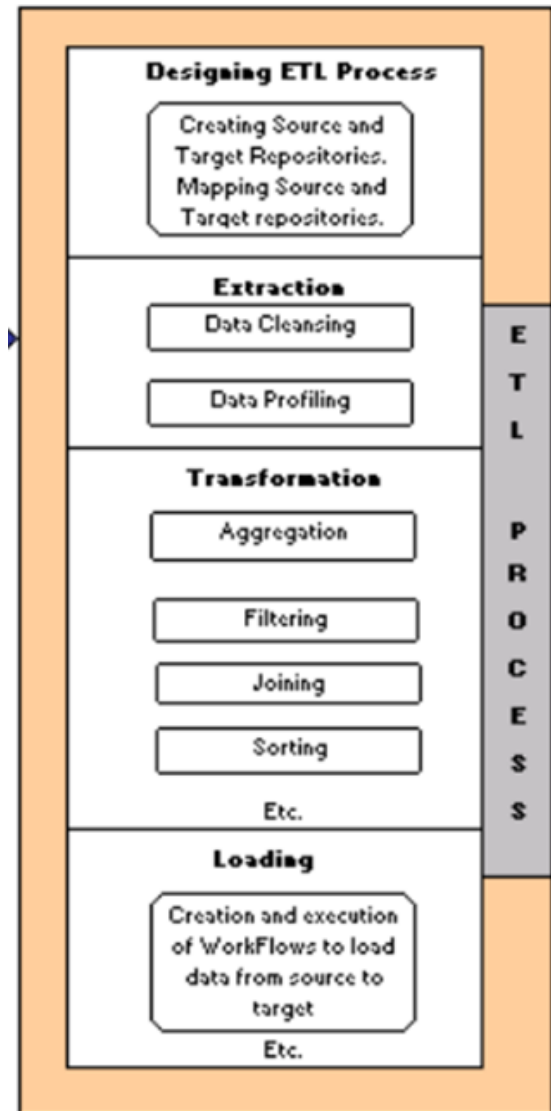
Извлечение данных из источников данных

Transformation

Может включать в себя очистку, фильтрацию, валидацию, применение бизнес логики

Load

Загрузку уже подготовленных данных в warehouse.



Integration Services



Source

ETL

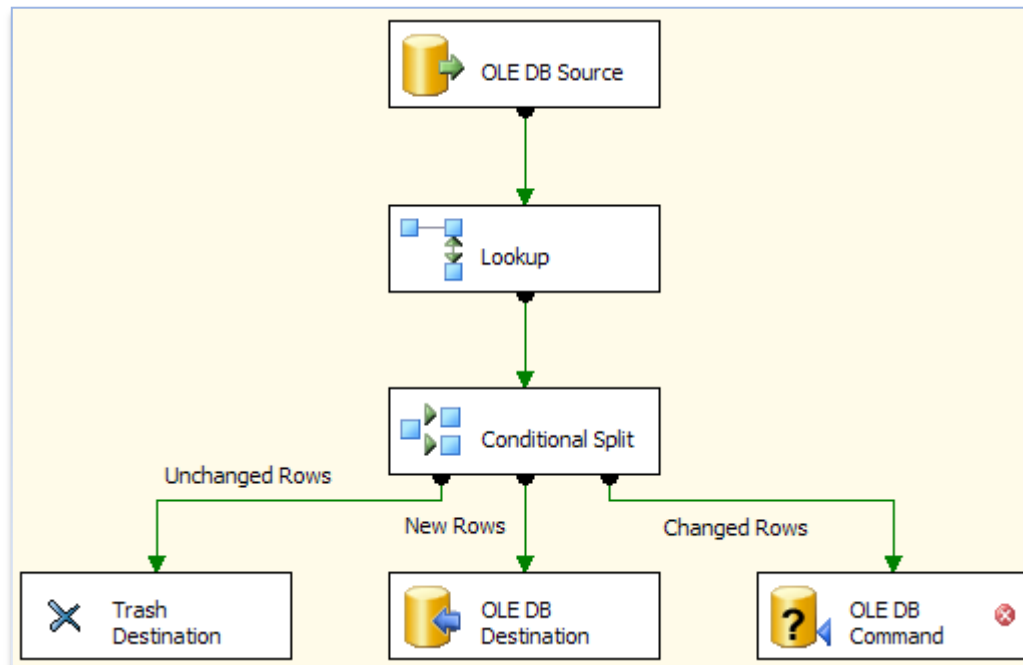
Data
Warehouse

Integration Services

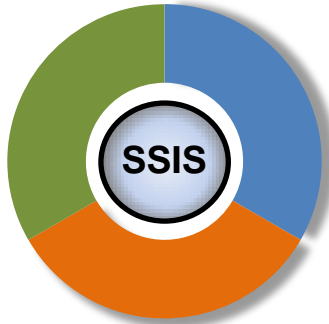


Development:

- Drag-&-Drop programming
- .Net language scripts support
- SQL scripts support



Integration Services. Особенности



Как ведется разработка ?

Что тестировать ?

Какой существует тест-инструментарий ?

Development:

- Performed in SSIS Designer
- Results stored in database

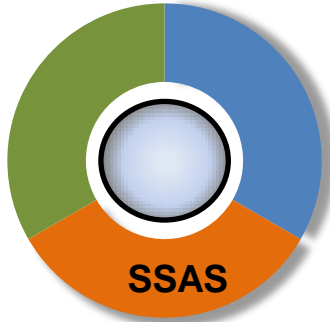
Test tools:

- No tools available
- No compatibility between ETL vendors

What to test:

- ETL logic*
- Integration testing
- System testing
- Performance
- ...

Analytical Services. Особенности



Как ведется разработка ?

Что тестировать ?

Какой существует тест-инструментарий ?

Development:

- Multi Dimensional Query (MDX) language
- Results stored in SSAS database

Analytical Processing



```
SELECT
    Years.MEMBERS ON COLUMNS,
    Regions.Continent.MEMBERS ON ROWS
FROM Sales
```

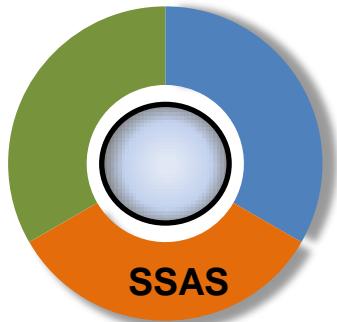
The result of this query visualized through the table will look like the following:

	1994	1995	1996	1997
N. America	120,000	200,000	400,000	600,000
S. America	-	10,000	30,000	70,000
Europe	55,000	95,000	160,000	310,000
Asia	30,000	80,000	220,000	200,000

Microsoft Excel - Book1

		Calendar Year			
		CY 2001	CY 2002	CY 2003	CY 2004
5	Accessories	20235.3646	92735.3534	590242.5866	568844.5832
6	Order Count	135	356	8082	10950
7	Reseller Ratio to All Products	0.002508899	0.003840859	0.009208332	0.010088147
8	Internet Gross Profit Margin			0.625998354	0.625998321
9	Bikes	10661722.28	26486358.2	34910877.69	22561568.03
10	Order Count	1358	3527	6944	6529
11	Reseller Ratio to All Products	0.916918721	0.826526655	0.793467608	0.8354652
12	Internet Gross Profit Margin	0.401548019	0.405315607	0.409584727	0.405355088
13	Clothing	34376.3353	485587.1546	1010112.167	587537.8026
14	Order Count	242	644	3973	5012
15	Reseller Ratio to All Products	0.00426218	0.020111767	0.027074283	0.024068566
16	Internet Gross Profit Margin			0.401639468	0.401373922
17	Components	615474.9788	3610092.472	5482497.289	2091011.918
18	Order Count	205	702	1138	601
19	Reseller Ratio to All Products	0.0763102	0.149520719	0.170249777	0.130378087
20	Internet Gross Profit Margin				
21	Total Sales Amount	11331808.96	30674773.18	41993729.72	25808962.34
22	Total Order Count	1379	3692	12440	13944
23	Total Reseller Ratio to All Products	1	1	1	1
24	Total Internet Gross Profit Margin	0.401548019	0.405315607	0.415964461	0.414464851
25					

Analysis Services. Особенности



Как ведется разработка ?

Что тестировать ?

Какой существует тест-инструментарий ?

Development:

- Multi Dimensional Query (MDX) language
- Results stored in SSAS database

Test tools:

- No tools available
- No compatibility between ETL vendors

What to test:

- Facts-to-Dimensional Links
- Aggregation rules validation
- Calculation rules validation

Тестирование в BI проектах

Часть 1. Что такое Business Intelligence

Microsoft Business Intelligence Technology Stack

- Extraction-Transformation-Load
- Analytical Services
- Reporting Services

Тестирование в BI проектах

- Задачи
- Особенности
- Сложности тестирования

Основные вызовы стоящие перед QA в BI

Часть 2. Пример построения Framework для юнит тестирования

Заключение

Тестирование в BI проектах

Часть 1. Что такое Business Intelligence

Microsoft Business Intelligence Technology Stack

- Extraction-Transformation-Load
- Analytical Services
- Reporting Services

Тестирование в BI проектах

- Задачи
- Особенности
- Сложности тестирования

Основные вызовы стоящие перед QA в BI

Часть 2. Пример построения Framework для юнит тестирования

Заключение

Особенности

- Загрузка данных в несколько этапов
- Часто видимый результат только в конце
- Поэтапное тестирование для раннего и точного обнаружения
- Автоматизация тестирования – необходима
- Часто тестирование проводится самими разработчиками
- До 60% тестирования приходится на ETL

Тестирование в BI проектах

Часть 1. Что такое Business Intelligence

Microsoft Business Intelligence Technology Stack

- Extraction-Transformation-Load
- Analytical Services
- Reporting Services

Тестирование в BI проектах

- Задачи
- Особенности
- Сложности тестирования

Основные вызовы стоящие перед QA в BI

Часть 2. Пример построения Framework для юнит тестирования

Заключение

Тестирование в BI проектах

Часть 1. Что такое Business Intelligence

Microsoft Business Intelligence Technology Stack

- Extraction-Transformation-Load
- Analytical Services
- Reporting Services

Тестирование в BI проектах

- Задачи
- Особенности
- Сложности тестирования

Основные вызовы стоящие перед QA в BI

Часть 2. Пример построения Framework для юнит тестирования

Заключение

Основные вызовы которые стоят перед QA

Технические:

1. Большой объем данных.
В среде BI введен термин – «Big Data».
2. Не устоявшиеся стандарты разработки.
У BI нет своего «SQL-92» стандарта
3. Отсутствие инструментов тестирования, а использование существующих не всегда продуктивно

Организационные:

1. Недостаток ресурсов. Как человеческих так и технических.
2. Недостаток четкости требований в ходе разработки

Тестирование в BI проектах

Часть 1. Что такое Business Intelligence

Microsoft Business Intelligence Technology Stack

- Extraction-Transformation-Load
- Analytical Services
- Reporting Services

Тестирование в BI проектах

- Задачи
- Особенности
- Сложности тестирования

Основные вызовы стоящие перед QA в BI

Часть 2. Пример построения Framework для юнит тестирования

Заключение

Тестирование в BI проектах

Часть 1. Что такое Business Intelligence

Microsoft Business Intelligence Technology Stack

- Extraction-Transformation-Load
- Analytical Services
- Reporting Services

Тестирование в BI проектах

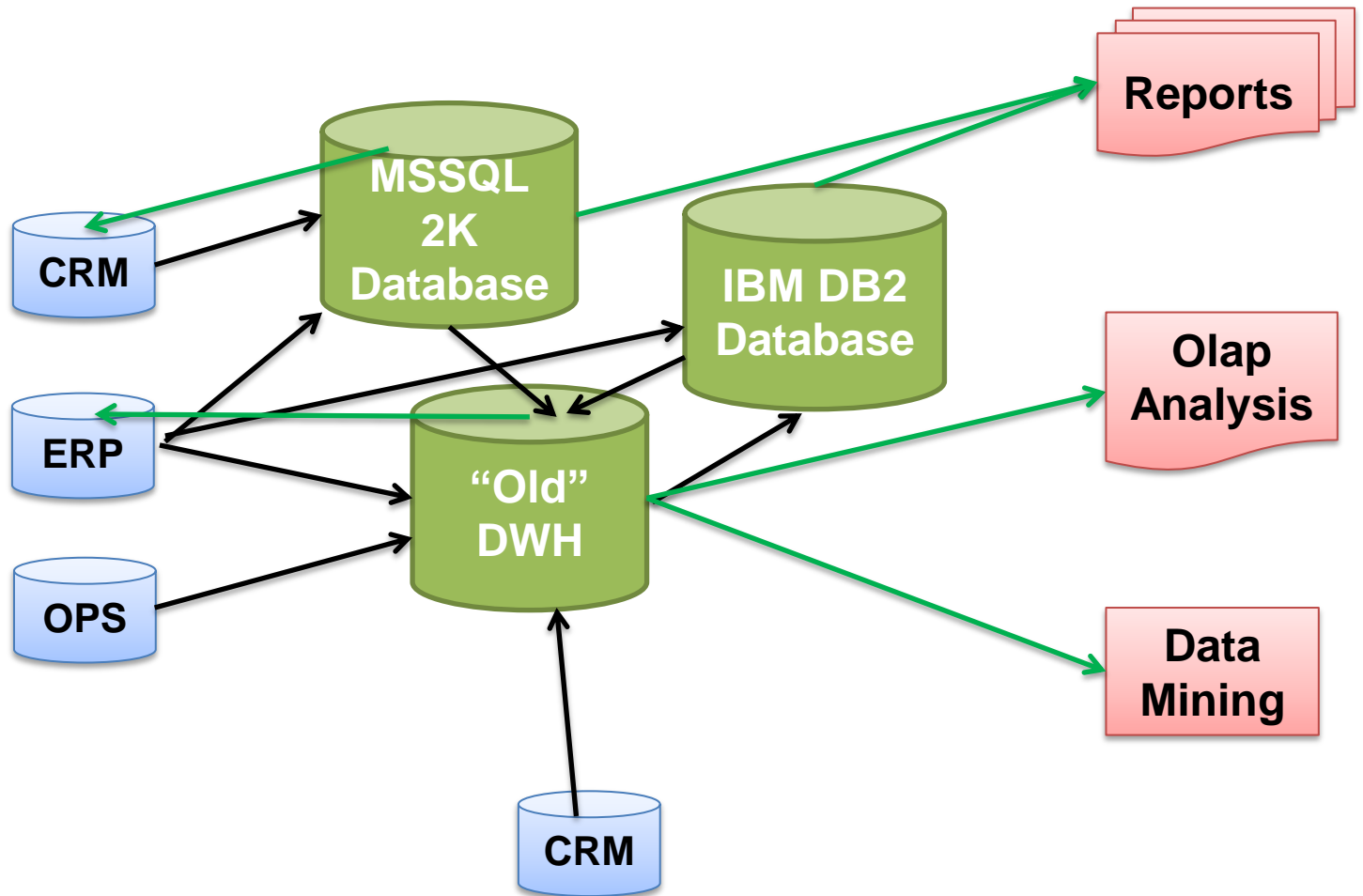
- Задачи
- Особенности
- Сложности тестирования

Основные вызовы стоящие перед QA в BI

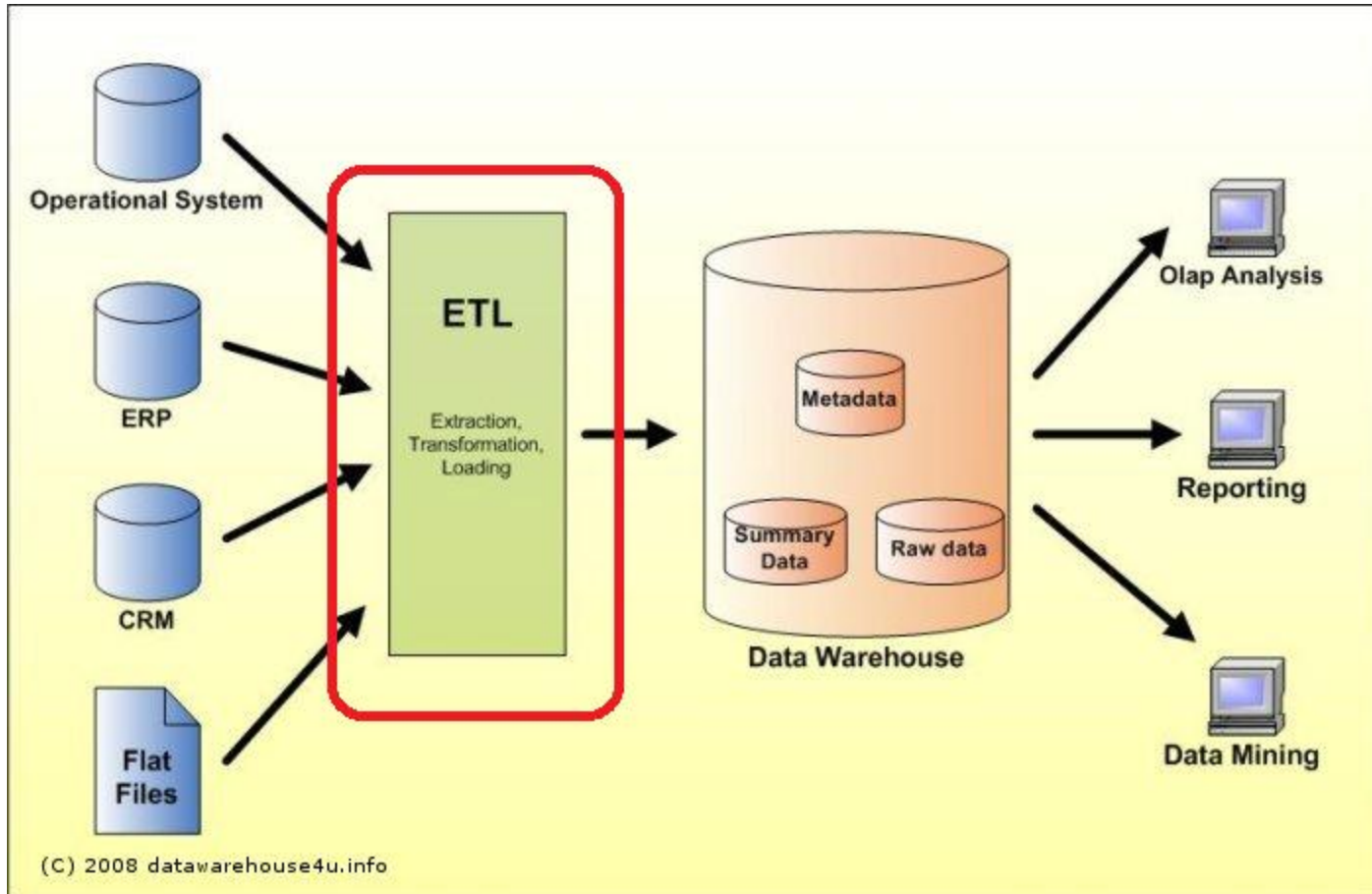
Часть 2. Пример построения Framework-а для тестирования в BI проектах

Заключение

**Пример построения
фреймворка для тестирования
в BI проекте**



Business Intelligence Project



Business Intelligence Project

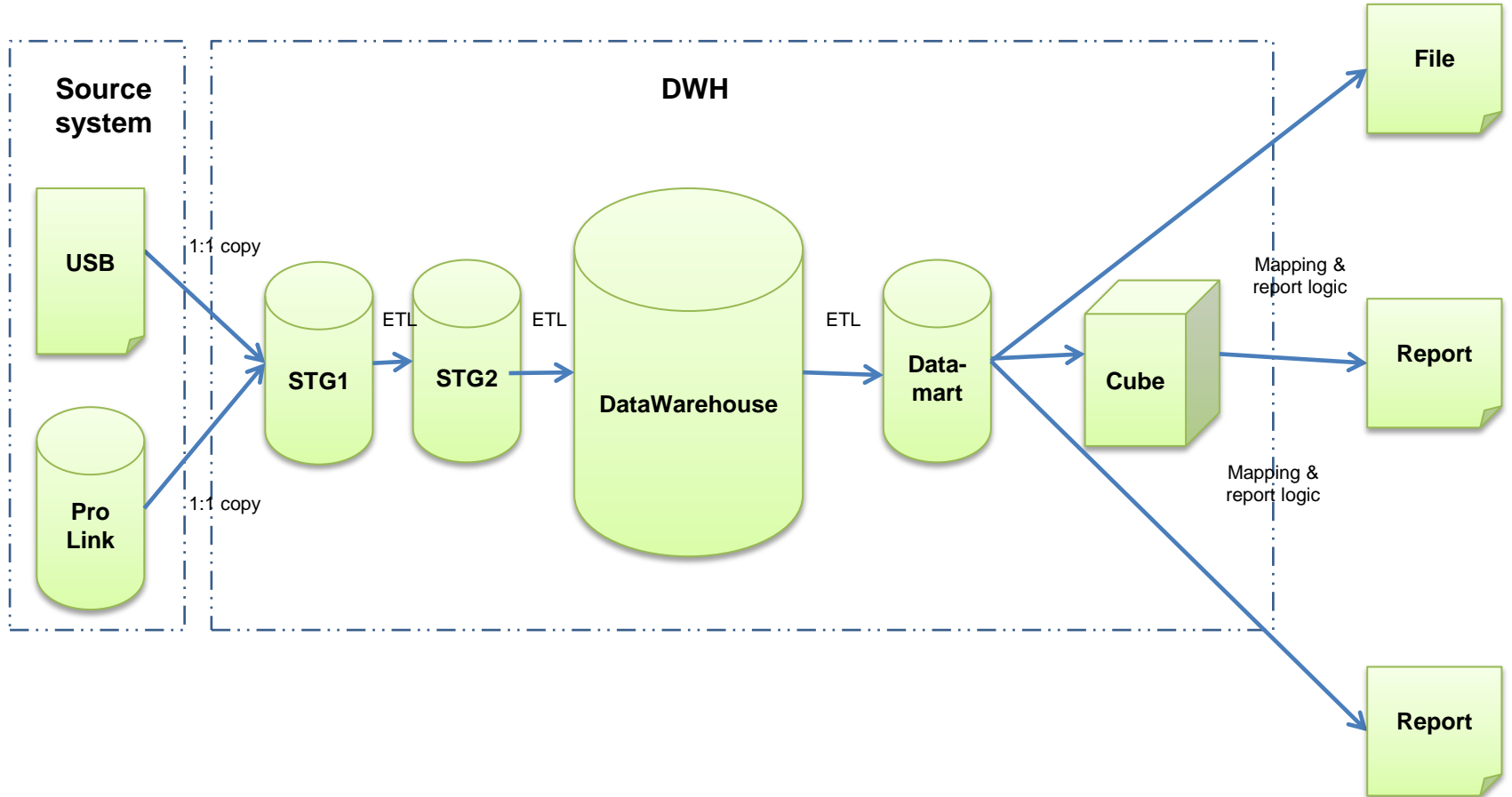
Initial Project state:

- Lots of running applications and businesses relayed on them
- We are focused on ETL development for enterprise level DWH
- Old systems should be switched off and dataflows should be switched to enterprise DWH.

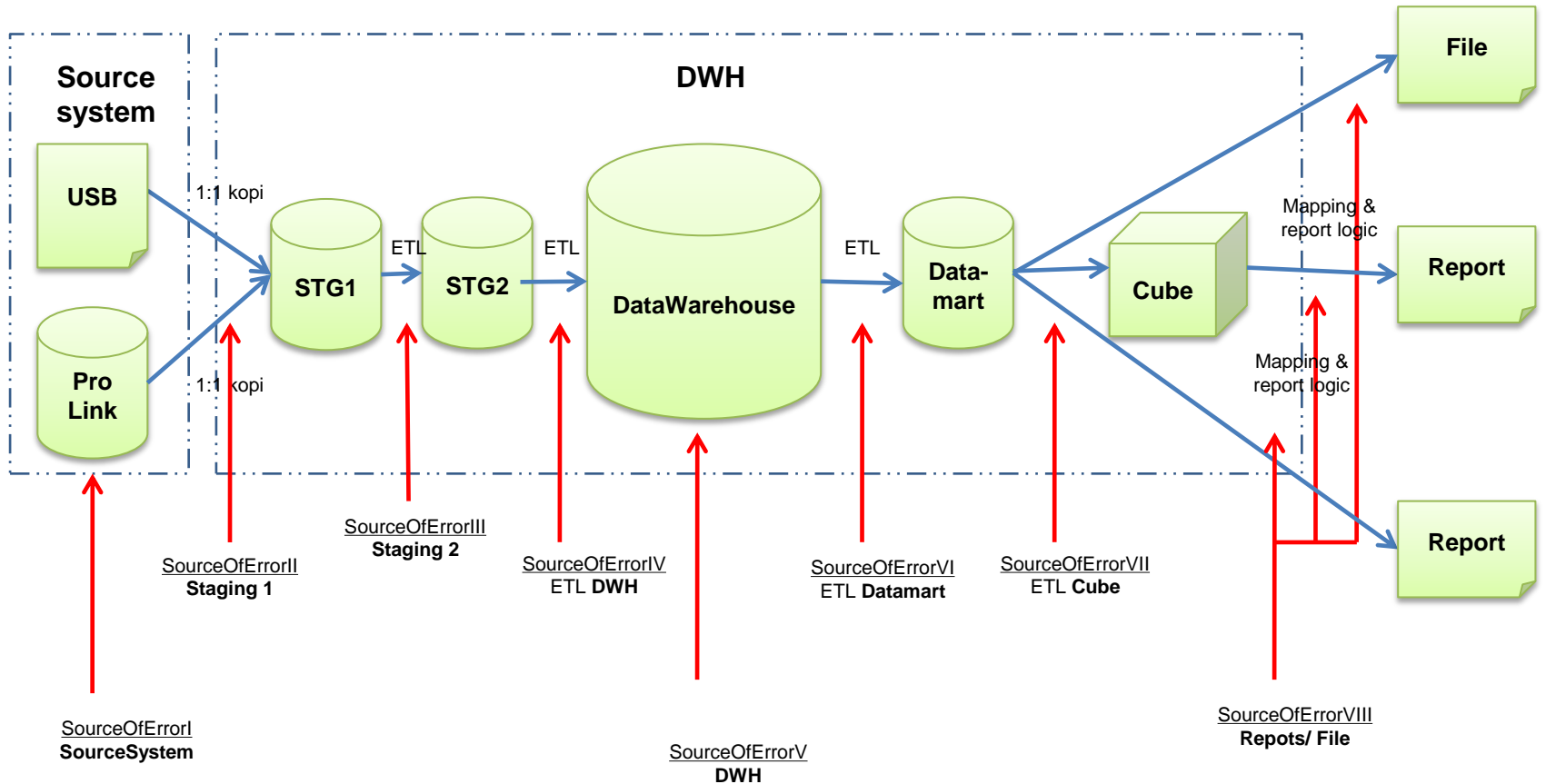
Challenges:

- Lots of legacy systems, different product versions, solution vendors, lack of documentation
- Time-shared test environments => time limitations for QA stage.
- Migration should be “transparent” for running businesses

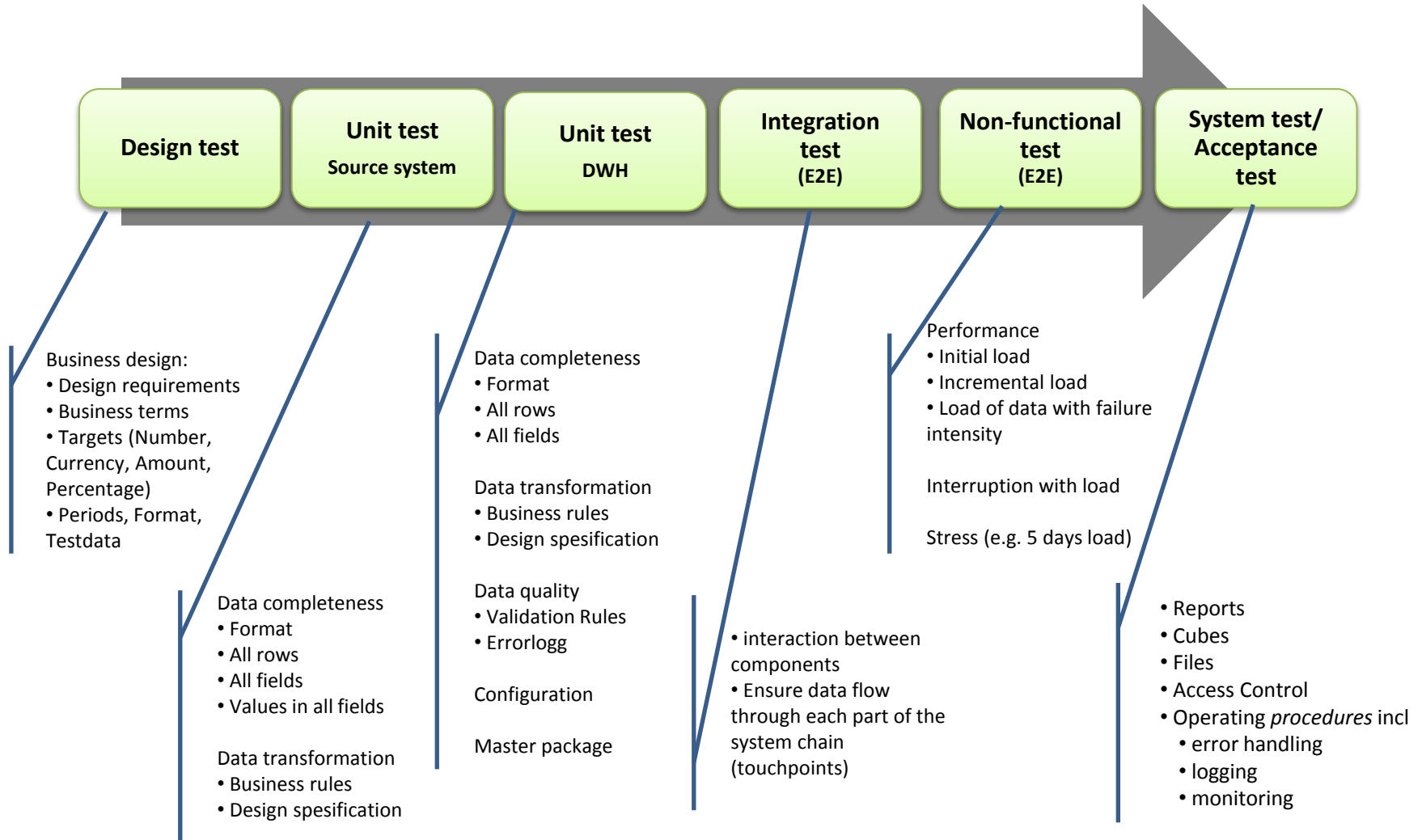
General Data Flow



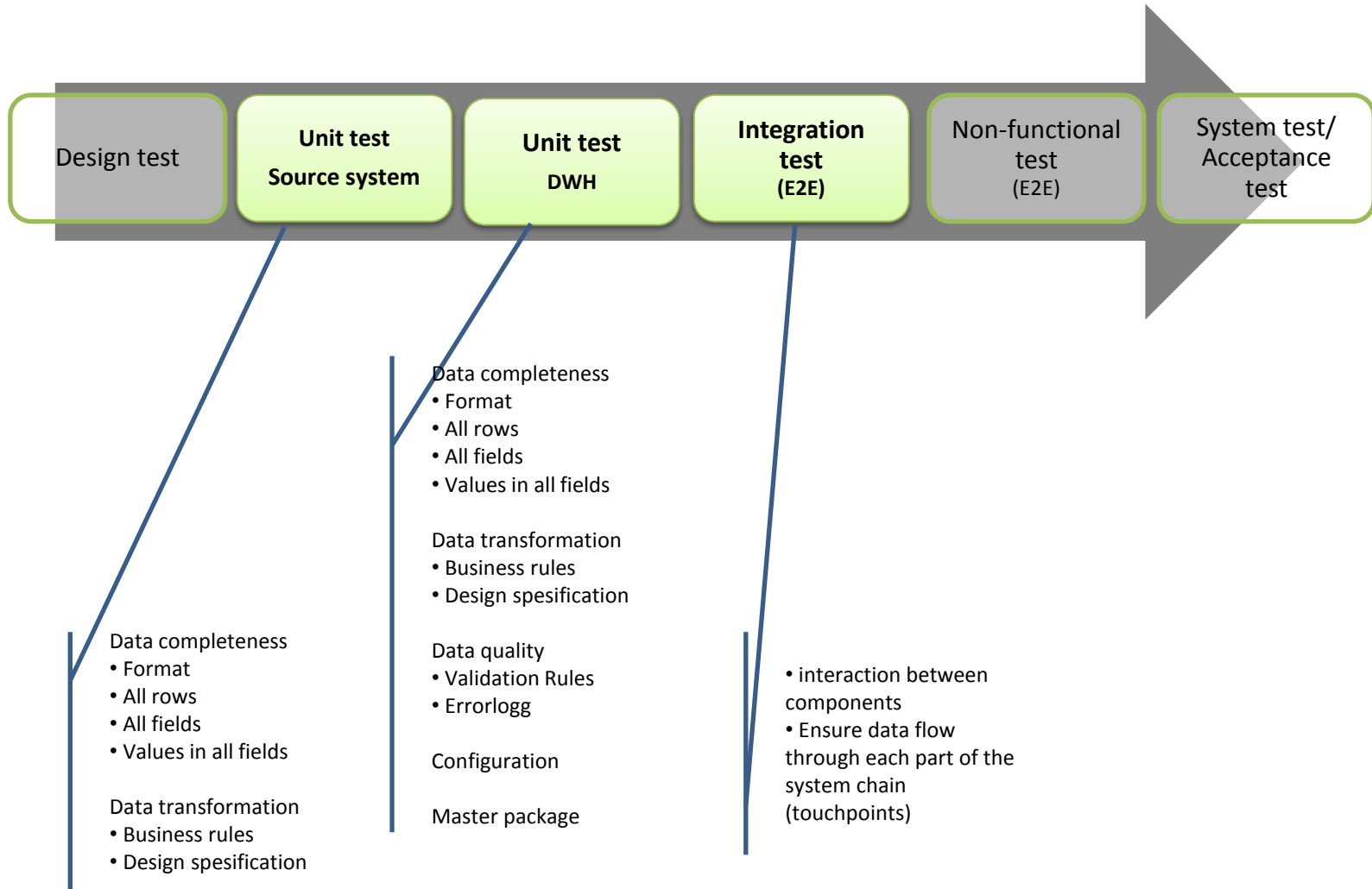
Error Sources



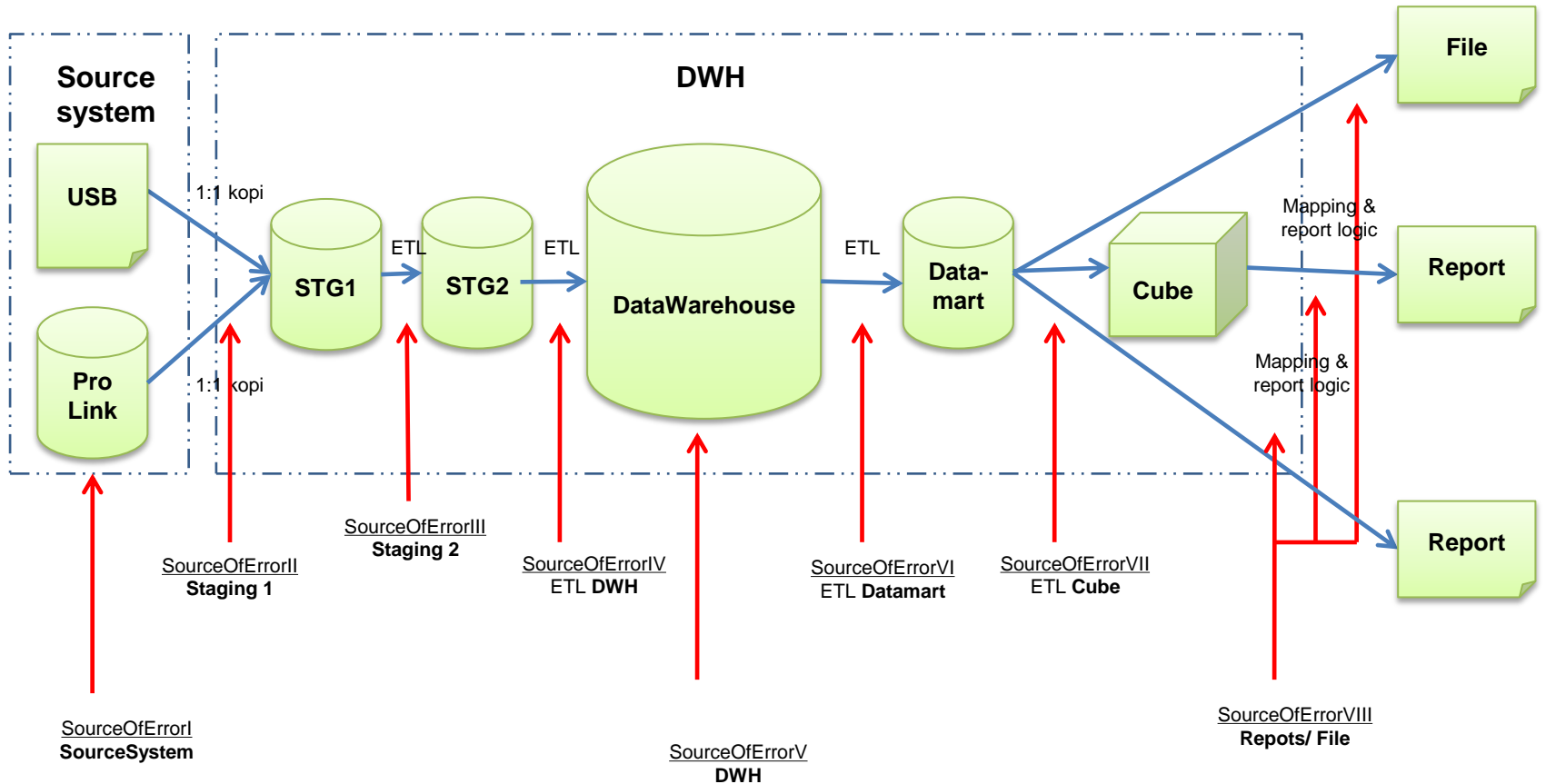
Test Level approach



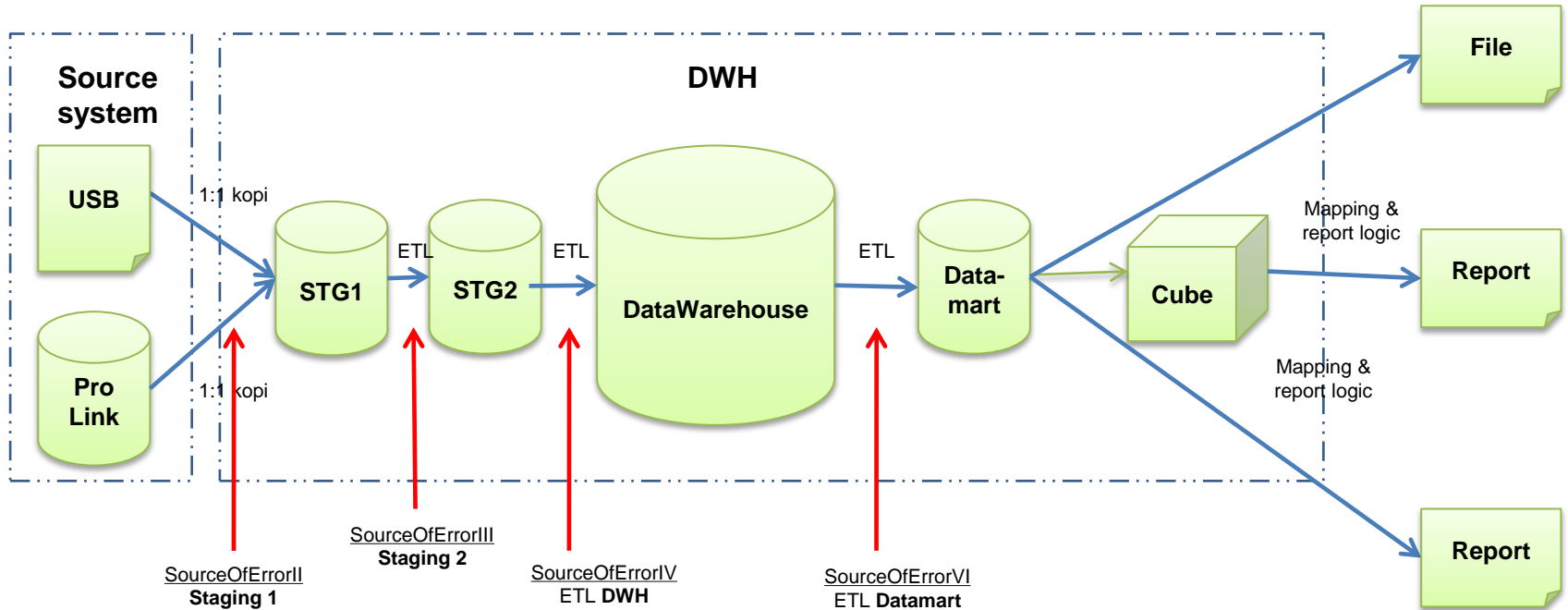
Where ETL Test Automation starts



Error Sources

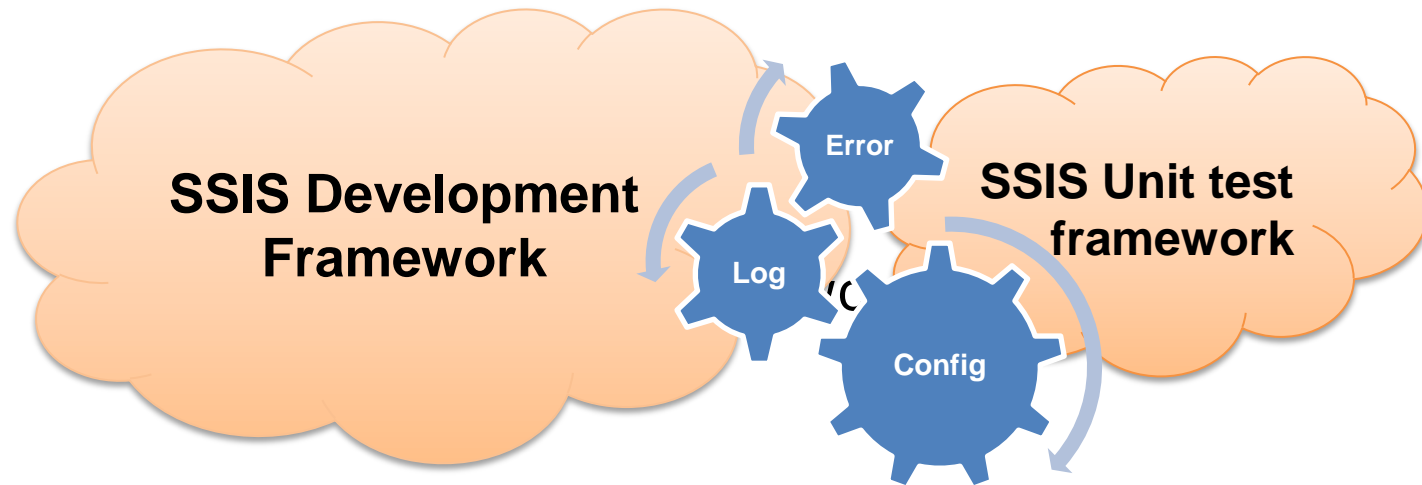


Error Sources from ETL perspective

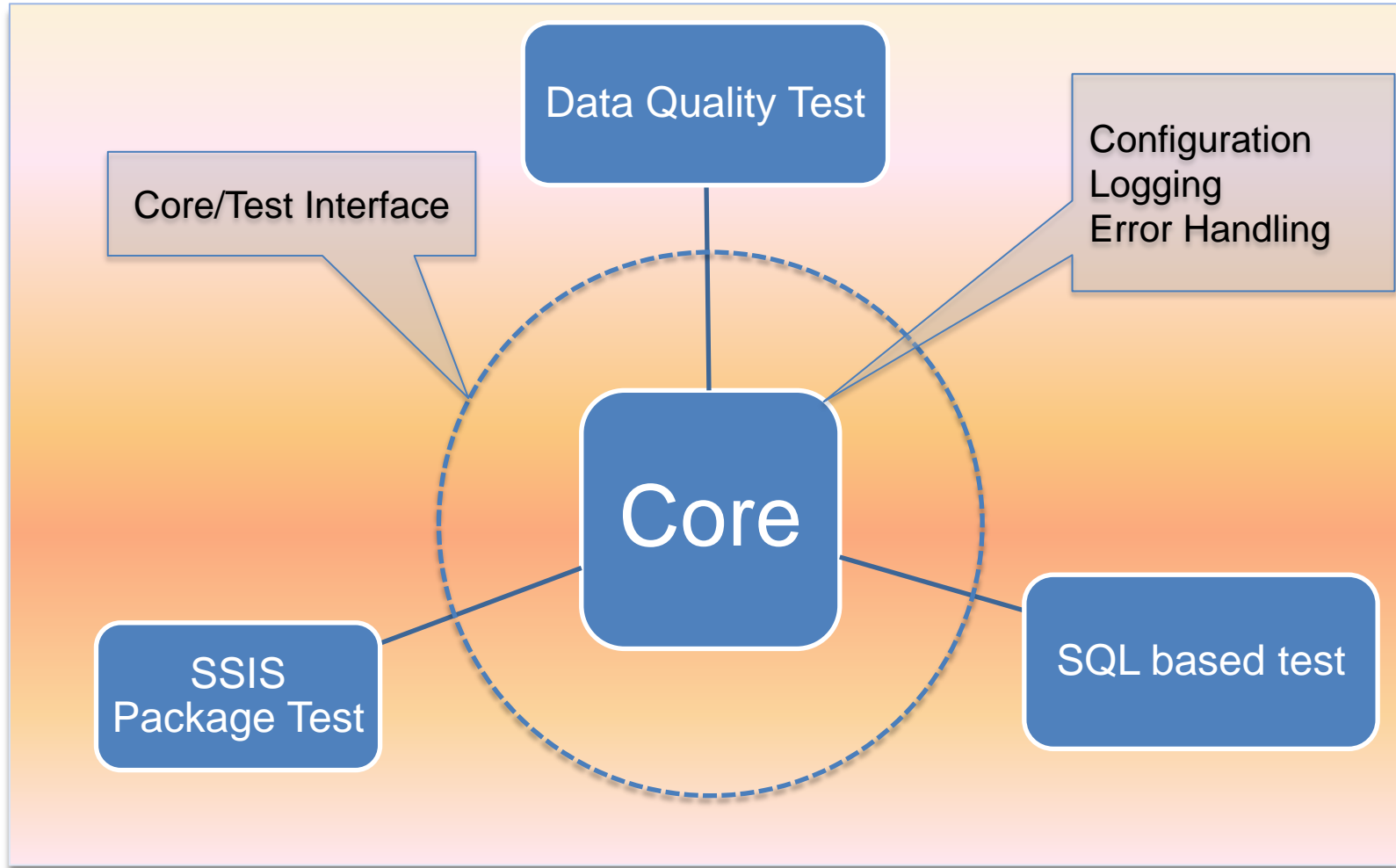


Test automation framework. Tools selection

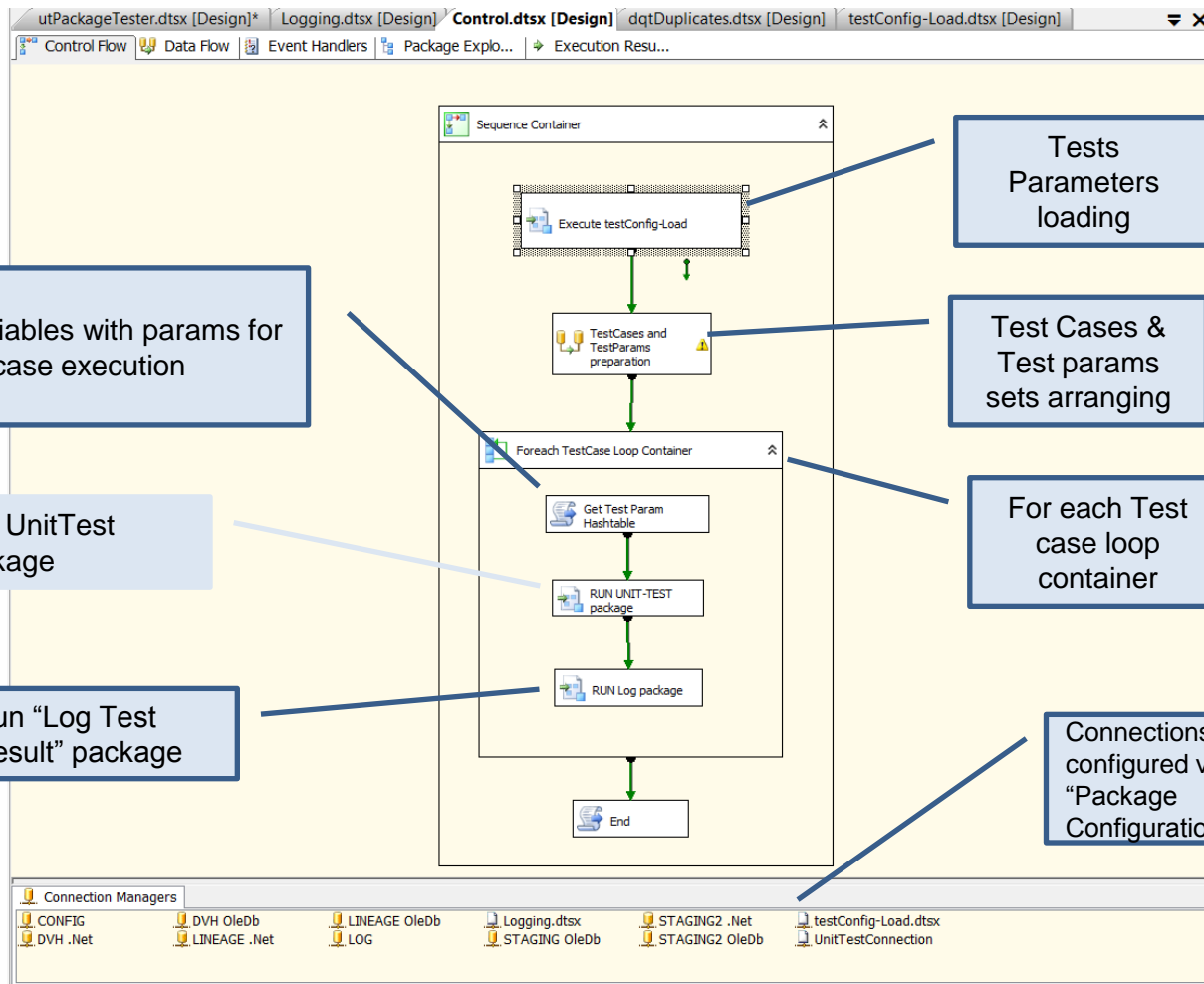
- Keep things simple for customer. BI developers would not learn C# or other programming language
- Reuse existing approaches for configuration, Logging, error handling



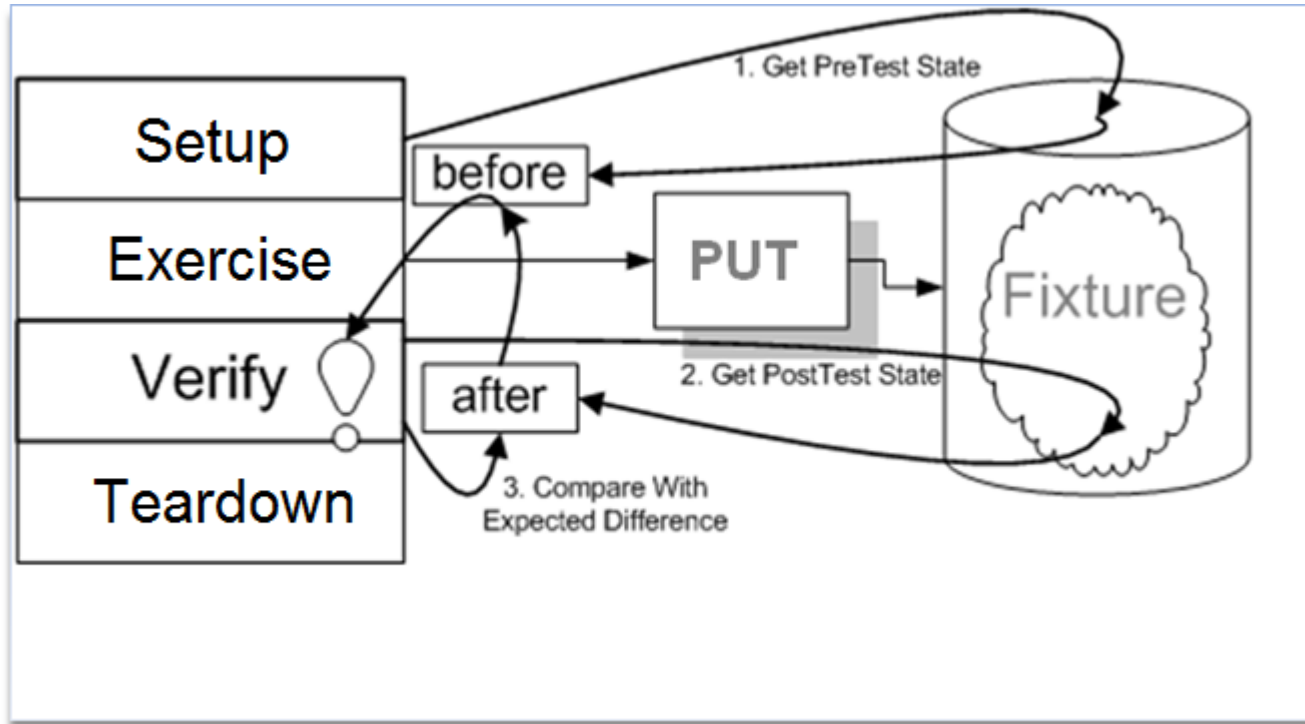
Test automation framework. Structure



Test automation framework. Control package



Four-phase test



Main idea:

- Prepare initial data
- Execute package under test (PUT)
- Validate transferred data
- Clean up database

Four-phase test. Test scenarios

Test Type	Description
Row Count	Ensures that count of tranfered rows are as expected
Boundary values	Min & Max values for numeric fields. Min/max length for varchar fields
NULL value	Try NULL values where its possible
Out-Of-Type boundary value	Misplaced data
Business Logic	Test Joins, Keys generation, Mapping, Cleansing
Random values load	Loading randomly generated values test

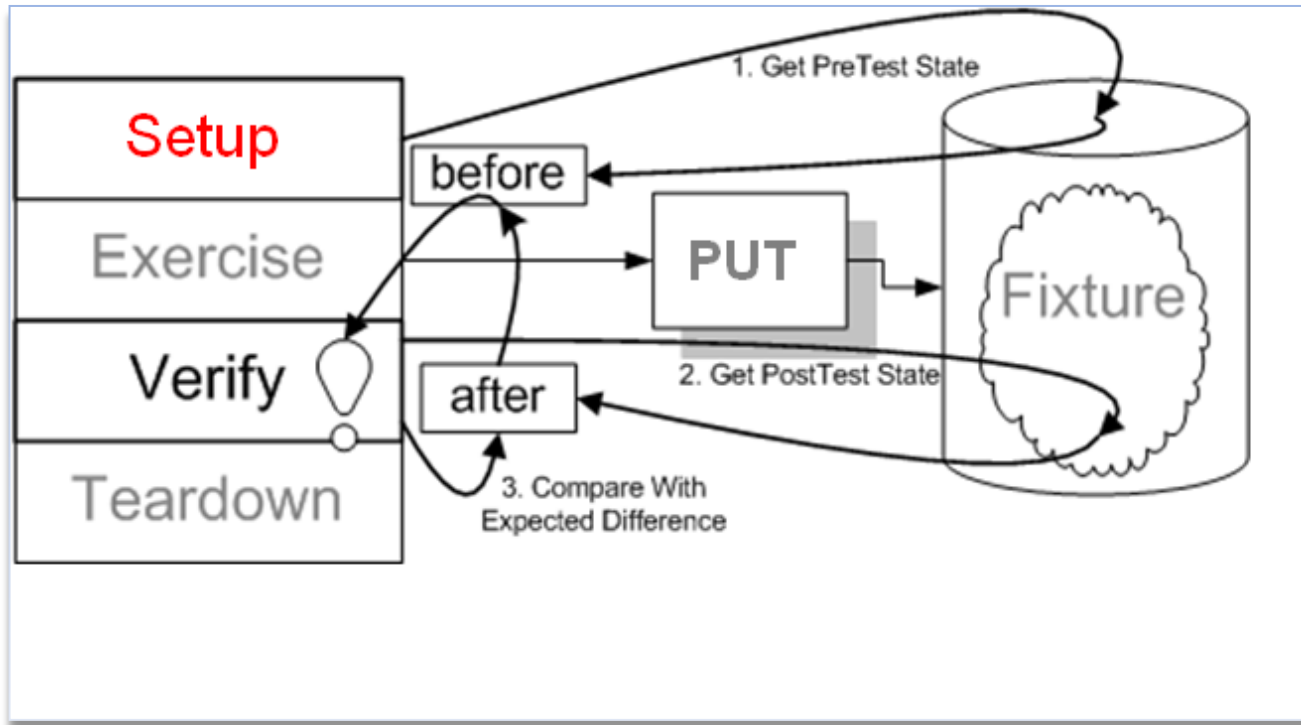
Row Count test design example

Purpose: Ensures that number of arrived rows is exactly as expected



Step-by-step test construction. “Setup” phase

Purpose: Place test data into source table

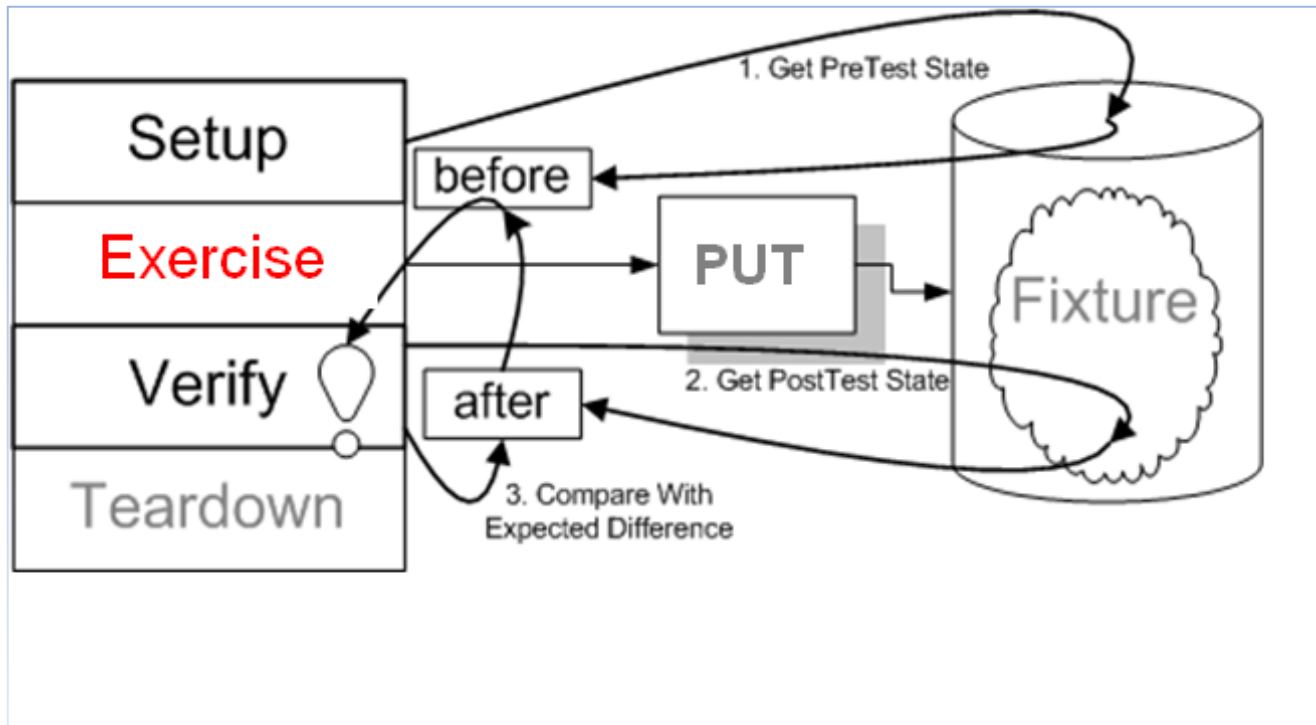


“Setup” phase. PRE- execution file example

```
DELETE from PG.STG1_Tegning
GO
insert
PG.STG1_Tegning(ENGMENT_NUM,DISTRIBUTOR_KDE,ISIN_NUM,EMISJON_NUM,VPS_KTO_NUM,TREKK_KTO_NUM,UTB
YTTE_KTO_NUM,TEGNING_DAT,INVESTERT_KAP_BEL,TEGN_OMKOSTN_PST,STATUS_KDE,SELGER_SGN,SELGER_NVN,
SELGER_EMAIL_ADR,SELGER_2_SGN,SELGER_2_NVN,SELGER_2_EMAIL_ADR,PROVISJ_SELSK_NUM,PROVISJ_KJORT_
DAT,REG_SGN,REG_TMS,SIST_ENDRET_SGN,SIST_ENDRET_TMS,KOMMENTAR_TXT,SELGER_FNR_NUM,SELGER_2_FN
R_NUM,Lastet_dat) values('11111111111111', '06', 'NO0010291933', 6, '091000836907', 15942971696, 15942971696, 'Nov 14 2006
12:00AM', 150000.00000000, 2.700, '2', 'AEA', 'XXXXXXXXXX', null, null, null, null, 1, 'Jan 9 2007 12:00AM', 'UCJ', 'Nov 14 2006
4:53PM', 'ZRR', 'Jan 9 2007 8:13AM', null, null, null, 'Jan 13 2012 11:41AM')
insert
PG.STG1_Tegning(ENGMENT_NUM,DISTRIBUTOR_KDE,ISIN_NUM,EMISJON_NUM,VPS_KTO_NUM,TREKK_KTO_NUM,UTB
YTTE_KTO_NUM,TEGNING_DAT,INVESTERT_KAP_BEL,TEGN_OMKOSTN_PST,STATUS_KDE,SELGER_SGN,SELGER_NVN,
SELGER_EMAIL_ADR,SELGER_2_SGN,SELGER_2_NVN,SELGER_2_EMAIL_ADR,PROVISJ_SELSK_NUM,PROVISJ_KJORT_
DAT,REG_SGN,REG_TMS,SIST_ENDRET_SGN,SIST_ENDRET_TMS,KOMMENTAR_TXT,SELGER_FNR_NUM,SELGER_2_FN
R_NUM,Lastet_dat) values('11111111111111', '06', 'NO0010291933', 6, '091000823988', 91001854181, 91001215421, 'Nov 15 2006
12:00AM', 150000.00000000, 2.700, '2', 'FUS', 'XXXXXXXXXX', null, null, null, null, 1, 'Jan 9 2007 12:00AM', 'EI7', 'Nov 15 2006
9:27AM', 'ZRR', 'Jan 9 2007 8:13AM', null, null, null, 'Jan 13 2012 11:41AM')
insert
PG.STG1_Tegning(ENGMENT_NUM,DISTRIBUTOR_KDE,ISIN_NUM,EMISJON_NUM,VPS_KTO_NUM,TREKK_KTO_NUM,UTB
YTTE_KTO_NUM,TEGNING_DAT,INVESTERT_KAP_BEL,TEGN_OMKOSTN_PST,STATUS_KDE,SELGER_SGN,SELGER_NVN,
SELGER_EMAIL_ADR,SELGER_2_SGN,SELGER_2_NVN,SELGER_2_EMAIL_ADR,PROVISJ_SELSK_NUM,PROVISJ_KJORT_
DAT,REG_SGN,REG_TMS,SIST_ENDRET_SGN,SIST_ENDRET_TMS,KOMMENTAR_TXT,SELGER_FNR_NUM,SELGER_2_FN
R_NUM,Lastet_dat) values('11111111111111', '81', 'NO0010291933', 6, '091000836915', 10901034414, 10901034414, 'Nov 15 2006
12:00AM', 100000.00000000, 2.700, '2', 'FWS', 'XXXXXXXXXX', null, null, null, null, 1, 'Jan 9 2007 12:00AM', 'EI7', 'Nov 15 2006
9:33AM', 'ZRR', 'Jan 9 2007 9:24AM', 'selger: helge johansen', null, null, 'Jan 13 2012 11:41AM')
Insert
.....
GO
update statistics PG.STG1_Tegning
GO
```

Step-by-step test construction. “Exercise” phase

Purpose: Perform data transfer by invoking Package Under Test (PUT)



Step-by-step test construction. “Exercise” phase

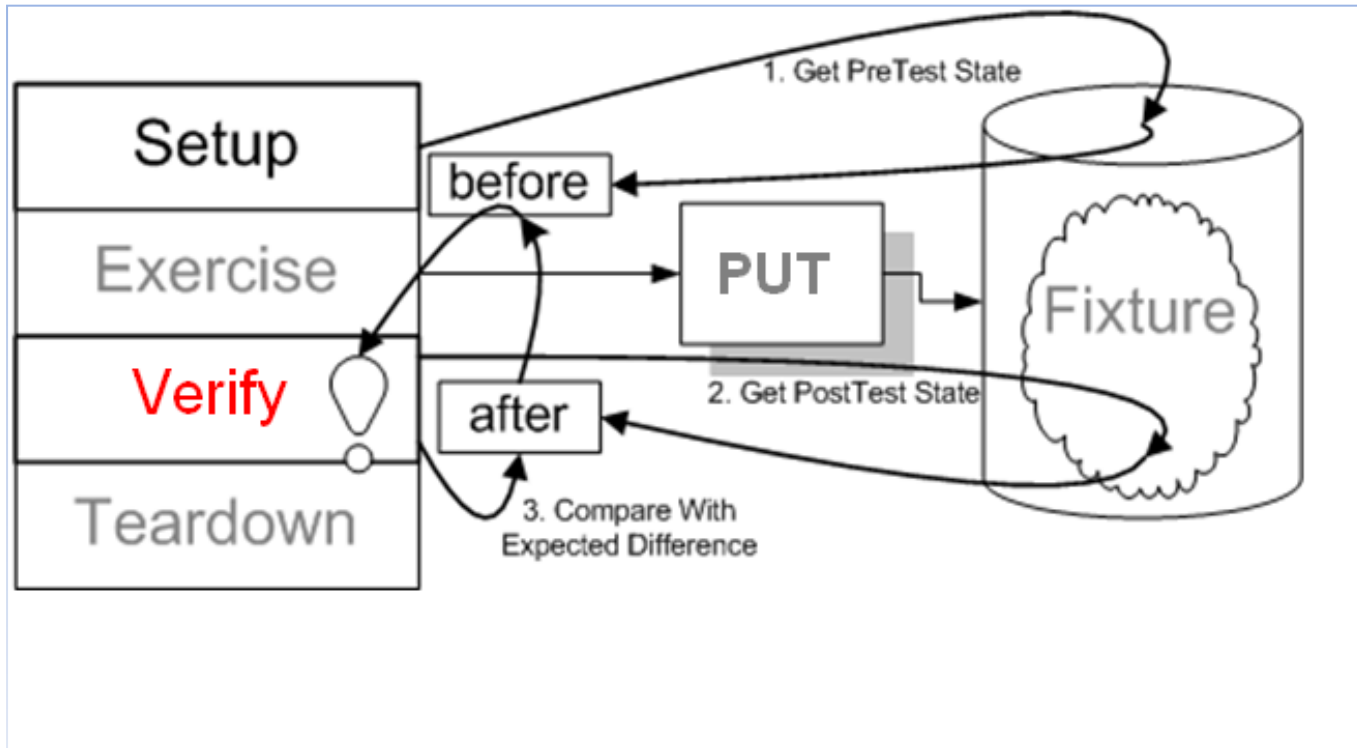
Package Under Test (PUT) launching from test runner

All we need is correct setup:

```
<test>
  <name>
    [STG2_FAK_Beholdning_128.dtsx].[Row count]
  </name>
  <description>
    RowCount ETL unit test
  </description>
  <params>
    <param name='TestType' type='String' value="unitTest"/>
    <param name='UnitTestPackage' value='UnitTestRunner.dtsx'/>
    <param name='ConnectionManager' type='String' value='STAGING OleDb'/>
    <param name='schema' type='String' value='dbo'/>
    <param name='table' type='String' value='STG2_FAK_Beholdning_128'/>
    <param name='PackageUnderTest' type='String' value='STG2_FAK_Beholdning_128.dtsx'/>
    <param name='PackagesPath' type='String' value='\\Main\ETL\SSIS\Financial Instruments\EiendomOptimer'/>
    <param name='tag' type='String' value='staging2'/>
    <param name='dummy'/>
  </params>
</test>
```


Step-by-step test construction. “Verify” phase

Purpose: Run assertion statement in T-SQL



Step-by-step test construction. “Verify” phase

1. Put names of schema and table in test parameters.

All we need is correct setup:

```
<test>
    ...
    <params>
        ...
        <param name='schema' type='String' value='dbo'/>
        <param name='table' type='String' value='STG2_FAK_Beholdning_128'/>
        ...
    </params>
</test>
```

2. Put expected row count in CheckRows template.

-- Test SQL logic

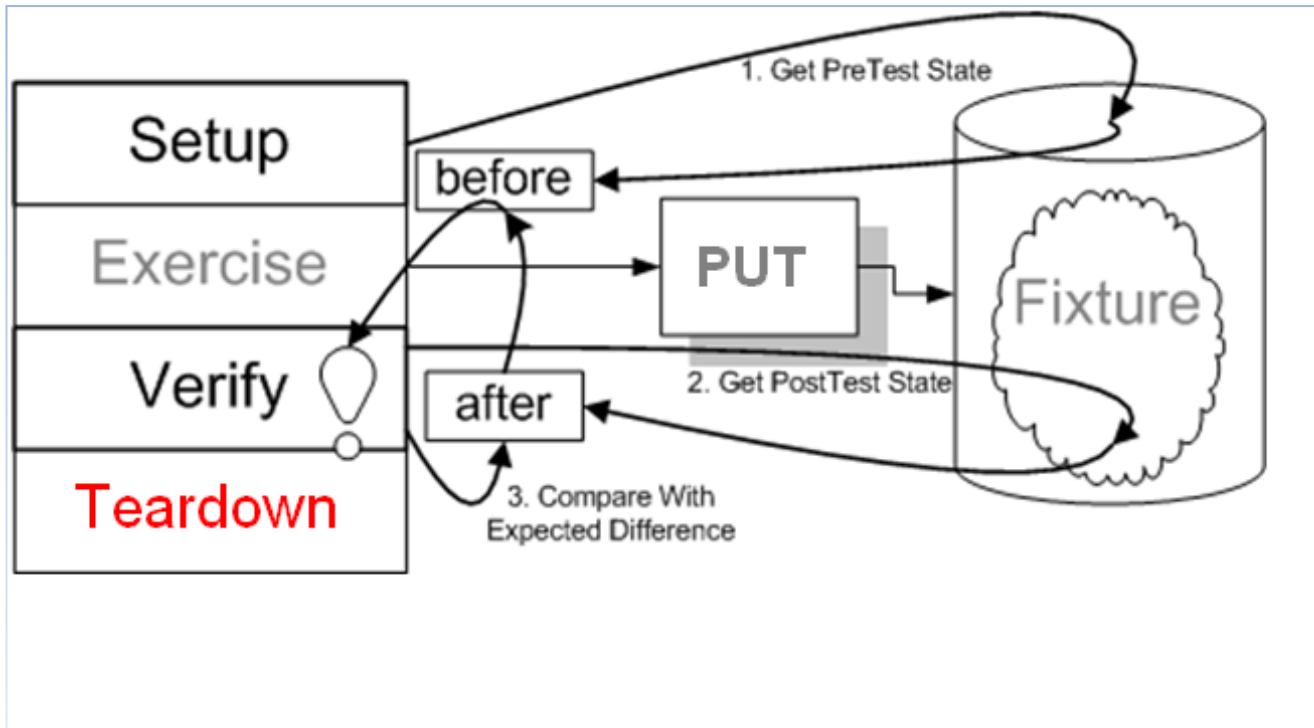
```
SET @Statement = '
DECLARE @Test_Result varchar(8000)
DECLARE @Test_Severity INT
```

```
DECLARE @exp_value int
DECLARE @act_value int
```

```
SET @exp_value = 20 --place expected value there
```

Step-by-step test construction. "Teardown" phase

Purpose: Clean tables



Тестирование в BI проектах

Business Intelligence - интересная область в разработке программного обеспечения.

Тестирование здесь особенно критично, т.к. данные предоставляемые BI напрямую влияют на принятие решений

Работая в условиях BigData, хороший тестировщик просто обречен использовать автоматизацию в своей работе

На сегодня готовых фреймворков для помощи в такой работе нет

Мы рассмотрели пример создания такого фреймворка для тестирования ETL от Microsoft с использованием SSIS

Что дальше ? Развитие фреймворка, создание новых тестов, ...

Портирование фреймворка под BI системы других производителей

Часть 1. Что такое Business Intelligence

Microsoft Business Intelligence Technology Stack

- Extraction-Transformation-Load
- Analytical Services
- Reporting Services

Тестирование в BI проектах

- Задачи
- Особенности
- Сложности тестирования

Основные вызовы стоящие перед QA в BI

Часть 2. Пример построения Framework для юнит тестирования

Тестирование в ВІ проектах





Тестирование в Business Intelligence проектах

Дмитрий Романов. Itera Consulting
dmitriy.romanov@gmail.com