SeETL

The Worlds Most

Productive ETL Tool

www.InstantBI.com

1/1/2012

A Message from Our President

Hello,

Thank you for downloading this presentation.

For many years now my colleagues and I have spent countless thousands of hours, probably tens of thousands of hours now, searching for better, faster, cheaper and more effective ways of bringing the benefits of business intelligence to our clients and to 'all companies everywhere'.

We are very proud to announce the rebranding of our software to SeETL (See-TL) for 'See ETL'.

We have rebranded our products to be called SeETL to clearly communicate our view that all ETL should be transparent and fully visible to the ETL architect. (Not opaque as is often the case!!)

SeETL marks a huge milestone in business intelligence. It is the first product that allows the generation of ALL required ETL 'objects' from the mapping spreadsheet (stored in Excel) with complete control in the hands of the BI Architect. And it then provides a complete run time environment to execute the ETL designs developed by the BI Architect.

We expect that this approach will transform the ETL portion of the business intelligence beyond all recognition over the next few years. No longer will BI projects be tied down in the morass of coding and code changes!!

We expect that in the future more and more products will be able to define their complete ETL needs in a spreadsheet workbook that a BI Architect can easily complete.

We have now been developing and integrating the SeETL DesignTime component with the SeETL RunTime component for over a year.

SeETL RunTime has proven itself on many projects over many years in many versions. The latest version proving to be the most productive yet by far. With SeETL DesignTime further reducing the time it takes for a BI Architect to define the ETL subsystem we see SeETL continuing to increase it's lead in ETL Subsystem productivity.

Now we have completed the task of rebranding to one product called SeETL we are free to focus on adding more functions into it as quickly as our customers need them.

The price and productivity benefits of SeETL are 'compelling reasons' for companies to use our software.

We look forward to your feedback on what you would like in future releases of SeETL.

instant business intelligence

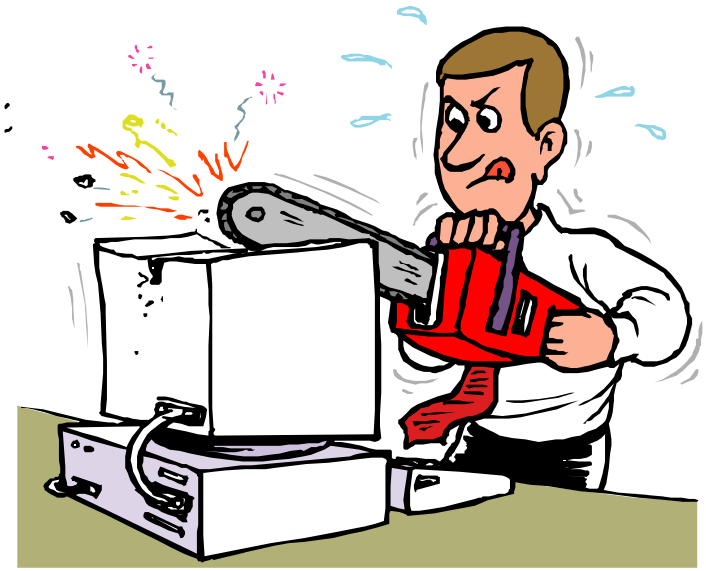
Introduction

- Let's talk about ETL
- What is SeETL?
- Who is SeETL intended for?
- SeETL DesignTime - Details
- SeETL RunTime – Details
- How much is SeETL?
- Why would I buy SeETL?
- How do I find out more?
- Summary

Let’s Talk about ETL

How Much ETL in a BI Project?

- You will spend between 50% and 70% of your 5 year BI budget on ETL
  - At least that is our experience
  - Most of it will be people time
- ETL tools are bringing this percentage down
- Higher priced BI tools are also bringing this % down!!



instant business intelligence

Issues on a BI Project

- There are many issues on a BI project
- One of the main issues is
  - As more source systems are introduced
  - As the ETL becomes more complex
  - As there is more data to understand and move
  - The effort to do this work increases dramatically
- The ETL work on a BI project can be one of the major points of failure
- Hence the success of ETL tools
- As the ETL load slows down the delivery schedule you get…

instant business intelligence

Expectation vs Delivery

- …a big difference between what is expected and what is delivered
- It's like the picture to the right
  - This is what the business thought they were promised
- And this is what was delivered
  - Result?
  - Unhappy business people

Hell
HELL
GODS - EXPEDITION
HELL

I don't want to "Go to Hell"

What are My Options?

instant business intelligence

What are My Options?

- Buy an ETL Tool from one of the vendors
  - DataStage, Informatica, Sagent, Abinitio
  - Microsoft DTS, Oracle Warehouse Builder, Cognos Decision Stream, Business Objects Data Integrator, SAS Warehouse Administrator, Talend
  - There are some others around as well
  - This will cost you more €/$/£ now and less €/$/£ later
- Develop ETL Code yourself
  - Which language?
  - C/C++, PL/SQL, stored procedures, even Cobol?
  - If you can, get some sample code and learn
  - And now, you can get ***SeETL***

What is SeETL?

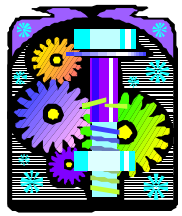
instant business intelligence

What is SeETL?

- ***The world's most productive ETL Tool***
- SeELT DesignTime
  - DesignTime Spreadsheet to Capture all ETL Rules
    - Source to target mappings
    - Definitions of jobs, commands, keys, aggregation etc.
  - DesignTime software to load spreadsheet data to the SeETL control tables
- SeETL RunTime
  - A scheduler and run time programs to run the ETL subsystem
  - Logs all processing to SeETL Reports

instant business intelligence

How Does SeETL Work?



- SeELT DesignTime
  - Enter all source to target ‘mapping’ definitions into the DesignTime spreadsheet.
  - Set the values on the processing tab for what actions you want SeETL DesignTime to load into the SeETL control database or which views you want to generate
  - Run SeETL DesignTime to perform actions
  - If you generated views, run the views on database
- SeETL RunTime
  - The scheduler waits until a batch’s pre-requisites are satisfied. It then runs the batch.
  - All commands can be run at the command line

instant business intelligence

Why is SeETL So Productive?

- All BI Projects must define source to target mappings
- This is usually done in a spreadsheet
- SeETL provides such a mapping spreadsheet
- SeETL generates the ETL subsystem from the SeETL DesignTime spreadsheet
- There is no 'coding' or 'translation' from the mapping spreadsheet to the ETL tool
- SeETL is productive because it does this translation automatically
- Therefore SeETL will remain
  - ***the world's most productive ETL tool***

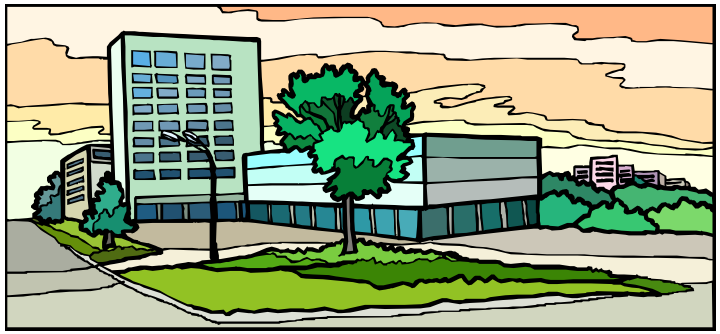
- ***SeETL is 10x more productive and 10% of the price***

Who is SeETL Intended for?

instant business intelligence

Who is SeETL Intended For?

- SeETL is intended for companies who fit into one or more of the following categories



  - You are building a BI System on a tight budget and you want to spend as little money as possible
    - Which is most companies at the moment!!
  - You have decided not to purchase an ETL tool
  - You have bought an ETL tool and want to augment the ETL tool with features of SeETL
  - You would prefer to minimise the amount of code you have to write to build the BI System

instant business intelligence

What Will SeETL Do For Me?

- SeETL will save you time and money using existing company resources. How?
  - Enabling you to develop BI Systems faster, cheaper, more easily, more reliably than ever before

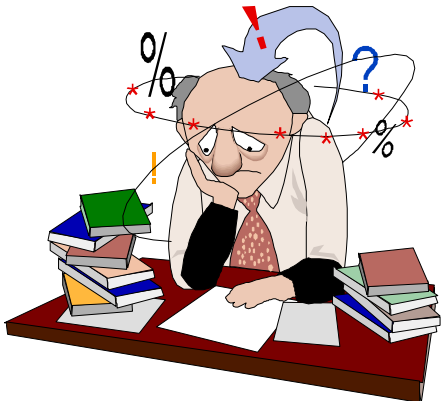
- SeETL will dramatically reduce:
  - The elapsed time to making data available
  - The costs of making data available
- SeETL will dramatically improve:
  - Your ability to run ‘specialised’ short term analysis
  - Your ability to add new summaries of information to improve performance of analyses you run on a regular basis

SeETL DesignTime
Details

instant business intelligence

What is the Mapping Problem?

- Every BI project must define source->target mappings
  - The usual way of defining these is in spreadsheets
- Every project must ‘convert’ the ‘documentation’ to a suite of code in various tools.
  - Tables, views, indexes in a modelling tool
  - Commands in scripts and schedulers
  - Transformations in ETL tools or code
  - Stored procedures? Functions? Code? Etc…



- Every project then suffers
  - Excel spreadsheet documentation and code out of sync
  - Which leads to undocumented code
  - Which leads to errors, patches and more undocumented code
  - Which leads to endless maintenance of undocumented code
  - This is true even if you have a ‘metadata tool’!!!

instant business intelligence

What is the Cost of this Problem?

- It is hard to say exactly, but it is a sizable proportion of the overall DW budget
- The 'coding' of mapping documentation to tools is easily 20% of the total cost of the DW project
  - Not considering the cost of purchasing an ETL tool
- The maintenance of code is easily 20% of the annual running costs of a large DW including license fees
- There is significant frustration in the DW team and the end user team that 'bad code' keeps happening
- We would be pleased to hear other people's opinions on the cost they experience with this problem
- On a USD1M DW project this cost will be in the order of USD200K during the project and USD100K per annum
  - The costs would be relatively large on more complex projects.

What is the Solution to this Problem?

- The solution has always been and will always be:

  Generate as much ETL code as possible from definitions developed as early as possible

- We have wondered for many years

  Is it possible to use the mapping spreadsheet as the source for generating ETL code?

- We have never seen that this was feasible
- Until now…

What is the Solution to this Problem?

- With SeETL 'Is it possible to use the mapping spreadsheet as the source for generating ETL code?'

Yes!!!

- We have seen the future: it is generated ETL
- Once source for ALL ETL + ALL code is generated
  - Use mapping spreadsheet workbook to define all ETL code
  - Generate the ETL code from the spreadsheet
  - Mapping spreadsheet is permanently 'in sync' with code
  - Only the DW Architect needs to define/understand 'mappings'
  - Define commands and job streams in the spreadsheet
  - Spreadsheet can be 'analysed/reported on' because it can be saved as an XML document

What is the Value of this Solution?

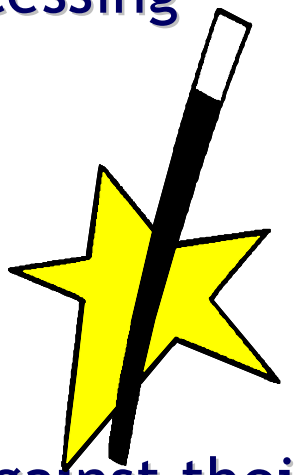
- The cost savings are HUGE
  - Mapping spreadsheet is permanently 'in sync' with code
  - Eliminate 'ETL programming' efforts
  - Eliminate 'ETL maintenance' efforts
  - Eliminate confusion between documentation and code

  - Ability to write 'any ETL Report' from the mapping spreadsheet
  - Know that there are 'no bugs in the mapping code'
    - Bugs can only be in the spreadsheet
- The productivity gain is also huge
  - The prototype can be built as the mapping takes place
  - Changes to the mappings are as easy as changing a spreadsheet!!
  - There is no delay for 'coding, testing, changing ETL and models'
  - Once the DW architect is happy with the spreadsheet the DW can be built with a 'press of the button'

instant business intelligence

How Does SeETL Work?

- The mapping spreadsheet contains well defined worksheets that contain well defined contents
- Fields have specific meanings and are interpreted in specific ways
- The User Guide defines the control data for worksheets
- The worksheet defines all other field values
- The ‘processing’ worksheet defines what processing should be performed next time SeETL is run
- SeETL then presents a trivial interface
  - Open the worksheet
  - Process the worksheet
  - Close SeETL
- Most outputs are generated ready to be run against their respective tools. Eg. Views to run against a database



instant business intelligence

SeETL Development Project

- SeETL was developed on a real world project to reduce the overall cost/elapsed time of the project
- The project has the following 'headlines'
  - One of the largest 'Version 1.0' Sybase IWS efforts to date
  - 4,000 fields in the staging area
  - 2,500 are mapped through to the EDW
  - 100+ Dimension tables, 80+ Fact tables
  - 10M rows in the 'customer' table
  - Memory mapped IO files totalling 1.7GB
  - AIX 5.1 / Oracle 9.2
  - Achieved sustained rates of 1,400 transactions/second/processor
    - Equivalent of 40M transactions in an 8 hour batch on 8 CPU machine
- SeETL = SeETL DesignTime + SeETL RunTime
- Up to the challenge of major EDW Projects!!

instant business intelligence

SeETL Usage Project

- SeETL was then further developed on a second real world project The project has the following ‘headlines’
  - 18,000 fields in each of the staging area and ODS
  - Probably only a few thousand fields in first cut EDW
  - 220+ Tables in each of the staging area and ODS
  - Around 200M rows in the ‘prototype’ on linux
  - Red Hat Linux / Oracle 9.2
- SeETL = SeETL DesignTime + SeETL RunTime
- Up to the challenge of major EDW Projects!!

Current State of Development

- We have added most ‘required’ items to SeETL.
- SeETL supports 95%+ of all ETL requirements
- Today it is possible to store:
  - All source to target mappings
    - Except association tables
  - All control tables required by SeETL^{RT}
  - The job schedule to be run by the SeETL^{RT} scheduler
  - All for SQL Server and Oracle on win2003+
  - Store table/view/index definitions that can be generated
  - Store or copy stored procedures/functions from files to database
  - And to many more to mention
- Components to be developed
  - We will survey customers on what they want

We are hard at work

SeETL RunTime
Details

instant business intelligence

What is SeETL RunTime?

- A miracle of modern programming
- An ETL tool written by BI professionals for BI professionals
- Delivers the 10x productivity at 10% price
- SeETL DesignTime captures all ETL definitions in Excel and loads them into your database
- SeETL RunTime reads these definitions and executes your ETL Subsystem
- SeETL is:
  - A complementary product to ETL tools
  - An alternative to ETL tools for smaller companies

What is SeETL RunTime?

Details

What is SeETL RunTime?

- The best way to learn is by browsing the manual
  - 200+ pages of detailed definition of SeETL RunTime
- In summary it is:
  - A scheduler that checks batch pre-requisites every N seconds
  - A set of parameterised commands that the scheduler runs
  - One program per type of processing
  - Can also be run as command line programs
    - Easy to integrate to other scheduling packages
  - Performs all required processing even for large BI Systems

- Note that SeETL RunTime may also run SQL Statements
- Let us look at the features of SeETL RunTime

instant business intelligence

What are the Key Features?

- There are three main features:
  - Codeless
  - Mappingless
  - Typeless
- Let's discuss each of these in detail

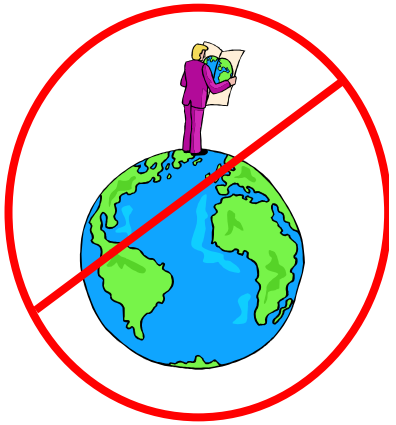
Key Feature 1 - Codeless

- The most important feature is:
  - You write ***NO CODE*** to load data from ANY data sources into a BI System.
  - We know of no other tool that will do this for you.
  - All other tools require you to drop into some development environment to implement a large complex BI System.
    Eg Stored procedures, DataStage basic, C++

- If you want new functions
  - We will write code that is common to many companies
  - The source code is C++ and the development environment is Visual Studio
- ***We have separated data structures and processing***
- One program for each 'type' of processing
- Parameters defined to control processing

Key Feature 2 - Mappingless

- You define ***NO MAPPINGS*** to map source to target
  - Again, this is unique to our knowledge
  - Mappings are only ever defined in the mapping spreadsheet
- SeETL 'discovers mappings' at run time:
  - Eg. Source and target tables passed as parameters are opened and the column names discovered. An internal map is then developed to map columns by column name from source to target.
  - Equivalent to the 'move corresponding' from cobol language



- SeETL DesignTime ensures columns have matching names
- To add/remove columns
  - You change the tables (required anyway) and SeETL Spreadsheet
  - There is no code to change because there is no code
- 'Mappingless' is the foundation for the order of magnitude productivity gains

instant business intelligence

Key Feature 3 - Typeless

- SeETL is unaware of data types
  - It moves columns based on name
  - If ODBC can move the data to the target column based on implicit data conversions the data will be moved regardless of data type
  - If the ODBC driver cannot do the data conversion it will issue an error message and stop
  - Well, almost completely, there are some specific cases where it knows the data type of a column

- In SeETL everything is a character string
- There are tools provided to check data types to detect data movements that will cause errors
- No need to check the data types inside ETL code
- No need to get rid of data type mismatch messages from ETL jobs

instant business intelligence

Other Features

- SeETL is not ‘case sensitive’
  - Parameters, column names etc are compared ‘case insenstive’
  - Data is compared ‘case insensitive’
- Because all processing is performed via views there are no ‘restrictions’ on your underlying tables
- All ODBC data types <32K are supported
  - Chunking is not supported, so no graphics loading is supported
- SeETL is null aware
  - Virtually any field can be null. (Portions of keys cannot be null)
  - Change detection for Type 2 dimension fields is null aware. If a field goes from having a value to null it is detected as a change.

instant business intelligence

Other Features

- SeETL includes auditability:
  - The parameters used to invoke a program are stored. This includes userid/password, date time, tables accesses.
  - The number of records read/written are logged to an audit trail
  - In the future you will be able to specify field names for hash totals

- SeETL includes extensive error checking and reporting
  - There is a separate message table and an extensive suite of error conditions are checked for and reported

How Much is SeETL?

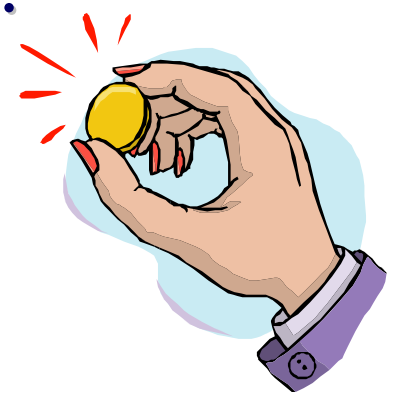
instant business intelligence

How Much is SeETL?

- SeETL DesignTime is free
  - We do insist that users register to our web site for the free software that is available there.
- SeETL RunTime is purchased at two levels:
- Executables Only

  or

- Executables + All Source Code

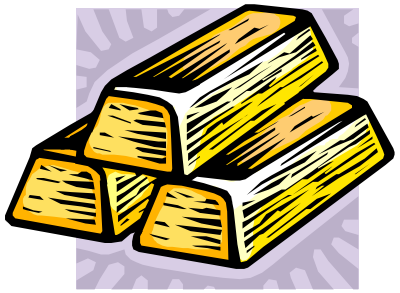


10x productivity
10% of the price

- Current prices are available on request
- Volume purchase agreement applies for multiple servers

What Do I Get For My Money?

- Complete documentation of how to use SeETL
  - Available as downloads from our web site
- Executables Only
  - The suite of executable programs



Solid Gold !!

- Executables + All Source Code
  - You bring SeETL into your organisation as a piece of software you own and can further develop for your own needs
  - You will be able to see the details of how SeETL actually builds star schema data warehouses
  - The source code is heavily documented
- For the annual maintenance you will receive:
  - Q&A support via email (Phone/on-site support will be chargeable)
  - Enhancements to SeETL as they are developed

What Do I Get For My Money?

- Much more than the code, you also get
  - Access to 17+ years of star schema implementation embedded in code
  - Vastly reduced time and effort to implement star schema databases
  - Executable code that you **KNOW** works
    - When you use ETL tools you have to test every job in detail
    - With SeETL you just put data in at the front end and check it at the back end. If it checks out ok the code executed correctly.
    - SeETL moves data based on column names. So once the views are set up properly there is very little testing to do.
  - Low cost maintenance because the majority of configuration is via views
  - All at negligible cost compared to today's ETL tools

Why Would I Buy SeETL?

instant business intelligence

Why Would I Buy SeETL?

- ✓ I am doing my BI project on a tight budget and I don't have the money to buy an ETL tool right now.
- ✓ We are new to BI and our project sponsor wants proof of the value of BI in our company before he/she will spend the money to buy the infrastructure for the real BI System. (eg. Hardware, ETL Tool, Query Tools etc)
- ✓ We are new to BI and we want to build our first few stars and learn how to do it properly before we spend the money to buy the infrastructure for the real BI System. (eg. Hardware, ETL Tool, Query Tools etc)
- ✓ We have decided to use a combination of C/C++ and DTS to build our BI System. We are looking for some utilities that help us in this environment.

instant business intelligence

Why Would I Buy SeETL?

- ✓ I have the money to buy an ETL tool….But my IT people already know C++ and perhaps we can get our early work done without the ETL tool.
  - ✓ I am concerned that if I buy an ETL tool now we will just slow down the initial implementation while we learn/train on the tool.
  - ✓ I am concerned how much I might have to pay for external consultants to help us use the tool properly.
  - ✓ I'm thinking we should get the first models up and running, prove the project, and then purchase the ETL tool when we know more and are more able to prove the value of it.
- ✓ We recognise the value of the aggregate keys for summaries and the incremental update of summaries. We are trying to do this in our ETL tool and found that it is difficult. We'd like to use the IDW to support allocating keys and building summaries.

instant business intelligence

Why Would I Buy SeETL?

- ✓ Our fact tables are changing on a regular basis. It takes a long time to add the new columns in the ETL tool we have. We are interested in doing the bulk of the calculation work in our ETL tool but populating the fact tables using SeETL because it automatically adapts to new columns with no code changes.

- ✓ We already have our BI System up and running. We built it without an ETL tool. Now we are making changes and adding new data we are looking around at what tools might be available to help us.

instant business intelligence

How Do I Find Out More?

- A sample mapping spreadsheet is on our products page
- It works on the sample database on our products page
- SeETL is on our products page
- You are welcome to download all components and try them out for yourself!!
- When you try it out for yourself you will see just how easy it is to define source to target mappings

Try SeETL for yourself.
See how fast you can write your ETL now!!!

instant business intelligence

Summary

- Let's talk about ETL
- What is SeETL?
- Who is SeETL intended for?
- SeETL DesignTime - Details
- SeETL RunTime – Details
- How much is SeETL?
- Why would I buy SeETL?
- How do I find out more?
- Summary

Thank You for Your Time!