

DW Architecture

More than just Facts & Dimensions

Martin Schoombee



#131 | PHOENIX 2012

About Me (**NOT**)

- I have **NOT** presented at a SQL Saturday
- I have **NOT** published any books
- I have **DO NOT** have a Ferrari
- I **DO NOT** have a Boston accent

About Me

- Senior BI Consultant at Insight
- MCITP – **BI Developer 2008**
- Blog – www.did-i-say-that.com
- Twitter – **@sqlmartin**
- SpeakerRate – **Martin Schoombee**

The Facts of Life



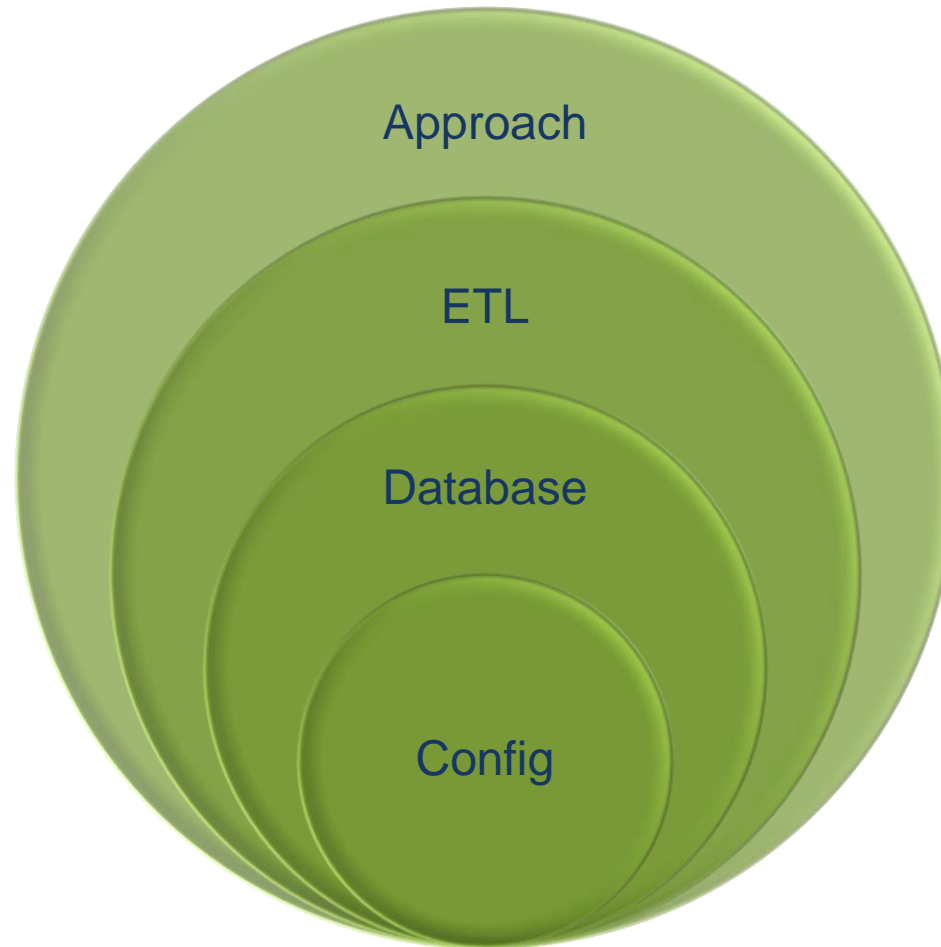
The Facts of BI

- Requirements will be **vague** (at best)
- Change is **inevitable**
- Growth will be **exponential**

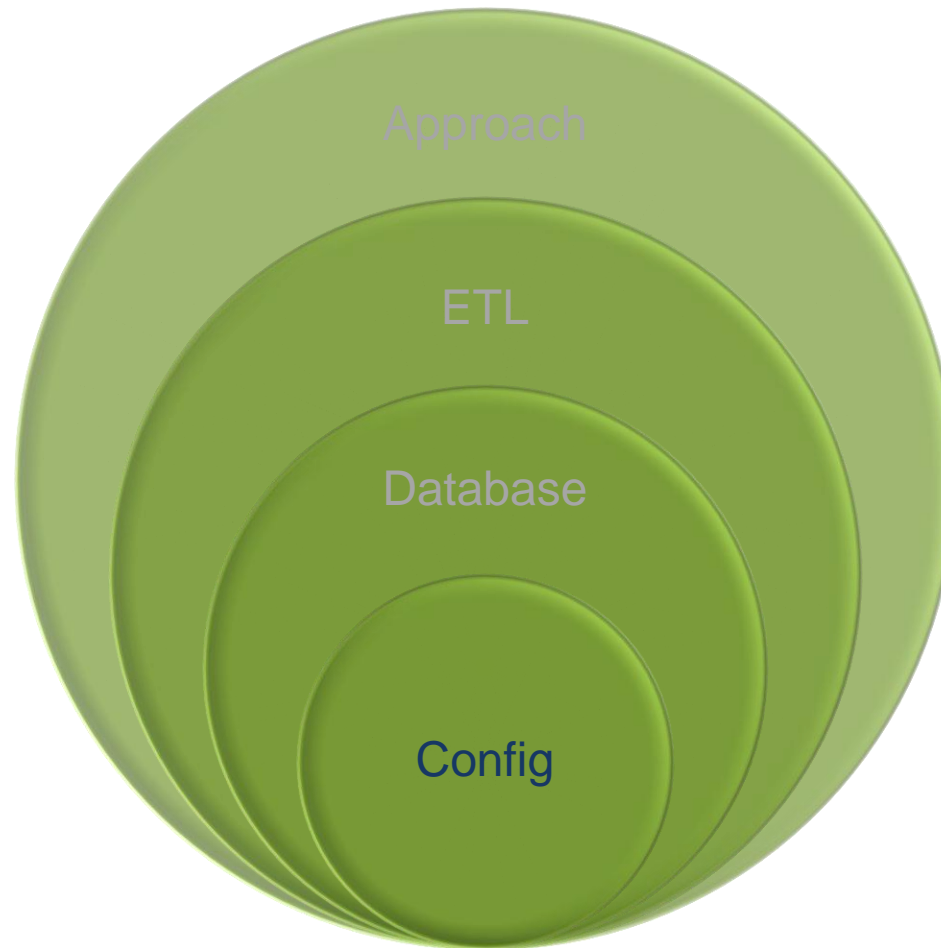
The Key to Success

- Do the things that **nobody** has asked you to do...

Agenda



SQL Server Configuration



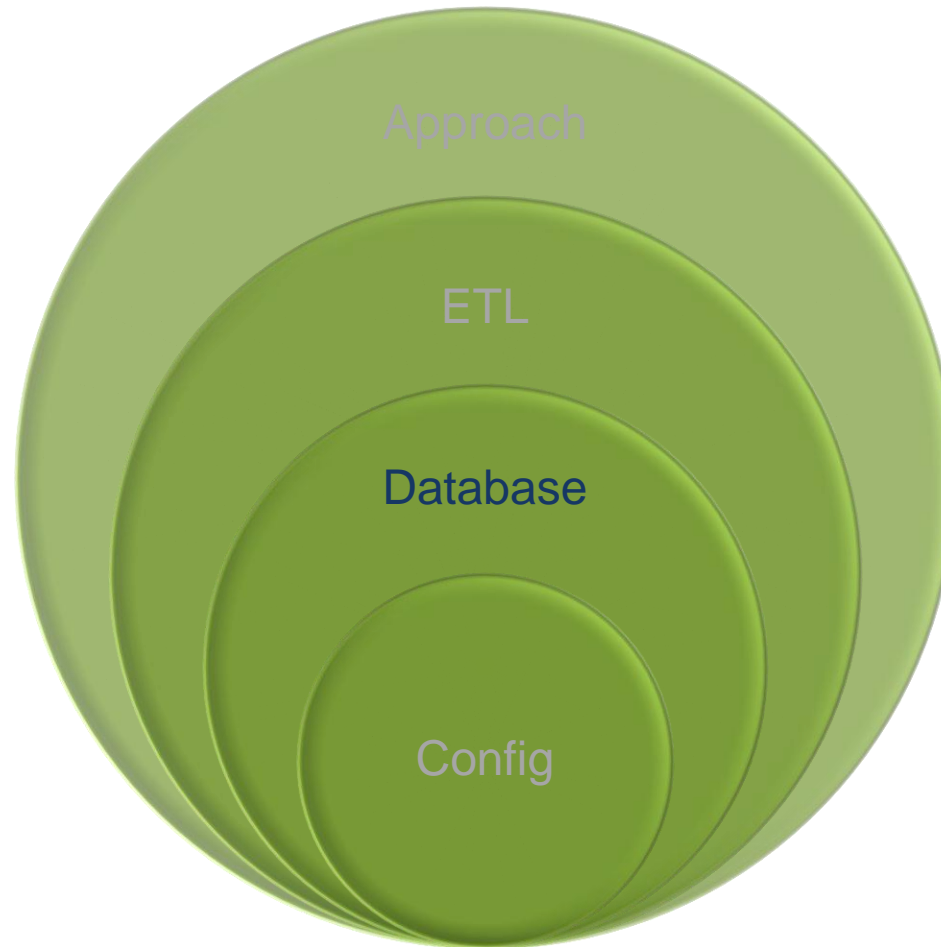
SQL Server Configuration

- File Autogrowth Settings
 - MB, **NOT** %
 - Fail-safe Only
- Log Files
 - Separate Drives
 - RAID 1+0 Ideal

SQL Server Configuration

- Tempdb
 - $\frac{1}{4}$ to $\frac{1}{2}$ of Processor Cores (Paul Randal)
 - Autogrowth: MB, NOT %
- MAXDOP
 - Less than 8
 - Never More than Physical Processors
 - Monitor CX-Packet Wait-Types
 - Test with Typical Workload

Database Areas



Database Areas

- Admin
- Staging
- Reference
- Data Mart

Admin Area

- Separate Database? **Yes**
- Why?
 - Flexibility; Backup Strategy; Recovery
- Contents
 - Metadata
 - Logs
 - ETL Parameters

Staging Area

- Separate Database? **Yes**
- Why?
 - Flexibility
 - No Data Backups Required

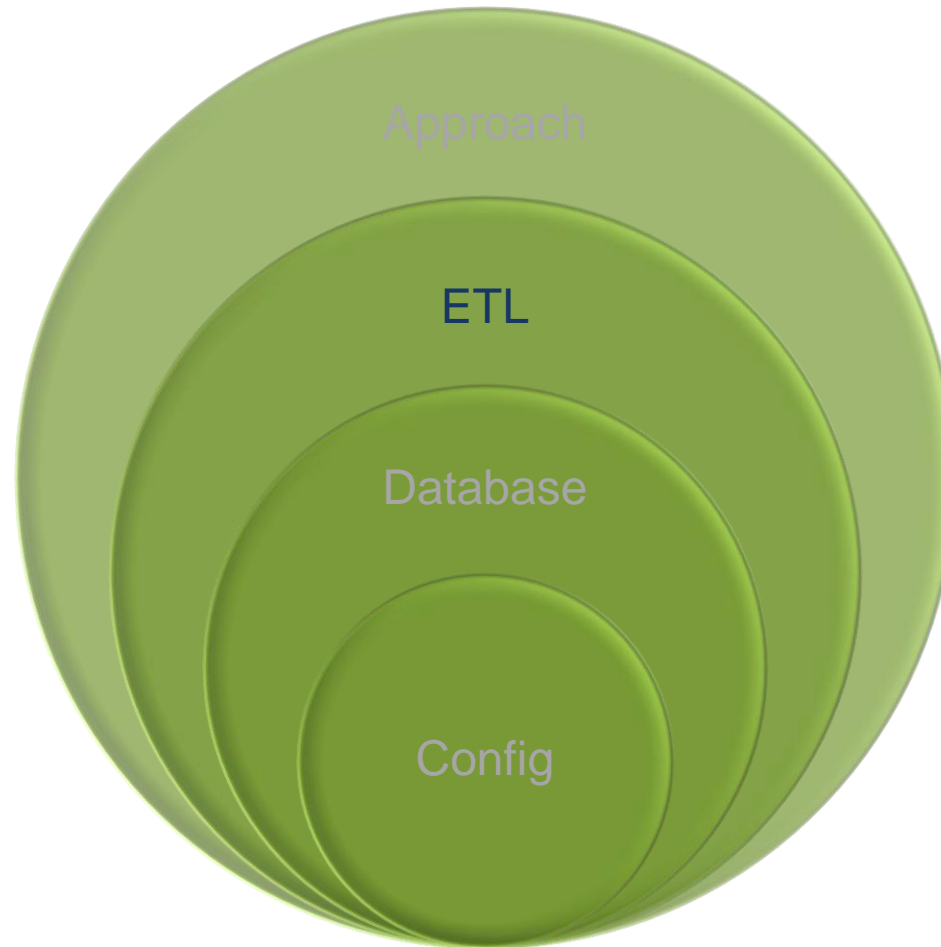
Reference Area

- “Sandbox” Area for Advanced Users
- Separate Database? **Yes**
- Why?
 - Flexibility; Backup Strategy; Recovery
- Contents:
 - Maintained Tables
 - Generic Lookup Table
 - Tally Table

Data Mart Area

- Separate Database? **Yes**
- Contents
 - Facts
 - Dimensions
 - Snapshots
 - Helper/Bridge Tables
- Use Schema's for Logical Separation

ETL Process



ETL Process

- “Engine Room” of Data Warehouse
- Spend 80% of Your Time on ETL
- Extract Data Quickly
- Parameter-Driven
- Sliding Window Approach
- Automate Fully

ETL Process: SSIS

- Choose Configuration/Logging Options Early
- Create Templates
- Use Queries Against Data Sources
- Oracle/Teradata: Attunity Connectors
- SCD Component: **DO NOT USE IT**
- Experiment with Buffer Sizes
- Parallel Processing

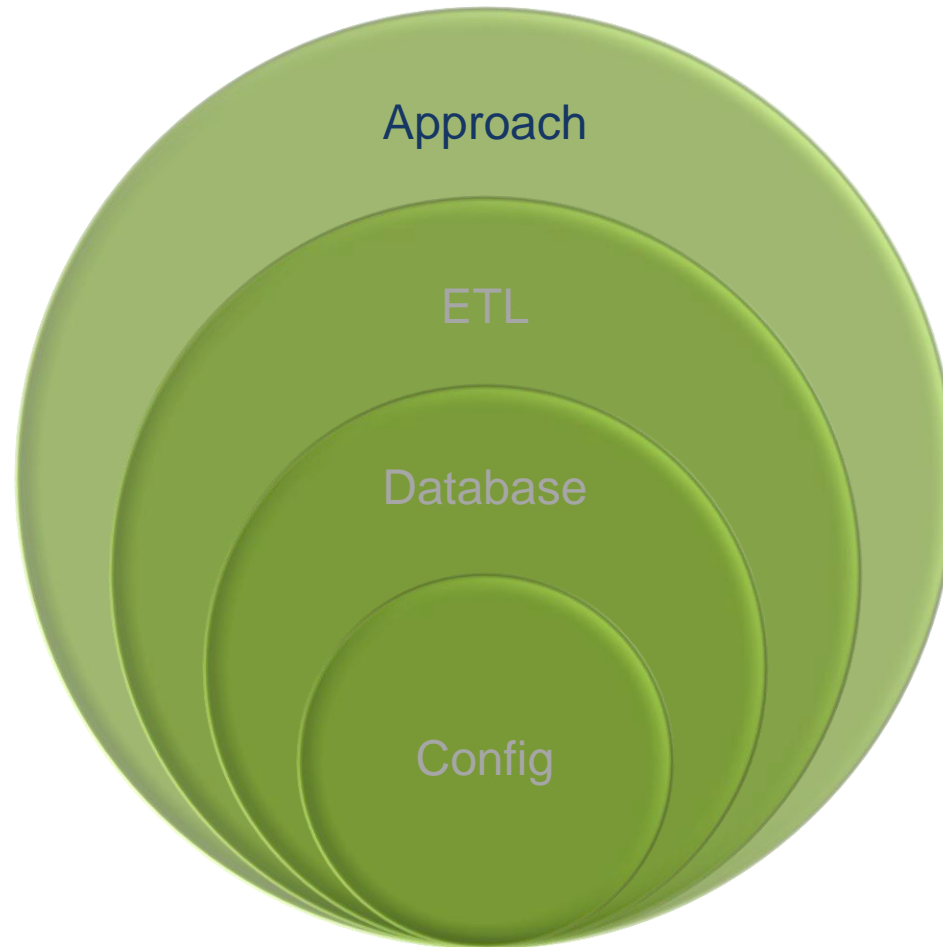
ETL Process: Staging Tables

- Utilize Schema's
- Drop & Recreate Indexes
- Mirror Source System Names
 - Map Source Objects 1:1
- Varchar for All Source Fields
- The Use of Identity Fields
- Persist Lookup Values
- Experiment with Computed Columns

ETL Process: Considerations

- Stored Procs vs. SSIS Components
- Consistency
 - Naming Conventions
 - Framework
- Operational Data Store (ODS)

Approach



Data Mart Objects

- Date Dimension
 - Surrogate Key: <yyyymmdd>
 - Comprehensive
- Dimensions
 - Numeric vs. Varchar for Codes etc.
 - Type 1,2 for the Same Subject
- Facts
 - Measures: Compatible Data Types

Reporting Layer

- Cubes
 - Qualify Necessity
 - Calculated Measures
- Reporting Services
 - Create Templates
 - Admin Reports
 - Use Stored Procedures
- Excel

Measure of Success

- Automation
 - Manual Interventions
 - Failures
- Performance
- Reports/Output
 - Data Manipulation
 - Intuitive Interface
 - End-User Adoption

The Key to Success

- Do the things that **nobody** has asked you to do...
- Anticipate, Prepare, Deliver, Repeat

Martin Schoombee

- Senior BI Consultant at Insight
- MCITP – **BI Developer 2008**
- Blog – **www.did-i-say-that.com**
- Twitter – **@sqlmartin**
- SpeakerRate – **Martin Schoombee**