About Me

- Business Intelligence Consultant, in IT for 28 years
- Owner of Serra Consulting Services, specializing in end-to-end Business Intelligence and Data Warehouse solutions using the Microsoft BI stack
- Worked as desktop/web/database developer, DBA, BI and DW architect and developer, MDM architect, PDW developer
- Been perm, contractor, consultant, business owner
- Presenter at PASS Business Analytics Conference and PASS Summit
- MCSE for SQL Server 2012: Data Platform and BI
- SME for SQL Server 2012 certs
- Contributing writer for SQL Server Pro magazine
- Blog at JamesSerra.com
Agenda

- Why use a Data Warehouse?
- Fast Track Data Warehouse (FTDW)
- Appliances
- Data Warehouse vs Data Mart
- Kimball and Inmon Methodologies
- Populating a Data Warehouse
- ETL vs ELT
- Surrogate Keys
- SSAS Cubes
- SQL Server 2012 Tabular Model
- End-User Microsoft BI Tools
Why use a Data Warehouse?

- All these reasons are for data warehouses only (not OLTP):
  - Reduce stress on production system
  - Optimized for read access, sequential disk scans
  - Integrate many sources of data
  - Keep historical records (no need to save hardcopy reports)
  - Restructure/rename tables and fields
  - Use Master Data Management, including hierarchies
  - No IT involvement needed for users to create reports
  - Improve data quality and plugs holes in source systems
  - One version of the truth
  - Easy to create BI solutions on top of it (i.e. SSAS Cubes)
Why use a Data Warehouse?

Legacy applications + databases = chaos

<table>
<thead>
<tr>
<th>Production Control</th>
<th>MRP</th>
<th>Inventory Control</th>
<th>Parts Management</th>
<th>Logistics</th>
<th>Shipping</th>
<th>Raw Goods</th>
<th>Order Control</th>
<th>Purchasing</th>
<th>Finance</th>
<th>Marketing</th>
<th>Sales</th>
<th>Accounting</th>
<th>Management Reporting</th>
<th>Engineering</th>
<th>Actuarial</th>
<th>Human Resources</th>
</tr>
</thead>
</table>

Enterprise data warehouse = order

- Continuity
- Consolidation
- Control
- Compliance
- Collaboration

Single version of the truth

Enterprise Data Warehouse

Every question = decision
Hardware Solutions

- Fast Track Data Warehouse - A reference configuration optimized for data warehousing. This saves an organization from having to commit resources to configure and build the server hardware. Fast Track Data Warehouse hardware is tested for data warehousing which eliminates guesswork and is designed to save you months of configuration, setup, testing and tuning. You just need to install the OS and SQL Server.

- Appliances - Microsoft has made available SQL Server appliances that allow customers to deploy data warehouse (DW), business intelligence (BI) and database consolidation solutions in a very short time, with all the components pre-configured and pre-optimized. These appliances include all the hardware, software and services for a complete, ready-to-run, out-of-the-box, high performance, energy-efficient solutions.
Fast Track Data Warehouse

FT Version 4.0

Benefits:
- Pre-Balanced Architectures
- Choice of HW platforms
- Lower TCO
- High Scale
- Reduced Risk
Appliances

- HP Business Data Warehouse Appliance (FT 3.0, 5TB)
- HP Business Decision Appliance (BI, SharePoint 2010, SQL Server 2008R2, PP)
- HP Database Consolidation Appliance (virtual environment, Windows2008R2)
- HP Enterprise Data Warehouse Appliance (1st PDW, SQL2008R2, 610TB)
- Dell Quickstart Data Warehouse Appliance 1000 (FT 4.0, 5TB)
- Dell Quickstart Data Warehouse Appliance 2000 (FT 4.0, 12TB)
- Dell Parallel Data Warehouse Appliance (2nd PDW, SQL2008R2, 600TB)
- IBM Fast Track Data Warehouse (FT 4.0, 3 versions: 24TB, 60TB, 112TB)

- V2 of PDW by HP released, uses SQL Server 2012, quarter-rack (75TB)
Appliances - Future

• HP Business Data Warehouse Appliance (FT 4.0, 2nd half 2013, likely with flash or SSD)
• HP Business Decision Appliance (BI, SharePoint 2013, SQL Server 2012, PP, 2nd quarter 2013)
• HP Database Consolidation Appliance (virtual environment, Windows Server 2012, System Center 2012 SP1)
Data Warehouse vs Data Mart

- **Data Warehouse**: A single organizational repository of enterprise-wide data across many or all subject areas
  - Holds multiple subject areas
  - Holds very detailed information
  - Works to integrate all data sources
  - Feeds data mart

- **Data Mart**: Subset of the data warehouse that is usually oriented to specific subject (finance, marketing, sales)
  - The logical combination of all the data marts is a data warehouse

In short, a data warehouse contains many subject areas, and a data mart contains just one of those subject areas.
Kimball and Inmon Methodologies

Two approaches for building data warehouses
Kimball and Inmon Myths

- Myth: Kimball is a bottom-up approach
  - Really top-down: BUS matrix (choose business processes/data sources), conformed dimensions, MDM

- Myth: Inmon requires a ton of up-front design that takes a long time
  - Inmon says to build DW iteratively, not big bang approach (p. 91 BDW, p. 21 Imhoff)

- Myth: Star schema data marts are not allowed in Inmon’s model
  - Inmon says they are good for direct end-user access of data (p. 365 BDW), good for data marts (p. 12 TTA)

- Myth: Very few companies use the Inmon method
  - Survey had 39% Inmon vs 26% Kimball. Many have a EDW

- Myth: The Kimball and Inmon architectures are incompatible
  - They can work together to provide a better solution
Kimball and Inmon Methodologies

- Relational (Inmon) vs Dimensional (Kimball)
- Relational Modeling:
  - Entity-Relationship (ER) model
  - Normalization rules
  - Many tables using joins
  - History tables, natural keys
  - Good for indirect end-user access of data
- Dimensional Modeling:
  - Facts and dimensions, star schema
  - Less tables but have duplicate data (de-normalized)
  - Easier for user to understand (but strange for IT people used to relational)
  - Slowly changing dimensions, surrogate keys
  - Good for direct end-user access of data
Relational Model vs Dimensional Model
Kimball and Inmon Methodologies

• Kimball:
  ▪ Logical data warehouse (BUS), made up of data marts
  ▪ Business Driven, users have active participation
  ▪ Decentralized atomic tables
  ▪ Independent dimensional data marts optimized for reporting/analytics
  ▪ Integrated via Conformed Dimensions (provides consistency across data sources)
  ▪ 2-tier, less ETL, no data duplication

• Inmon:
  ▪ Enterprise data model (CIF) that is a enterprise data warehouse (EDW)
  ▪ IT Driven, users have passive participation
  ▪ Centralized atomic normalized tables (off limit to end users)
  ▪ Later create dependent data marts that can be used for multiple purpose
  ▪ Integration via enterprise data model
  ▪ 3-tier, duplication of data
Kimball Model

DW Bus Architecture

Data Warehouse (DN Star Schema)

Staging

OLTP Data Sources

Data Mart 1 (3NF)

Data Mart 2 (3NF)

Data Mart 3 (3NF)

Atomic Data

Dimensionalized View

Cube Process

Multi-dimensional

Multi-dimensional

Reporting Layer
Inmon Model
Reasons to add a Enterprise Data Warehouse

- Single version of the truth
- May make building dimensions easier using lightly denormalized tables in EDW instead of going directly from the OLTP source
- Normalized EDW results in enterprise-wide consistency which makes it easier to spawn-off the data marts at the expense of duplicated data
- Less daily ETL refresh and reconciliation if have many sources and many data marts
- Easier to add a new data mart
- Reason not to: If have a few sources that need reporting quickly
Which model to use?

- Models are not that different, having become similar over the years, and can compliment each other
- Boils down to Inmon creates a normalized DW before creating a dimensional data mart and Kimball skips the normalized DW
- With tweaks to each model, they look very similar (adding a normalized EDW to Kimball, dimensionally structured data marts to Inmon)
- Bottom line: Understand both approaches and pick parts from both for your situation – no need to just choose just one approach
- BUT, no solution will be effective unless you possess soft skills (leadership, communication, planning, and interpersonal relationships)
Use SQL Server Views to interface between each level in the model
Kimball defines a development lifecycle, where Inmon is just about the data warehouse (not “how” used)
Populating a Data Warehouse

- Frequency of data pull
- Full Extraction – All data
- Incremental Extraction – Only data changed from last run
- Determine data that has changed
  - Timestamp - Last Updated
  - Change Data Capture (CDC)
  - Partitioning
  - Triggers
  - MERGE
  - Column DEFAULT value
- Online Extraction – Data from source. First create copy of source:
  - Replication
  - Database Snapshot
  - Availability Groups
- Offline Extraction – Data from flat file
ETL vs ELT

• Extract, Transform, and Load (ETL)
  ▪ Transform while hitting source system
  ▪ No staging tables
  ▪ Processing done by ETL tools (SSIS)

• Extract, Load, Transform (ELT)
  ▪ Uses staging tables
  ▪ Processing done by target database engine (SSIS: Execute T-SQL Statement task instead of Data Flow Transform tasks)
  ▪ Use for big volumes of data
  ▪ Use when source and target databases are the same
  ▪ Use with Parallel Data Warehouse (PDW)

ELT is better since database engine is more efficient than SSIS
 ▪ Best use of database engine: Transformations
 ▪ Best use of SSIS: Data pipeline and workflow management
**Surrogate Keys**

Surrogate Keys – Unique identifier not derived from source system

- Embedded in fact tables as foreign keys to dimension tables
- Allows integrating data from multiple source systems
- Protect from source key changes in the source system
- Allows for slowly changing dimensions
- Allows you to create rows in the dimension that don’t exist in the source (-1 in fact table for unassigned)
- Improves performance (joins) and database size by using integer type instead of text
SSAS Cubes

Reasons to use instead of data warehouse:

- Aggregating (Summarizing) the data for performance
- Multidimensional analysis – slice, dice, drilldown
- Hierarchies
- Advanced time-calculations – i.e. 12-month rolling average
- Easily use Excel to view data via Pivot Tables
- Slowly Changing Dimensions (SCD)
Data Warehouse Architecture

1. Gather
2. Clean/Master
3. Store
4. Model/Present
5. Analyze
6. Share

Sources
- CRM
- Operational Systems
- ERP
- External Data

Data Platform
- Staging Tables
- ODS
- Dimensional SSIS
- DQS
- MDM
- SSIS

3NF Data Warehouse
- Sales Schema
- Finance Schema
- Marketing Schema
- Star Schema

Data Mart
- Multi-dimensional

SSAS
- Tabular
- PowerPivot For SharePoint

SharePoint
- SQL Server Reporting Services
- PerformancePoint Services
- Excel Services
- Power View

Publish
- PowerPivot for Excel
SQL Server 2012 Tabular Model

- New xVelocity in-memory database in SSAS
- Build model in Power Pivot or SQL Server Data Tools (SSDT)
- Uses existing relational model
- No star schema, no extra SSIS
- Uses DAX (Excel-like formula language to create custom calculations)
- Faster and easier to use than multidimensional model
- In Hybrid Model, DW Bus Architecture goes away, but only if you don’t need surrogate keys, SCD, conformed dimensions, and all reporting will be off of the cubes
End-User Microsoft BI Tools

Many options on Microsoft tools to use to go against the DW/BI:

- Excel PivotTables (Cube, 3NF)
- SQL Server Reporting Services (SSRS) (Cube, 3NF)
- Report Builder (Cube, 3NF)
- PowerPivot (3NF)
- PerformancePoint Services (PPS) (Cube)
- Power View (Cube)
- Data Mining (Cube)
Resources

Questions?

James Serra, Business Intelligence Consultant
Email me at: JamesSerra3@gmail.com
Follow me at: @JamesSerra
Link to me at: www.linkedin.com/in/JamesSerra
Visit my blog at: JamesSerra.com
Thank You for Attending