



# SmartArrays for Data Integration

July10, 2010

James Wheeler





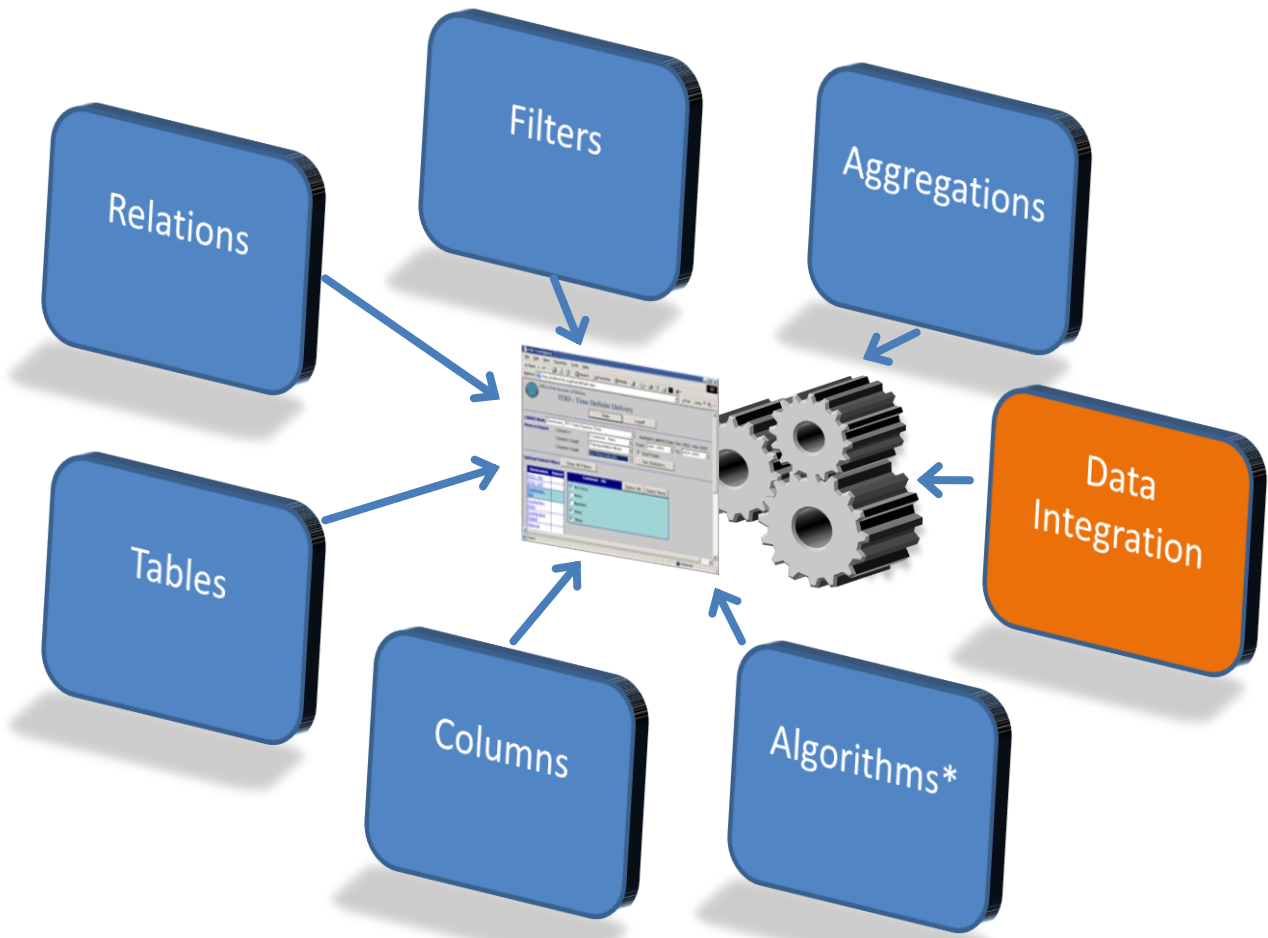
## SmartArrays Technology

- A software library for use in .NET and Java applications.
- Holds data in a high-speed memory-resident structures, which include relational tables.
- Computation oriented, including searching, matching, filtering, sorting, math, etc. More than 200 operations.
- Data adapters for external sources: ODBC, JDBC, .NET ADO, flat files, or custom formats and protocols.





## Building Blocks of Analytical Software





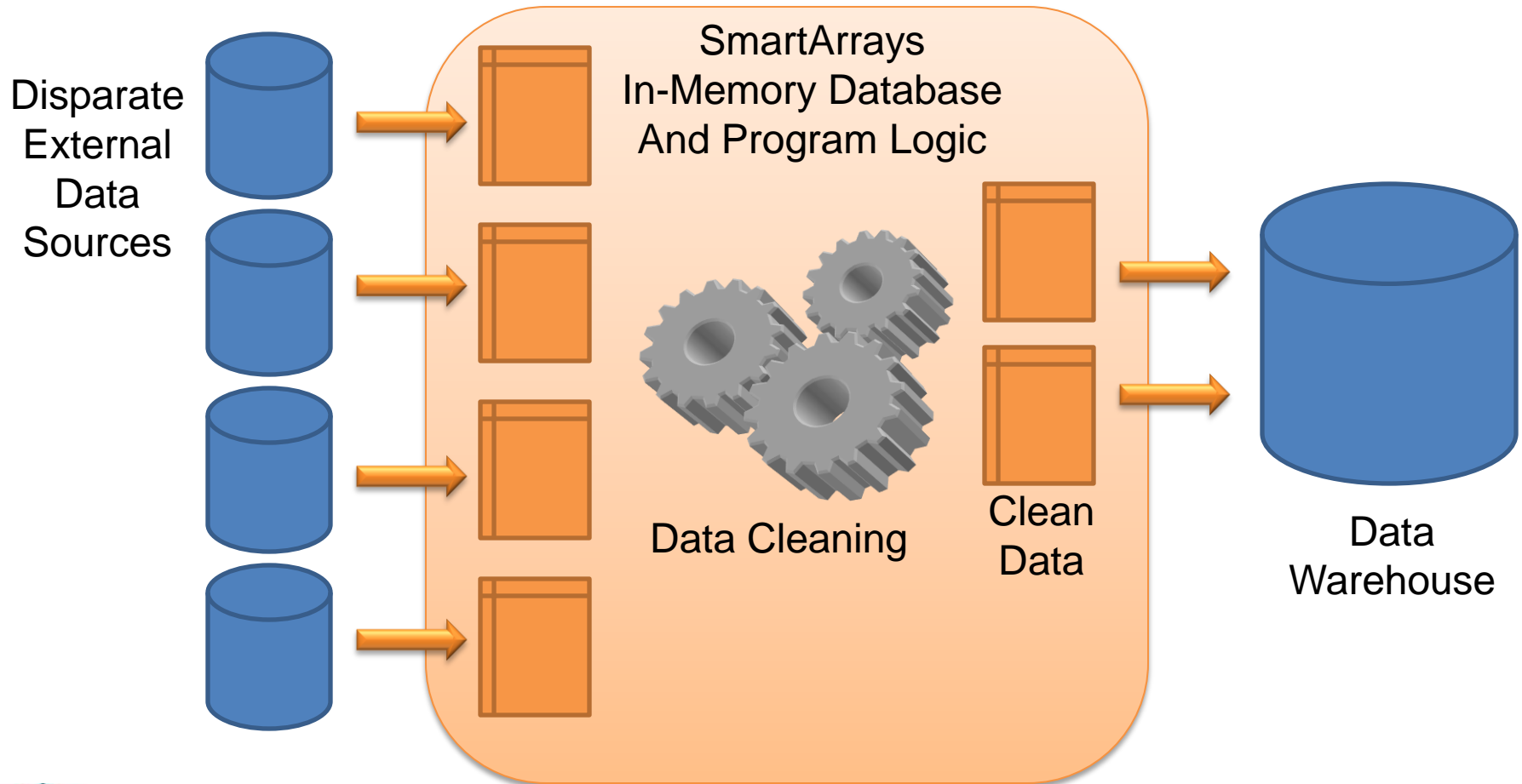
## Data Integration with SmartArrays

- Pull data from multiple sources (databases, flat files, live data feeds, web services)
- Store data in “Smart Table” format (extremely fast to write, and fast to compute)
- Relate data from disparate sources
- Validate, adjust, consolidate
- Post to data warehouse, etc.





## Sample Data Flow





## Database Adapters

- ODBC: Most legacy databases, plus Excel, flat files. Supported in both .NET and Java platforms
- ADO.NET: Microsoft data sources: SQL Server, Access, MySQL, XML, others. For .NET Platform apps.
- JDBC: Virtually every database. For Java Platform apps.

## File Based I/O

- RecordParse Tool: Parses text data into records/fields.
- Legacy data (EBCDIC, Zoned Decimal)
- Date/Time Formats





## Data Loading Toolkit

- Bulk data import driven off XML-based rules
- Easily scripted into standalone batch tasks, or packaged with a user interface if appropriate.
- Not a standalone ETL tool (like Informatica, etc), but able to program flexible, specific operations that are awkward or missing with COTS package.





## Example: Insurance Industry

- Accessed legacy IBM mainframe data. Packed decimal files dumped from legacy systems (ISAM).
- Multiple transformations, validations, cleansing. Fast because based on SmartArrays.
- Result is a relational table – bulk-exported to SQL Server.







## Non-Disruptive Approach for Optimizing Data Integration

- Identify bottlenecks in current process.
- Select stage in the process that can be isolated and optimized.
- Rest of process / solution remains unchanged.





## How To Begin

- Briefing on the client's current process. Locate bottlenecks and opportunities to improve.
- Short pilot project to test feasibility and estimate performance improvements.

