



# DQM Technical Summary

## Data Quality Manager

Easy to use, powerful data quality tool, built on experience and scalable database technologies.

### THE DQM DIFFERENCE

DQM is a powerful, simple to use systems data quality application, that is built on industry strength database technology. It has

- Powerful parse & match engine
- Scalability and high throughput performance
- Industry experience packaged in templates providing a quick start solution

### THE SOLUTION

DQM allows you to restructure and cleanse data, and to assemble a set of data representing the best of breed information from all available sources.

DQM is a server based system, that leverages database engine functionalities to provide superior performance and scalability. It is designed to process massive volumes of data.

DQM engine can be invoked from a number of environments - in batch, real time or part of a SOA environment. Running DQM online, gives you instant analysis. Batch operated DQM, is best suited for very large datasets. When DQM is set up as a real-time service, provides front end data quality control capabilities.

### FUNCTION / FEATURE LIST

DQM provides a full range of data quality management features. These include :

#### Data Profiler & investigation

- Quick Profile that gives you a instant snap shot with minimal set up
- Detailed Investigation focusing content, pattern and semantic meaning of data
- Report library management system that track full history of investigations
- Presentation using Microsoft Excel providing desktop integration

#### Data Cleansing & matching

- Prebuilt cleanse and match templates
- Parse, Validate domain, Standardize, Reformat source data
- Match, Household dedupe records
- Scoring & Weighting, Override to fine tune match rate
- Split, Filter, Link data functions to manipulate data streams
- Enhancement, Append data
- Survive, Consolidate, Lineage tracking to provide a best of breed output

#### Management dashboard

- Scorecard, graphs and traffic lights to illustrate current situation
- Business rules clearly listed with data quality score
- Drill down from dashboard to review data issues
- Rol templates to manage overall project

#### Job Management

- Execution engine that runs and monitor progress in real time
- Job scheduler that submit jobs to server
- SOA interface, that publishes DQM as an independent services
- Library system that manage the record of the status, parameters, source, and output

#### Data Services

- Import and output capability data into database environment
- Database Attachment functionalities to link databases as sources for DQM
- Data Browser function to review data and data structures
- Data Summary function that counts and displays patterns in data column

### PERFORMANCE

DQM provides a scalable performance solution for complex and high data volume environments. The application leverages the hardware scalability of the database engine, and is able to provide excellent throughput, through a range of data volume.

A number of performance trails have been conducted to illustrate the throughput and scalability. Sample data of consumers with name, address and phone numbers were presented to DQM for Profiling, Parse, Match and Append.

These processes test the performance of the DQM application under progressive more complex and demanding conditions. The source data has 30 columns with a record length of 250 bytes.

The servers used for these trails were a modest and standard Intel® Quad Core 2.8GHz CPU, with 4 GB of RAM, and 600 GB SATA HDD, running XP Professional.

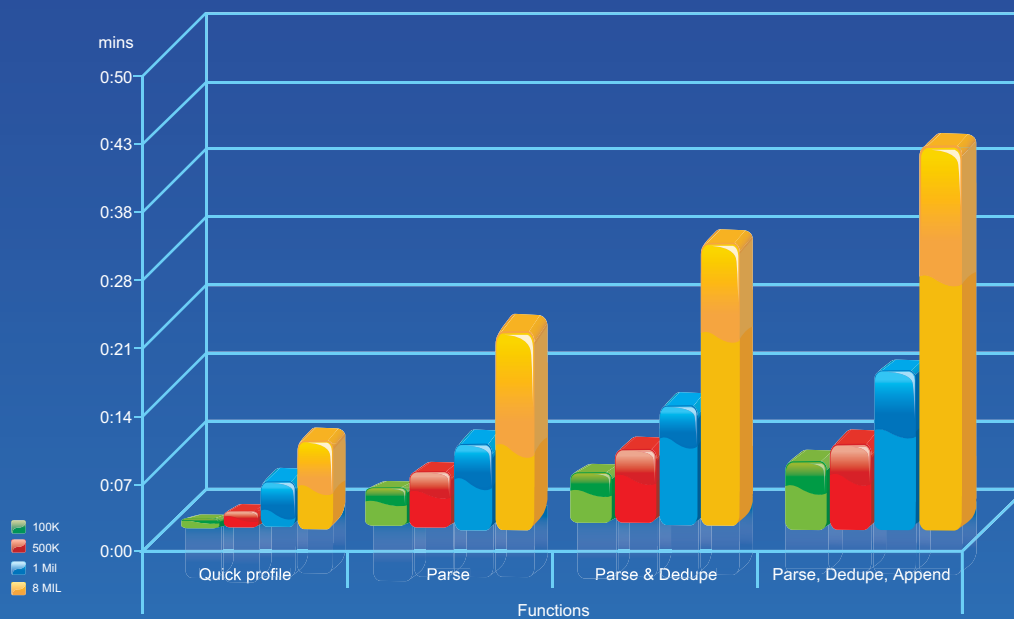
DQM will Parse and Profile 1 million records in less than 30 mins, using modest hardware set up. (Figure 1)

(Figure 1)

Rows processes	Functions			
	Quick profile	Parse	Parse & Dedupe	Parse, Dedupe, Append
100K	0:01:00	0:03:00	0:05:00	0:06:00
500K	0:01:30	0:04:30	0:07:00	0:08:00
1 mil	0:03:30	0:07:30	0:11:00	0:15:00
8 mil	0:15:00	0:25:00	0:35:00	0:45:00

The servers used for these trails were a standard Intel® compatible 1 CPU Windows server configured with direct attached SATA storage.

### DQM Performance



### TECHNICAL REQUIREMENTS

Server hardware Standard Intel P4 Dual 2GHz or above  
4GB RAM  
200GB HDD

Server OS Windows 2003 Server  
Windows (x86-32)  
Windows NT 4 Server,  
Windows 2000 Server,  
Windows XP Professional  
Windows Server 2003 (64-bit on Itanium)

Client desktop Standard Intel® compatible PC  
Minimum 1GB memory and 3 GB of HDD

Client OS Recommended : Windows® XP Professional  
with SP2 or later  
Minimum : Windows® 2000 with SP4

Database Recommended : SQL Server 2005  
Minimum : SQL Server 2000

## Data Integration Specialist

LEVERAGE ON THE POWER OF DQM  
Visit [www.dqm.com.au](http://www.dqm.com.au) or call +61 3 9819 9591 to speak to a Data Quality Specialist.