



LESS DISRUPTION. MORE INSIGHTS.

HAMR is a big data analytics engine that builds upon expertise in data flow technologies. HAMR software is less of a disruption and more of a source of efficiency for enterprises with big data problems. HAMR enables enterprises to do more with less. HAMR can be used by Data Scientists as well as IT staff.

Gartner projects that the Master Data Management (MDM) market will reach \$3.2 B in 2015, with a CAGR of 21%. This includes data cleansing to correct errors in large data sets, such as:

- Data entry errors
- Measurement errors
- Distillation errors
- Data integration errors

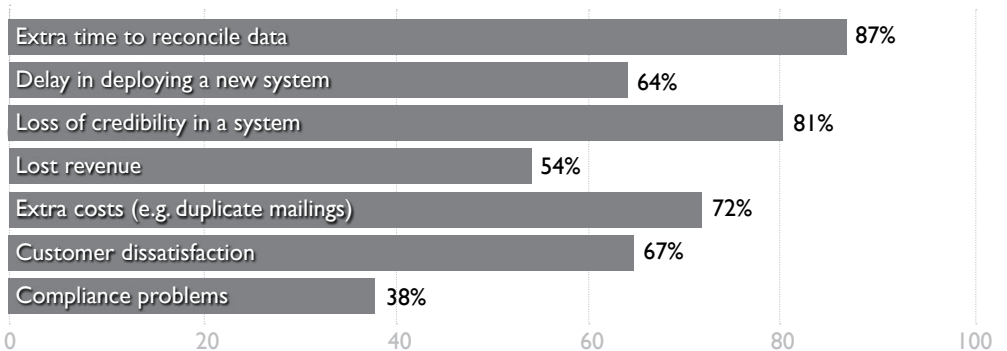
Data Cleansing:

The act of detecting, removing and/or correcting a database's dirty data, also known as "data scrubbing," "monging" or "wrangling."

The challenge of Big Data is to create consistency among different sets of data. Sophisticated applications are available to clean data using algorithms, rules and look-up tables – making it semi-automated. However, many of today's systems cannot keep up with the velocity and variety of new data sources.



PROBLEM Some data-driven organizations receive more than a terabyte (TB) of new data each day. This big data can include online clickstreams, transactions, log files, sensor data, feeds, CRM files, emails, business records, contracts, IP and operational information. According to a survey of MDM professionals by The Data Warehouse Institute (TDWI), poor data quality impacts organizations in at least seven ways (multiple responses allowed):



TDWI calculated that 10-25% of revenue in data-driven enterprises are impacted by data quality. The top line is impacted by customer dissatisfaction. The bottom line is influenced by delays, lost productivity, and additional expenses.

With current data warehouse & business intelligence (DW/BI) systems facing a growing volume, velocity and variety of data, a huge team is required to process a fraction of the data. There is an urgent need to deploy machine learning techniques in near real-time to make the process more automated and accurate.



LESS DISRUPTION. MORE INSIGHTS.

Extract-Transform-Load (ETL) is a key process used in data cleansing. ETL tools use a repository built on a relational database to manage tags in a uniform way, that focuses on script programs, mappings, target schemas, and data resources.

SOLUTION HAMR is a Big Data analytics engine designed to pull data from multiple sources, transform the data, and make it conform to a given taxonomy – all in one intuitive interface. HAMR presents a compelling value proposition, on strategic and operational levels, because it reduces the number of steps, accelerates the process, and enables programmers and data scientists to use machine learning.

5 FIVE STEPS FOR THE NEW HAMR DATA CLEANSING REVOLUTION

- External / Foreign Data Source
- E** Data Resource Reader] **HAMR**
- T** Flowlets (cleansing, transforms, routing, analytics)
- L** Data Resource Writer]
- File

4 HAMR SAVES FOUR STEPS

9 NINE STEPS FOR THE CURRENT HADOOP ETL PARADIGM

- External / Foreign Data Source
- E** Extract] **ETL Tool**
- T** Transform (cleansing)
- L** Load]
- Staging Database]
- Extract]
- HDFS (Hadoop)] **Hadoop**
- Hadoop MR]
- HDFS]

9 NINE STEPS FOR THE CURRENT DW/BI ETL PARADIGM

- External / Foreign Data Source
- E** Extract] **On-Board**
- T** Transform (cleansing)]
- L** Load]
- Staging Database]
- SQL Execution/ Data Normalization] **DW**
- Data Warehouse]
- BI Execution] **BI**
- File]

HAMR provides a simpler five step process to cleanse data. Flowlets process data much faster by enabling data scientists to iterate & adjust algorithms – taking full advantage of machine learning techniques in close to real-time. This is not true of Hadoop because it processes data in batch mode.

Strategic Value: Increased throughput allows BI vendors to begin to process Big Data. Instead of processing 260 GB per day, they can process 2,080 GB per day. That is more than 2TB. This will allow BI vendors to compete with Hadoop.

Operational Value: HAMR saves 1,820 hours of processing in one year. If five consultants are working on an ongoing project, that can save \$2,730,000 (based on \$300/hour).

HAMR REDUCES PROCESSING TIME FROM DAYS TO HOURS

DAY 1	DAY 2	DAY 3

DW/BI ETL can take over two days
Hadoop batch ETL can take two days
HAMR can take a few hours