IBM InfoSphere Discovery

IBM

IBM is an American multinational technology and consulting corporation, with headquarters in New York but with a global presence. It manufactures and markets computer hardware and software and offers infrastructure, hosting and consulting services in areas ranging from mainframe computers to nanotechnology. The company was founded in 1911 and over the years has grown through acquisition into its present state. IBM initially made its profits largely from selling hardware but, these days, is more of a software solutions and services company covering storage, security, process, development, operations etc. - and IBM Global Services. It currently stresses the delivery, from smarter computing, technology and process, of better business outcomes for the smarter planet.

It has considerable original research capabilities and is responsible for many computer science innovations from relational database theory (see for example, E. F. Codd's 1981 ACM Turing Award Lecture) to advanced semi-AI computers that offer cognitive computing capabilities. IBM has recently described cognitive computing as "the third era of computing".

IBM InfoSphere Discovery is a data profiling and discovery tool. However, there is another InfoSphere product that also offers data profiling, known as Information Analyzer. In our view Discovery is much the superior product (because it offers extended discovery capabilities) for processes such as data migration and in supporting master data management (MDM) initiatives. There are plans in place to merge these two product lines.

Major elements in the Discovery platform are the Discovery Engine and Discovery Studio. The former is the component that does the actual process of discovering business rule transformations, data relationships, data inconsistencies and errors, and so on. Where appropriate it generates cross-reference tables that are used within the staging database, it creates metadata reports either in HTML format or Excel, and it generates appropriate SQL, XML and ETL scripts. Discovery Studio is the graphical user interface employed by data analysts or stewards to view the information discovered by the engine and to edit, test and approve (via guided analysis capabilities) relationships and mappings from a business perspective.

Company details

IBM
New Orchard Road
Armonk, New York, 10504
USA
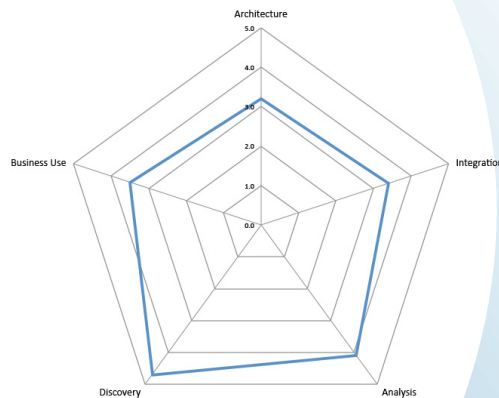
Tel. +1 914 499 1900
Web. www.ibm.com

Figure 1: Requirements (scale 0 to 5)

Markets and Channels

InfoSphere Discovery is an enabling tool rather than a solution in its own right so it is horizontally applicable across all sectors. It will be particularly useful where it is necessary to understand business entities (for example, a customer with his orders, delivery addresses, service history and so on) and process those business entities as a whole. Notable environments that require such an approach include application and database-centric data migrations, master data management and archival.

Users

We have no doubt that IBM has many successful users of InfoSphere Discovery. However, you wouldn't know that to judge by its web site, which includes just three case studies (including videos) that discuss customers using the product.

Technology

In addition to providing conventional data profiling capabilities, Discovery supports the discovery of orphaned rows, scalar relationships (simple mappings, substrings, concatenations and the like), arithmetic relationships between columns, relationships based on inner and outer joins, and correlations for which cross-reference tables are generated. Cross-source data analysis is available both to discover attribute supersets and subsets, and to identify overlapping and unique attributes. In the latter case there is a visual comparison capability that allows you to compare record values from two different sources on a side-by-side basis. In addition there are automatically generated source rationalisation reports that compare data sources to one another. Further features include support for filtering, aggregations and if-then-else logic, amongst others.

There is also a Unified Schema Builder designed specifically to support new master data management, data warehousing and similar implementations that includes precedence discovery, and empty target modelling and prototyping. There are also facilities for cross-source data preview, automated discovery of matching keys (that is, a cross-source key for joining data across sources), automated discovery of business rules and transformations across two or more data sets with statistical validation, and automated discovery of exceptions to the discovered business rules and transformations. In support of MDM, migration, archival, and similar environments, Discovery is clearly the leading product in the market.

Services

In addition to the normal sorts of training and support services you would expect from any vendor, IBM offers business services, IT services, outsourcing services, asset recovery, hardware and software financing, IT lifecycle financing, and commercial financing.

Product Sheet

Bloor

2nd Floor
145–157 St John Street
LONDON EC1V 4PY
United Kingdom

Tel: **+44 (0)20 7043 9750**
Web: **www.Bloor.eu**
email: **info@Bloor.eu**