

# InBrief

Philip Howard – Research Director, Information Management  
 Daniel Howard – Senior Analyst, Information Management and DevOps



www.iri.com

2194 Highway A1A, 3rd Floor  
 Melbourne, FL 32937-4932, USA  
 Tel: +1 321 777 8889  
 Email: info@iri.com

## IRI Voracity

### The Company

The company is a privately owned ISV founded in 1978. Its offices are in Florida and it relies on a partner network of resellers for international coverage (in 40 locations throughout the world).

The company's first product, CoSort, is a high performance data transformation utility that was first designed to offload JCL sort/merge steps to CP/M. Needless to say, this has been extended and ported to other environments but it remains at the heart of IRI's offerings, including Voracity.

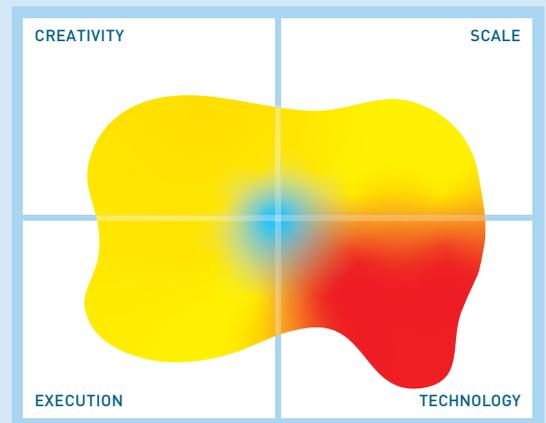
### What is Voracity?

Powered by CoSort (or Hadoop), and built on Eclipse, IRI Voracity is a multi-purpose data management platform designed to perform, speed, and consolidate common work in 5 general areas:

- 1. Data Discovery** – data profiling, classification, search, and metadata redefinition
- 2. Data Integration** – high volume ETL, change data capture, slowly changing dimensions
- 3. Data Migration** – file/data/database type conversion, replication, and federation
- 4. Data Governance** – data quality, PII masking, re-ID risk scoring, test data synthesis
- 5. Analytics** – embedded BI, tie ups to Datadog, KNIME and Splunk, wrangling for the rest

“ We sought a reliable tool that would quickly sort and transform very large files... we see the Voracity platform as a much more cost-effective (and higher-performing) alternative to legacy ETL tools. ”

Optum



The image in this Mutable Quadrant is derived from 13 high level metrics, the more the image covers a section the better. Execution metrics relate to the company, Technology to the product, Creativity to both technical and business innovation and Scale covers the potential business and market impact.

As can be seen in **Figure 1**, Voracity drives solution depth by including standalone products in both the IRI Data Manager Suite and the IRI Data Protector Suite, each of which have various sub-components that support multiple capabilities.

Voracity is an integrated platform with metadata shared across the whole environment, which supports the provision of data lineage. A formal data catalogue is missing, though the product does have inherent data classification capabilities and its central metadata stores are easy to understand, share, and use across the above applications, or create for Collibra.

A similar consideration applies to data governance whereby there are some capabilities provided, mostly related to data privacy and quality, but not a general-purpose capability for which the company would rely on integration with partners like Erwin. Most notable of ancillary governance capabilities in Voracity are test data management, with options for synthetic data generation, database subsetting, and static and dynamic data masking (with the option to combine both).

Illustrated in **Figure 1** but not discussed is the IRI Workbench IDE, which supports graphical metadata creation, conversion, discovery, and application wizards to create, deploy, and manage data rules, job scripts, data definition files (DDF), and the XML workflows common to all IRI software. In the same

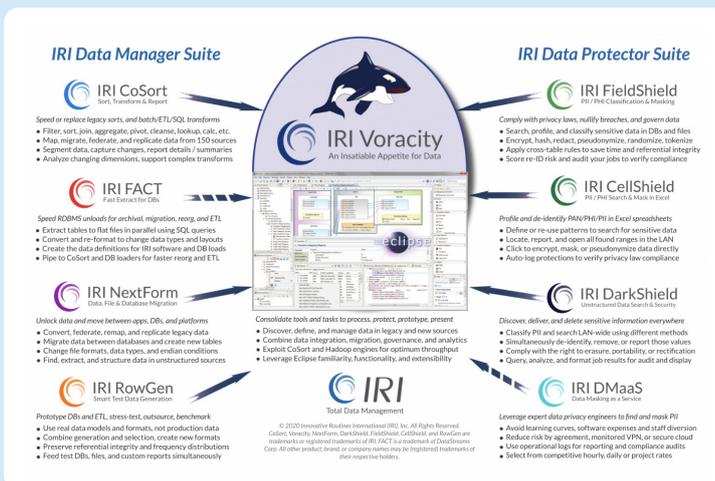


Figure 1 – IRI solutions

Architecture	★★★★	Data Movement	★★★★★
Automation	★★★★	Data Privacy	★★★★★
Data Catalogue	★★★★	Data Quality	★★★★
Data Governance	★★★★	Product Integration	★★★★

“CoSort accurately and quickly processes billions of rows of data and allows us to join and analyze this information in connection with our other data warehouse processes. No other tool gives us this much speed and flexibility.”  
Comcast

pane of glass, you can also administer your databases and develop or use applications in other languages and any plug-in supported in Eclipse. As an alternative to the wizards you can also develop jobs using diagrams, dialogs, or IRI’s domain specific language (a 4GL), called SortCL.

### What Does Voracity Do?

IRI CoSort is the default Voracity data integration engine. Unlike other such products, it is not confined to ETL (extract, transform and load) operations but also performs data replication (change data capture), federation, masking, cleansing, and reporting. Another key point to note about it is that it does not have to transform data in separate steps. You can define jobs that way, but

heritage, it also supports mainframe sources that use COBOL copybooks, EBCDIC and so on. While it does not run on z/OS it does support mainframe databases as sources and will itself run on z/Linux. While IRI Voracity does not offer a module called “data quality”, it does provide substantial relevant capability, as illustrated in **Figure 2**.

A major strength of IRI Voracity is clearly in its Data Protector Suite. To begin with, IRI has deployed machine learning (including within IRI DarkShield) to support the identification of sensitive data (though we are disappointed that M/L has not yet been implemented more widely across the platform). It also uses natural language processing for this purpose. Once discovered, as mentioned, the company offers significant capabilities when it comes to masking. In particular, dynamic data masking may be proxy-based, run in situ or driven by APIs, and can be mixed and matched with static masking. It is also worth mentioning that Voracity supports the ability to search, parse and protect multiple sources containing semi- and unstructured data.

Finally, given the current predilection for companies to migrate from on-premises data warehouses to cloud-native data warehouses such as Snowflake or Google BigQuery, it is worth noting the availability of IRI FACT and IRI NextForm, which bolster high volume database migration operations.

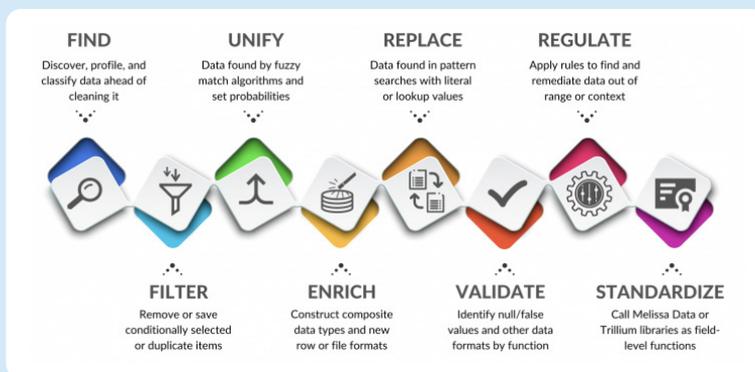


Figure 2 – Data quality functions in IRI Voracity

at run time the engine consolidates multiple steps to reduce I/O. Added to the fact that the run-time engine is a 2MB, multi-threaded C executable and loads only the libraries it requires, and you will appreciate why CoSort has a performance advantage over its competitors.

Note that IRI also offers a Hadoop-based option that does not have the same footprint advantages of CoSort but otherwise runs in a similar fashion. Moreover, many jobs developed for native CoSort implementations will run without change in Map Reduce 2, Spark, Spark Streams, Storm or Tez. Dataflows are actually stored in files and can be executed from anywhere.

The company offers an extensive range of native connectors (including MQTT and Kafka) plus JDBC support. Not surprisingly given its

### Why Should You Care?

IRI Voracity is close to being a complete data management platform. It only lacks a formal data catalogue and some extensions to its policy and governance capabilities, which are in development. On the other hand, it is much more advanced when it comes to ETL performance and sensitive data protection than many of its competitors. The company’s data migration capabilities will also be a boon in the current environment, as will its relatively attractive price points and licensing options.

### The Bottom Line

The key features of IRI Voracity are the performance that the CoSort engine offers, and the depth of capability it provides in extending its data management platform into the identification and management of sensitive data. If these are important issues for you, then you should seriously consider IRI Voracity.

[FOR FURTHER INFORMATION AND RESEARCH CLICK HERE](#)