



IRI Voracity
An Insatiable Appetite for Data

Platform Introduction



Who We Are

Specialists in data management and data protection

Known since 1978 for 'big data' transformation speed

A 'top big data provider' (CIO Review & Insights Success)

Trusted by data-driven customers in every industry

Partners to resellers and consultants worldwide

Our Mission

To support a wide range of data management solutions through software which uniquely combines:

- Speed and scalability
- Functional versatility
- Familiarity and usability
- Licensing flexibility and affordability



Selected Customers



How We Help

Rapidly integrate and prepare/clean data on premise or in the cloud for DW ETL & BI ops

Find, classify, and mask PII for privacy law compliance and breach nullification

Proven CoSort engine lowers cost, system impact, and risk of mission-critical projects

Seamless Hadoop integration eases the transition to grid storage and processing

What is Voracity?

A modern, end-to-end data lifecycle management platform for data discovery, integration, migration, governance, analytics, and curation, PLUS...



A Big Data Solution Stack

Package, protect, and provision data in legacy and modern repositories

Migrate, transform, and mask data in Eclipse using CoSort or Hadoop MR2, Spark, Storm, or Tez without coding



A Data Stewardship Portal

Search, profile, and classify data

Validate, cleanse, enrich, and unify

Encrypt, pseudonymize, and redact

Manage metadata and master data



A Faster ETL & BI Alternative

CoSort and Hadoop engines for data preparation and integration

- 6x faster than legacy ETL tools
- 10x faster than SQL
- 12x faster than BI tools



A Database Ops Environment

Speed VLDB unloads, loads, and reorgs

Offload SQL transformation and reporting

Profile, classify, subset, mask, and generate DB test data

Platform Product Components

IRI Data Manager Suite



IRI CoSort

Sort, Transform & Report

Speed or replace legacy sorts, and batch/ETL/SQL transforms

- Filter, join, aggregate, pivot, cleanse, lookup, calc, etc.
- Map, migrate, federate, and replicate data from 150 sources
- Segment data, capture changes, report details / summaries
- Analyze changing dimensions, support complex transforms



IRI FACT

Fast Extract for DBs

Speed RDBMS unloads for archival, migration, reorg, and ETL

- Extract tables to flat files in parallel using SQL queries
- Convert and re-format to change data types and layouts
- Create the data definitions for IRI software and DB loads
- Pipe to CoSort and DB loaders for faster reorg and ETL



IRI NextForm

Data, File & Database Migration

Unlock data and move between apps, DBs, and platforms

- Convert, federate, remap, and replicate legacy data
- Migrate data between databases and create new tables
- Change file formats, data types, and endian conditions
- Find, extract, and structure data in unstructured sources

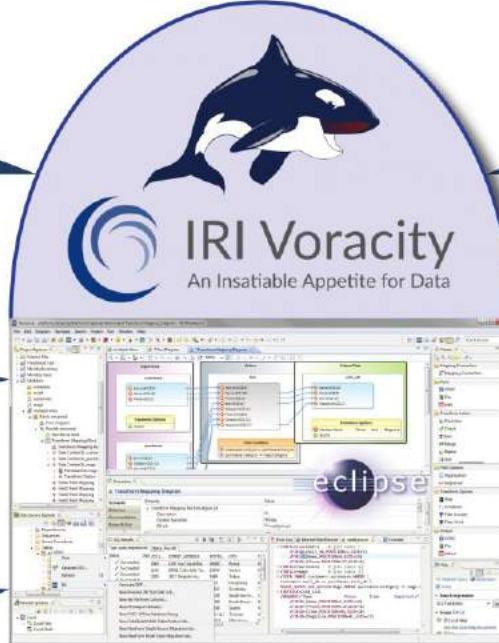


IRI RowGen

Smart Test Data Generation

Prototype DBs and ETL, stress-test, outsource, benchmark

- Use real data models and formats, not production data
- Combine generation and selection, create new formats
- Preserve referential integrity and frequency distributions
- Feed test DBs, files, and custom reports simultaneously



Total Data Management

© 2019 Innovative Routines International (IRI), Inc. All Rights Reserved.
CoSort, Voracity, NextForm, DarkShield, FieldShield, CellShield, and RowGen are
trademarks or registered trademarks of IRI. FACT is a trademark of DataStreams
Corp. All other product, brand, or company names may be (registered) trademarks
of their respective holders.

IRI Data Protector Suite



IRI FieldShield

PII / PHI Classification & Masking

Comply with privacy laws, nullify breaches, and govern data

- Search, profile, and classify sensitive data in DBs and files
- Encrypt, hash, redact, pseudonymize, randomize, tokenize
- Apply cross-table rules to save time and referential integrity
- Score re-ID risk and audit your jobs to verify compliance



IRI CellShield

PII / PHI Search & Mask in Excel

Profile and de-identify PAN/PHI/PII in Excel spreadsheets

- Define or re-use patterns to search for sensitive data
- Locate, report, and open all found ranges in the LAN
- Click to encrypt, mask, or pseudonymize data directly
- Auto-log protections to verify privacy law compliance



IRI DarkShield

Unstructured Data Search & Security

Discover, deliver, and delete sensitive information in dark files

- Classify PII and search LAN-wide using different methods
- Simultaneously de-identify, remove, or report those values
- Comply with the right to erasure, portability, or rectification
- Query, analyze, and format job results for audit and display



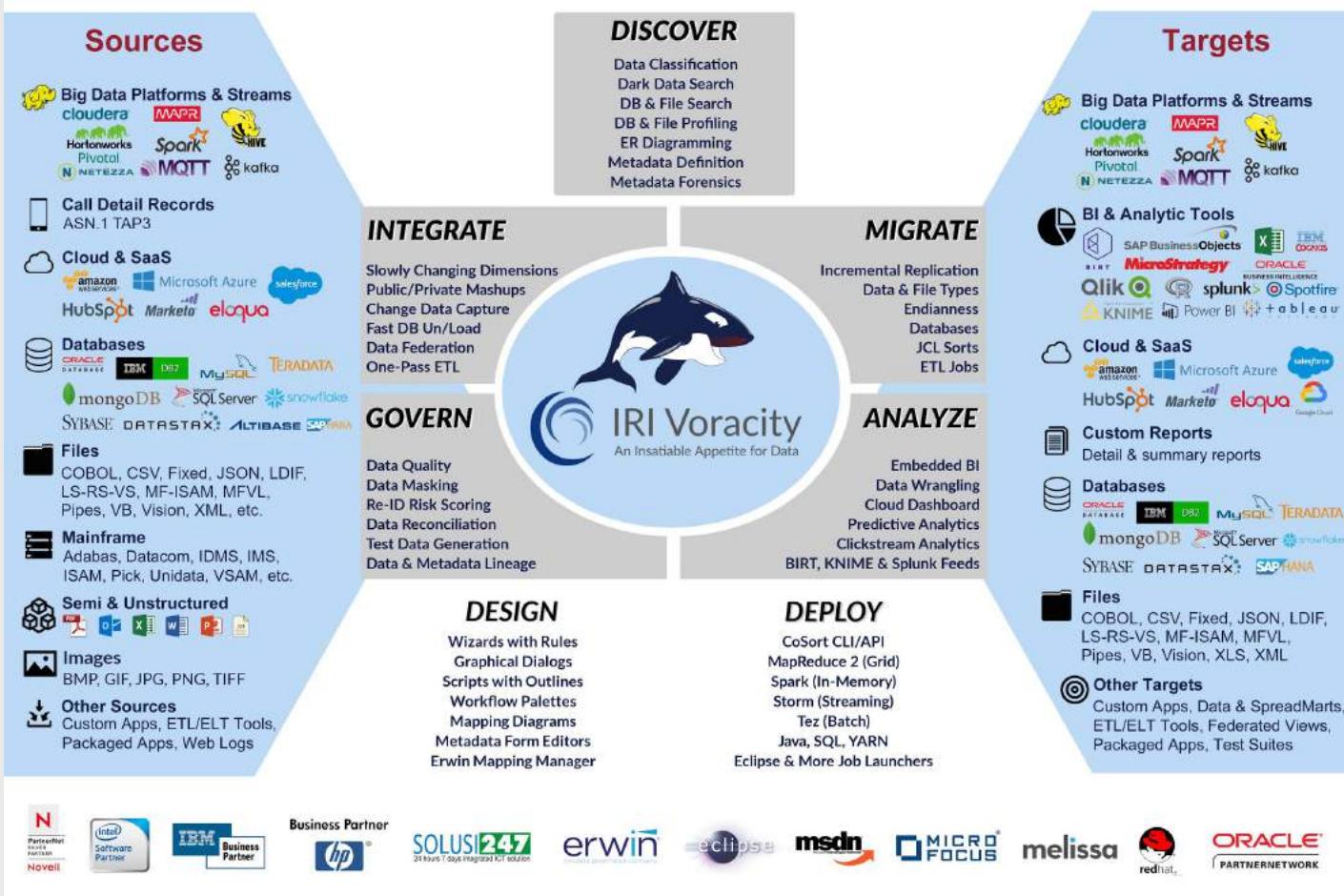
IRI DMaaS

Data Masking as a Service

Leverage expert data privacy engineers to find and mask PII

- Avoid learning curves, software expenses and staff diversion
- Reduce risk by agreement, monitored VPN, or secure cloud
- Use operational logs for reporting and compliance audits
- Select from competitive hourly, daily or project rates

Base Included Capabilities



Voracity Architecture

The default Voracity stack uses IRI Workbench for client-side design of data-driven jobs defined in portable CoSort SortCL scripts.

Many of the same scripts also run interchangeably in Hadoop.

The scripts are fully supported in the Workbench data model and by erwin Mapping Manager, for graphical creation, modification, and management.

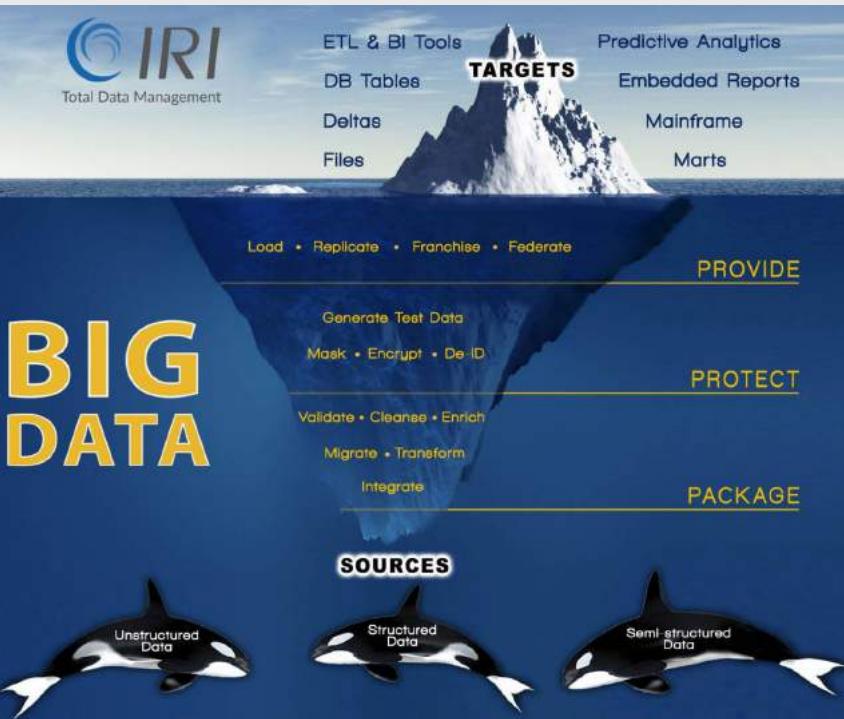


Newest Data Sources and Targets

Amazon EMR Hive	JSON	Marketo	Pivotal Greenplum
Apache Cassandra	Force.com apps	MongoDB	Pivotal HD Hive
Apache Hadoop Hive	Hortonworks Hive	MS Dynamics CRM	Salesforce.com
Cloudera CDH Hive	Hubspot	MS SQL Azure	ServiceMAX
Cloudera Impala	Lightning Connect	Oracle Eloqua	Spark SQL
Database.com	MapR Hive	Oracle Service Cloud	Veeva CRM

Voracity also supports a [large list](#) of structured and semi-structured database and file sources, and via DarkShield within, many unstructured document and image file sources, too.

Voracity's Big Data Functions & Advantages



Using Voracity for Data:

Discovery Integration Migration Governance Analytics



Data Classification



Dark Data Discovery



DB & File Profiling



ER Diagramming



Metadata Definition



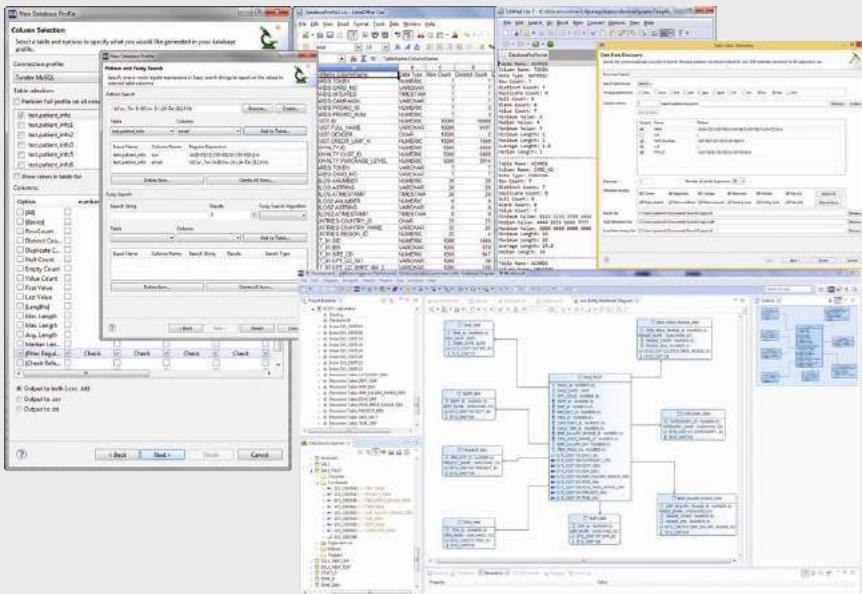
Metadata Forensics



Multi-Method Search

Data Discovery Features

Voracity has data (e.g., PII) discovery facilities to: 1) **classify** and **diagram** multiple sources; 2) **search** by string (literal or in-dictionary), pattern, fuzzy-match, or machine-learned NER; 3) **report** on statistical profiles; and, 4) **parse** and **re-define** all metadata needed. It includes



[fit-for-purpose wizards](#) for:

- Data classification, with rule matcher libraries
- DB profiling and ER diagramming
- Inter- and intra-schema pattern and data class searches
- Dark data discovery and extraction (structuring), and reporting, including file-specific metadata
- Flat-file statistical reporting and value searching
- Structured & semi-structured metadata creation
- Metadata sharing, lineage, version control, etc.

Using Voracity for Data:

Discovery **Integration** Migration Governance Analytics



Slowly-Changing
Dimensions



Public/Private
Mashups



Change Data
Capture



Fast DB Un/Load



Data Federation

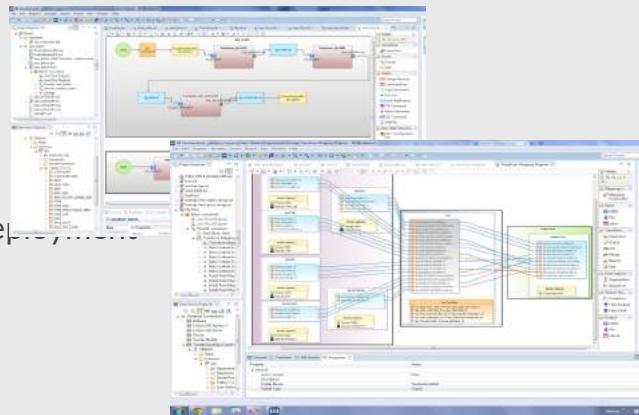


One-Pass ETL

Why Voracity for Data Integration

Fast and Easy Onboarding and Multiple Ways to Speed ETL

- 1) Voracity's free, familiar Eclipse environment has more job design and deployment options than any other data integration platform.
- 2) Support every DI architecture: ODS/EDH, EDW/LDW, data lakes, and the DW/lake hybrid 'Production Analytic Platform' [paradigm](#)



Speed New ETL Jobs

Extract VLDBs in parallel via FACT, or stream web, brokered, or piped data

Transform with CoSort or Hadoop engines (interchangeably), without coding!

Load bulk DB targets pre-sorted

Speed Other ETL Tools

"Push down" sort, join, and aggregation steps in ODI, DataStage, Informatica, SSIS, Talend, or Pentaho to Voracity via command-line calls, and get ETL job results back 2-20X faster (and cheaper!)

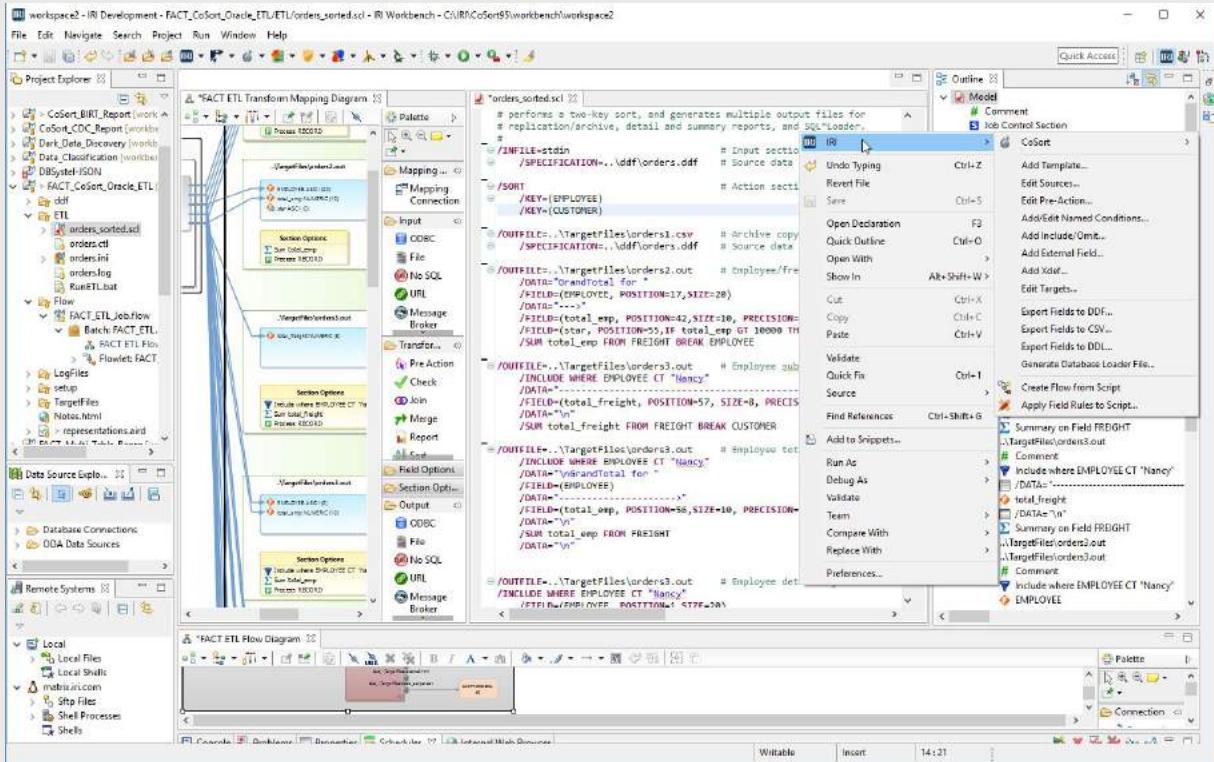
Replace Other ETL Tools

Replatform to save big money in a few weeks. Voracity is supported by erwin Mapping Manager so you can automate the conversion of legacy ETL tool mappings to Voracity jobs.

Voracity's 7 Job Design Options

Only Voracity gives you seven ways to create and modify metadata, jobs, and workflows in the same UI:

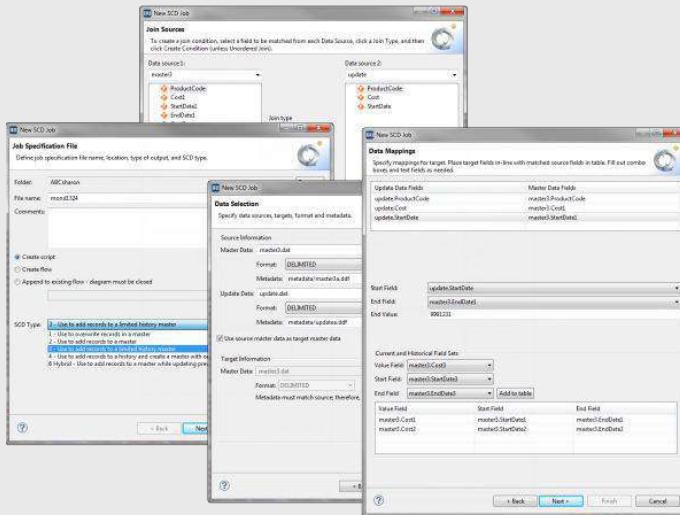
- 1) Wizards
- 2) Scripts w/ outlines
- 3) Form Editors
- 4) Dialogs
- 5) Diagrams
- 6) erwin Mapping Manager
- 7) IRI 'Gulfstream' Java API



Voracity Includes Purpose-Built Wizards for...



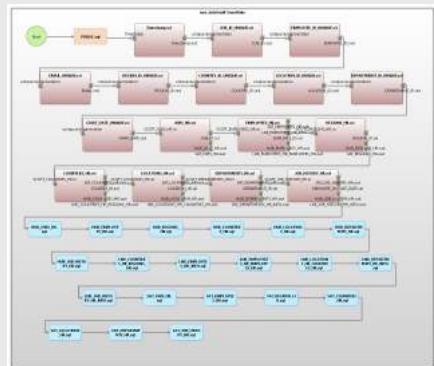
Change Data Capture



Slowly Changing Dimensions



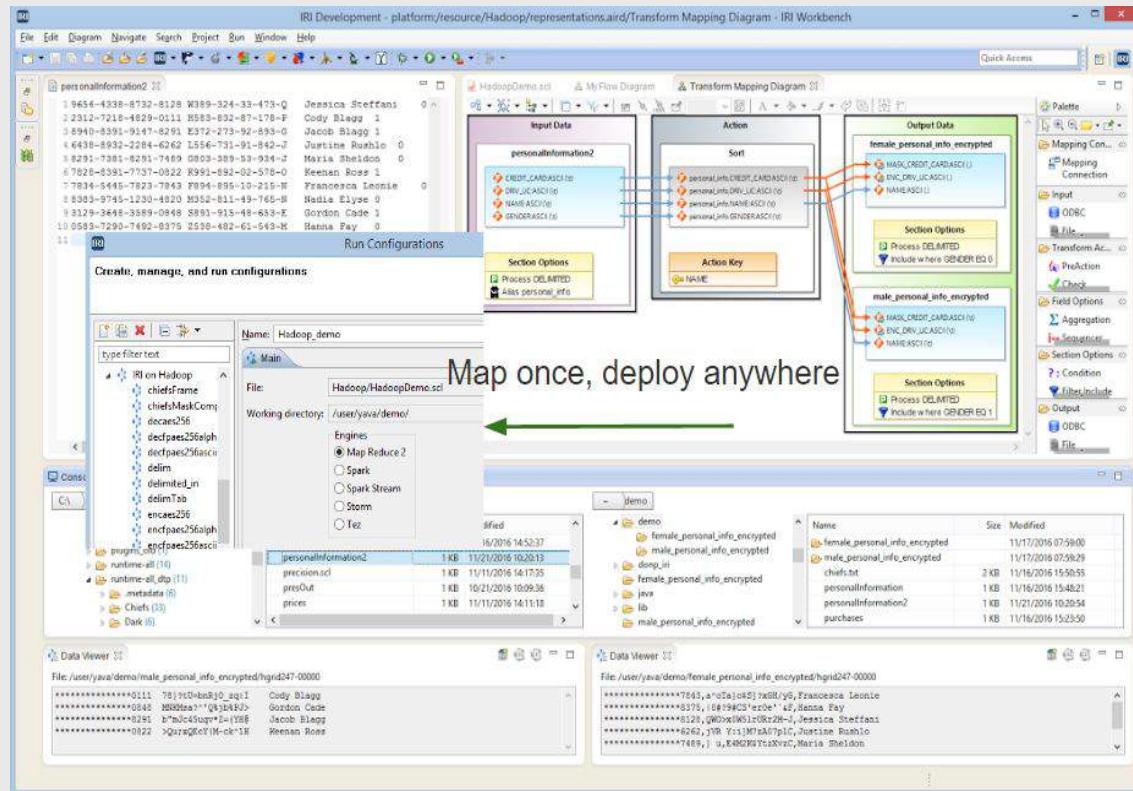
Pivot/Unpivot



Data Vault Creation & Test

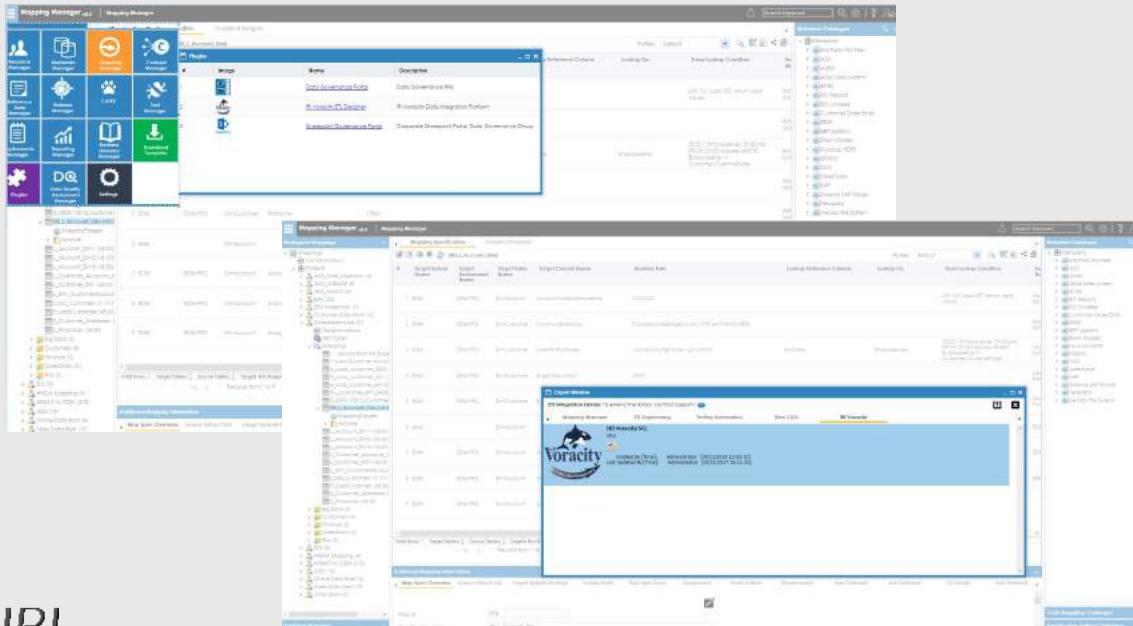
Voracity's 7 Job Deployment Options

- 1) Run 4GL scripts on the command line or in batch.
- 2) Use 3rd party automation tools like Stonebranch UAC, cron, AutoSys, Oracle job scheduler, etc.
- 3) Launch jobs from KNIME in Eclipse, or via Splunk apps, *as* you run or index them.
- 4) Execute seamlessly in Hadoop with MR2, Spark, Spark Stream, Storm or Tez. →
- 5) Use graphical run configurations and/or the built-in task scheduler to launch local, remote, or HDFS jobs from IRI Workbench
- 6) Make web service or 3GL program calls to Voracity's sortcl_routine() API
- 7) Invoke as SQL or COBOL system actions



Tie-In to erwin Metadata Mapping & Governance

Voracity is plug-compatible with the erwin metadata-driven automation and data governance platform. Create new, or convert legacy ETL tool, mappings for Voracity; plus assess data quality, set up workflows, track data lineage and impacts graphically, etc.



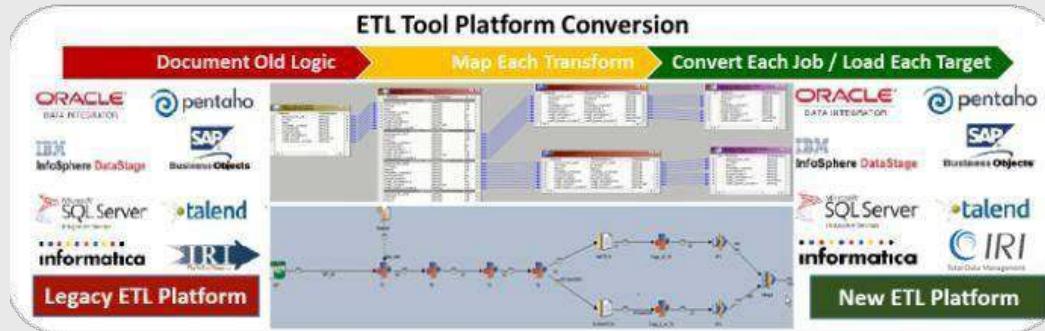
Ideal for:

- Data Integration Teams
- Business Users
- Regulatory & Compliance Officers
- Governance & Information Architects

How & Why You'd Leave Your Legacy ETL Tool

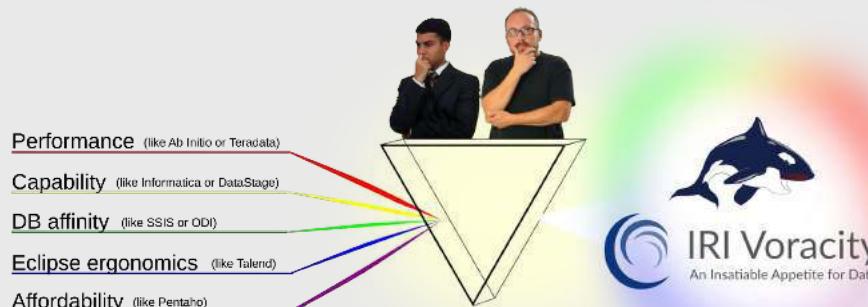
How

Through Erwin, legacy ETL tool and SQL users can convert their existing mappings to Voracity workflows automatically.



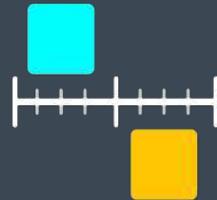
Why

Voracity workflows are faster, simpler, and far less expensive, allowing these users to re-platform and save 5-7 figures.



Using Voracity for Data:

Discovery Integration **Migration** Governance Analytics



Incremental Replication



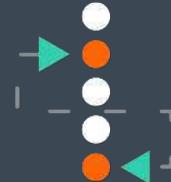
Data & File Types



Endianness



Databases



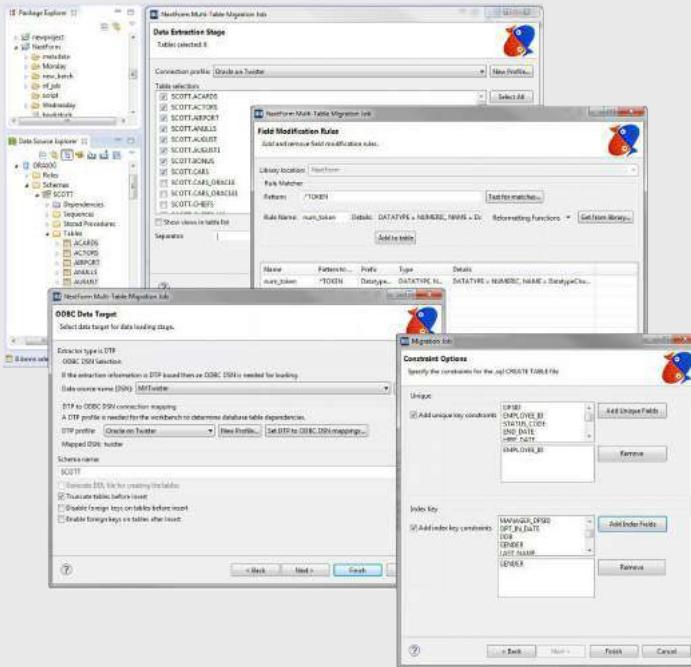
JCL Sorts



ETL Jobs

Why Voracity for Data Migration

Voracity converts, replicates, and reformats data from mainframe datasets, relational and NoSQL databases, index and sequential files, dark data documents, and cloud apps.



- Change data types, record layouts, file formats, and endianness
- Migrate column values and layouts, and relationships (constraints) between DBs
- Copy or refresh data from one or more sources to one or more targets
- Federate, or virtualize, data by mashing it up from disparate sources and creating custom, ad hoc views

Using Voracity for Data:

Discovery Integration Migration Governance Analytics



Data Quality



Data Masking



Data Lineage



Data Reconciliation



Test Data Generation



Metadata Management

Why Voracity for Data Governance

Search, Categorize, Cleanse, Enrich, Unify, Mask, and Track Data

- 1) Voracity data discovery wizards help you locate and classify data based on pattern searches, fuzzy matches, ML-NER, or value lookups, and then apply transformation or masking rules to data classes.
- 2) Disparate values can be reconciled and consolidated (mastered), while also being checked and fixed to comply with data formatting, data privacy, and business rules.

Use Voracity to acquire and govern data in a central marshalling area, and to achieve these outcomes:

Quality	Security	Lineage	Assurance
Validate, cleanse, enrich, and unify data for better apps, ETL, and BI results.	Find, classify, and rule-mask PII, or build test files/DBs and masked DB subsets.	See forward and reverse views of data changes through time, and analyze impacts.	Use query-ready audit logs and re-ID risk measurement to verify compliance.

Data Quality Features

Voracity has multiple ways to improve data quality in the data warehouse or data lake, and thus improve the accuracy of operations and the reliability of analyses and decisions.

- ✓ **Find** - discover, profile, and classify data from a quality standpoint
- ✓ **Filter** - remove or save conditionally selected or duplicate items
- ✓ **Unify** - data found by fuzzy match algorithms and set probabilities
- ✓ **Replace** - data found in pattern searches with literal or lookup values
- ✓ **Validate** - identify null values and other data formats by function
- ✓ **Regulate** - apply rules to find and fix data out of range or context
- ✓ **Synthesize** - custom composite data types and new row or file formats
- ✓ **Standardize** - use field-function APIs for Melissa Data or Trillium



PII Masking via Voracity-included FieldShield / DarkShield / CellShield EE!

- Connect and interact with **multiple sources** and targets, on-prem or cloud
- **Discover and classify** data in RDB, flat-file, Excel, and unstructured sources
- **Separate or combine** searching and masking operations
- Mask **static** or **streaming** inputs and data in Amazon S3, MQTT, Kafka, etc.
- Select from **14 masking categories** (e.g., encrypt, hash, pseudonymize, redact, blur)
- **Address multiple** protections, targets and recipients all in one job, one I/O
- Apply consistent, cross-table masking rules for **referential integrity**
- **Score re-ID** risk for FERPA & HIPAA EDM compliance and **anonymize** quasi-IDs
- **Condition** your masking based on data classes, patterns, values, or ranges
- Specify your target protections and formats in **Eclipse**, or in reusable **scripts**
- Integrate with **DB apps** via ODBC, or API via .NET/Java SDK
- Retain data **realism** via FPE and pseudonymization for testing or outsourcing
- **Mask during** Voracity ETL, DB migration, sub-setting, reporting or wrangling jobs
- **Log runtime** details to XML audit files, and manage user identities through **RBACs**



MongoDB Masking

Supported Access Methods:

1. Export, mask CSV, import
2. Progress J&ODBC drivers
3. Native BSON (shown)

The screenshot shows the IRI Workbench interface with the 'Transformation Mapping Diagrams' tab selected. On the left, there are two tables: 'db.list' and 'db.chiefs.find()'. The 'db.list' table has columns: _id, objects, device, never, nsize, dataRatio, index, extent, address, collect, nsize. The 'db.chiefs.find()' table has columns: id, listId, system, userId, random, count, never, totals, size, partition, status. A transformation mapping diagram is centered, connecting fields from both tables. On the right, there is a 'Section Options' panel with sections for 'Action' and 'chiefmask'. The 'Action' section contains 'Process MONGODB' and 'Mask Keys'. The 'chiefmask' section contains 'Process MONGODB' and 'Mask Keys'.

The screenshot shows the IRI Workbench interface with the 'Workbench' tab selected. The 'Project Explorer' pane shows a workspace structure with 'Mongo5-18' and 'chiefmask' projects. The 'Document List' pane shows a file named 'chiefout.sld' containing BSON code for masking. The 'Outline' pane shows the schema structure. The 'Terminal' pane shows the MongoDB command 'db.chiefs.find()' and its output, which is a list of presidents with their names, parties, and states. The 'Output Data' pane shows the masked data with redacted fields.

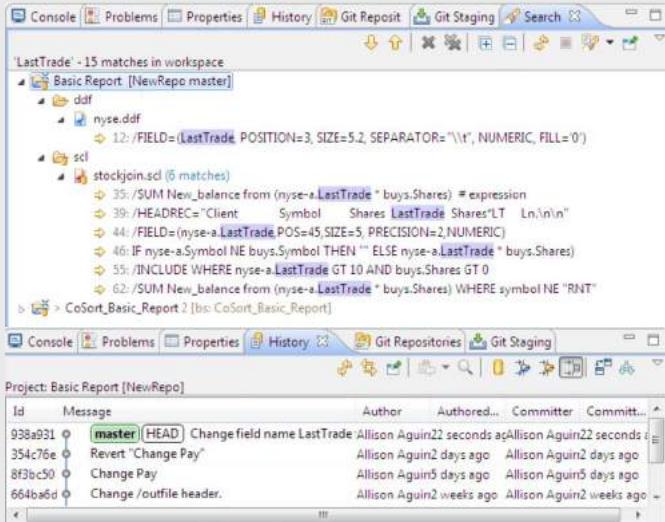
```
president party state
Adams, John FED MA
Adams, John Quincy D-R MA
Arthur, Chester A. REP VT
Buchanan, James DEM PA
Bush, George H.W. REP TX
Bush, George W. REP TX
Carter, James E. DEM GA
Cleveland, Grover DEM NJ
Clinton, William J. DEM AR
Coolidge, Calvin REP VT
Eisenhower, Dwight D. REP TX
Fillmore, Millard WHG NY
Ford, Gerald R. REP NB
Garfield, James A. REP OH
Grant, Ulysses S. REP IL
Harding, Warren G. REP OH
Harrison, Benjamin REP OH
Harrison, William Henry WHG VA
Hayes, Rutherford B. REP OH
Hoover, Herbert C. REP IA
Jackson, Andrew DEM SC
Jefferson, Thomas D-R VA
```

```
db.chiefs.find()
{
  $query: {},
  $fields: {
    president: 1,
    party: 1,
    state: 1
  }
}

[{"president": "Adams, John", "party": "FED", "state": "MA", "term": {"start": 1797, "end": 1801}, "id": 1, "Object": "Object#5d6b698f6b91a2000000c1"}, {"president": "Adams, John Quincy", "party": "D-R", "state": "MA", "term": {"start": 1825, "end": 1829}, "id": 2, "Object": "Object#5d6b698f6b91a3000000c2"}, {"president": "Arthur, Chester A.", "party": "REP", "state": "VT", "term": {"start": 1881, "end": 1885}, "id": 3, "Object": "Object#5d6b698f6b91a3000000c3"}, {"president": "Buchanan, James", "party": "DEM", "state": "PA", "term": {"start": 1857, "end": 1861}, "id": 4, "Object": "Object#5d6b698f6b91a3000000c4"}, {"president": "Bush, George H.W.", "party": "REP", "state": "TX", "term": {"start": 1869, "end": 1877}, "id": 5, "Object": "Object#5d6b698f6b91a3000000c5"}, {"president": "Bush, George W.", "party": "REP", "state": "TX", "term": {"start": 1889, "end": 1893}, "id": 6, "Object": "Object#5d6b698f6b91a3000000c6"}, {"president": "Carter, James E.", "party": "DEM", "state": "GA", "term": {"start": 1977, "end": 1981}, "id": 7, "Object": "Object#5d6b698f6b91a3000000c7"}, {"president": "Cleveland, Grover", "party": "DEM", "state": "NJ", "term": {"start": 1885, "end": 1889}, "id": 8, "Object": "Object#5d6b698f6b91a3000000c8"}, {"president": "Clinton, William J.", "party": "DEM", "state": "AR", "term": {"start": 1889, "end": 1893}, "id": 9, "Object": "Object#5d6b698f6b91a3000000c9"}, {"president": "Coolidge, Calvin", "party": "REP", "state": "TX", "term": {"start": 1893, "end": 1897}, "id": 10, "Object": "Object#5d6b698f6b91a3000000ca"}, {"president": "Eisenhower, Dwight D.", "party": "REP", "state": "IL", "term": {"start": 1897, "end": 1901}, "id": 11, "Object": "Object#5d6b698f6b91a3000000cb"}, {"president": "Fillmore, Millard", "party": "WHG", "state": "NY", "term": {"start": 1850, "end": 1853}, "id": 12, "Object": "Object#5d6b698f6b91a3000000cc"}, {"president": "Ford, Gerald R.", "party": "REP", "state": "NB", "term": {"start": 1861, "end": 1865}, "id": 13, "Object": "Object#5d6b698f6b91a3000000cd"}, {"president": "Garfield, James A.", "party": "REP", "state": "OH", "term": {"start": 1881, "end": 1885}, "id": 14, "Object": "Object#5d6b698f6b91a3000000ce"}, {"president": "Grant, Ulysses S.", "party": "REP", "state": "IL", "term": {"start": 1865, "end": 1869}, "id": 15, "Object": "Object#5d6b698f6b91a3000000cf"}, {"president": "Harding, Warren G.", "party": "REP", "state": "OH", "term": {"start": 1889, "end": 1893}, "id": 16, "Object": "Object#5d6b698f6b91a3000000cg"}, {"president": "Harrison, Benjamin", "party": "REP", "state": "OH", "term": {"start": 1881, "end": 1885}, "id": 17, "Object": "Object#5d6b698f6b91a3000000ch"}, {"president": "Harrison, William Henry", "party": "WHG", "state": "VA", "term": {"start": 1889, "end": 1893}, "id": 18, "Object": "Object#5d6b698f6b91a3000000ci"}, {"president": "Hoover, Herbert C.", "party": "REP", "state": "IA", "term": {"start": 1893, "end": 1897}, "id": 19, "Object": "Object#5d6b698f6b91a3000000cj"}, {"president": "Hayes, Rutherford B.", "party": "REP", "state": "OH", "term": {"start": 1877, "end": 1881}, "id": 20, "Object": "Object#5d6b698f6b91a3000000ck"}, {"president": "Jackson, Andrew", "party": "DEM", "state": "SC", "term": {"start": 1881, "end": 1885}, "id": 21, "Object": "Object#5d6b698f6b91a3000000cl"}, {"president": "Jefferson, Thomas", "party": "D-R", "state": "VA", "term": {"start": 1885, "end": 1889}, "id": 22, "Object": "Object#5d6b698f6b91a3000000cm"}]
```

Data Lineage & Impact Analysis

Track changes in column use over time for free through Eclipse searches, and metadata asset management utilities like Git:



Or get graphical column-level forward and reverse lineage and impact analysis for Voracity in erwin Mapping Manager:

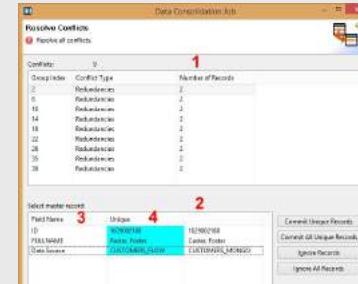


IRI is also working on an internal, encrypted IAM and granular logging system for reports on specific data value changes. See also [CDC](#).

Data Reconciliation / MDM

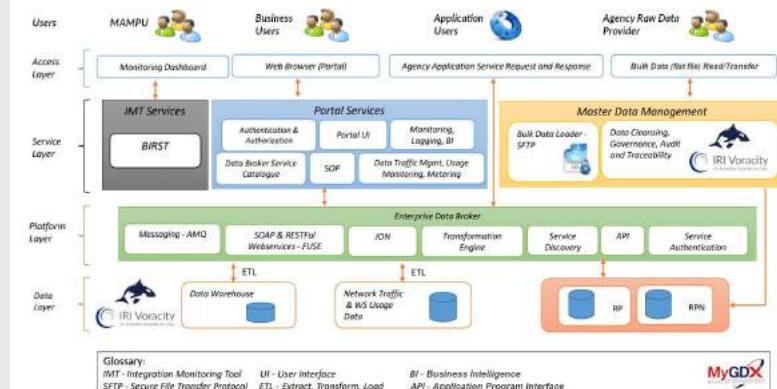
Voracity supports identification, matching, standardization, and protection of master customer and product information. Users can:

- Search, extract, profile, and classify
- Identify, unify, and bucket values
- Create and template values and formats
- Select and standardize from transactional data
- Deposit master data in tables or set files
- Extract, transform, load, virtualize, and report
- Cleanse and mask values
- Team-share, version-control, and lineage-track



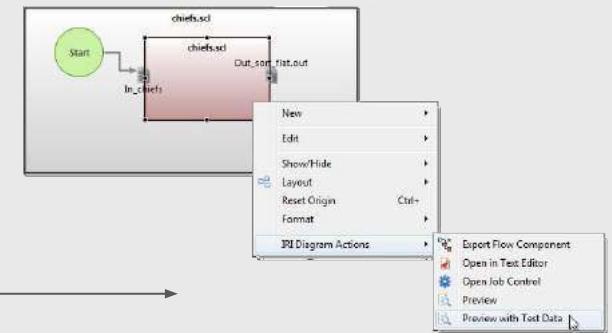
Voracity is also an ideal platform for building custom master data management applications, like the inter-agency government data exchange portal for Malaysia called MyGDX.

[Read the use case here.](#)



Test Data Management

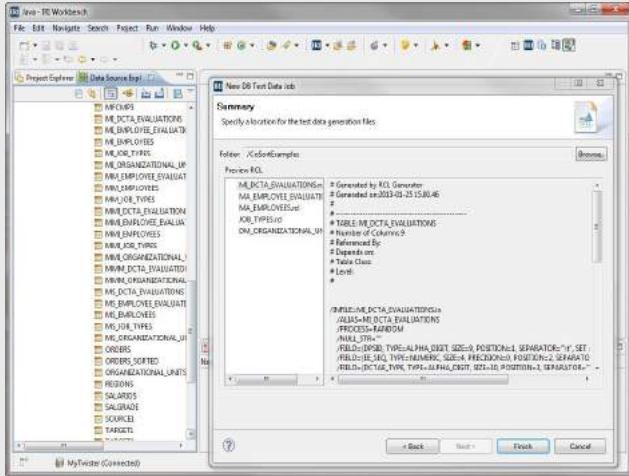
- Create synthetic but realistic **random** and **random-real** test data simultaneously
- Improve **DB prototypes**, application quality, benchmarking, and devops
- Leverage DDL, production file, and/or custom metadata
- Preserve structural and **referential integrity**
- Produce data in any type, structure, volume, value range, and “if” condition
- Synthesize **composite values** and custom (master) data formats
- Generate computationally valid and invalid NID, SSN, or CC#
- Set and graph test data **value distributions** (linear, normal, random, etc.)
- Apply common attribute rules (e.g., lookups) for pattern-matched field names
- **Filter, transform, and pre-sort** test data as you generate it
- Write loader metadata, and perform the loading, automatically
- Build test flat-file and custom detail and summary reports
- **Subset and mask** databases automatically as an alternative approach
- Use Java SDK functions to generate test data in apps and Hadoop
- Preview Voracity ETL jobs with immediate test data



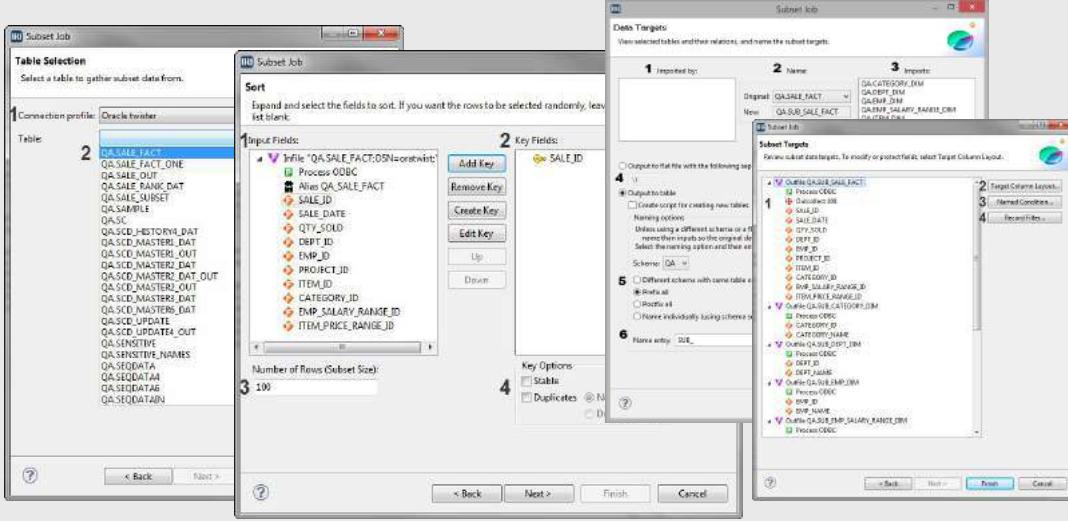
Versatile, Realistic Test Data from scratch, or masked subsets

Target Formats

- Files & Reports
- Mainframe
- RDBs
- Cloud/SaaS Apps
- ETL Ops/Tools
- Software Dev
- Benchmarking
- Demos & Outsourcing



Target Uses



In addition to data masking, Voracity also includes robust test data generation/population and DB subsetting wizards to facilitate DB, ETL, and BI prototyping. Either way, the test data is realistic, referentially-correct, and privacy-law compliant. And thanks to IRI RowGen within, Voracity users can even transform reformat, and report on data as it is (randomly) generated.

Metadata Management

Voracity leverages the same, simple 4GL metadata for data layout and manipulation.

IRI's data definition file (.ddf), mapping tasks/scripts, data class and rule libraries, and workflow metadata are all explicit, portable, and common across all data sources and platforms, including Hadoop.



Create or Acquire



Modify



Use Rules



Save and Reuse



Repurpose



Track, Audit, Analyze



Standardize and Save



Manage and Share

Using Voracity for Data:

Discovery Integration Migration Governance Analytics



Embedded BI



Cloud Dashboard



Data Wrangling



Predictive Analytics



BIRT, KNIME & Splunk



Clickstream Analytics

Why Voracity for BI & Analytics

Immediate Displays, or Prepared Data for Decision Tools

- 1) Simultaneously prepare raw data and present it in 2D reports, BIRT, cloud dashboards, or Splunk or ...
- 2) Hand off filtered, transformed, cleansed, and masked subsets to BOBJ, Cognos, Microstrategy, Oracle DV, Power BI, QlikView, R, SpotFire and Tableau so they can display results 2-20X faster than if they self-stage.

Either way, analytic data quality and speed improve dramatically. Additional advantages are:

Efficiency

Design effort and I/O drop significantly if data prep tasks and reporting jobs run at the same time and place.

Consistency

Homogenize and centralize data so it can be reliably re-used in multiple reporting scenarios.

Compliance

Apply field-level data masking and cleansing functions directly in reports or handoffs as they are produced.

Cost

Voracity subscriptions are priced lower than data preparation tools. BIRT in IRI Workbench is free.

From its one IRI Workbench (Eclipse IDE), Voracity supports multiple analytic approaches ...

Voracity Analytic Option 1: Embedded BI

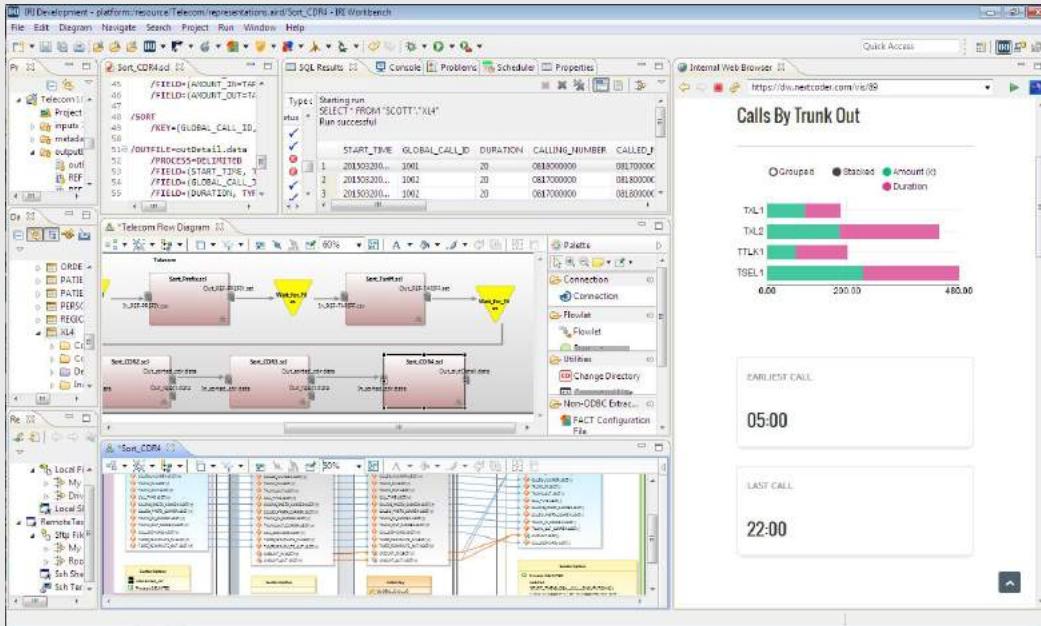
Unlimited 2D reporting
in custom-formatted,
detail and summary files,
XML, HTML, etc.

The screenshot shows the IRI Workbench interface with multiple windows open:

- Project Explorer:** Shows a project named "CoSort_Basic_Report" with various source files like "buy.csv", "nyse-a.csv", and "stockjoin.scl".
- Console:** Displays a "SortCI job" command and its output, which includes a table of stock data with columns "Symbol", "Client", "Shares", and "LastTrade".
- Data Source Expl:** Shows a tree view of data sources including SCOTT, ODBC, and MongoDB.
- HTML produced by:** A preview window showing a table of names and companies.
- TradingA.htm:** An XML file containing an Action Statement for sorting and joining data from "buy.csv" and "nyse-a.csv". It also defines output files for "buyers" and "New_Balance".
- StockJoin.scl:** An SQL script for joining "buy.csv" and "nyse-a.csv" based on "Symbol".
- Targets:** A list of targets including "TradingA.htm", "TradingA.xml", and "Notes.html".
- Output Files:** A list of generated files including "buyers", "New_Balance", and "TradingA.xml".
- Summary:** A summary table showing the total value of trades.

Voracity Analytic Option 2: Cloud Dashboard

Leverage drill-down, browser-based dashboard applications, like this one in [DWDigest](#), or others like iDashboards



Voracity Analytic Option 3: Data Blending

Prepare CSV, XML, or table subsets
to speed time-to-display 2-20X,
and to improve data quality,
privacy, and storage space



Option 3 Example: Data Blending for R



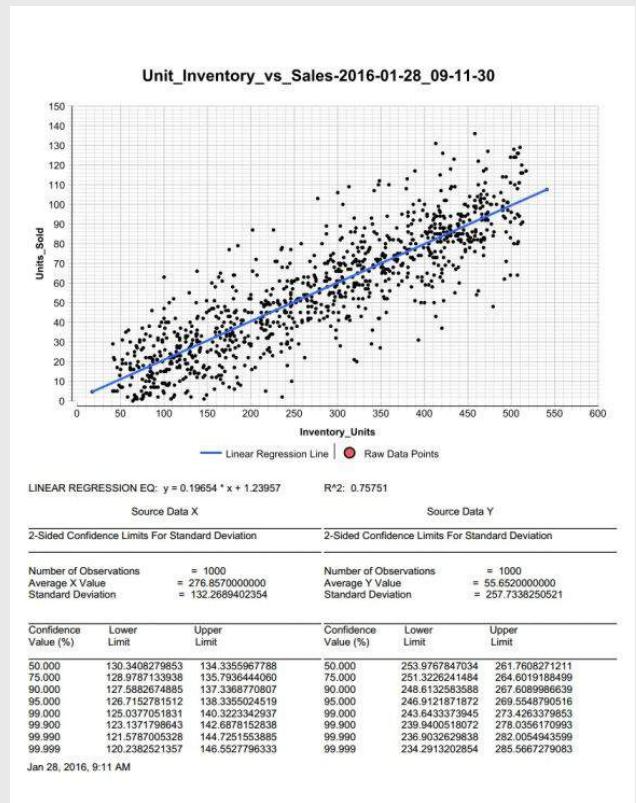
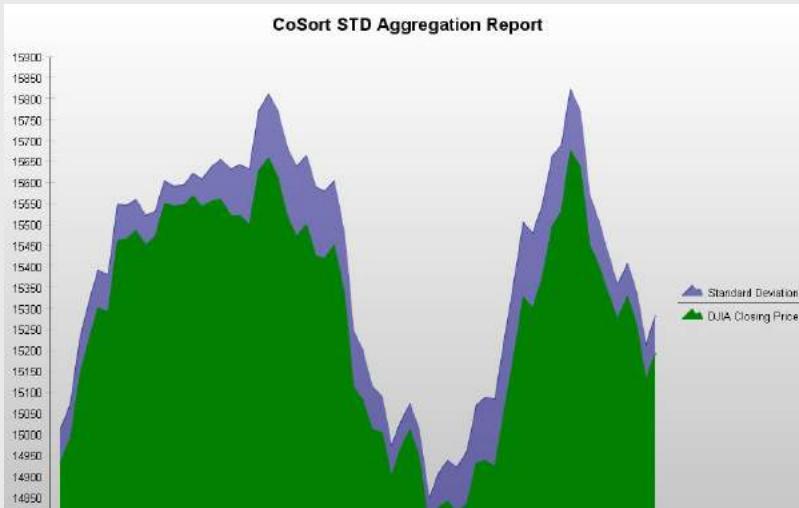
The image shows four windows from the RStudio IDE. The top-left window displays an R script named `t_commands.R` which reads transaction records from a database, merges them by store, and then writes the results to a file. The top-right window shows another R script named `t_commands_sgDataR.R` which performs similar operations using a different data source. The bottom-left window is a terminal showing the execution of the R script, with output indicating a total elapsed time of 475.38 seconds. The bottom-right window shows the command-line interface for Voracity, with a job named `SetC1.job` running, which involves reading serial numbers from a file and writing them to a database.

On a PC with 6GB of RAM, R could only process 30MB of data in 3MB chunks. R needed 11 jobs or nodes to break down the data and merge the results ...

... The same data prep in Voracity happens in just one sort-join-aggregate program (and I/O pass) that runs 45% faster than R in this small case.

Voracity Analytic Option 4: Predictive Analytics

Use statistical functions and fuzzy lookup logic native to CoSort, or regression libraries from Boost. Simultaneously display trends and other predictive information in 2D reports and/or [BIRT displays](#).

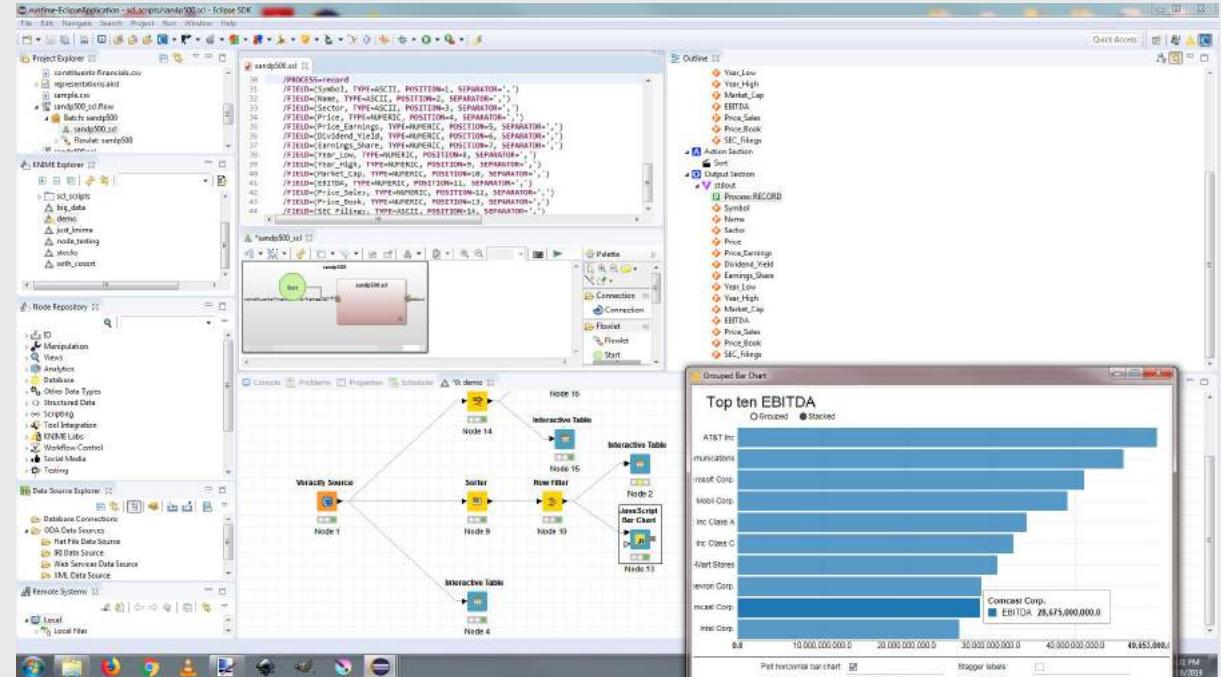


Voracity Analytic Option 5: KNIME Integration

Feed KNIME Analytic

Platform targets in memory with data prepared for predictive analytics, deep learning, machine learning, and other data mining and science nodes.

Speed time to insight in the same pane-of-glass ...

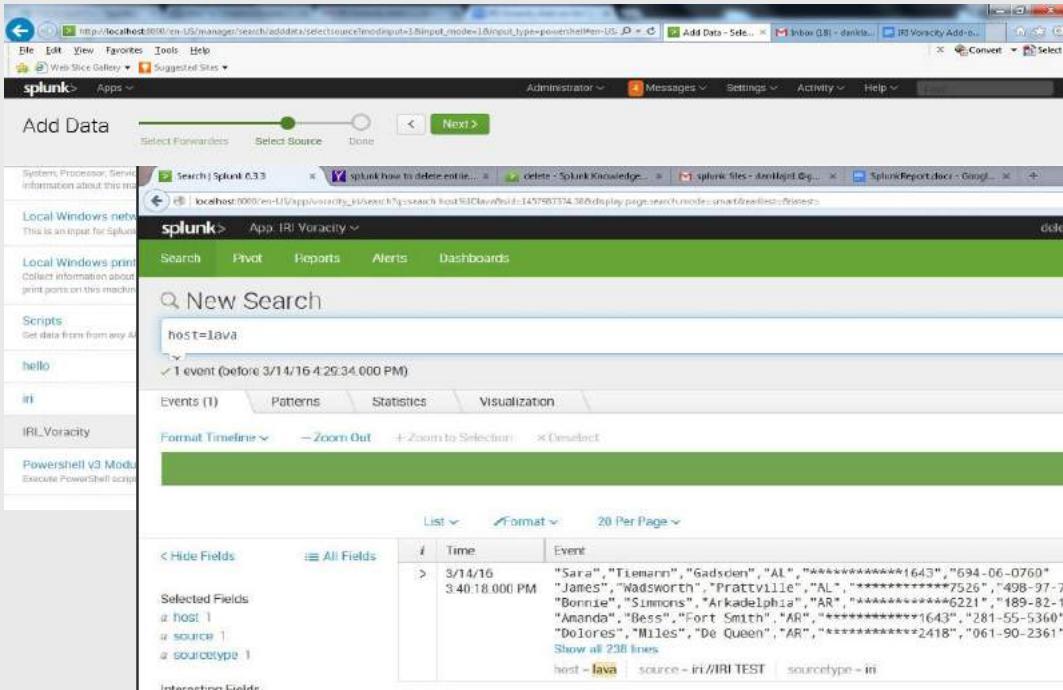


Voracity Analytic Option 6: Splunk

Prepare and index data for
Splunk simultaneously.

There is both a Voracity
app and add-on for Splunk.

Voracity also supports
operations through the
Splunk Universal
Forwarder and Splunk
Phantom Playbooks.

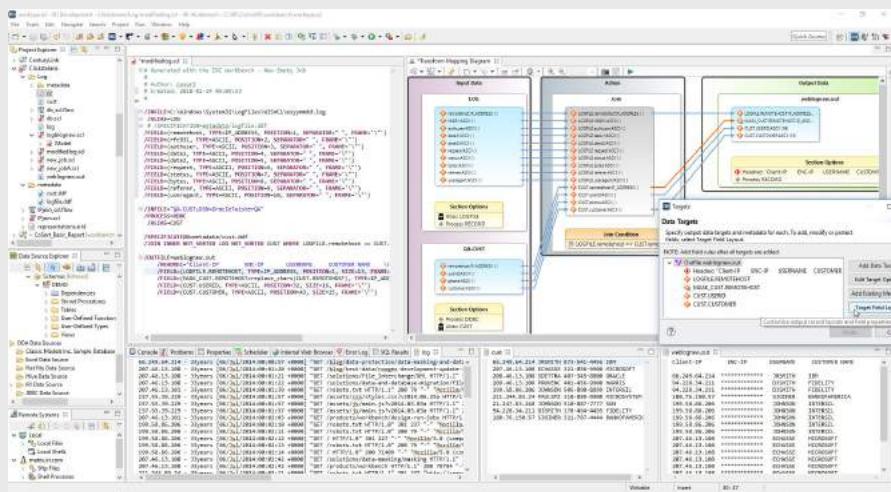


The screenshot shows the Splunk web interface. In the top navigation bar, there is a link to 'IRI Voracity'. The main search bar contains the query 'host=java'. Below the search bar, it says '1 event (before 3/14/16 4:29:34.000 PM)'. The results table has columns for 'Time' and 'Event'. One visible event is: "Sara", "Tiemann", "Gadsden", "AL", "*****1643", "694-06-0760". The bottom of the table shows 'Show all 238 lines' and filters for 'host = java', 'source = iri/IRI TEST', and 'sourcetype = iri'.

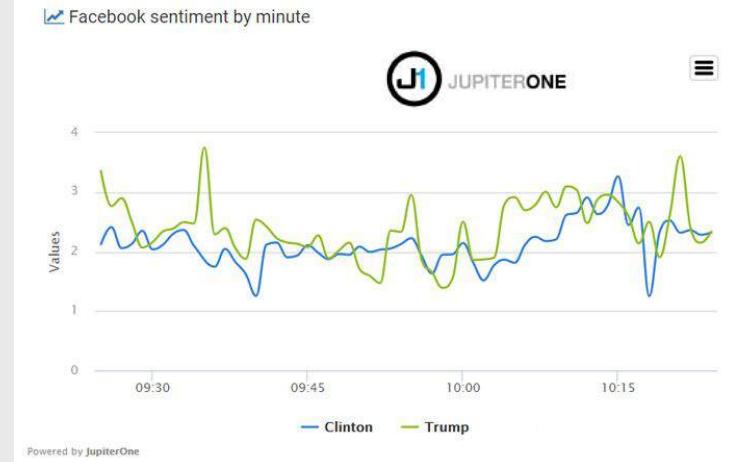
Time	Event
3/14/16 3:40:18 000 PM	"Sara", "Tiemann", "Gadsden", "AL", "*****1643", "694-06-0760" "James", "Wadsworth", "Prattville", "AL", "*****7526", "498-97-75 "Bonnie", "Summons", "Arkadelphia", "AR", "*****6221", "189-82-17 "Amanda", "Bess", "Fort Smith", "AR", "*****1643", "281-55-5360" "Dolores", "Miles", "De Queen", "AR", "*****2418", "061-90-2361"
Show all 238 lines	

Voracity Analytic Option 7: Clickstream Analytics

Native support for CLF and ELF log formats facilitates integration and reporting with other sources



Kafka support enables big data push and pull with NLP-enabled / social media sentiment analytic platforms



Voracity Summary / Data Curation Functions

Profile & Acquire

Discover and extract data and metadata in disparate sources. Define custom structures, mask formats, and build test data.

Cleanse & Unify

Filter, enrich, scrub and standardize data in multiple sources. Find and merge reference data into master sets.

Process & Provide

Integrate, migrate, govern, and analyze data in the same job and I/O pass. Visualize and feed test or real targets.

Protect & Audit

Mask data at the field level as you acquire, transform, report, or blend it. Log activity granularly and score re-ID risk.

Express & Predict

Aggregate, cross-calc, and format data in detail, summary and trend reports. Or, hand-off results to your analytic tool or BIRT/Splunk in memory.

Convert & Replicate

Migrate legacy databases, or files and data types -- or specify new record layouts. Copy or subset (and mask) data in any structured format or schema.

Publish & Share

Federate, save, or populate multiple targets at once. Connect to sources and their metadata in secure repositories for change tracking, etc.

Why Voracity is Better

Voracity users do more, run faster, and pay less than users of legacy ETL platforms and specialty/Apache tools

Speed

Voracity has the best E, T, and L performance without Hadoop (via CoSort), plus multiple Hadoop options for unlimited scalability.

Ease

Voracity uses a simple, open 4GL metadata and familiar Eclipse™ GUI for everything, and includes more job design options than any other tool.

Versatility

Voracity combines data discovery, integration, migration, governance, and analytic functionality so IT architects, business users, and governance teams can work together and adapt to change.

Value

Voracity unifies data and enterprise information management, delivers what ETL and Hadoop users want, and bends big data's cost-benefit curve in your favor. \$30K and up for unlimited users per year.



Total Data Management



IRI Voracity

An Insatiable Appetite for Data

Use Cases



Retail



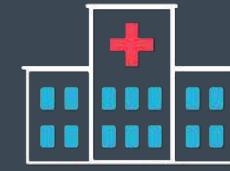
Energy &
Transport



Telco &
Media



BFSI



Healthcare



Banking, Financial Services & Insurance (BFSI)

Assess Credit Risk

Use CoSort and Hadoop engines in Voracity to blend traditional credit data with sources like utility bill and rental payments to improve score accuracy, facilitate lending, marketing, etc.

Optimize Loan Performance

Use Voracity to blend and prepare internal and external data points (borrower history, industry repayment stats, social/market forces, etc.) for visual analytics on risk factors vs. loan rates.

Expose Insurance Fraud

Use Voracity to rapidly sort, filter, and expose claim data outside normal parameters to identify suspicious behavior, and feed it to visualization and notification apps in the same IDE.



Healthcare

Improve Treatment Outcomes

Flow IOT data through slowly changing dimension or change data capture processes in Voracity to compare patient data with diagnostic values to spot, alert, and correct for abnormalities.

Individualize Drug Therapies

Rapidly integrate genetic data into single-node-type networks, gene-set libraries, and bi-partite graphs to help reveal new relationships between patient genes, drugs, and phenotypes.

See the Whole Patient

Use Voracity search, join, consolidate, and masking features to unify and de-identify protected health information (PHI) in family, provider, demographic, diagnostic, and treatment data silos.



Energy & Transport

Conserve & Troubleshoot

Use the IoT edge aggregation and hub analytics in Voracity on smart meter and thermostat data to identify peak uses, or on grid sensors and weather data to re-route power, inspect, repair, etc.

Improve Traffic Flow

Combine data from street cameras and sensors, cell phone apps, and weather data in Voracity and feed it directly into BIRT-connected Integeo geospatial reports to warn drivers.

Optimize Fleet Performance

Use IoT analytics and alerting features in Voracity to predict and prevent equipment failures, and its DW/BI prowess against historic O&D and pricing data to maximize passenger revenues.

((o)) Telco & Media

Monetize Calls & Clicks

Use Voracity to process CDRs and clickstream data for billing and analytics, and to sell that data to marketing affiliates and others who can permissibly use it.

Anticipate Spending Trends

Use Voracity to extract string and pattern-matching values from social data from Hubspot, etc., and munge it with transaction and demographic data to identify and predict content preferences.

Throttle & Enforce

Use Voracity to identify excessive bandwidth usage or illegal activity from network traffic or web logs, and tie it to analytic and notification mechanisms in the same IDE.



Retail

Micro-Target Customers

Use Voracity to segment purchase groups for targeted marketing and to create holistic, unified views of each customer that help you customize service and build loyalty.

Leverage Consumer Psychology

Use Voracity to integrate consumer behavior and sentiment data against seasonal, regional, and other factors, and mine it with regression analyses that reveal trends.

Price Smarter

Use Voracity to integrate preference and pricing data from retail data brokers, public data, your own pricing history, and competitive research.

Voracity Partnering Opportunities

IRI aligns with consulting companies across multiple disciplines and industries, and through many different commercial models (referral, resale, and value-added support and training services). IRI never imposes quotas or “partner fees” ... please see iri.com/partners or email partners@iri.com. Some the companies trained (or training now) on Voracity or its components for their clients are:





iri.com

[IRI Blog](#)

[IRI Voracity Data Management Group on LinkedIn](#)

