
IRI Voracity
An Insatiable Appetite for Data
2024 Platform Introduction

IRI

Total Data Management

Who We Are

Specialists in big data manipulation and governance

Known since 1978 for JCL sort migration and ETL speed

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

A multi-discipline data masking and TDM industry leader

Partners to resellers and consultants worldwide

Our Mission

To continue supporting our wide range of data management and security solutions through software which uniquely combines:

- Speed and scalability
- Versatility and interoperability
- Familiarity and usability
- Flexibility and affordability

IRI
Total Data Management

Our Bigger Data Users

How We Help Them

Rapidly integrate and wrangle big data on premise or in the cloud for ETL and analytics

Validate, cleanse, enrich, and migrate data for replication and legacy modernization

Find, classify, and mask PII for privacy law compliance, DevOps (test data management) and data breach protection

Lower learning curves, licensing costs, and system impacts with software reputed for its ease, affordability and speed in volume

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

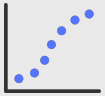
A Data Stewardship Portal

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

IRI
Total Data Management

Platform Product Components

IRI
Total Data Management
www.iri.com
info@iri.com
+1.321.777.8889

IRI Data Manager Suite

IRI CoSort
Sort, Transform & Report

Speed or replace legacy sorts, 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, reports, and DevOps simultaneously

IRI Voracity
An Insatiable Appetite for Data

Consolidate tools and tasks to process, protect, prototype, present

- Discover, define, and manage data in legacy and new sources
- Combine data integration, migration, governance, and analytics
- Use IRI Ripcurrent to replicate or mask changed data in real-time
- Leverage the familiarity of Eclipse and the power of CoSort

IRI Data Protector Suite

IRI FieldShield
PII / PHI Classification & Masking

Static and dynamic masking of structured data sources

- 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

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

- Define or 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 everywhere

- Find PII in LAN and cloud souces using multiple methods
- Simultaneously de-identify, remove, or report those values
- Mask text, MS, PDF, Parquet & image files + LOBs & NoSQL
- Comply with the right to erasure, portability, or rectification

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

DESIGN

Wizards with Rules | Graphical Dialogs
Scripts with Outlines | Form Editors
Workflow & Mapping Diagrams
Erwin Mapping Manager
DataSwitch No-Code

SOURCES

- Hadoop & Streams
- ASN.1 CDRs
- Flat & EDI Files
- Cloud & SaaS
- Relational DBs
- NoSQL DBs
- Text & Images
- Mainframe
- Logs, Excel, etc.

DISCOVER

Data Classification
Dark Data Search
DB & File Search
DB & File Profiling
ER Diagramming
Multi-Source Metadata

INTEGRATE

Slowly Changing Dimensions
Public/Private Mashups
Change Data Capture
Fast DB Un/Load
Data Federation
One-Pass ETL

MIGRATE

Incremental Replication
Database Platforms
Data & File Types
Legacy Sorts
Endianness
ETL Tools

IRI Voracity
An Insatiable Appetite for Data

GOVERN

Data Quality
Data Masking
DB Subsetting
Re-ID Risk Scoring
Test Data Synthesis
Data & Metadata Lineage

ANALYZE

IoT Feeds
In Datadog
Embedded BI
Data Wrangling
KNIME & Splunk
Predictive Analytics

TARGETS

- Kafka & MQTT
- BI & Analytic Tools
- Cloud Stores
- Relational DBs
- NoSQL DBs
- Custom Reports
- DevOps
- Flat & EDI Files
- Logs, Excel, Images

DEPLOY

GUI, CLI, API | MapReduce 2 (Grid)
Spark (In-Memory) | Storm (Streaming)
Tez (Batch) | CI/CD | Java | SQL | YARN
Eclipse or Any Scheduler

Data Sources (Standard)

Acucobol (MF) Vision	ESDS	MF- & RM-ISAM	Tibero (FACT)
Altibase (FACT)	Excel	MF Var. Length	Teradata
ASN.1 CDRs	HL7 (DS)	MySQL	Text
C-ISAM	HSQLDB (WB)	Oracle	TSV
CLF web logs	IDX 3, 4 & 8	PDF (DS)	UTF-8 & 16
CSV	Informix	PostgreSQL	Variable Block
DB2 (UDB)	Ingres	Record Sequential	Variable Sequential
DB2 for i5/OS	LDIF	RTF (WB)	VSAM MVS (UniKix)
DB2 for z/OS	JSON	SQL Anywhere	Web Services
Delimited	Line Sequential	SQL Server	Word (DS)
Derby (WB)	MariaDB	SQLite	X12 (DS)
ELF web logs	MaxDB	Sybase ASA/E & IQ	XML

***FACT**: requires IRI Fast Extract (FACT)* ***DS**: requires IRI DarkShield*
***WB**: requires IRI Workbench, the free Eclipse GUI for FieldShield, etc.*

IRI
Total Data Management

Data Sources (Legacy)

Access	D3	GA-Power 95, R91	K-ISAM	Pathway	RMS
Adabas	Datacom	Gemstone	Knowledgeman	PDS	Reality/X
Advanced Pick	Dataflex	GENESIS	KSDS	PervasiveSQL	RRDS
ALLBASE	Db4o	Gigabase	Lotus	Pick/Pick64+	Sequoia
Alpha5	dBase	H2	Manman	PI-Open	SFS (VS*)
Amazon RDS	Desktop Adapter	IDMS	Mentor / pro	Powerflex	Sharebase
Azure	DL/1	IDS	MO	Powerhouse	Supra
BizTalk	DSM	Image	Model 204	Progress	Terracotta
Cache	Enscribe	IMS	Mumps	QueryObject	Total
Clipper	Enterprise Adapter	Interbase	MyBase	rBase	Ultimate
Codasyl	FileMaker	Intersystems	Netezza	R83	UltPlus
CorVision	Firebird	ISM	NonStop SQL	Rdb	Unidata
ConceptBase	Focus	Jasmine	ObjectStore	REALITY	Universe
D-ISAM	FoxPro	JBase	Paradox	Red Brick	VSAM VSE

These sources are typically only accessible via IRI partner (SoftwareAG-CONNX) J/ODBC drivers.
**IBM/Encina SFS files should be supported when written in COBOL using RECORDING MODE IS VARIABLE*

IRI
Total Data Management

Data Sources (Modern)

Amazon EMR Hive	DynamoDB	Redis & Solr	Parquet files
Amazon RDS	ElasticSearch	MarkLogic (XML)	Pivotal Greenplum
Apache Cassandra	Google BigQuery	MongoDB	Pivotal HD Hive
Apache Hadoop Hive	Google BigTable	MS Dynamics CRM	SAP HANA
Azure CosmosDB	Hortonworks Hive	MS SQL Azure	Salesforce.com
Cloudera CDH Hive	Hubspot	Oracle Eloqua	Snowflake DB
Cloudera Impala	Kafka Connect	Oracle Cloud DB	Spark SQL
Database.com	MapR Hive	Netezza	Vertica DB

IRI FieldShield finds and masks structured RDB data on-premise, or in HDFS, AWS, Azure, GCP or OCI, **plus**: data in flat files (which can also be in S3 buckets or Hadoop), as well as ASN.1 CDRs, MF-ISAM or Vision fies, and Excel sheets. ***IRI DarkShield*** supports RDB and flat file data, too, **plus**: semi- and unstructured data in static or streaming text, log and EDI formats like JSON, HL7, X12 and XML; C/LOB columns in RDBs; Excel, PDF, and Word documents (including PII in their embedded images); NoSQL DBs; and, image files (BMP, DICOM, GIF, JPG, PNG and TIFF). The DarkShield API can run on premise or in the cloud, and read/mask/write PII from/to files in AWS S3 buckets, Azure BLOB, GCP Storage, or SharePoint (OneDrive). ***IRI CellShield*** only supports Excel (XLS/X), and works inside Excel, on-premise or in Office 365.

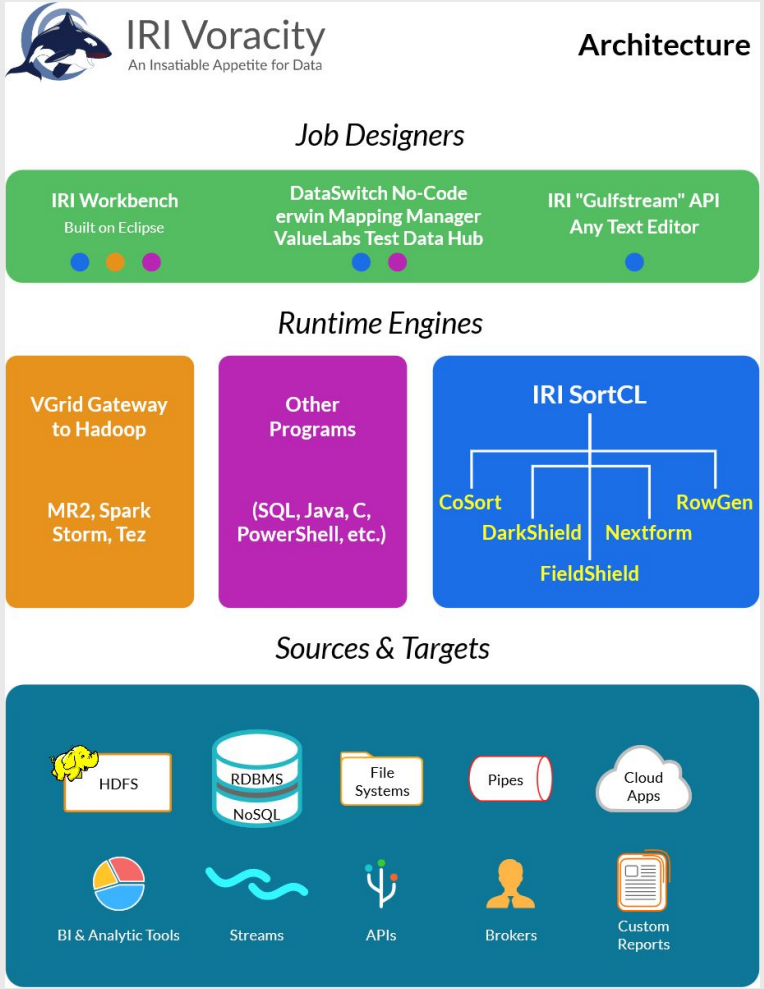
IRI
Total Data Management

Voracity 2-Tier Architecture

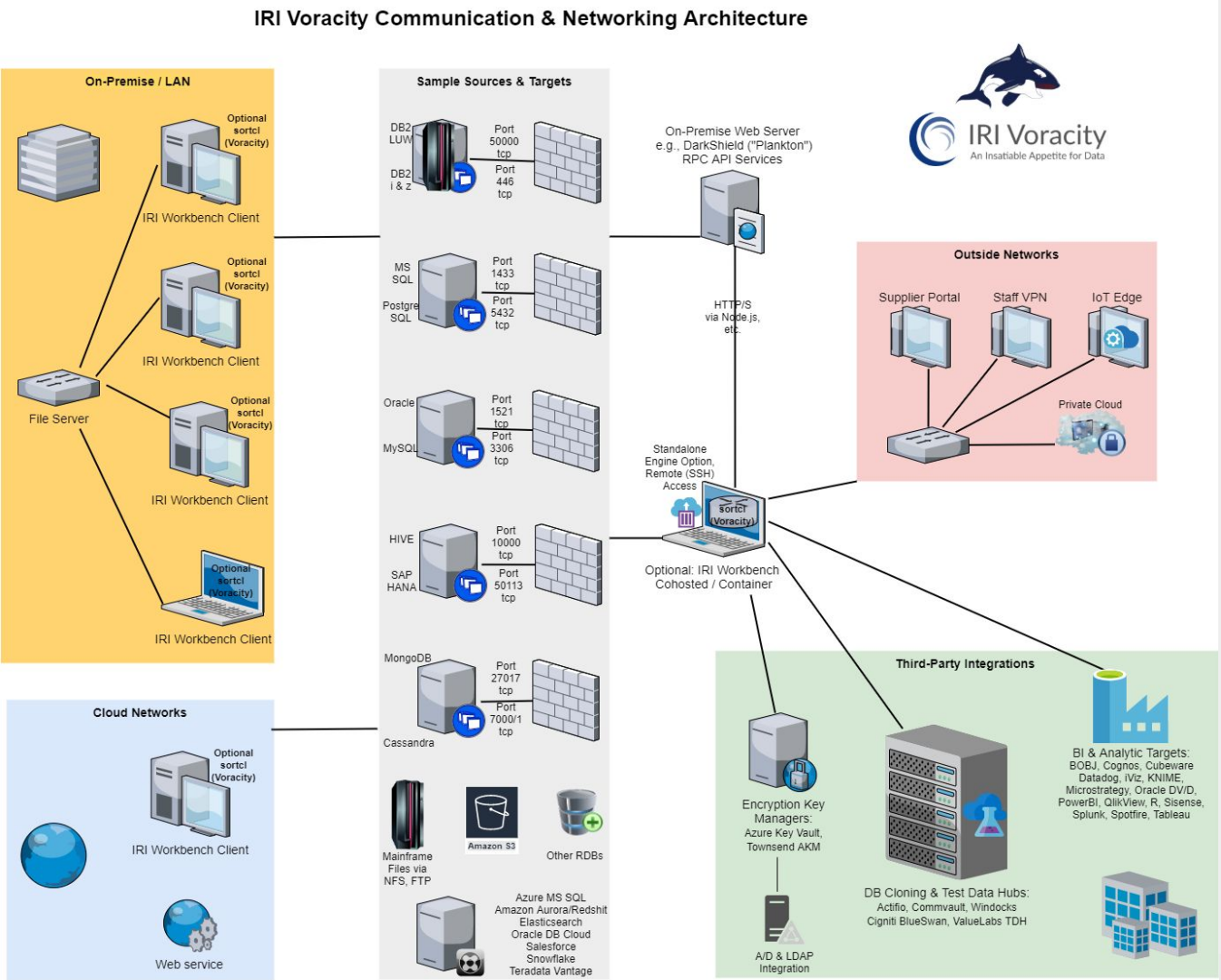
The default Voracity stack uses both the ***front-end*** IRI Workbench graphical IDE for client-side design of data-driven jobs defined in portable CoSort SortCL scripts.

The scripts are fully supported in the Workbench data model, a Java API called Gulfstream, and in web applications like erwin Mapping Manager and DataSwitch for seamless job creation, modification, and management.

The ***back-end*** SortCL engine which runs these jobs is the default, C-language executable supporting Windows, Linux and Unix systems ranging from a Raspberry Pi to a z/Series mainframe. Many of the same scripts also run interchangeably in Hadoop.



IRI
Total Data Management



Hardware Prerequisites

For x86 systems, a minimum configuration for Workbench would be 4GB of RAM and 10GB of free disk space, after the installation of any VMs, DBs, etc.. However, 6GB and up works best for each system to accommodate multiple database connections and table parsing for metadata and job definition. For schemas with hundreds of tables to enumerate, as much as 64GB of RAM could be appropriate for the Workbench machine(s) where RDB-related jobs are built.

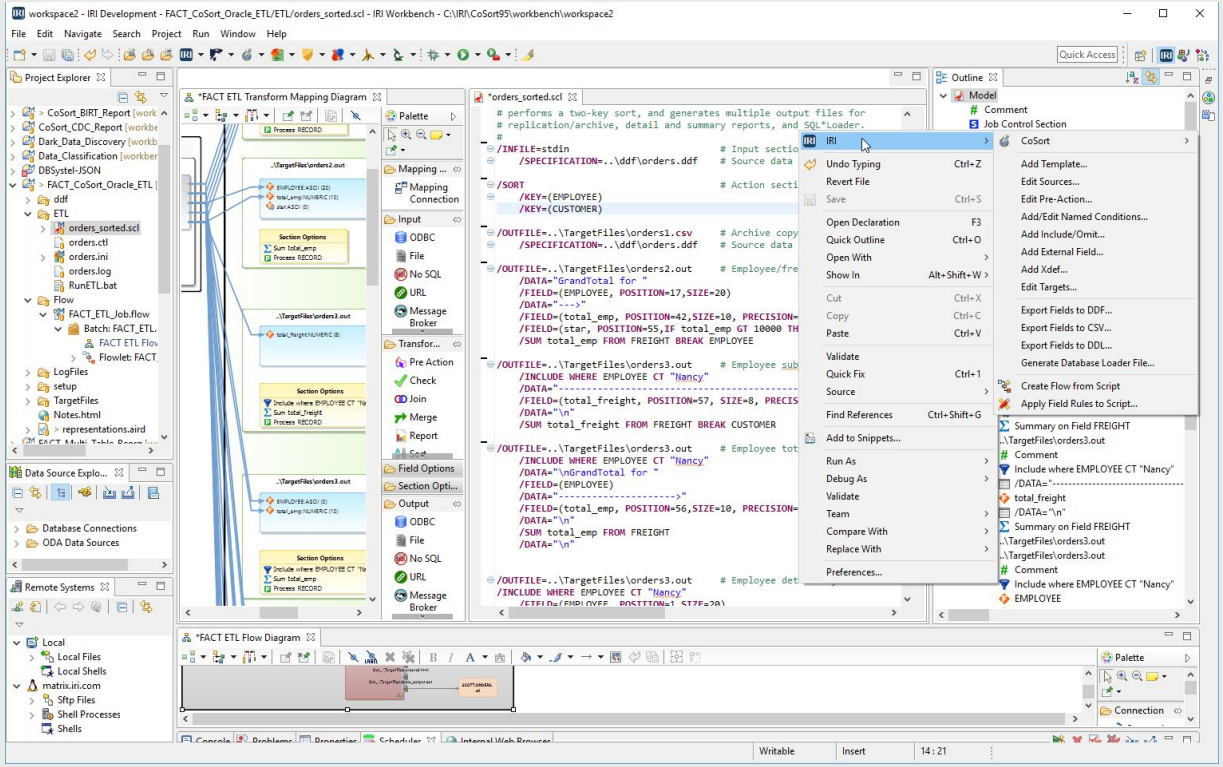
IRI also recommends where possible the co-location of the licensed back-end (SortCL executable) on or within close network proximity to database source or target servers for performance reasons, particularly if there are known network bottlenecks.

Data maps, masks, munges, and mines essentially at movement speed, so consider network and I/O resources.

Voracity's 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) DataSwitch no-code app
7) erwin Mapping Manager
8) IRI 'Gulfstream' Java API



Voracity's Job *Deployment* Options

1) 4GL scripts on the command line or in batch

2) 3rd-party automation tools like Stonebranch UAC, cron, AutoSys, & Oracle job scheduler

3) Launch jobs from a KNIME node, or from a Splunk app which indexes the target data

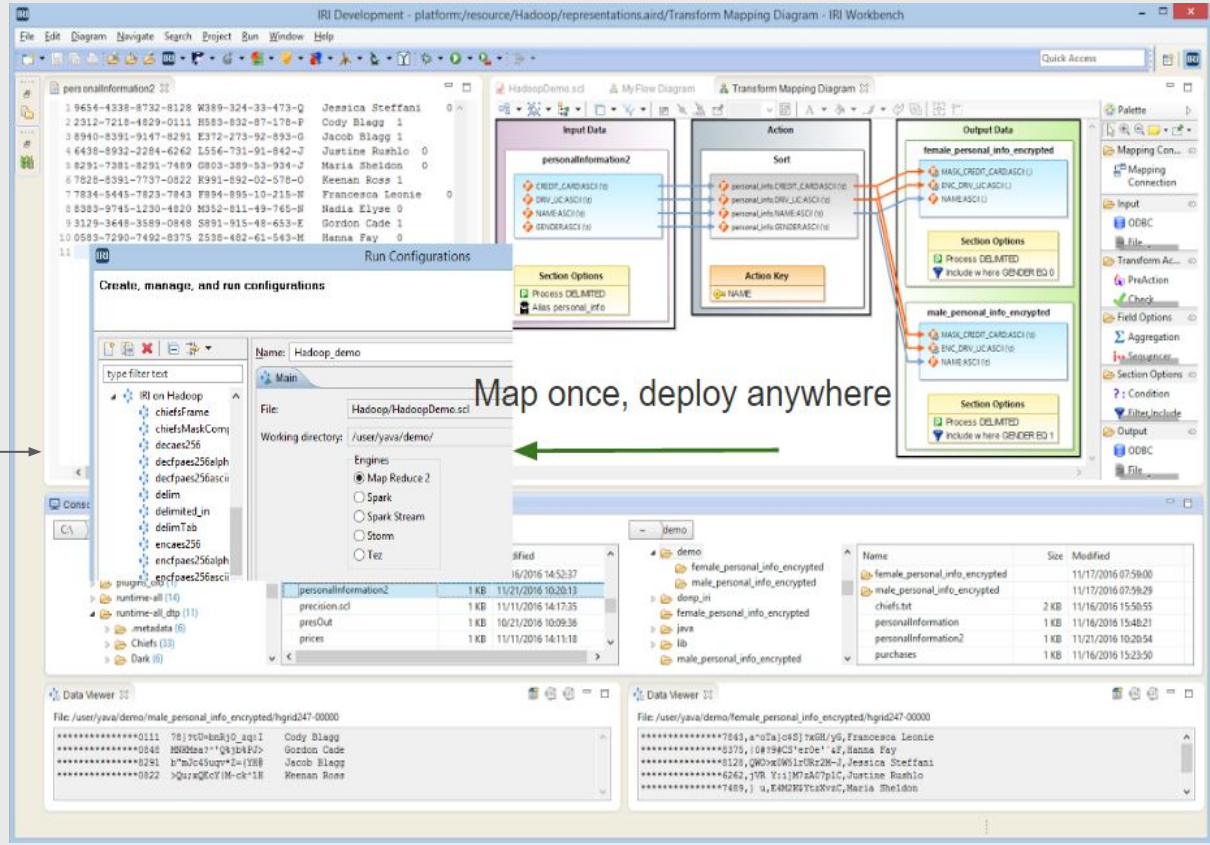
4) Execute seamlessly in Hadoop with MR2, Spark, Spark Stream, Storm or Tez.

5) From graphical run configurations and/or the built-in task scheduler for local, remote, or HDFS jobs directly from IRI Workbench

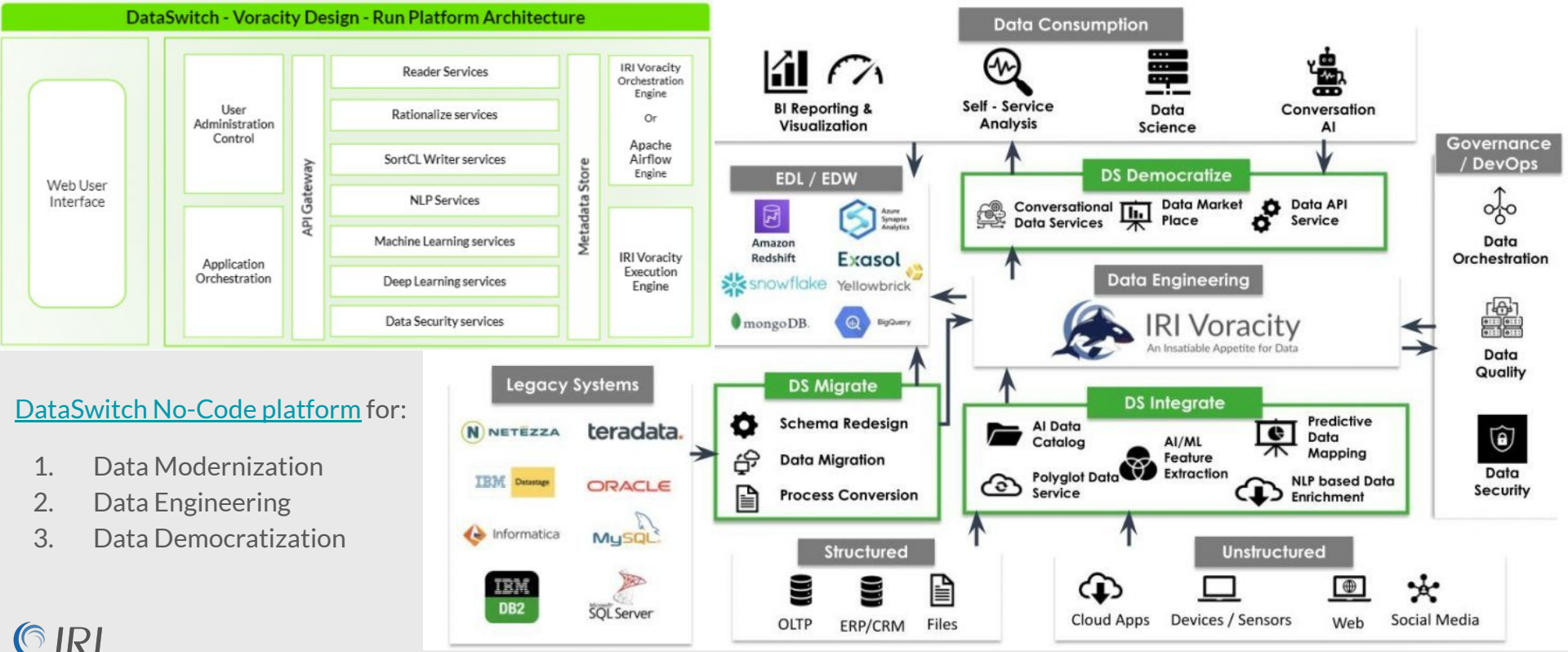
6) Make web service (e.g. node.js) or 3GL calls to Voracity scripts or the sortcl_routine() API

7) Invoke as SQL or COBOL system actions, or as CI/CD command tasks from GitLab, Azure DevOps, Amazon CodePipeline or Jenkinsfile

8) Run tasks during DB cloning in Actifio, Commvault, or Windocks, or from test data management hubs like Value Labs or Cigniti



Optional DataSwitch No-Code Web Platform Option for Voracity Job Design

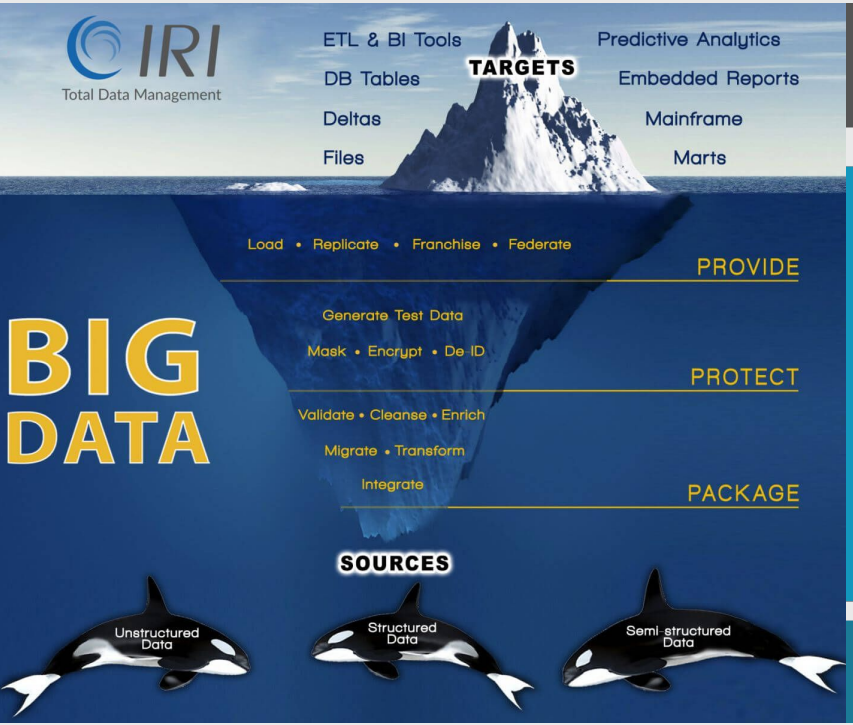


[DataSwitch No-Code platform](#) for:

1. Data Modernization
2. Data Engineering
3. Data Democratization

IRI
Total Data Management

Voracity's Big Data Functions & Advantages



Volume | Variety | Velocity | Veracity | Value

All-in-One Functionality

Unlimited Users

DISCOVER | INTEGRATE | MIGRATE | GOVERN | ANALYZE | CURATE

Using Voracity for Data:

Discovery Integration Migration Governance Analytics

Data Classification

Dark Data Discovery

DB & File Profiling

ER Diagramming

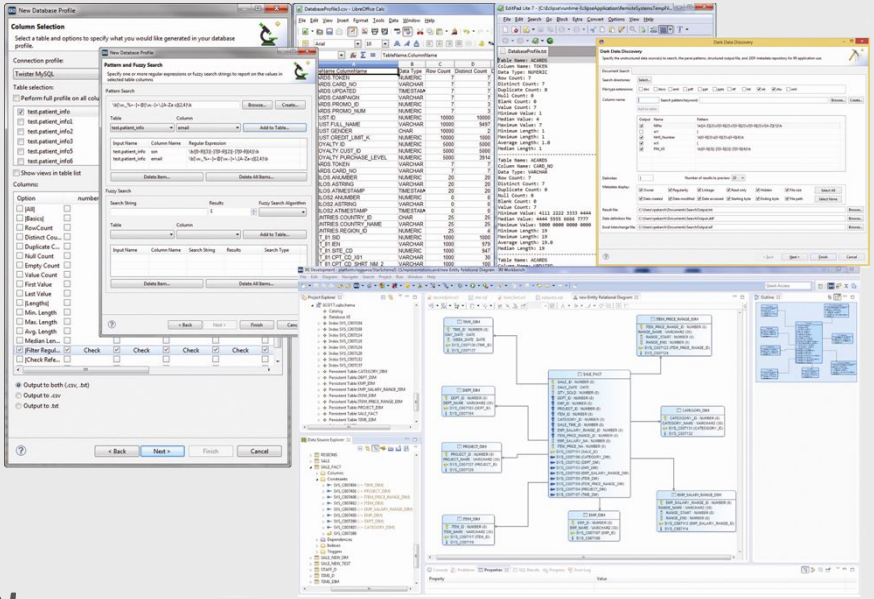
Metadata Definition

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.

IRI
Total Data Management

Using Voracity for Data:

Discovery Integration Migration Governance Analytics

Slowly-Changing Dimensions

Public/Private Mashups

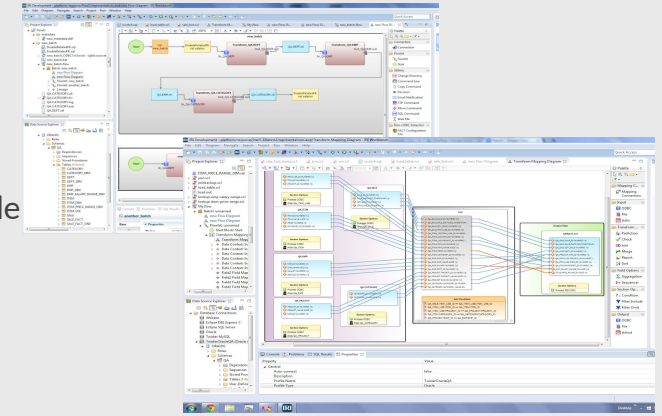
Change Data Capture

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 de 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!)

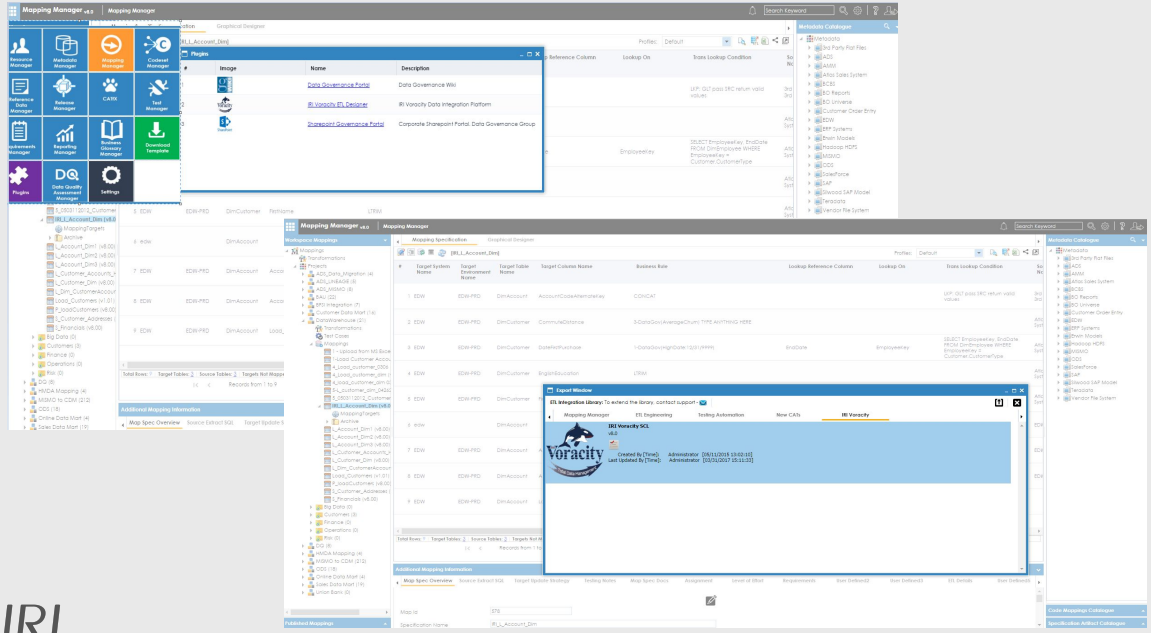
Replace Other ETL Tools

Replatform to save big money in a few weeks. Voracity is supported by erwin and DataSwitch so you can automate the conversion of legacy tool mappings to faster, more affordable Voracity ETL.

IRI
Total Data Management

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:

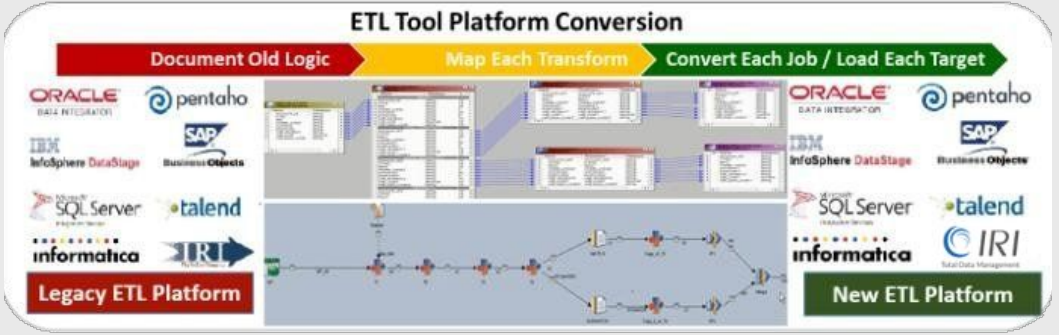
- Data Integration Teams
- Business Users
- Regulatory & Compliance Officers
- Governance & Information Architects

IRI
Total Data Management

How & Why You'd Leave Your Legacy ETL Tool

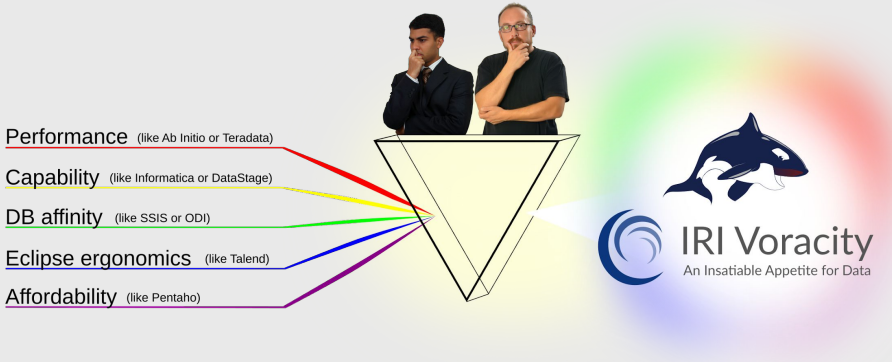
How

Through Erwin or DataSwitch, legacy ETL tool and SQL users can **convert** most existing mappings to Voracity workflows automatically.



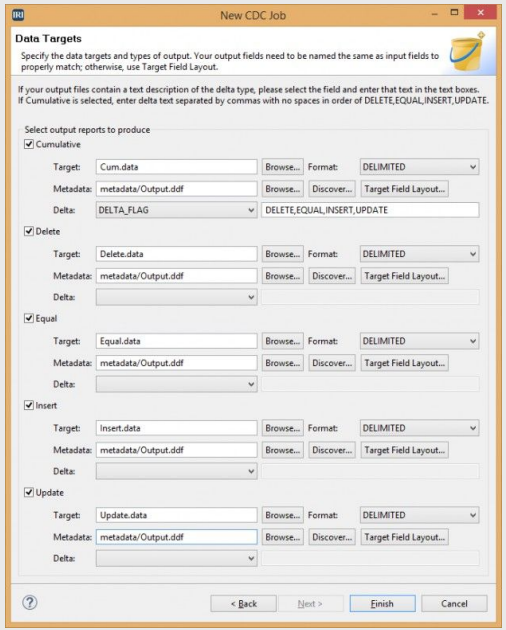
Why

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

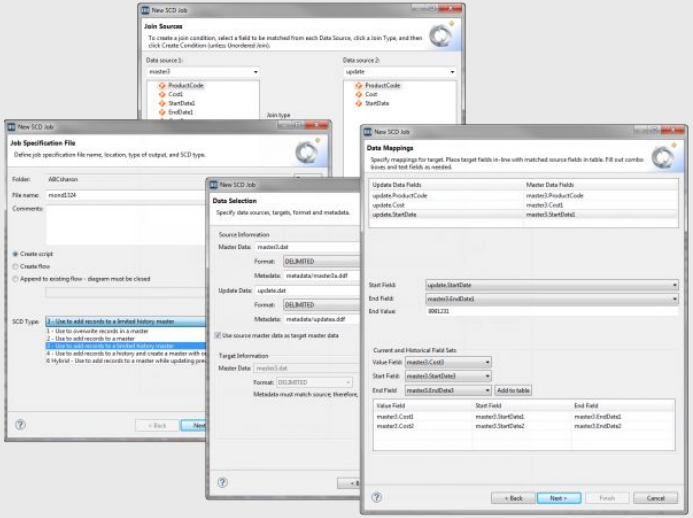


IRI
Total Data Management

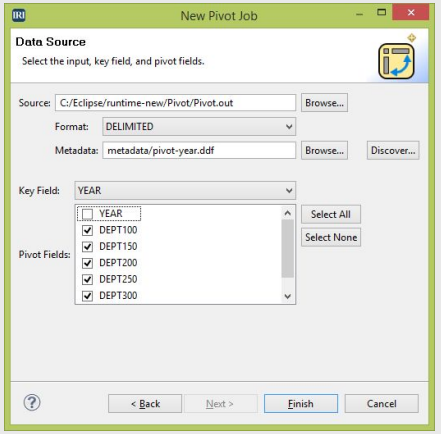
More Voracity Purpose-Built Wizards for...



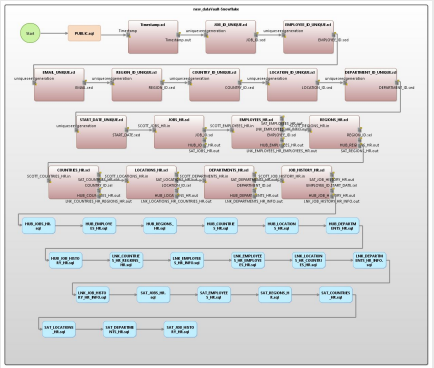
Change Data Capture



Slowly Changing Dimensions



Pivot/Unpivot

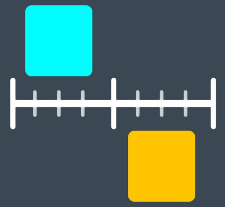


Data Vault Creation & Test

IRI
Total Data Management

Using Voracity for Data:

Discovery Integration Migration Governance Analytics

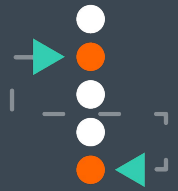


Incremental Replication

Data & File Types

Endianness

Databases

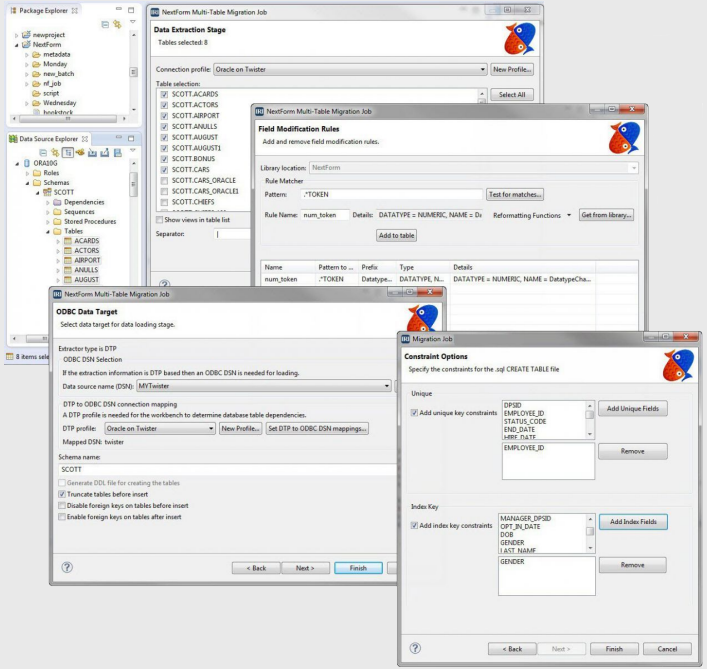


JCL Sorts

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

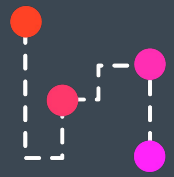
IRI
Total Data Management

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

Validate, cleanse, enrich, and unify data for better apps, ETL, and BI results.

Security

Find, classify, and rule-mask PII, or build test files/DBs and masked DB subsets.

Lineage

See forward and reverse views of data changes through time, and analyze impacts.

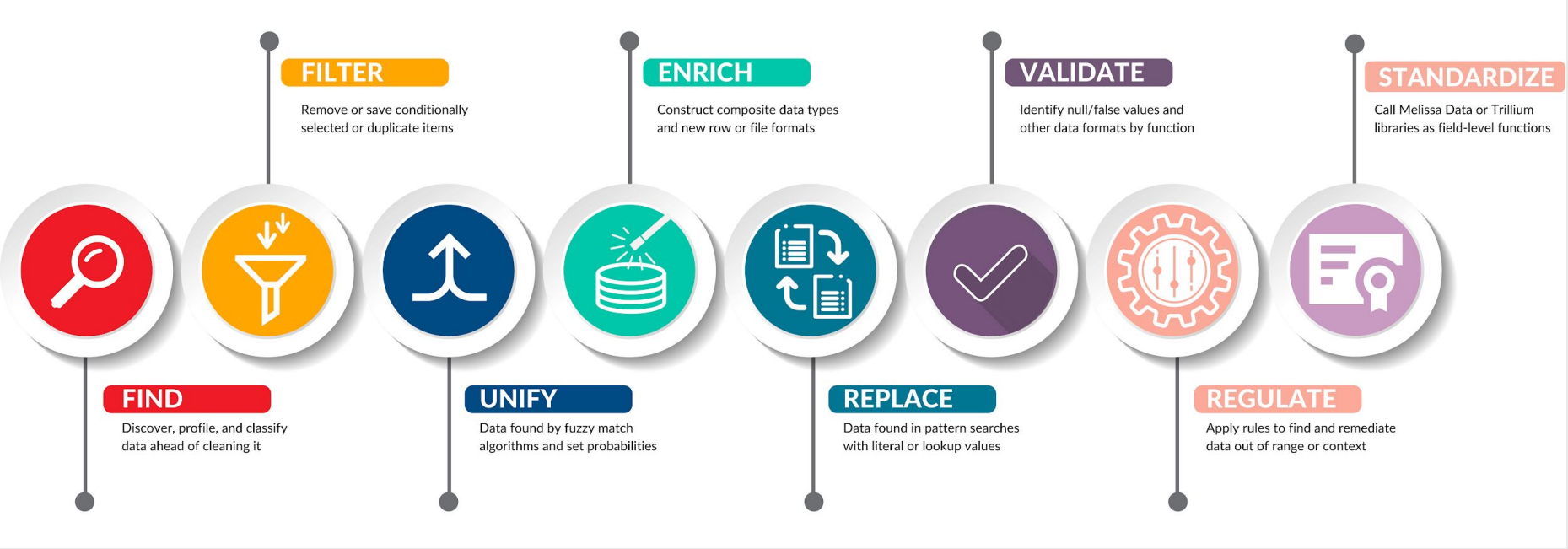
Assurance

Use query-ready audit logs and re-ID risk measurement to verify compliance.

IRI
Total Data Management

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.



IRI
Total Data Management

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 RDBs/NoSQL, MS and PDF docs, images, text, EDI files, etc.
- **Comply** with GDPR / CCPA / KVKK / PIPEDA, POPI, HIPAA, PCI-DSS and more
- **Separate** or **combine** searching and masking operations, and run in your infrastructure
- Mask **static** or **streaming** data to/from Amazon S3, FTP, HDFS, Kafka, MQTT, etc.
- Select from **15 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, proxy-based DDM, 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 audit files, and manage user identities through **RBAC**s

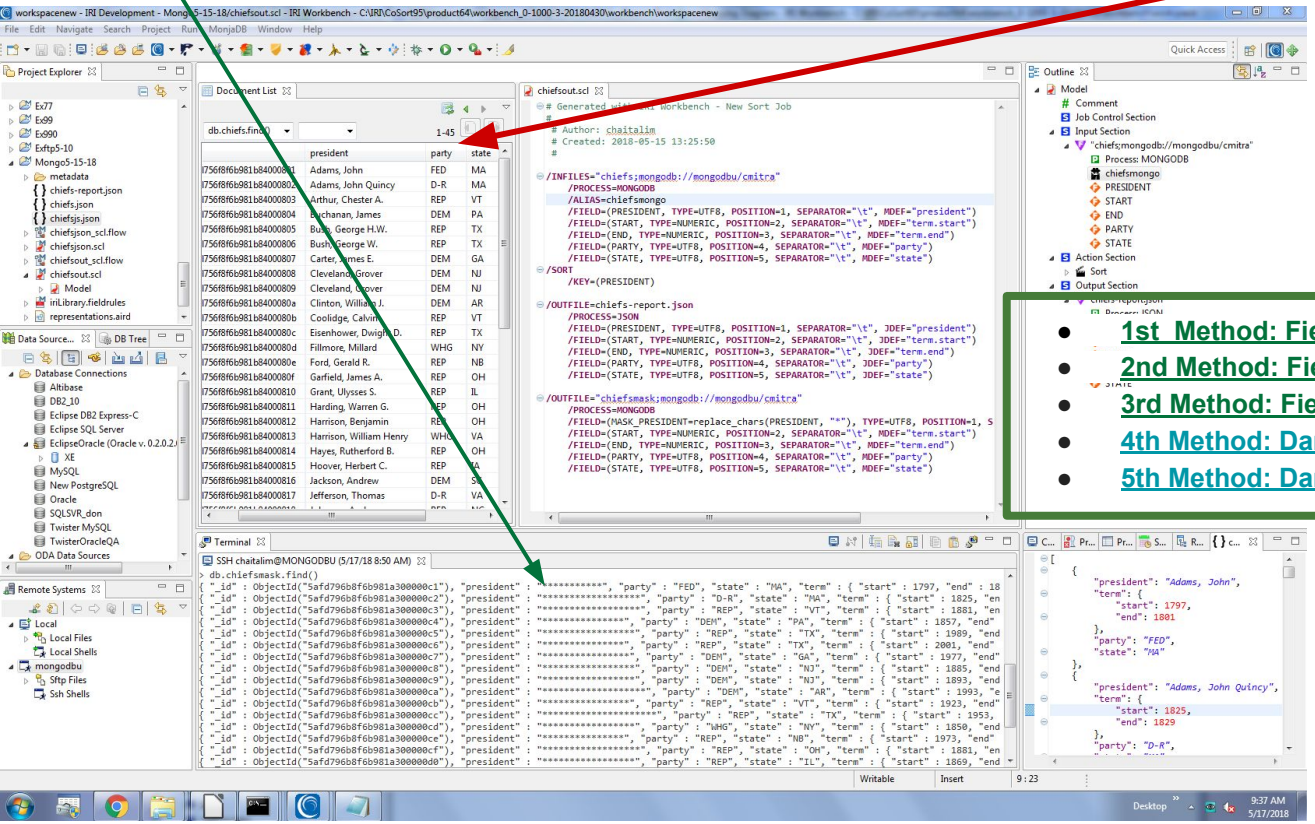
IRI
Total Data Management

IRI Data Masking Customers

Both earlier and current sites need to find and de-identify PII & PHI in RDBs, flat files and Excel sheets on premise, or in the cloud. The more recent engagements *also* involve NoSQL DBs, and semi- and unstructured data in documents, images, and EDI and log files like JSON, HL7, X12 and XML. Streaming and Hadoop data sources, plus faces, and cloud storage silos may also require masking.

MongoDB masked

unmasked



- 1st Method: FieldShield w/CSV export & import
- 2nd Method: FieldShield w/CData O/JDBC drivers
- 3rd Method: FieldShield w/BSON driver (for Mongo)
- 4th Method: DarkShield GUI
- 5th Method: DarkShield API

DarkShield also supports:

- Cassandra and Elasticsearch
- Solr, Redis, Couchbase
- BigTable, CosmosDB, DynamoDB

IRI
Total Data Management

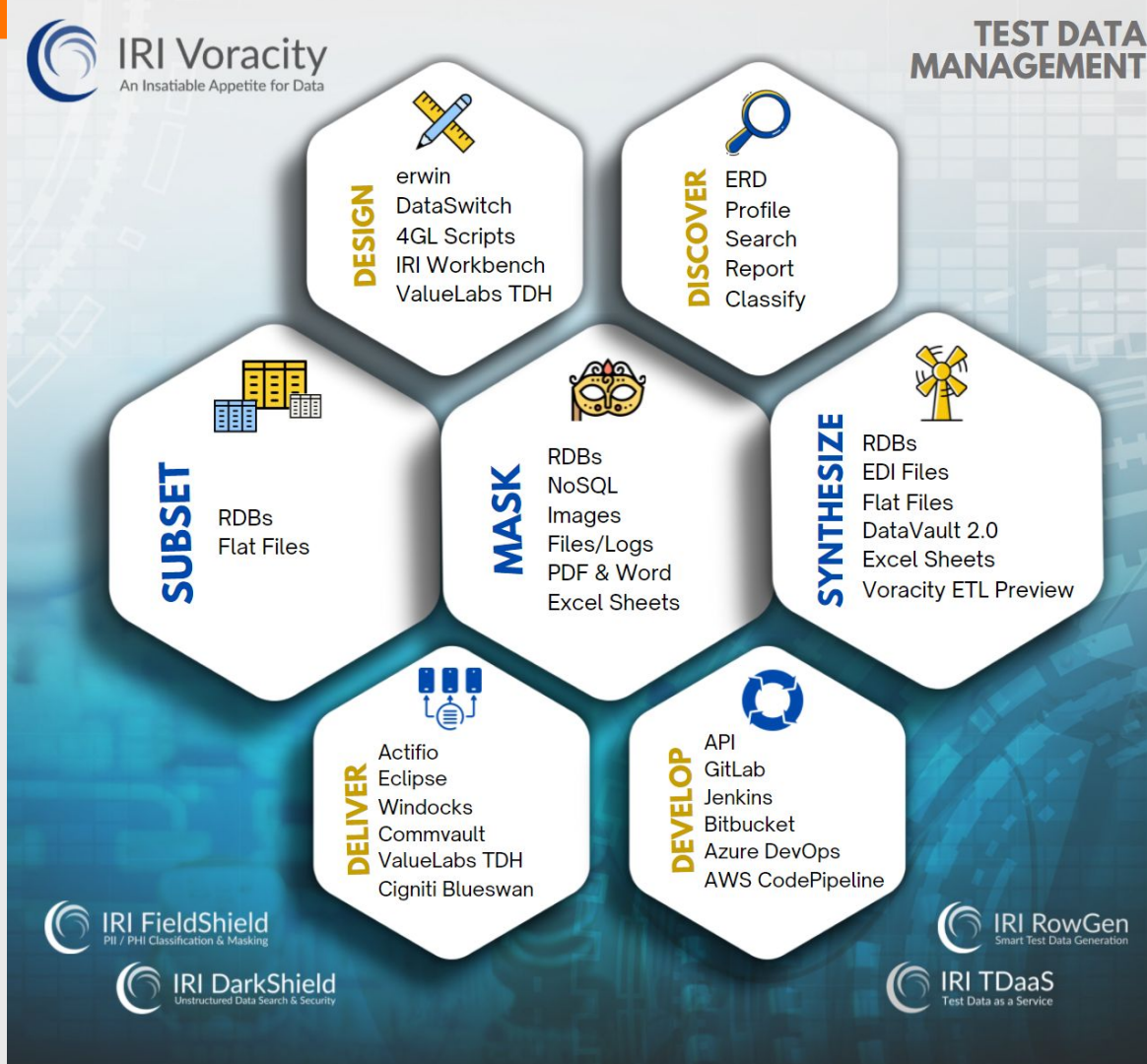
TDaaS & TDM Options

1. Test Data as a Service (TDaaS), a remotely provided professional engagement leveraging RowGen or any of the masking and subsetting features described above to provide highly customized test data without licensing or learning new technology.

2. Run in CI/CD workloads like Azure DevOps, AWS CodePipeline, GitLab, Jenkins, etc.

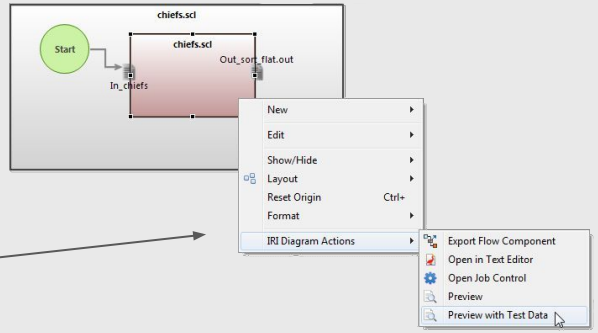
3. Three DB virtualization tools call IRI scripts:
   a. Actifio
   b. Commvault
   c. Windocks

4. Two on-demand Test Data Management (TDM) portals are tightly integrated with IRI:
   a. Cigniti BlueSwan
   b. ValueLabs Test Data Hub (TDH)



Test Data via Synthesis & Subsetting w/Masking, too

- 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
- Cloud-share and manage **test data on-demand** in ValueLabs Imagine TDH
- Preview Voracity ETL jobs with immediate test data



IRI
Total Data Management

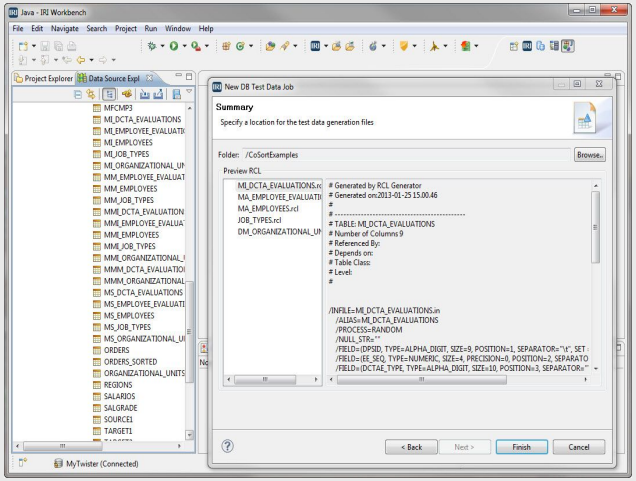
Versatile, Realistic Test Data *from* ***scratch***, ***or*** *masked* ***subsets***

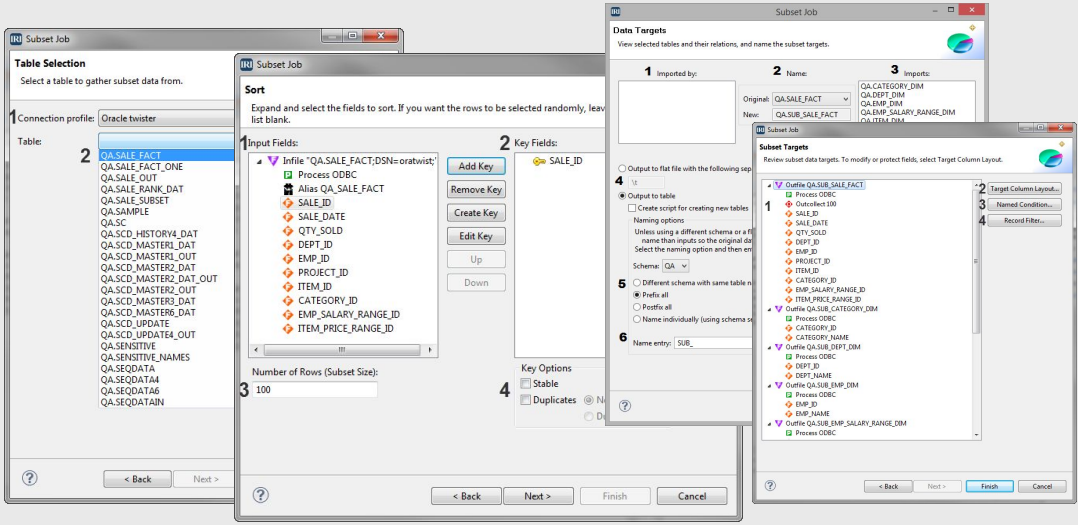
Target Formats

- Files & Reports
- Mainframe
- RDBs
- Cloud/SaaS Apps

Target Uses

- ETL Ops/Tools
- Software Dev
- Benchmarking
- Demos & Outsourcing



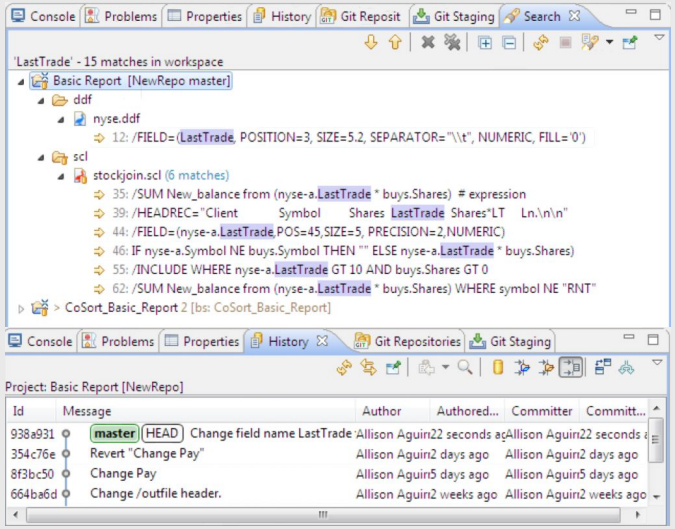


In addition to data masking, Voracity also includes robust test data generation/population and DB subsetting wizards to facilitate DB, ETL, and BI prototyping. Either way, the test data is realistic, referentially-correct, and privacy-law compliant. And thanks to IRI RowGen within, Voracity users can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

IRI
Total Data Management

Data Lineage & Impact Analysis

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



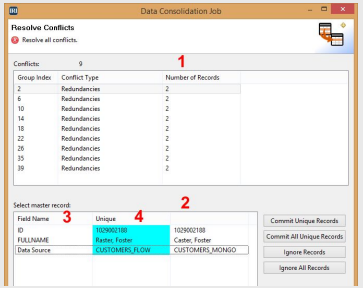
IRI is also working on an internal, encrypted IAM and granular logging system for reports on specific data value changes. See also CDC.

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

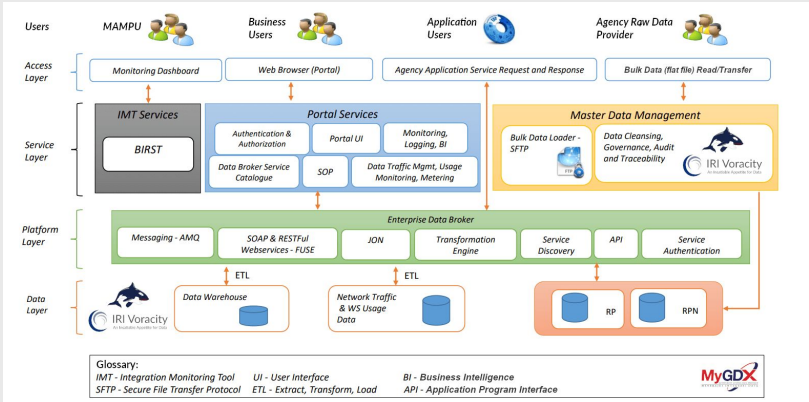
Data Reconciliation / MDM

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.]



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

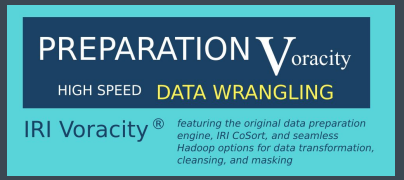
IRI
Total Data Management

Using Voracity for Data:

Discovery Integration Migration Governance Analytics

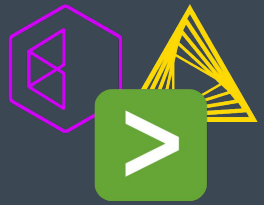
Embedded BI

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, KNIME, 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

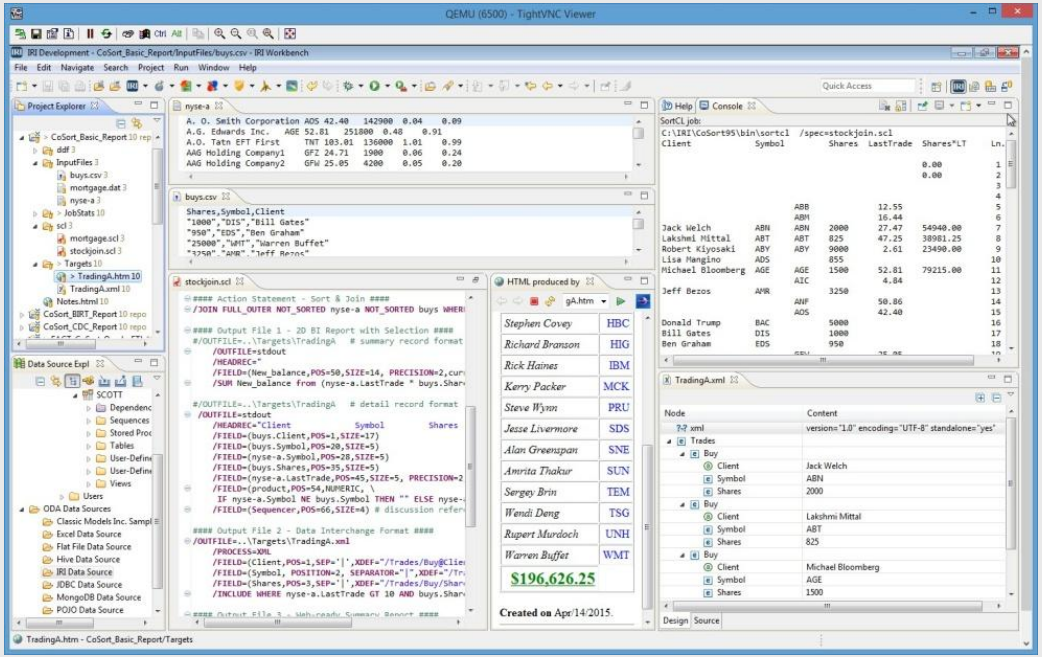
Voracity subscriptions are priced lower than data preparation tools. BIRT in IRI Workbench is free.

IRI
Total Data Management

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

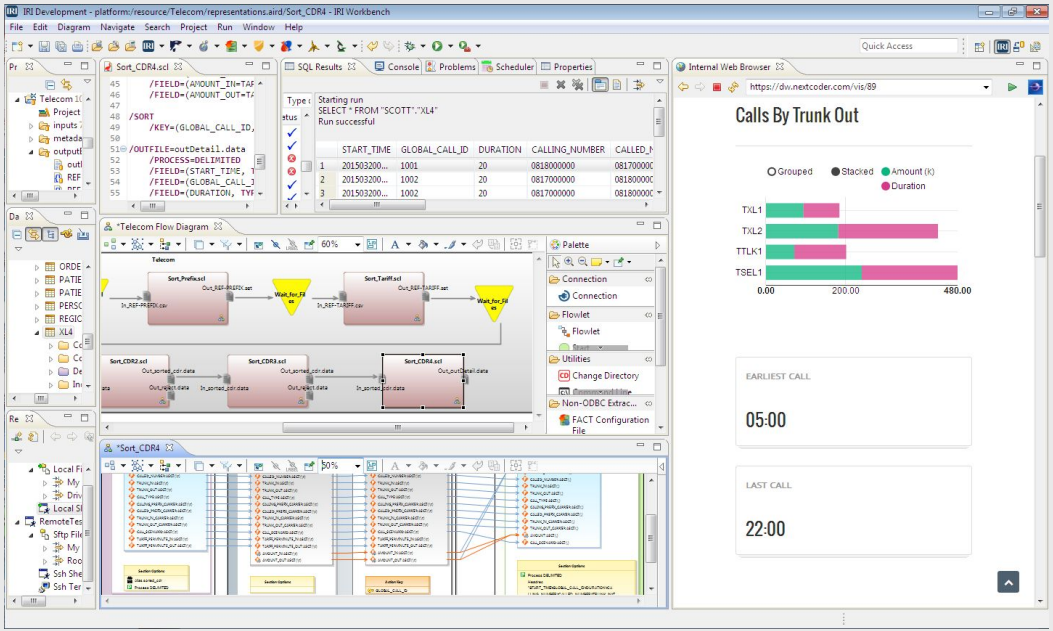
Voracity Analytic Option 1: Embedded BI

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



Voracity Analytic Option 2: **Cloud** Dashboards

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

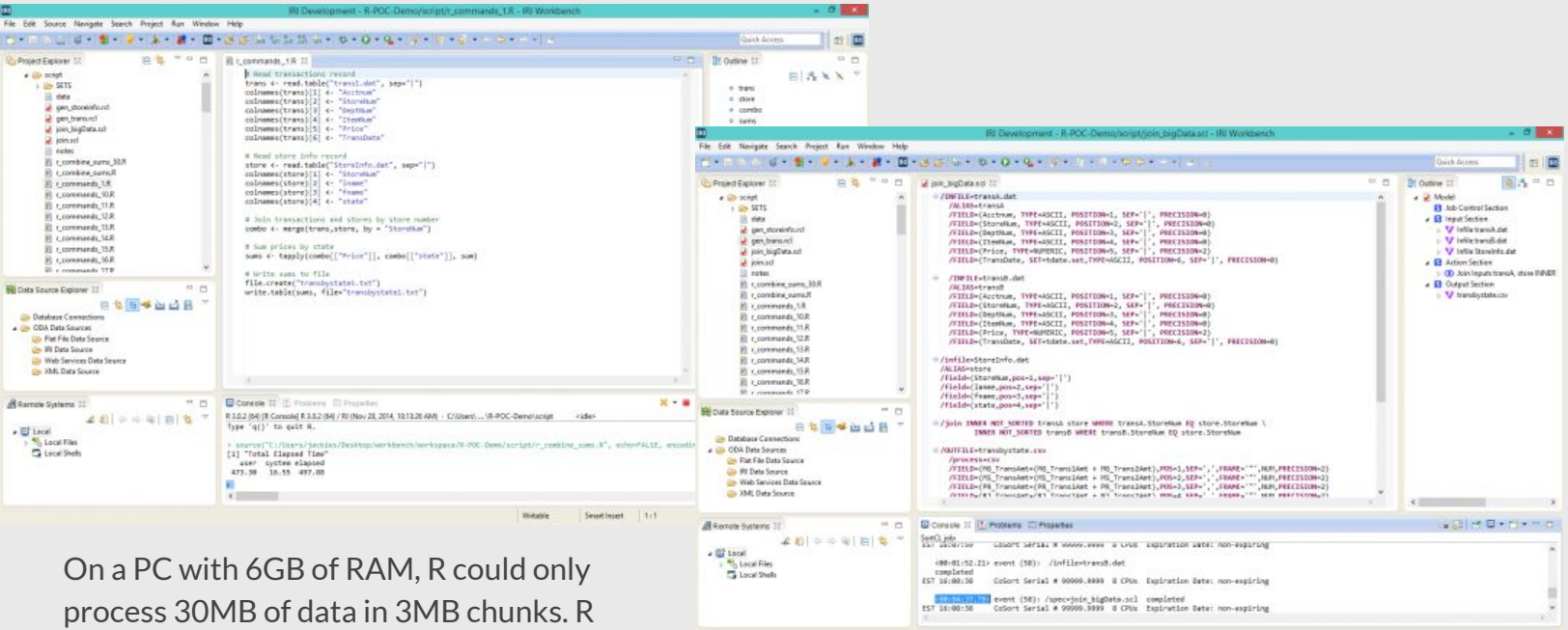


Voracity Analytic Option 3: **Data Wrangling**

Rapidly blend (prepare) data into CSV, XML, or table subsets for any BI/analytic tool. This process can speed time-to-display 2-20X, and improve data quality, privacy, and storage space

IRI
Total Data Management

Option 3 Example: Data Wrangling for R



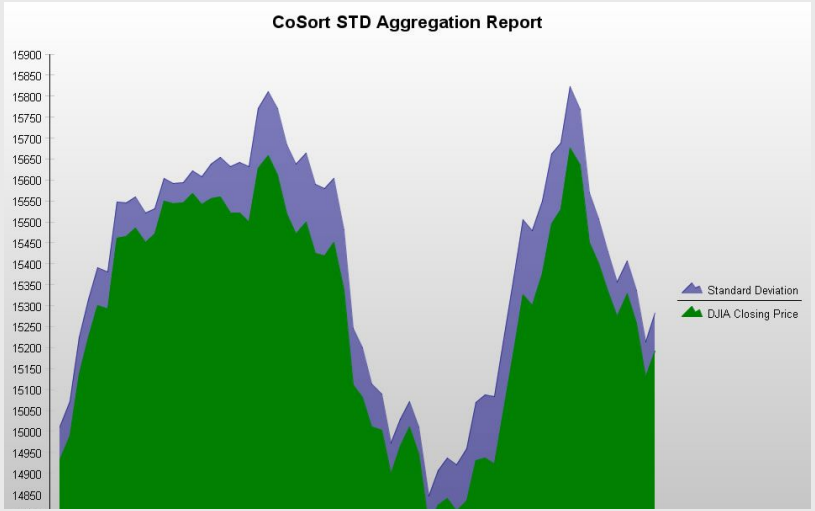
On a PC with 6GB of RAM, R could only process 30MB of data in 3MB chunks. R needed 11 jobs or nodes to break down the data and merge the results …

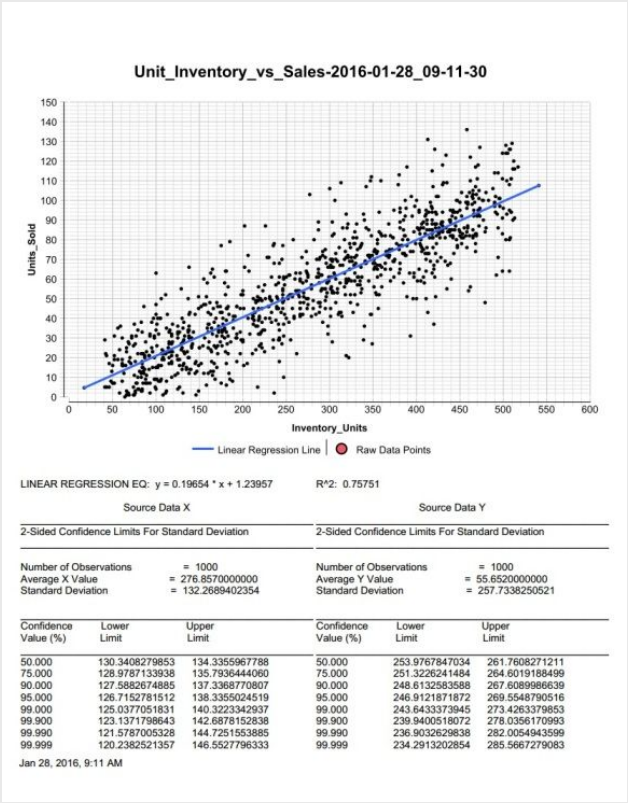
… The same data prep in Voracity happens in *just one* sort-join-aggregate program (and I/O pass) that runs 45% faster than R in this small case.

IRI
Total Data Management

Voracity Analytic Option 4: **Predictive**

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.



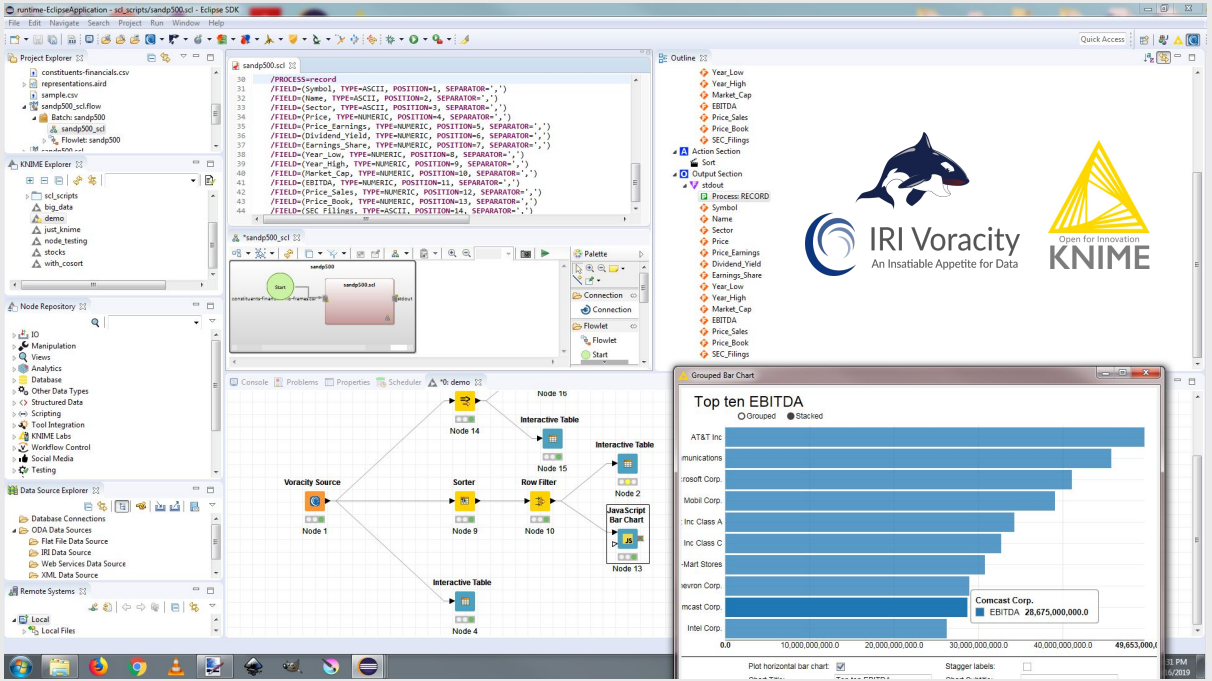


IRI
Total Data Management

Voracity Analytic Option 5: **KNIME** Data Mining / Science

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 ...

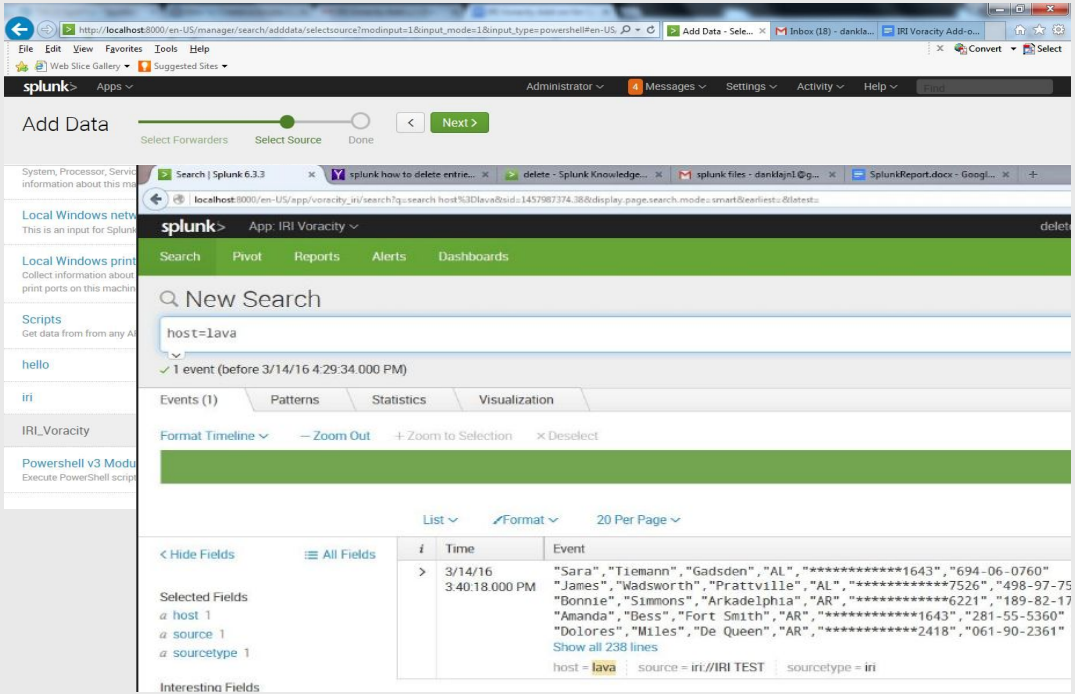


Voracity Analytic Option 6: **Splunk** Enterprise & Security

Prepare and index data for Splunk *simultaneously*.

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

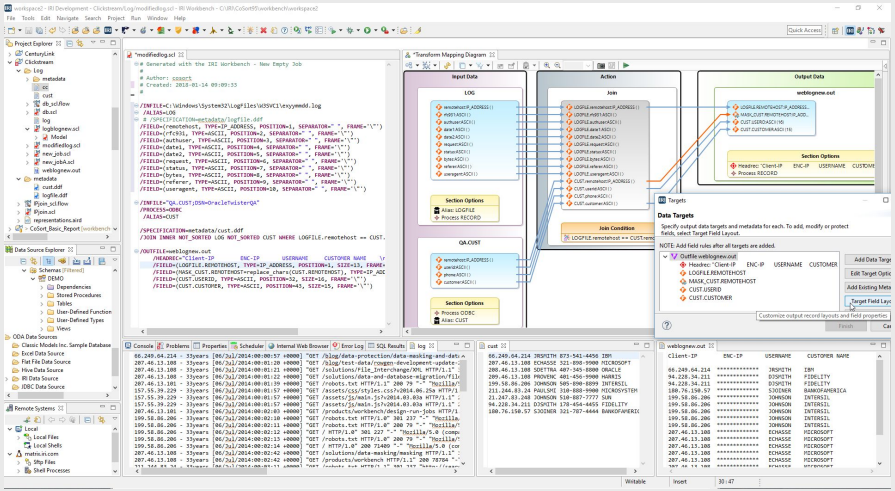
Voarcity also supports operations through the Splunk Universal Forwarder and Splunk Phantom Playbooks.



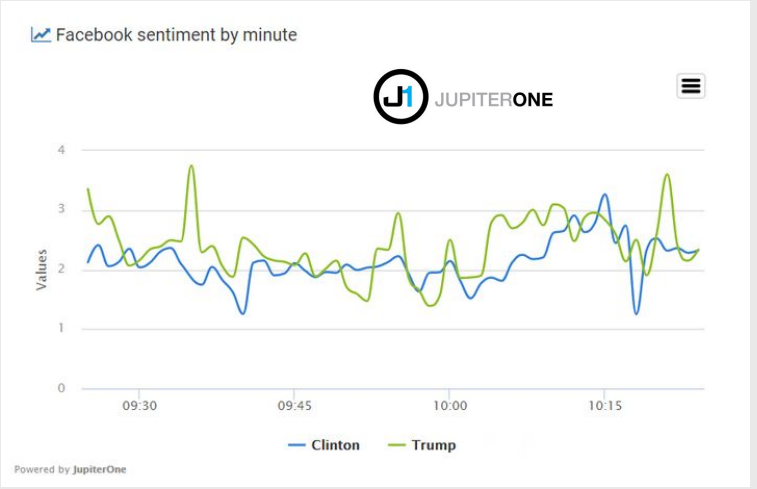
IRI
Total Data Management

Voracity Analytic Option 7: **Clickstream**

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



IRI
Total Data Management

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.

IRI
Total Data Management

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. $45K and up for unlimited users per year.

IRI
Total Data Management

IRI
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 (but also anonymize where needed) claim data outside normal parameters to identify suspicious behavior, and feed it to visualization and notification apps in the same IDE (and still protect applicable PII).

IRI
Total Data Management

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 reroute 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.

Telco & Media

Monetize Calls & Clicks

Use Voracity to natively process ASN.1-compatible Call Detail Records (CDRs) and clickstream data for billing and analytics applications, and to repackage and 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, CPG & e-commerce

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
Total Data Management

IRI Voracity

An Insatiable Appetite for Data

iri.com **blog.iri.com**

IRI Voracity Data Management Group on LinkedIn

IRI
Total Data Management