**A Rule-Based, Automated Data Masking Solution for IBM AFP Files via API Call**

# IRI, The CoSort Company

<u>Vendor Background</u>

- Specializing in fast data management and data-centric security

- Privately owned and profitable since 1978

- Sales and support in more than 40 cities worldwide

- Organically grown, shared metadata and Eclipse IP stack

- Featured in: CIO Review (top GRC and Compliance vendors); DBTA; Gartner Market Guide to Data Masking Tools; and in the QY, Markets & Markets, and Research & Markets *forecast reports* on Data Masking, DB Security, Data Classification, Data Governance

# IRI Data Masking Reference Customers & Partners

Most IRI data masking customers profile and protect PII in RDBs, flat files and Excel sheets on premise, or in the cloud. Recent engagements also involve NoSQL DBs, documents, images, and EDI and log files. Streaming and Hadoop data sources, plus faces, are also supported. Entities include:

# Selected Recognitions of IRI in the Data Governance and Security Industry

- BARC Technology Map for Data Governance 2021 (IRI Voracity)
- Bloor Research Data Searching and Masking InBrief 2022
- CIO Applications: Top 25 GRC Technology Providers
- CIO Review: 20 Most Promising Compliance Technologies
- Computerworld Germany (IRI FieldShield)
- DBTA: Trendsetting Products (IRI DarkShield)
- Forrester Research Now Tech: Data Masking Q1'2021
- Gartner Market Guide for Data Masking (2017-2022)
- Global Data Governance Market Size & Analysis through 2026
- Global Data Masking Market Size & Forecast verified 2020
- Insight Partners - Data Classification Market to 2027
- Insight Partners - Test Data Management Market Outlook to 2027
- Markets & Markets - Data Governance Market to 2024 - Visionary Leader
- Markets & Markets - Data Masking Market Forecast to 2022
- Outlook Series: The Case for Data Masking
- QY Research - Global Test Data Management Market to 2025
- Privacy & Data Security Law Journal
- Research & Markets - DB Security Market to 2022

IRI
Total Data Management

# IRI Data Manager Suite

## IRI CoSort
Sort, Transform & Report

*Speed or replace legacy sorts, and batch/ETL/SQL transforms*
- Filter, sort, 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

# IRI Voracity
## An Insatiable Appetite for Data
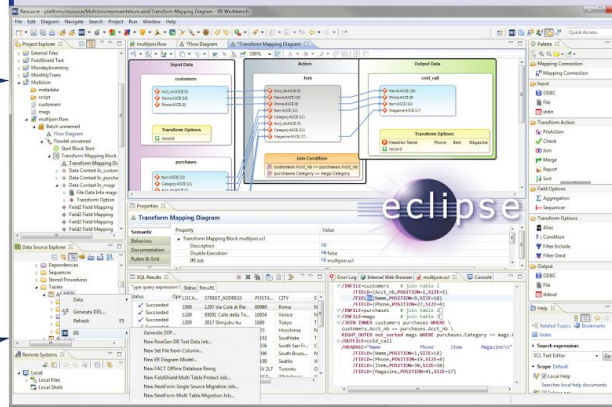
*Consolidate tools and tasks to process, protect, prototype, present*
- Discover, define, and manage data in legacy and new sources
- Combine data integration, migration, governance, and analytics
- Exploit CoSort or Hadoop engines for optimum throughput
- Leverage Eclipse familiarity, functionality, and extensibility

## IRI
Total Data Management

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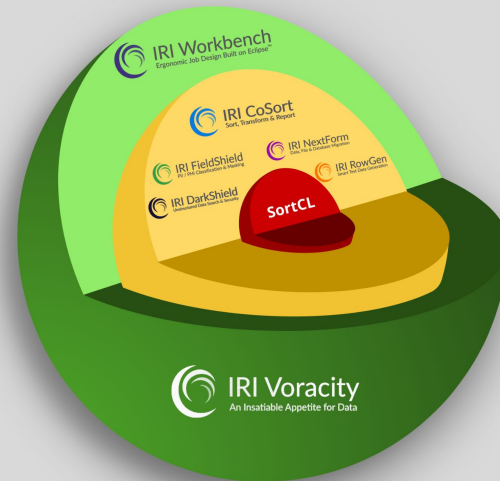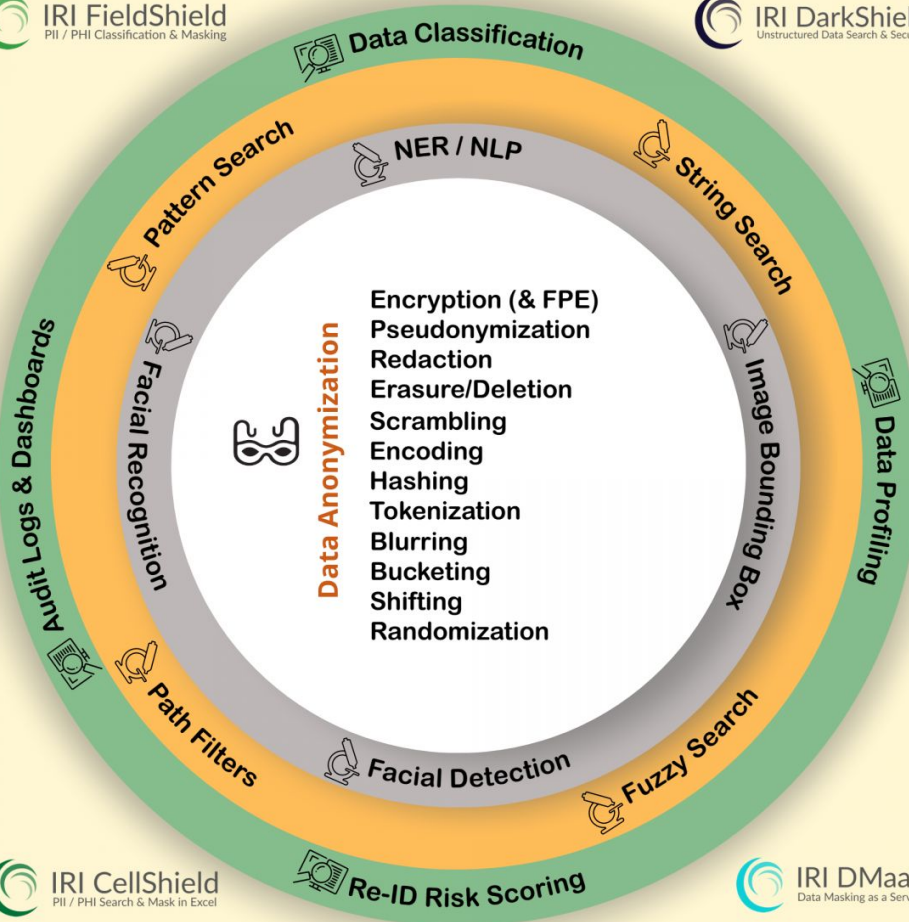- Comply with the right to erasure, portability, or rectification

## 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

**IRI DarkShield**
Unstructured Data Search & Security

**IRI FieldShield**
PII / PHI Classification & Masking

**IRI CellShield**
PII / PHI Search & Mask in Excel

**IRI RowGen**
Smart Test Data Generation

---

**IRI Voracity**
An Insatiable Appetite for Data

**DESIGN**
erwin
DataSwitch
4GL Scripts
IRI Workbench
ValueLabs TDH

**DISCOVER**
ERD
Profile
Search
Report
Classify

**SUBSET**
RDBs
Flat Files

**MASK**
RDBs
NoSQL
Images
Files/Logs
PDF & Word
Excel Sheets

**SYNTHESIZE**
RDBs
EDI Files
Flat Files
DataVault 2.0
Excel Sheets
Voracity ETL Preview

**DELIVER**
Actifio
Eclipse
Windocks
Commvault
ValueLabs TDH
Cigniti Blueswan

**DEVELOP**
API
GitLab
Jenkins
Bitbucket
Azure DevOps
AWS CodePipeline

**IRI FieldShield**
PII / PHI Classification & Masking

**IRI DarkShield**
Unstructured Data Search & Security

**IRI RowGen**
Smart Test Data Generation

**IRI TDaaS**
Test Data as a Service

---

**IRI FieldShield**
PII / PHI Classification & Masking

**IRI DarkShield**
Unstructured Data Search & Security

**Data Classification**
Pattern Search
NER / NLP
String Search
Facial Recognition
Image Bounding Box
Audit Logs & Dashboards
Data Profiling
Path Filters
Facial Detection
Fuzzy Search
Re-ID Risk Scoring

**Data Anonymization**
Encryption (& FPE)
Pseudonymization
Redaction
Erasure/Deletion
Scrambling
Encoding
Hashing
Tokenization
Blurring
Bucketing
Shifting
Randomization

**IRI CellShield**
PII / PHI Search & Mask in Excel

**IRI DMaaS**
Data Masking as a Service

---

**IRI Workbench**
Ergonomic Job Design Built on Eclipse

**IRI CoSort**
Sort, Transform & Report

**IRI FieldShield**
PII / PHI Classification & Masking

**IRI NextForm**
Data, File & Database Migration

**IRI RowGen**
Smart Test Data Generation

**IRI DarkShield**
Unstructured Data Search & Security

**SortCL**

**IRI Voracity**
An Insatiable Appetite for Data

---

**IRI**

Data Protector Suite

# IRI Data Masking Tool Architectures

## IRI FieldShield
PII / PHI Classification & Masking

**Architecture**

### Job Designers

| IRI Workbench Built on Eclipse Any Text Editor | DataSwitch, erwin ValueLabs TDH Commvault, Windocks | IRI "Sandkey" API IRI "Gulfstream" API |
|---|---|---|

### Runtime Engines

| VGrid Gateway to Hadoop MR2, Spark Storm, Tez | Other Programs (SQL, Java, C, PowerShell, etc.) | IRI SortCL |
|---|---|---|

**IRI SortCL**
CoSort — DarkShield — Nextform — RowGen
**FieldShield**

### Sources & Targets
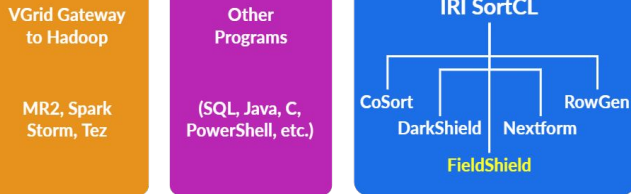
- HDFS
- RDBMS / NoSQL
- File Systems
- Pipes
- Cloud Apps
- BI & Analytic Tools
- Streams
- APIs
- Brokers
- Custom Reports

---
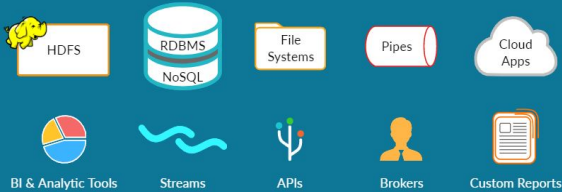
Both structured and unstructured data discovery functions -- including classification, search, and metadata creation -- are performed in IRI Workbench data discovery wizards.

Static data masking (SDM) jobs are usually built in IRI Workbench, while user-specific dynamic data masking (DDM) is available in multiple options (see matrix on slide #27).

Voracity data manipulation and masking jobs use the IRI CoSort (SortCL) engine on commodity LUW hardware, on premise or in the cloud. No database or cloud API is needed. This reduces runtime overhead, administrative complexity, and risk. The executable is also metadata-compatible with, and masks within data integration, cleansing, and reporting jobs, too.

Finally, no server framework is required to orchestrate jobs. This fosters more granular allocation and tuning of resources, and the sharing of metadata artifacts.

---

## IRI DarkShield
Unstructured Data Search & Security

**Architecture**

### Front-Ends

| IRI Workbench Built on Eclipse | JRE CLI For Batch Jobs, CI/CD (DevOps) | RPC API For Apps & Web Services |
|---|---|---|

### Back-End

| Search Matchers | IRI SortCL |
|---|---|

**Search Matchers**
Data Class — Discovery Methods
Data Rule — Masking Functions

**IRI SortCL**
CoSort — FieldShield — Nextform — RowGen
**DarkShield**

### Sources & Targets

- HDFS
- RDBMS / NoSQL
- Text Files
- Pipes
- Cloud Storage
- Images
- Streams
- APIs
- Web Apps
- Documents

---

IRI
Total Data Management
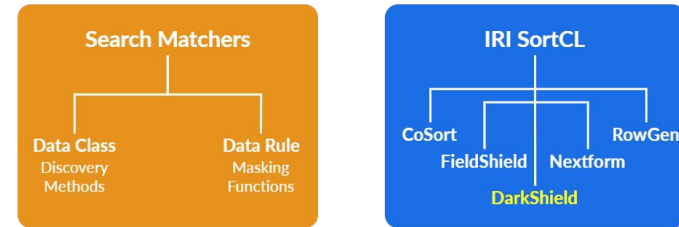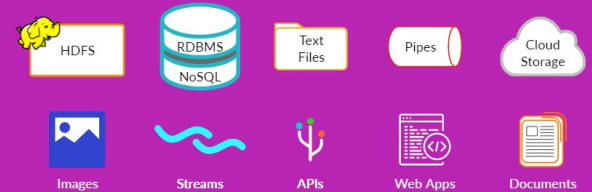
# IRI Voracity Communication & Networking Architecture



**IRI Voracity**
An Insatiable Appetite for Data

## On-Premise / LAN

Optional sortcl (Voracity)

IRI Workbench Client

Optional sortcl (Voracity)

IRI Workbench Client

Optional sortcl (Voracity)

IRI Workbench Client

Optional sortcl (Voracity)

IRI Workbench Client

File Server

## Cloud Networks

Optional sortcl (Voracity)

IRI Workbench Client

Web service

## Sample Sources & Targets

DB2 LUW — Port 50000 tcp

DB2 i & z — Port 446 tcp

MS SQL — Port 1433 tcp

Postgre SQL — Port 5432 tcp

Oracle — Port 1521 tcp

MySQL — Port 3306 tcp

HIVE — Port 10000 tcp

SAP HANA — Port 50113 tcp

MongoDB — Port 27017 tcp

Cassandra — Port 7000/1 tcp

Mainframe Files via NFS, FTP

Amazon S3

Other RDBs

Azure MS SQL
Amazon Aurora/Redshit
Elasticsearch
Oracle DB Cloud
Salesforce
Snowflake
Teradata Vantage

## On-Premise Web Server
e.g., DarkShield ("Plankton")
RPC API Services

HTTP/S via Node.js, etc.

Standalone Engine Option, Remote (SSH) Access

sortcl (Voracity)

Optional: IRI Workbench Cohosted / Container

## Outside Networks

Supplier Portal

Staff VPN

IoT Edge

Private Cloud

## Third-Party Integrations

Encryption Key Managers:
Azure Key Vault,
Townsend AKM

A/D & LDAP Integration

DB Cloning & Test Data Hubs:
Actifio, Commvault, Windocks
Cigniti BlueSwan, ValueLabs TDH

BI & Analytic Targets:
BOBJ, Cognos, Cubeware
Datadog, iViz, KNIME,
Microstrategy, Oracle DV/D,
PowerBI, QlikView, R, Sisense,
Splunk, Spotfire, Tableau

## 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.

IRI
Total Data Management

# 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 |
| AFP Files (via TST) | 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 (e.g., 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/BLOB 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

# IRI DarkShield
## Unstructured Data Search & Security

*IRI Data Protector Suite*

# What DarkShield Does
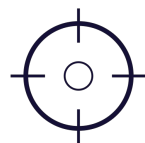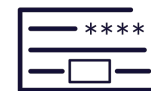
Search    Extract    Redact    Audit

- Simultaneously scans, extracts, and de-IDs or deletes PII (and audits actions) in all supported file formats

- Handles PII floating in semi/unstructured text, MS/PDF documents, BLOB columns, and NoSQL collections

- Finds faces or defined data classes tied to RegEx patterns, lookup sets, NER models, and/or image regions

- Builds, saves, and re-uses semi-supervised, machine learning models in project or cloud repositories

- Blacks-out PII in images, blurs faces, and applies of encryption (including FPE), pseudonymization, hashing, encoding, bit scrambling, redaction, or erasure functions for PII in text files and documents

- Writes masked files atop originals, or to different folders with the same file names and formats

- Shows search, remediation, and model training job status via real-time progress bars

- Shares search methods and masking functions with IRI FieldShield and IRI CellShield EE

- Generates logs of all values found or masked, along with IRI-compatible metadata for BI, queries, etc.

- Creates graphical, interactive displays of search and mask results, or hand-offs log files to Splunk, Datadog

- Runs in IRI Workbench with other IRI and Eclipse tools, from the command line, or via REST API

IRI
Total Data Management

# Image File Redaction or Value Replacement …

**Search**   **Extract**   **Redact**   **Audit**

… or Test Value (RowGen) Synthesis into Images or Documents …

AFP, BMP, DICOM GIF, JPx, PNG, and TIFF, alone or in docs like PDFs and Word!

IRI
Total Data Management

# Documented RPC APIs

Search    Extract    Redact    Audit

## Swagger.
Supported by SMARTBEAR

# DarkShield API  1.3.0  OAS3

./docs/openapi/darkshield.yaml

An RPC service for searching and masking unstructured text.

IRI, Inc. - Website
Send email to IRI, Inc.

## search

**POST**  /api/darkshield/searchContext.create

**POST**  /api/darkshield/searchContext.destroy

**POST**  /api/darkshield/searchContext.search

**POST**  /api/darkshield/searchContext.mask

## mask

**POST**  /api/darkshield/maskContext.create

**POST**  /api/darkshield/maskContext.destroy

**POST**  /api/darkshield/maskContext.mask

**POST**  /api/darkshield/searchContext.mask

---

Host a context which can be called whenever a search operation needs to be performed. The Context contains a list of search matchers that will be used to match on the text. Multiple contexts with the same name cannot be created.

**Parameters**    Cancel

No parameters

**Request body** required    application/json

The Search Context.

**Examples:**    Create a Search Context with a mix of matchers.

```
{
  "name": "MixedMatcherContext",
  "matchers": [
    {
      "name": "EmailMatcher",
      "type": "pattern",
      "pattern": "\\b[\\w._%+-]+@[\\w.-]+\\.[\\w]{2,4}\\b"
    },
    {
      "name": "NameMatcher",
      "type": "ner",
      "modelUrl": "http://opennlp.sourceforge.net/models-1.5/en-ner-person.bin"
    }
  ]
}
```

```
curl -X POST "http://localhost:8080/api/darkshield/searchContext.mask" -H "accept: application/json" -H "Content-Type: application/json" -d
"{\"searchContextName\":\"MixedMatcherContext\",\"maskContextName\":\"MixedMaskingContext\",\"text\":\"Hello, my name is John Doe and my email address
is johndoe@gmail.com.\"}"
```

**Request URL**

http://localhost:8080/api/darkshield/searchContext.mask

**Server response**

| Code | Details |
|------|---------|
| 200 | **Response body** |

```
{
  "failedResults": [],
  "maskedText": "Hello, my name is Wazz Edk and my email address is BqJA0RzCAWdtqXb3tJNBGB/RgNo3y+QKd0MsCjZsgMM=.",
  "results": [
    {
      "end": 26,
      "maskedResult": "Wazz Edk",
      "ruleName": "FpeNameRule",
      "start": 18
    },
    {
      "end": 95,
      "maskedResult": "BqJA0RzCAWdtqXb3tJNBGB/RgNo3y+QKd0MsCjZsgMM=",
      "ruleName": "HashEmailRule",
      "start": 51
    }
  ],
  "unmatchedAnnotations": []
}
```

Download

IRI
Total Data Management

# AFP Data Masking

## Pre-DarkShield



## Post-DarkShield

# DarkShield Options & Advantages

## Search Methods

- Java RegEx Pattern Search
- Exact Lookup File (Table) Match
- Fuzzy Lookup File Matching
- Multilingual Named Entity Recognition (NER) using trainable models under:
  - Open NLP
  - PyTorch
  - TensorFlow
- Area bounding boxes (X-Y)
- File-Type Specific Search Matchers (JSON path, XML path, column header pattern, Excel sheet/cell matchers)

## Masking Functions

- Full or partial string redaction
- Non/reversible pseudonymization
- Format-preserving encryption
- Public/private key encryption
- Hashing
- Scrambling
- Synthetic value generation
- Date/integer blurring (random noise)
- Fixed (constant string) replacement
- Deletion

## Additional Features

- PII (data value) search & report
- Encryption key management
- Image pre-processing (can improve accuracy)
- NGINX load balancing
- NGINX user authentication
- Multiple audit log options
- FieldShield-compatible masking functions
- Simultaneous outputs to other formats
- Support for many other file formats
- Support for relational and NoSQL DBs

## Additional Considerations

- Format-agnostic data rules cut job config time
- API-driven approach supports multiple:
  - application design and integration modes
  - source/target file formats & systems
  - calling languages to ease support
  - OpenAPI (Swagger) ergonomics
  - GUI wizards in Eclipse WB (November)
- Runs on or external to source/target systems
- Proven components and vendors

IRI
Total Data Management

# DarkShield Embedding StreamEDS for AFP Masking

TARGET**STREAM**
TECHNOLOGIES
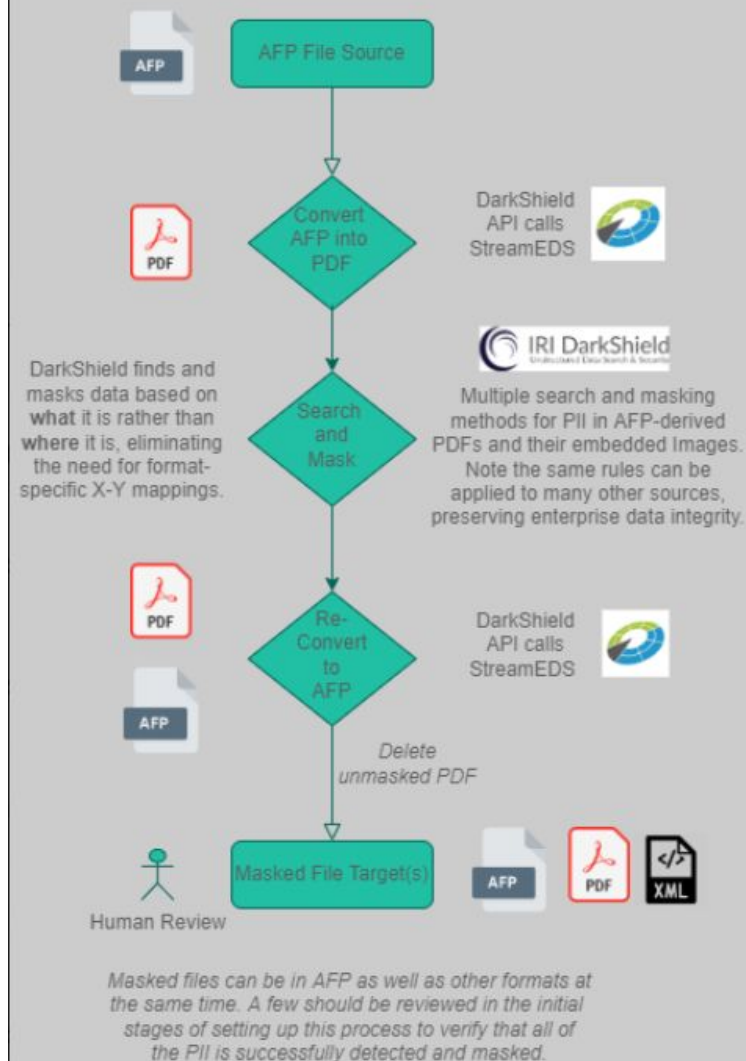


**Q&A:**

1) ***Does DarkShield mask a copy, or mask the original AFP file?***
   DarkShield can read and mask the original file or a copy of the file. It can also make a copy. The original AFP file can be retained. DarkShield can also overwrite the original with the masked result (but that's not advised).

2) ***Where does the DarkShield and TST software reside? Must they run on a separate server, thus requiring a copy and load to convert and mask the file, or is the software loaded onto the same system (and thus requiring more resources)?***
   Your call; it can run on the same or separate server, on-premise or in your cloud. Note that the closer DarkShield is to the AFP sources, intermediary PDFs, and target folders in I/O and network terms, the faster the masking jobs will run. As jobs are I/O bound, storage, disk, and bandwidth matter more than RAM and CPU. Sufficient storage is needed for source AFP, temporary PDF, and target AFP files. Intermediary files should be (and are by default) destroyed in the process, or at the latest, upon job completion. If running multiple jobs at the same time rather than sequentially, it may make sense to spread the software installation to multiple nodes if they all can access the source AFP files, in order to distribute the processing load.

3) ***Does the re-converted AFP file need to be stored to be read or is it only available in real-time?***
   Once re-converted to AFP, its being written (persisted) to the masked file target folder specified in the program calling DarkShield. Note that the masked PDF file can also be saved, as can XML et al.

IRI
Total Data Management

# Audit

Easily query, analyze, and format the results of search and mask operation through built-in reports and this graphical display.
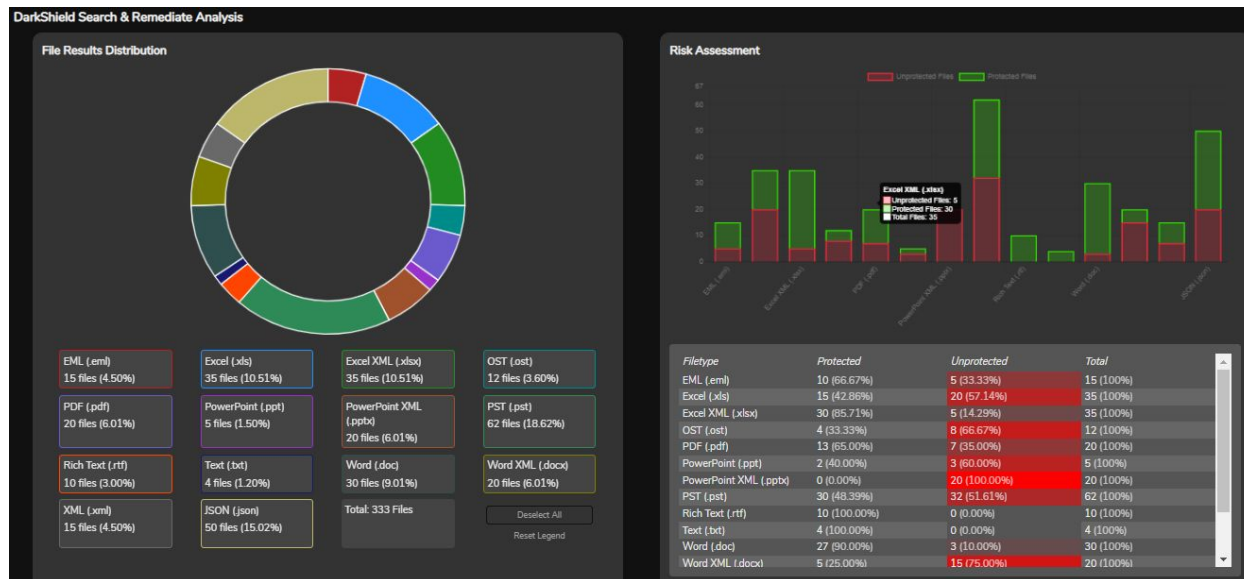
Or, export DarkShield log data for visualizations in BI tools like DataDog, or to SIEM environments like Splunk ES, shown here.

It is also then possible to take actions through the Splunk Adaptive Response Framework or a Splunk Phantom playbook.
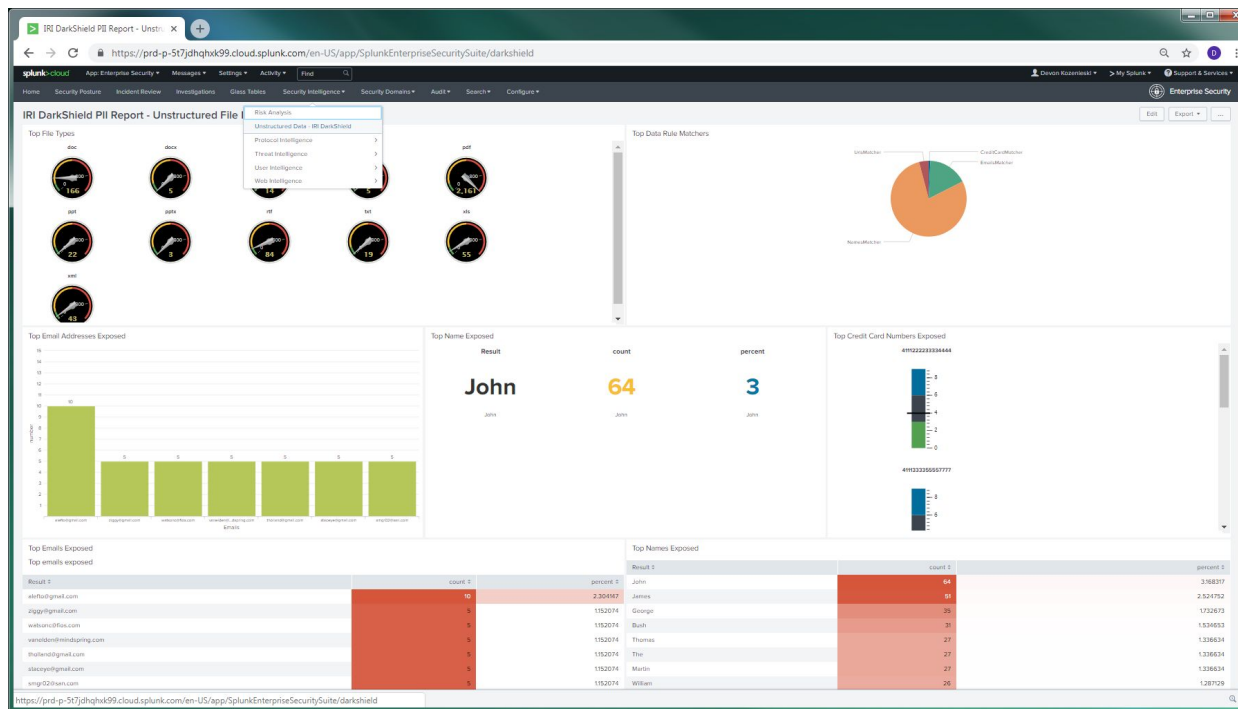
# DarkShield **Benefits**

1. Combines PII discovery, delivery, deletion, and reporting in multiple unstructured source formats into one or more ergonomic operations
2. Allows pattern reuse and stacked search methods for easier and more accurate results
3. Consolidates multiple *right to be forgotten* and *data portability* requests into the same find/fix operation through literal names or lookup-file matches
4. Supports multiple drives, nodes, and threads for searching and masking work
5. Can operates in the same Eclipse job design and metadata environment, IRI Workbench, with related data governance and management activities or via OpenAPI
6. Features affordable licensing options (standalone, bundled, or free in Voracity)
7. Compatibility with IRI FieldShield and CellShield data masking functions
8. Parameter serialization and modeling for easy job modification and batch execution
9. Integration with IRI RowGen to synthesize and insert test data into images, docs, etc.

# Development **Roadmap**

1. More unstructured format support, including A/V and proprietary/packaged apps
2. Additional ergonomic convergence with structured and embedded data sources
3. Plug-in integration with more SIEM tools *beyond* Datadog, Splunk ES, and Phantom Playbooks which are now supported, like IBM QRadar and SolarWinds
4. Retrofitting new API features into IRI Workbench wizards

IRI
Total Data Management

# IRI FieldShield
## PII / PHI Classification & Masking

*IRI Data Protector Suite*

# Sensitive Data Classification and Search Wizards

To facilitate data masking, IRI FieldShield includes: PII definition (cataloging through data classes); discovery through string (literal or dictionary), pattern, and fuzzy-logic searches; statistical reporting; and, automatic metadata creation.

Fit-for-purpose GUI wizards deliver:
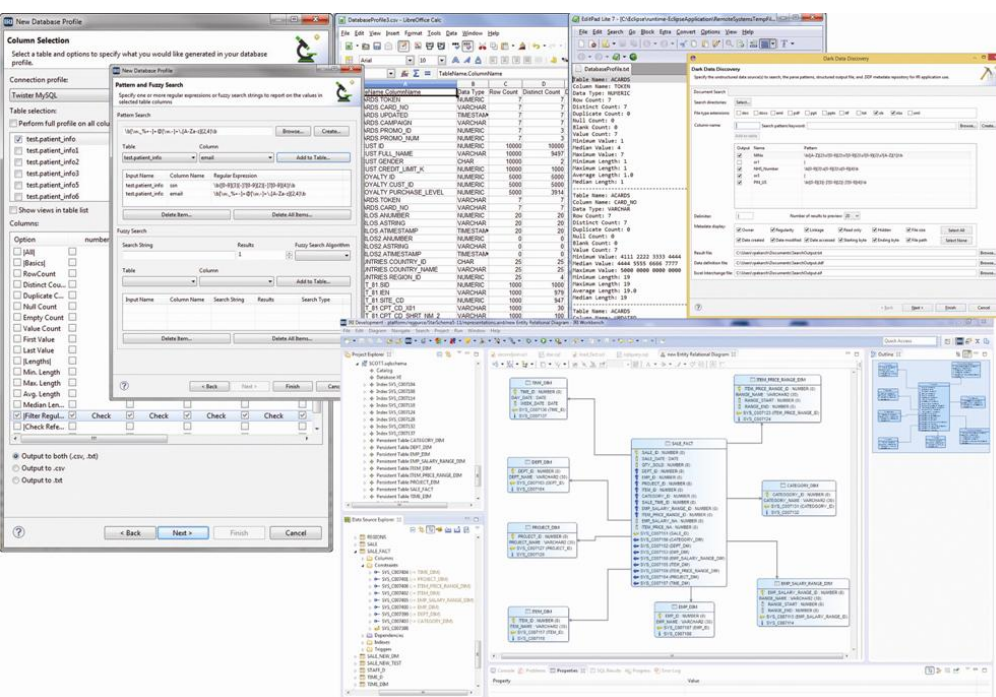
- DB and file data classification, with search and masking rule matchers
- DB profiling, ERDs, and table searches
- Flat-file profiling and value searches
- Data class searches through schema and directories for bulk discovery
- Metadata discovery and definition
- Dark data search and structuring, with metadata reporting (see DarkShield)

IRI
Total Data Management

# Search Result Reports, Dashboards & Exports

In addition to report-formatted and machine-readable outputs from PII search operations in IRI Workbench, FieldShield users can also see details at a glance in digital displays, or feed that data to tools like Splunk Enterprise Security and Datadog for analytics and action-taking:
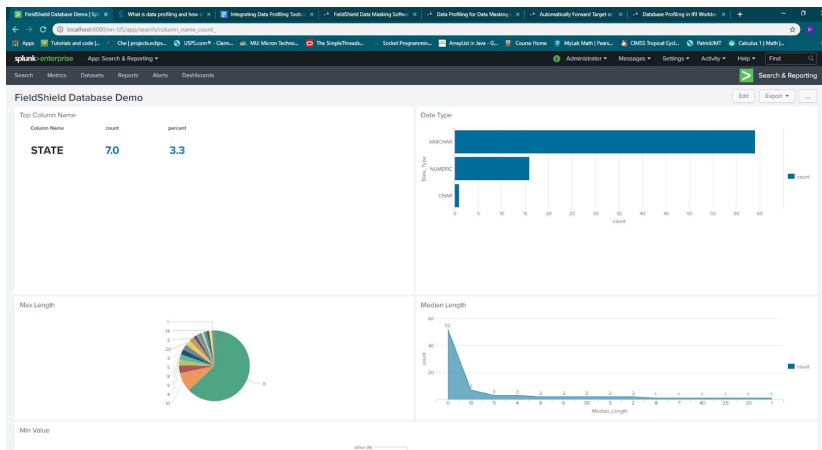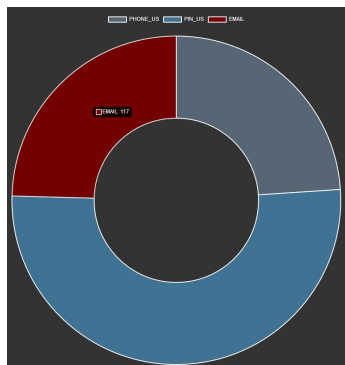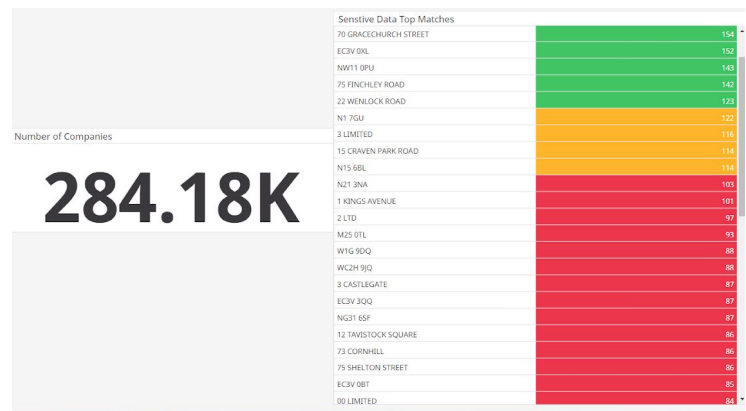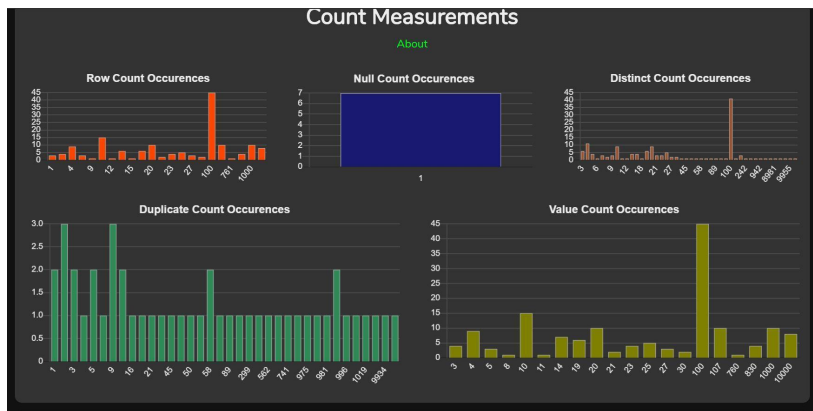
# Multiple Masking Job Design Options

IRI FieldShield and other Voracity data masking, cleansing, transformation, migration, reporting, and wrangling jobs can be created and run *inside or outside* of IRI Workbench.

Job design methods supported *inside*:

1) **Job creation wizards**

2) **Color-coded syntax-aware job script editor with outline**

3) **Form Editors**

4) **Graphical parameters Dialogs**

5) **Mapping Diagrams**

Job design methods supported *outside*:

6) **DataSwitch No-Code Web App**

7) **erwin Mapping Manager**

8) **Value Labs Test Data Hub**

9) **Any external text editor**

10) **3GL app (system or API calls)**

# And Multiple Job Deployment Options

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

2) **From 3rd party automation tools like Stonebranch UAC, cron, etc.**

3) **Directly from KNIME in Eclipse, or a Splunk add-on app, *as* you report or index**

4) **Some jobs run without code changes in Hadoop via MR2, Spark, Spark Stream, Storm or Tez**

5) **Use graphical run configuration dialogs or the built-in task scheduler to launch local, remote, or HDFS jobs from IRI Workbench**

6) **System or API calls from 3GL programs or web services**

7) **Value Labs Test Data Hub, Cigniti Blueswan TDM, GitLab, Azure DevOps, Amazon CodePipeline, and Jenkins CI/CD**

8) **Actifio, Commvault & Windocks cloning tools for virtualized DB images / containers**

9) **DataSwitch no-code data engineering app**

10) **IRI Ripcurrent facility in Voracity for real-time, incremental DB mask/refresh ops upon source table inserts, updates, or deletes**

# Re-ID Risk Determination



US HIPAA and FERPA regulations require that patient and student data sets used in research or marketing have a statistically certified "very small" chance of being re-identifiable.
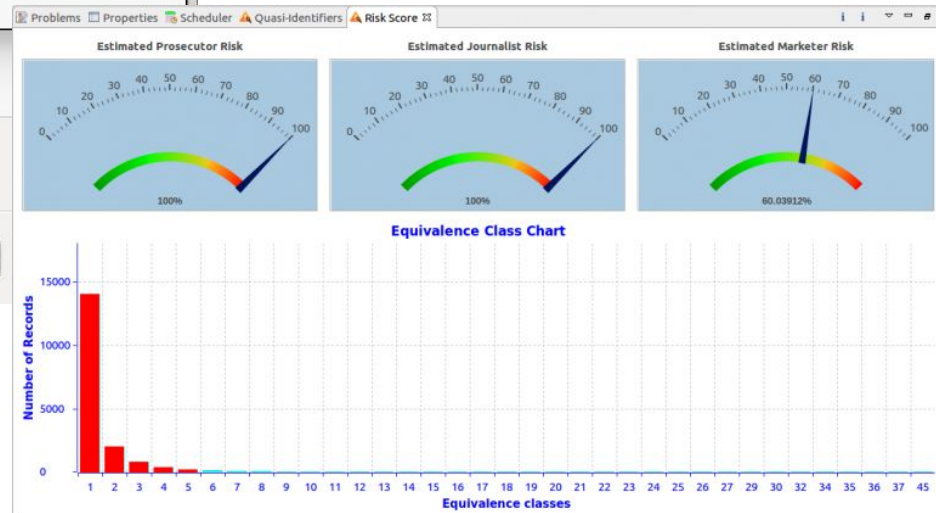
- IRI risk scoring wizard produces re-ID probability scores in 3 modes
- Analyzes quasi-identifiers with multiple, peer-reviewed functions
- Detail and graphed scoring reports

IRI
Total Data Management

# Query-Ready XML Audit Log

# Dynamic Data Masking Options

| Method | Operation |
|---|---|
| ODBC Select / Update | Apply FieldShield column masks to target (view) tables for specific users/rows |
| DB App Invocation | Use .NET or Java SDK library functions or system-call job scripts on the fly |
| In-Situ Redaction | User and SQL-specific full and partial column masking on query |
| Custom I/O Procedures | Drive real-time application data directly to/from FieldShield jobs in memory |
| Real-Time Processing | Hadoop Spark and Storm processing of dynamic input streams (via Voracity VGrid) |
| Proxy-based | New "JDBC SQL Trail" driver to intercept application queries with minimal impact |
| Governance Mode | New runtime facility tied to RBAC/IAM infrastructure masks fields for some users |

# Encryption Key Management Options

1. Passphrase (key string) embedded in script, in clear or encrypted text
2. Passphrase string as environment variable
3. Passphrase string in (securable) key file
4. MFA, HSM/VM etc. via Azure Key Vault
5. Townsend Security Alliance Key Manager

IRI
Total Data Management

# PII Encryption through Outside Key Vault

Azure Key Vault, Townsend
Alliance Key Manager, etc.

FieldShield Job Script or
DarkShield API Calling Program

Obtain Encryption Passphrase
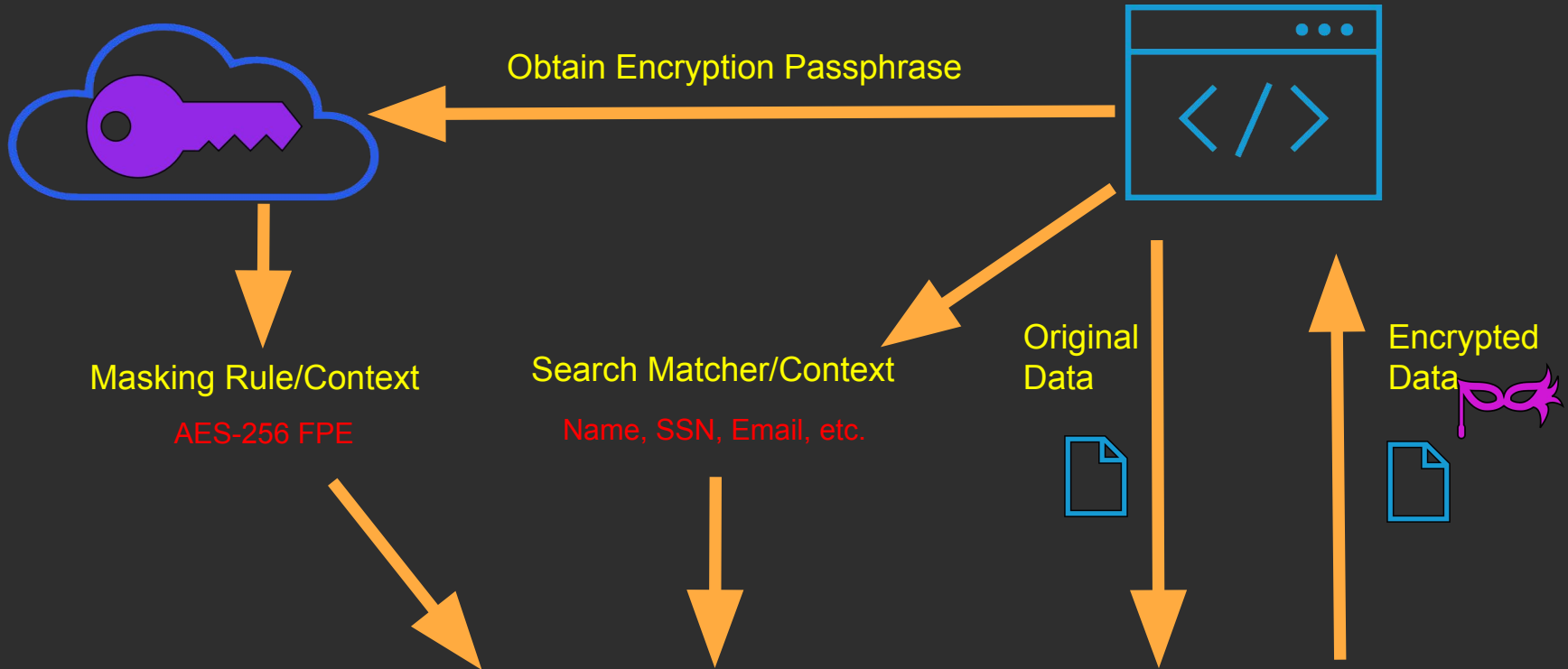
Masking Rule/Context

AES-256 FPE

Search Matcher/Context

Name, SSN, Email, etc.

Original Data

Encrypted Data

# IRI FieldShield
PII / PHI Classification & Masking
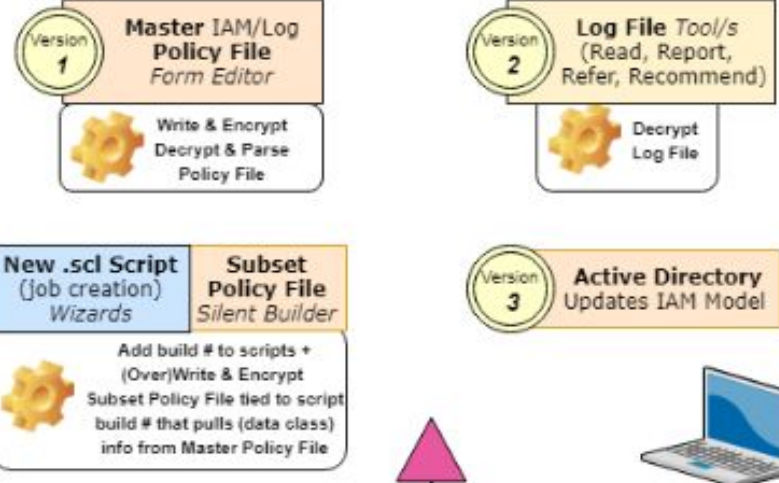
# IRI DarkShield
Unstructured Data Search & Security

# *IAM/RBAC Now & Later*

**Today**, you can assign permissions via encryption keys for select decryption, and: to data (file) sources, IRI masking programs (sortcl.exe), and the scripts they run (spec.fcl) in LUW file systems using central LDAP/AD settings. You can optionally control them via Apache Directory Studio in IRI Workbench:
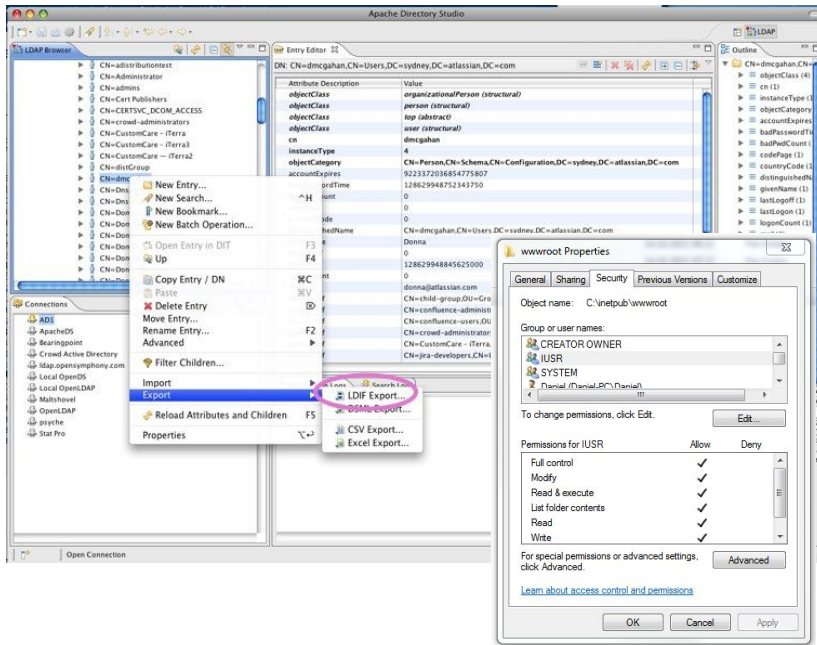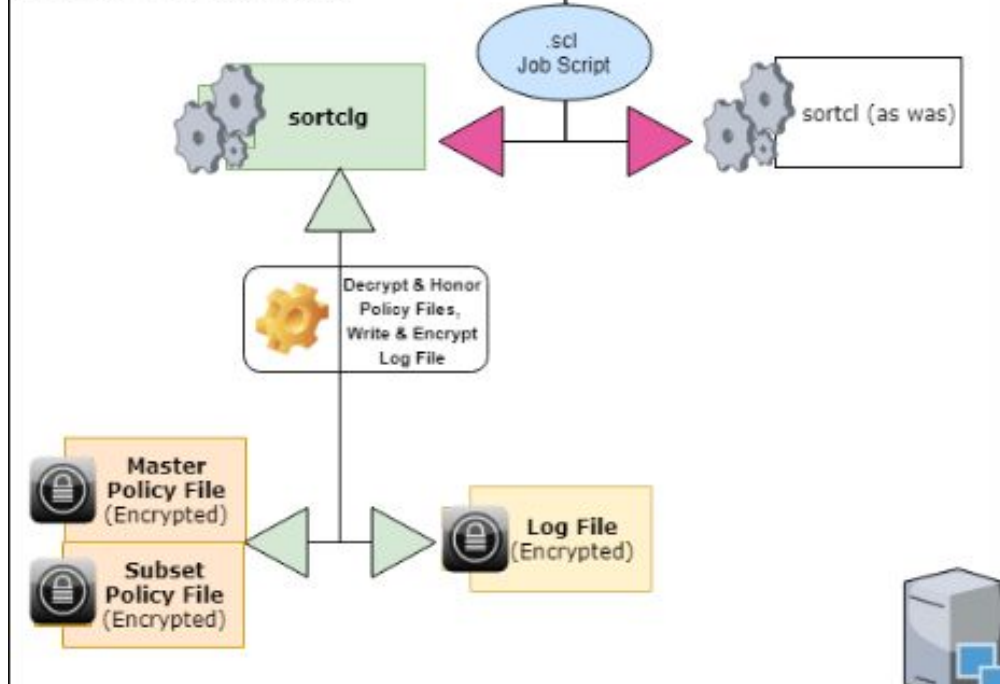
**Soon**, the IRI client/server governance system illustrated on the right will all you to assign and enforce RBACs to the same elements above, **and** to more granular elements like field names (mapped from data classes), functions, and perhaps specific data values (or ranges of values).

# Cloud Data & Systems Support

**FieldShield** can read/mask/write data in cloud DBs like Oracle 19c, Snowflake, MS SQL in Azure, AWS Redshift, etc. via J/ODBC, *plus* URLs & message queues.
**DarkShield** supports files in S3, GCP and Azure Storage, plus any RDB, 9 NoSQL DBs, and SMB-tied cloud drives like Dropbox and Sharepoint. **Both** operate on-premise or in the cloud, on bare metal shapes or VMs running Windows, Linux or Unix.
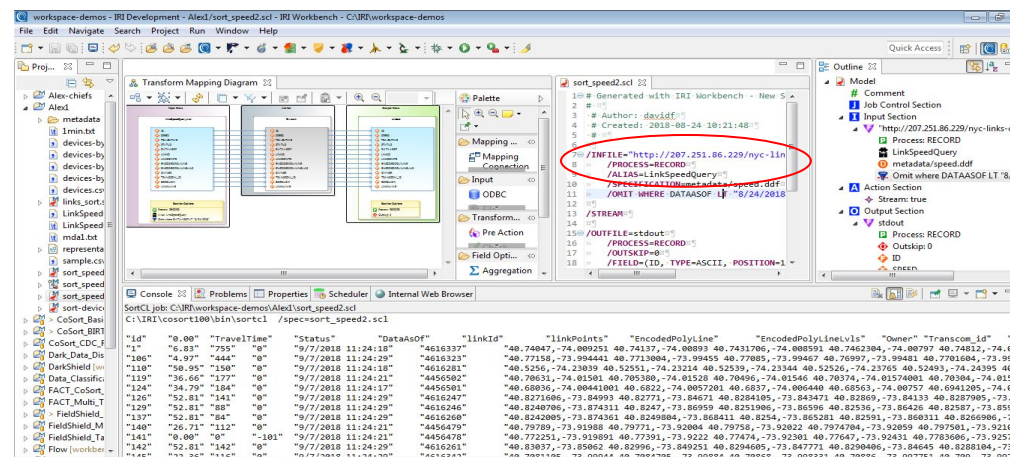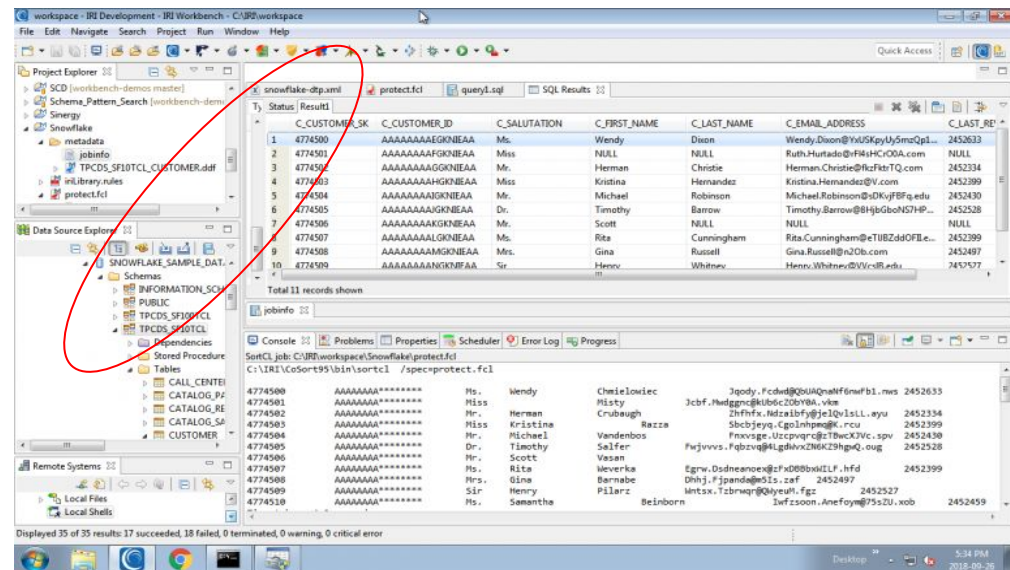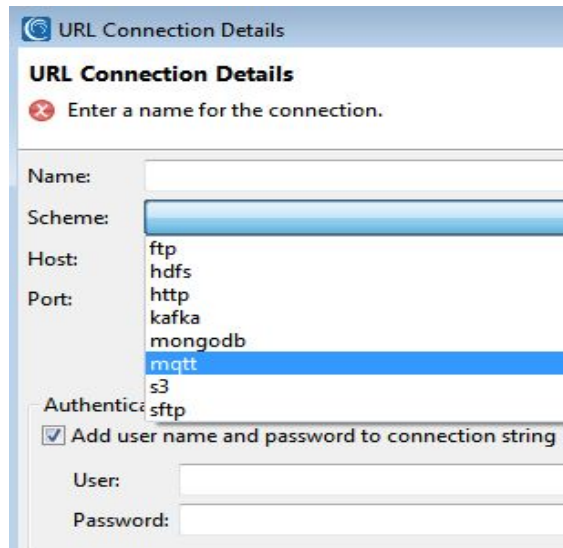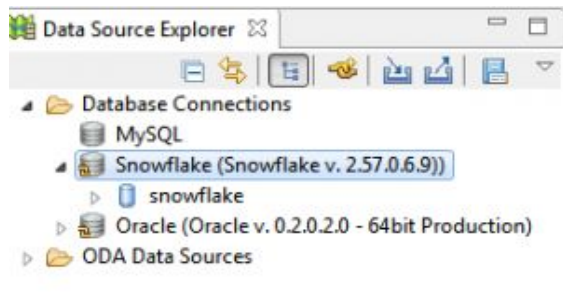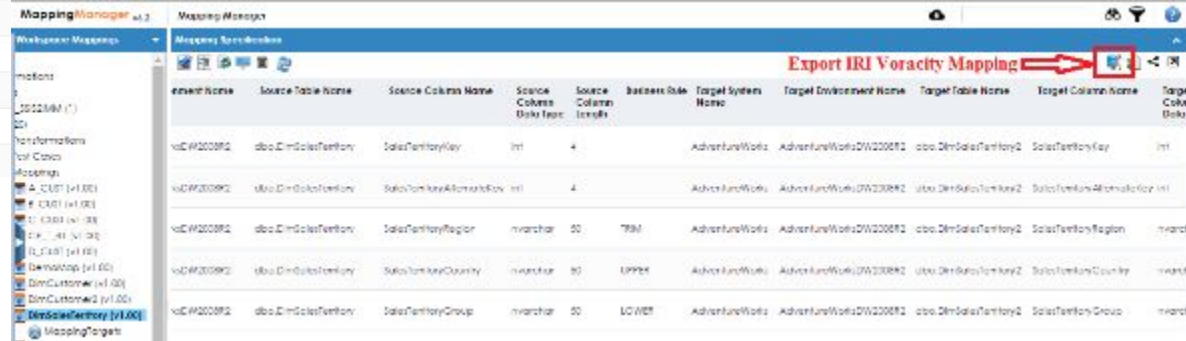
# Metadata Integrations

1. Voracity tooling *consumes* metadata from any structured source for data classification, profiling, search, de-ID, ETL etc.
2. FieldShield & RowGen job scripts also *produce* metadata for several **DB load utilities** in multi-DB masking & test data jobs.
3. Their data definition file metadata can also be *exported* (e.g. target field layouts) in CSV for catalog tools like **Collibra**.
4. DarkShield reads **attribute** metadata about source files, and produces artifactual metadata from its search and mask ops and it can auto-forward or populate Splunk ES with that information for analysis, dashboarding, or adaptive responses.
5. **MIMB, erwin, DataSwitch** and **ValueLabs TDH** hub and feed FieldShield DDF and .FCL specs from external metadata:



6. All IRI metadata -- including data source/target layouts, job/task speci and batch files, worfklows and metamodels, discovery configurations, search matchers and masking rules -- can also be team shared, secured and version controlled in **Git** et al
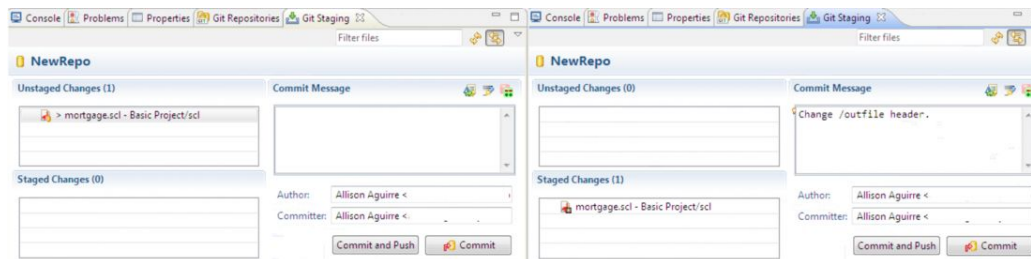
# Key Differentiators

## Developer Support

- Version controls
- Master data definition
- Secure key management
- Git project management (teaming)
- SDK supports .NET and Java calls
- Data profiling and metadata discovery
- XML (and soon JSON) job logs, IAM

## One-Stop-Shop

- Integrated data classification & search
- Includes re-ID risk scoring for HIPAA
- Use w/Voracity ETL, migrate, cleanse
- Metadata-compatible with RowGen TDM
- Used in DB subsetting wizard
- Also works in Voracity BI & KNIME jobs
- Runs w/Actifio DB clones, Splunk ES, etc.

## Price Performance

- The data-centric security tool with:
  - ➔ The most sources
  - ➔ The most protection functions
  - ➔ The most target file formats
- Fastest standalone protection software

## Ease-of-Use

- Familiar Eclipse GUI
- Self-documenting 4GL syntax
- Easy management and modification of jobs/metadata

IRI
Total Data Management

# Competitive Advantages

## vs. Oracle ([click](#))

### vs. IBM

- FieldShield scripts simpler than Optim interoperability model and Javascript options
- Seamless integration with more sources
- More functions
- Lower cost

### vs. Informatica

- FieldShield DDM inclusive with product (compared to Informatica's upgrade)
- More SDM protection functions
- Integration with Eclipse and Excel
- Access to 4GL scripts
- Lower cost

### vs. CA (Grid Tools)

- Built-in CoSort engine makes FieldShield faster than GT Fast Data Masking
- Tight integration with data profiling, ETL, data quality, and BI operations
- Multi-target/format options
- Lower cost
- Built-in re-ID risk determination wizard

### vs. Imperva (Camouflage)

- FieldShield has more masking and encryption functions
- Hash, decode, and pseudonymize functions
- Faster and more extensible in the IRI Workbench IDE
- Lower cost

IRI
Total Data Management

# Learn and Share

IRI Data Masking Solutions

Data Masking How-to Articles

LinkedIn Data Masking Group

IRI Mask/Test Tech Talk Videos

IRI Test Data Solutions

Test Data How-to Articles

LinkedIn Test Data Group

Voracity Platform Resources

IRI
Total Data Management