Global Big Data Conference

BIG DATA BOOTCAMP

Tampa

December 9th, 10th & 11th 2016

Tampa Convention Center, 333 S Franklin St, Tampa, FL 33602

www.globalbigdataconference.com

Twitter : @bigdataconf

IRI, The CoSort Company

Vendor Background

- ISV specializing in data management and data protection
- Known since 1978 for “big data” transformation speed
- 7 of 8 software products share 1 metadata and Eclipse GUI
- A ‘top big data provider’ (CIO Review & Insight Success)
- Headquartered 1 hour southeast of Orlando, FL
- Resellers in more than 40 international cities
- Customers in every industry with big and/or sensitive data

Selected IRI Customers

IRI customers process and protect data off the mainframe, for DW ETL/ODS ops, and in PII protection (privacy law compliance) initiatives. Hadoop use is optional. Most work with big and/or sensitive financial, call/click, or healthcare data.

BMO Bank of Montreal
RBS The Royal Bank of Scotland
WesBank
citi
Bank of America
KEB Hana Bank
CredibanCo
Fidelity INVESTMENTS
VISA
SUNGARD
EQUIFAX
Experian
EPSILON
Prudential
Aflac
BlueCross BlueShield
Allianz
The Hartford
HSBC
LG
SAMSUNG
AA AmericanAirlines
החברה לאוטומציה
במנהל השלטון המקומי
GENERALI GROUP
Hewlett Packard Enterprise
vivo
Mercedes-Benz
UNIVERSAL
comcast
NTTグループ
MEDICX digital
The Walt Disney Company
accenture High performance. Delivered.
1·800 flowers.com

Global Big Data Conference

IRI Data Manager Suite

FACT
IRI FAst extraCT

Speed DB 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

CoSORT
THE OPEN SYSTEMS STANDARD

Speed or replace batch, BI, ETL, sort, and SQL programs

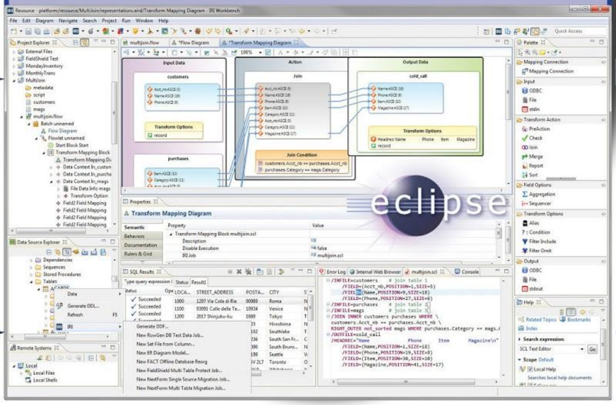
- Filter, sort, join, aggregate, pivot, cleanse, lookup, calc, etc.
- Map, migrate, federate and replicate data from 125 sources
- Segment data, capture changes, report details / summaries
- Analyze changing dimensions, support complex transforms

NextForm
Data & 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
- Search and structure data in "dark data" documents

IRI Voracity
Total Data Management



```
/INFILE=customers     # join table 1
    /FIELD=(Acct_nb,POSITION=1,SIZE=5)
    /FIELD=(Name,POSITION=9,SIZE=18)
    /FIELD=(Phone,POSITION=27,SIZE=8)
/INFILE=purchases     # join table 2
/INFILE=mags          # join table 3
/JOIN INNER customers purcheses WHERE \
customers.Acct_nb == purchases.Acct_nb \
RIGHT_OUTER not_sorted mags WHERE purchases.Category == mags.
/OUTFILE=cold_call
/HEADREC="Name              Phone     Item       Magazine\n"
    /FIELD=(Name,POSITION=1,SIZE=18)
    /FIELD=(Phone,POSITION=19,SIZE=8)
    /FIELD=(Item,POSITION=30,SIZE=10)
    /FIELD=(Magazine,POSITION=41,SIZE=17)
```

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

- Discover, define, and govern data in legacy and new sources
- Combine data integration, migration, protection, and analytics
- Exploit CoSort and Hadoop engines for optimum throughput
- Leverage Eclipse familiarity, functionality, and extensibility

IRI Data Protector Suite

RowGen
Safe Intelligent Test Data

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

FieldShield
Data Masking & Encryption Solutions

Comply with privacy laws, nullify breaches, govern data

- Select shields for each field per business rules
- De-ID, encrypt, hash, mask, pseudonym, random, token
- Apply cross-table rules to save time and referential integrity
- Create an XML audit log of each job to verify compliance

CellShield
Data Masking Add-In for Excel

Profile and protect PAN/PHI/PII in Excel spreadsheets

- Search and save patterns to discover 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

Chakra Max
Smart Data-Centric Audit & Protection

Define, monitor, block, and audit DB access

- **High-volume, data-centric audit and protection (DCAP)**
- **Monitor, block, alert, and log users in real-time**
- **Low-impact on DB performance and availability**
- **Classify and dynamically mask sensitive data with RBAC**

Embedded or callable analytics:
BIRT, JupiterOne, NextCoder, R

IRI
The CoSort Company

www.globalbigdataconference.com

Address the Challenges of Big Data

Volume

BI and analytic tools choke on high volumes; they drag, hang or crash

*Voracity blends and prepares data for analytic tools via **fast, combinatory transforms** like: filter, sort, join, aggregate and segment. Programs built on the CoSort SortCL language hand off digestible data chunks or cubes to BIRT, Qlik, R, SAS, Splunk, Tableau, etc.*

Variety

The myriad of structured and unstructured sources is beyond most tools

*Voracity either natively, or through partner drivers, connects to and integrates **>125 data sources** on premise or in the cloud. They can be structured, semi-structured, or unstructured, and static and streaming.*

Velocity

IOT logs, dark data, CDRs, etc. are generated too fast for analysis

*Voracity processes **streaming data** from: web services and brokers (MQTT, Kafka); pipes; in Hadoop Spark or Storm; SQL; and, through memory via input procedure calls to CoSort. Voracity's built-in task launcher can also run jobs in near-real-time.*

Veracity

Garbage in=garbage out: low quality data jeopardizes analytic value

*Voracity's data **discovery and quality** features let you: search for strings and patterns, do fuzzy matching, validate, scrub, enrich, and unify data for DW/BI, MDM, and analytics.*

Value

Without tackling the above, you won't get analytic value from big data

*Voracity runs with or without Hadoop on commodity hardware under an **affordable subscription** model based only on the number (not size) of servers. Its **Eclipse GUI** is free, familiar, and flexible, to speed learning and time-to-solution.*

Supported Data Sources/Targets:

Amazon EMR Hive	FinancialForce	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

… plus ‘legacy list’ on next 2 pages >>

- Acucobol Vision
- Altibase (FACT)
- ASN.1 TAP3
- BIRT DB (WB)
- BIRT Hive (WB)
- BIRT JDBC (WB)
- BIRT POJO (WB)
- C-ISAM
- CLF web logs
- CSV
- DB2 (UDB)
- DB2 for i5/OS (WB)
- DB2 for z/OS (WB)
- Delimited
- Derby (WB)
- ESDS
- Excel (WB)
- ELF web logs
- Fixed
- Heap / print
- HSQLDB (WB)
- IDX 3, 4 & 8
- Informix
- Ingres
- LDIF
- Line Sequential
- MaxDB
- Mongo (WB)
- MF-ISAM
- WF Var. Length
- MySQL
- Oracle
- Outlook (WB)
- PDF (WB)
- PostgreSQL
- Powerpoint (WB)
- Record Sequential
- RTF (WB)
- SQL Anywhere
- SQL Server
- SQLite
- Sybase ASA/E & IQ
- Tibero (WB)
- Teradata (WB)
- Text
- UTF-8 & 16
- Variable Block
- Variable Sequential
- VSAM MVS (UniKix)
- Web Services (WB)
- Word (WB)
- XML

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

Global Big Data Conference

Sources

Big Data
cloudera, MAPR, HIVE, Hortonworks, Spark, Pivotal, SAP HANA, NETEZZA

Call Detail Records
ASN.1 Formats

Cloud & SaaS
amazon web services, Microsoft Azure, salesforce, HubSpot, Marketo, eloqua

Databases
ORACLE DATABASE, IBM DB2, MySQL, TERADATA, TIBERO DataBase, mongoDB, Microsoft SQL Server, SYBASE, DATASTAX, ALTIBASE

Files & Pipes
COBOL, CSV, LDIF, LS-RS-VS, MFVL, Text, VB, Vision, XML

Mainframe
Adabas, Datacom, IDMS, IMS, ISAM, Pick, Unidata, VSAM, etc.

Semi & Unstructured
PDF, O, X, W, P

Other Sources
Custom Apps, ETL/ELT Tools, Packaged Apps, Web Logs

DISCOVER

Data Classification
Dark Data Search
DB & File Profiling
ER Diagramming
Metadata Definitions
Metadata Forensics
Multi-Method Search

INTEGRATE

Public/Private Mashup
Change Data Capture
Bulk DB Un/Load
Data Federation
One Pass ETL

MIGRATE

Data & File Types
Endianness
Databases
ETL Jobs
JCL Sorts

IRI Voracity
Total Data Management

GOVERN

Data Lineage
Data Masking
Data Quality
Metadata & MDM
Test Data Generation

ANALYZE

Embedded BI
BIRT & Splunk Feeds
Clickstream Analytics
Customer Segmentation
Slowly Changing Dimensions

DESIGN

ADS Mapping Manager
Form Editors
Graphical Dialogs
Outlines & Palettes
Script Editors
Visual Workflow
Wizards & Rules

DEPLOY

CoSort CLI/API (SMP)
Eclipse & Other Job Launchers
Java, Pequues, SQL
MapReduce (Grid)
Spark (In-Memory)
Storm (Streaming)
Tez (Batch)

Targets

Big Data
cloudera, MAPR, HIVE, Hortonworks, Spark, Pivotal, SAP HANA, NETEZZA

BI & Analytic Tools
BIRT, SAP BusinessObjects, X, IBM Cognos, MicroStrategy, ORACLE BUSINESS INTELLIGENCE, Qlik, R, splunk>, Spotfire, tableau software

Cloud & SaaS
amazon web services, Microsoft Azure, salesforce, HubSpot, Marketo, eloqua

Custom Reports
Detail & summary reports

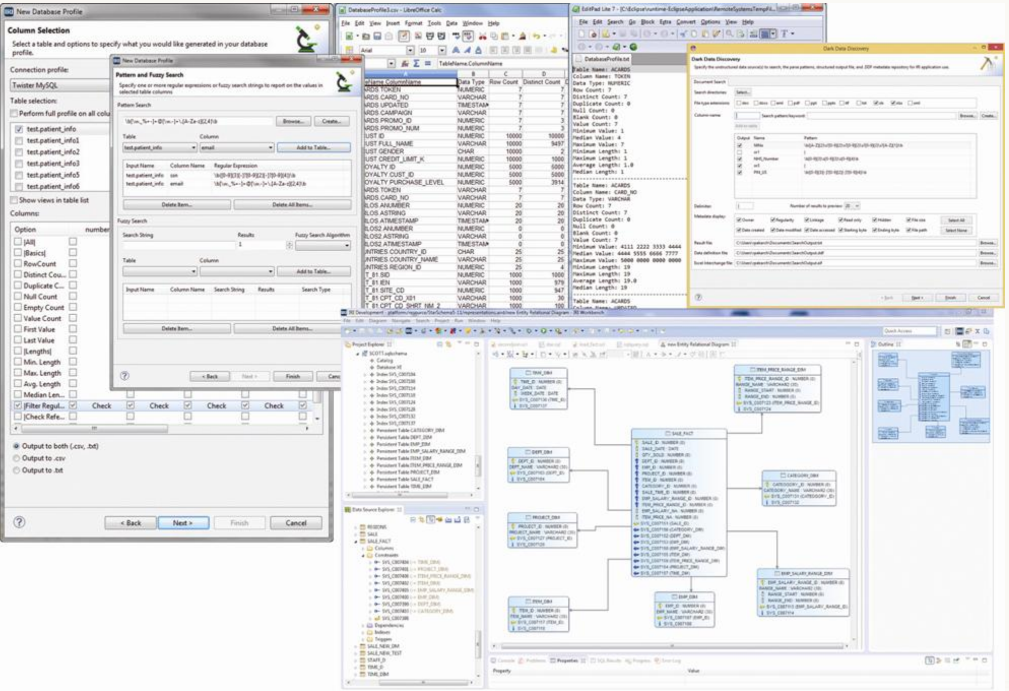
Databases
ORACLE DATABASE, IBM DB2, MySQL, TERADATA, TIBERO DataBase, mongoDB, Microsoft SQL Server, SYBASE, DATASTAX, ALTIBASE

Files & Pipes
COBOL, CSV, LDIF, LS-RS-VS, MFVL, Text, VB, Vision, XML

Other Targets
Custom Apps, Data & SpreadMarts, ETL/ELT Tools, Federated Views, Packaged Apps, Test Suites

N PartnerNet SILVER PARTNER Novell · intel Software Partner · IBM Business Partner · Business Partner hp · SOLUSI247 24 hours 7 days integrated ICT solution · AnalytiXDS Accelerate Project Delivery · eclipse · msdn Microsoft Developer Network · MICRO FOCUS · TRILLIUM SOFTWARE A Harte Hanks Company · redhat. · ORACLE PARTNERNETWORK

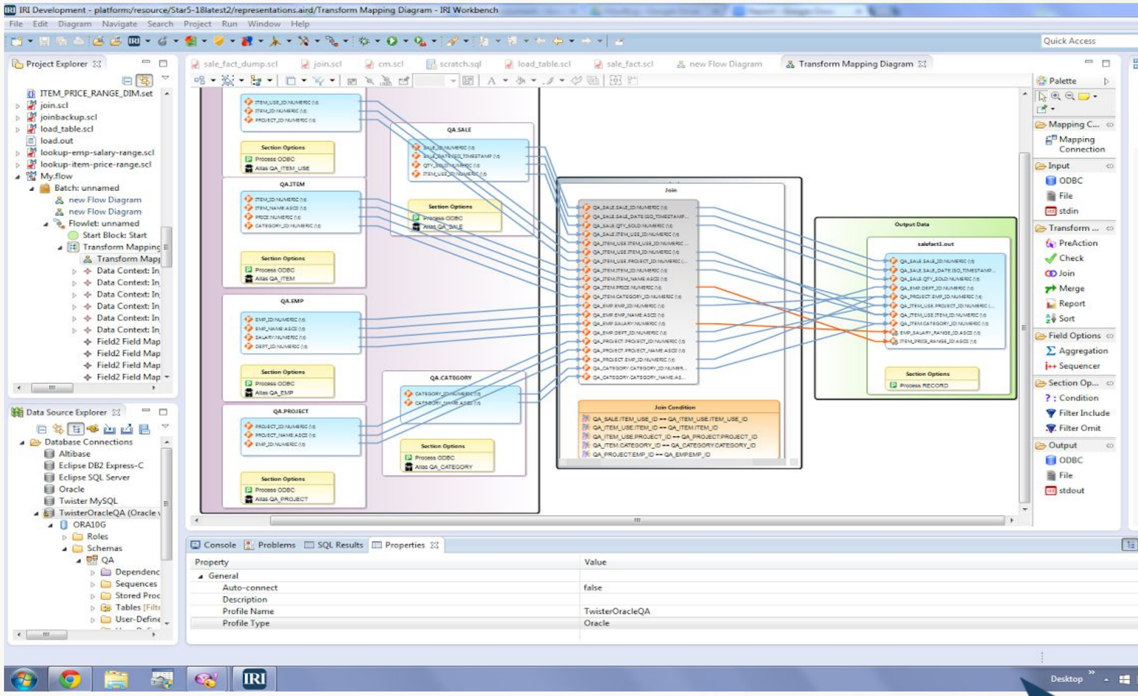
Voracity includes PII discovery facilities for multi-source data **classification**, string (literal or in-dictionary), pattern, and fuzzy-match **searches**, statistical **reports**, and automatic **metadata** creation. Fit-for-purpose wizards in Voracity perform:



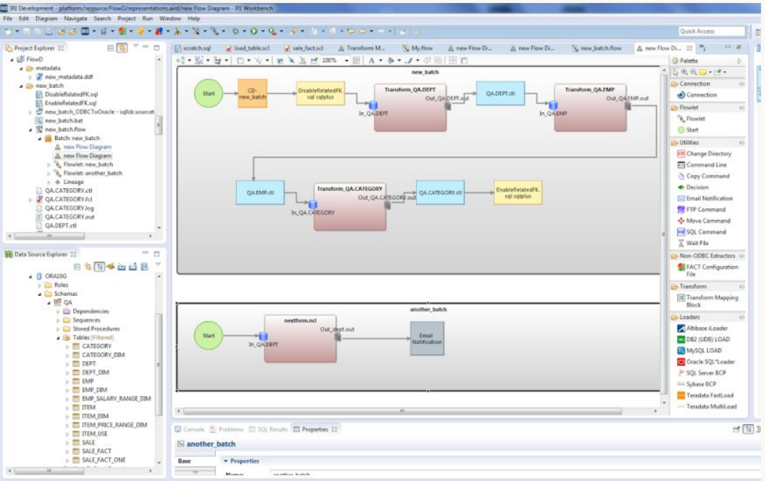
- Data classification, with rule matcher libraries
- DB profiling and E-R diagramming
- Dark data discovery and structuring, with forensic metadata display
- Flat-file statistical and value searching
- Metadata discovery and definition
- Metadata sharing, lineage tracking, etc.

Voracity combines fast ETL engines and task consolidation techniques with simple metadata in Eclipse that's shared by all IRI software and other products, like AnalytiX DS for ETL code conversion. You can use Voracity to speed or *re-platform* megavendor tools, and optimize:

- EDW, LDW, ODS, data lakes
- Data quality (cleansing)
- VLDB unload/reorg/load jobs
- SCD, CDC, pivoting, unification



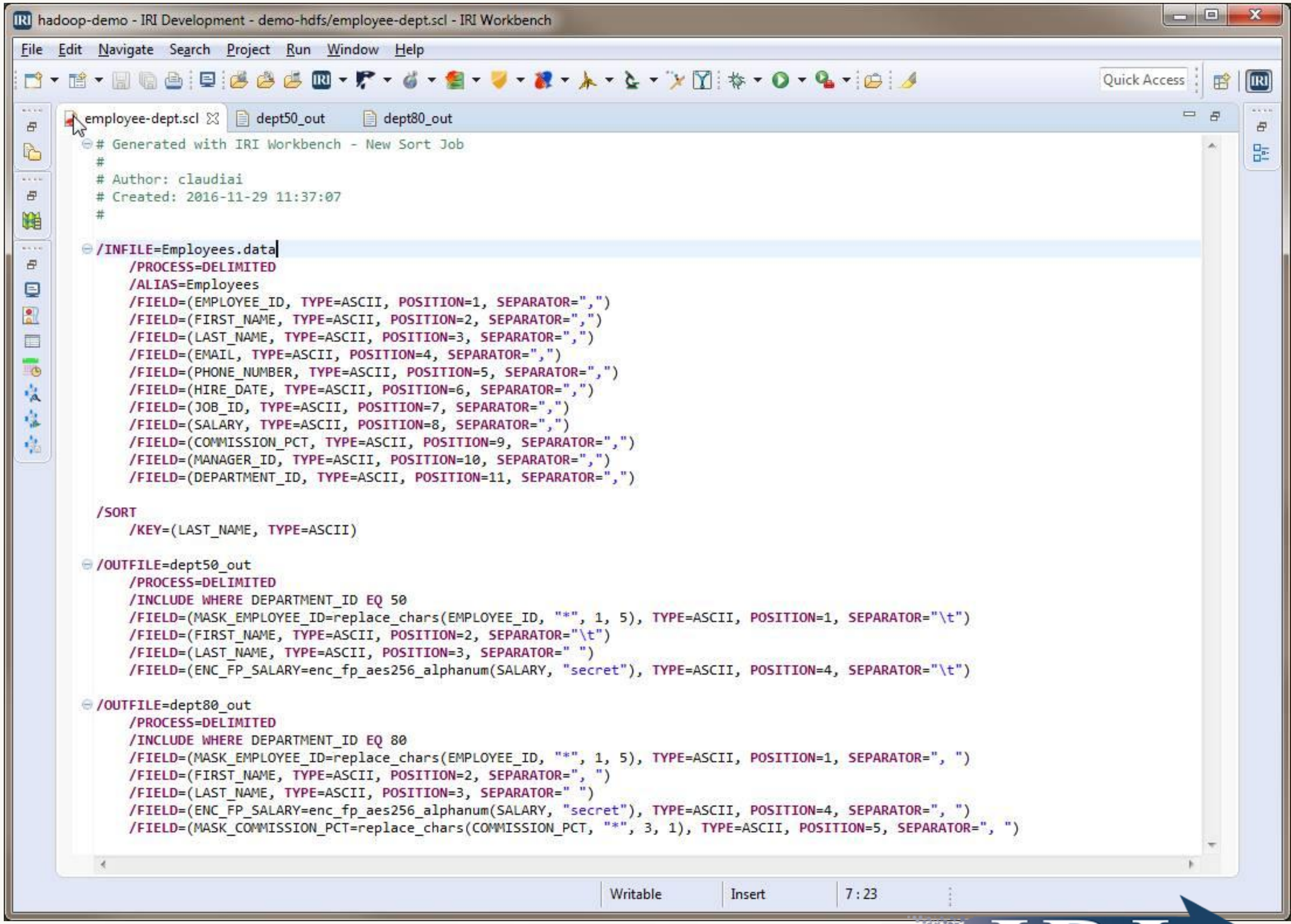
Property	Value
General	
Auto-connect	false
Description	
Profile Name	TwisterOracleQA
Profile Type	Oracle



Job Design …

In addition to GUI wizards, diagrams, and dialogs, you can also hand-code the underlying 4GL programs in Voracity's syntax-aware editor.

This job sorts and filters an employee CSV file into two target files, while also redacting ID #'s and commissions, and encrypting the salary.


hadoop-demo - IRI Development - demo-hdfs/employee-dept.scl - IRI Workbench
File Edit Navigate Search Project Run Window Help
Quick Access
employee-dept.scl
dept50_out
dept80_out
Writable
Insert
7 : 23

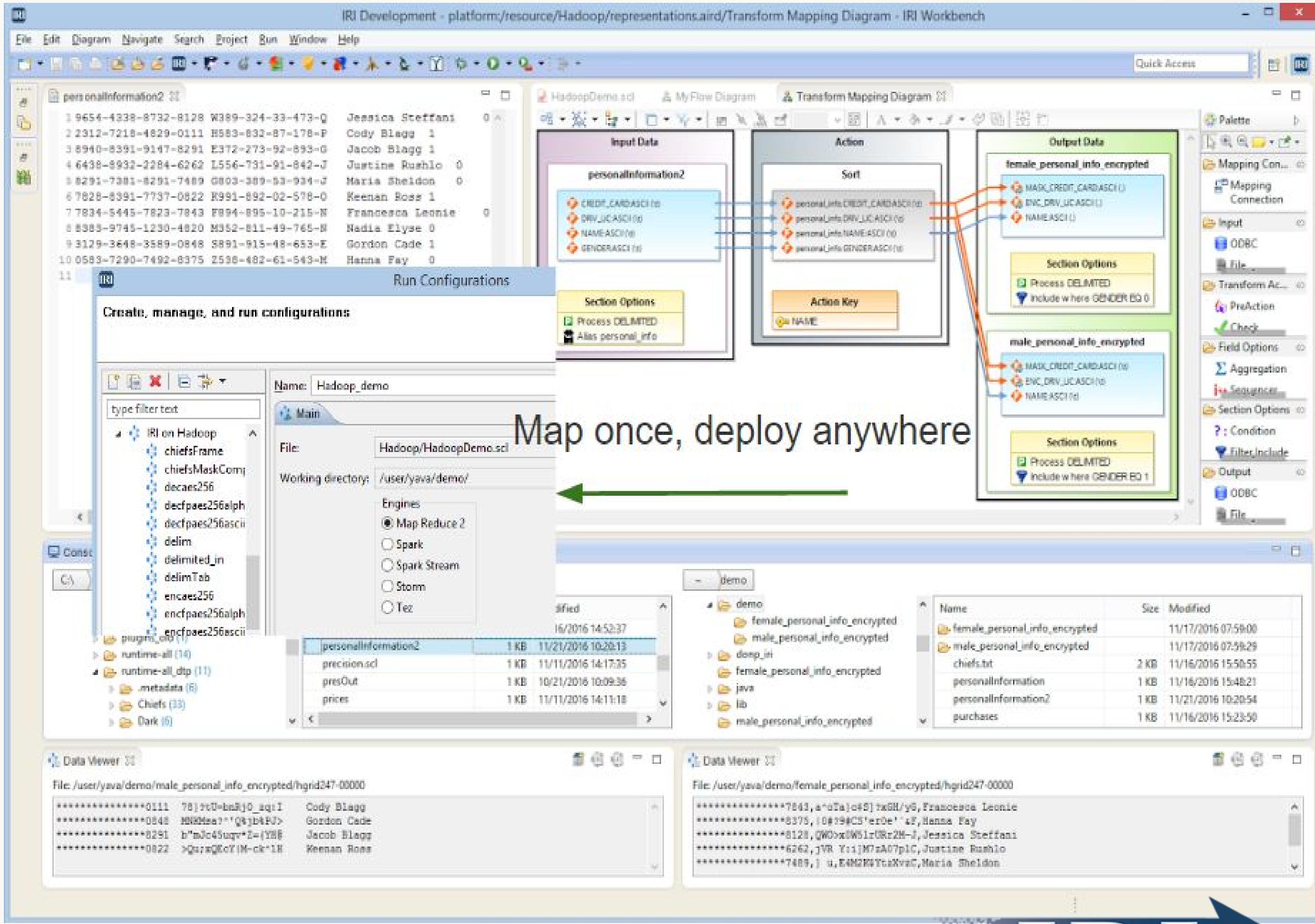
```
# Generated with IRI Workbench - New Sort Job
#
# Author: claudiai
# Created: 2016-11-29 11:37:07
#

/INFILE=Employees.data
    /PROCESS=DELIMITED
    /ALIAS=Employees
    /FIELD=(EMPLOYEE_ID, TYPE=ASCII, POSITION=1, SEPARATOR=",")
    /FIELD=(FIRST_NAME, TYPE=ASCII, POSITION=2, SEPARATOR=",")
    /FIELD=(LAST_NAME, TYPE=ASCII, POSITION=3, SEPARATOR=",")
    /FIELD=(EMAIL, TYPE=ASCII, POSITION=4, SEPARATOR=",")
    /FIELD=(PHONE_NUMBER, TYPE=ASCII, POSITION=5, SEPARATOR=",")
    /FIELD=(HIRE_DATE, TYPE=ASCII, POSITION=6, SEPARATOR=",")
    /FIELD=(JOB_ID, TYPE=ASCII, POSITION=7, SEPARATOR=",")
    /FIELD=(SALARY, TYPE=ASCII, POSITION=8, SEPARATOR=",")
    /FIELD=(COMMISSION_PCT, TYPE=ASCII, POSITION=9, SEPARATOR=",")
    /FIELD=(MANAGER_ID, TYPE=ASCII, POSITION=10, SEPARATOR=",")
    /FIELD=(DEPARTMENT_ID, TYPE=ASCII, POSITION=11, SEPARATOR=",")

/SORT
    /KEY=(LAST_NAME, TYPE=ASCII)

/OUTFILE=dept50_out
    /PROCESS=DELIMITED
    /INCLUDE WHERE DEPARTMENT_ID EQ 50
    /FIELD=(MASK_EMPLOYEE_ID=replace_chars(EMPLOYEE_ID, "*", 1, 5), TYPE=ASCII, POSITION=1, SEPARATOR="\t")
    /FIELD=(FIRST_NAME, TYPE=ASCII, POSITION=2, SEPARATOR="\t")
    /FIELD=(LAST_NAME, TYPE=ASCII, POSITION=3, SEPARATOR=" ")
    /FIELD=(ENC_FP_SALARY=enc_fp_aes256_alphanum(SALARY, "secret"), TYPE=ASCII, POSITION=4, SEPARATOR="\t")

/OUTFILE=dept80_out
    /PROCESS=DELIMITED
    /INCLUDE WHERE DEPARTMENT_ID EQ 80
    /FIELD=(MASK_EMPLOYEE_ID=replace_chars(EMPLOYEE_ID, "*", 1, 5), TYPE=ASCII, POSITION=1, SEPARATOR=", ")
    /FIELD=(FIRST_NAME, TYPE=ASCII, POSITION=2, SEPARATOR=", ")
    /FIELD=(LAST_NAME, TYPE=ASCII, POSITION=3, SEPARATOR=" ")
    /FIELD=(ENC_FP_SALARY=enc_fp_aes256_alphanum(SALARY, "secret"), TYPE=ASCII, POSITION=4, SEPARATOR=", ")
    /FIELD=(MASK_COMMISSION_PCT=replace_chars(COMMISSION_PCT, "*", 3, 1), TYPE=ASCII, POSITION=5, SEPARATOR=", ")
```

Job Deployment …

Voracity's 4GL scripts run on the command line or in batch from the GUI or shell.

BIRT or Splunk can also run them as they report or index.

Voracity can also schedule and run them seamlessly in MR2, Spark, Spark Stream, Storm or Tez.



IRI Development - platform:/resource/Hadoop/representations.aird/Transform Mapping Diagram - IRI Workbench

personalInformation2

```
9654-4338-8732-8128 W389-324-33-473-Q   Jessica Steffani    0
2312-7218-4829-0111 H583-832-87-178-F   Cody Blagg  1
8940-8391-9147-8291 E372-273-92-893-G   Jacob Blagg 1
6438-8932-2284-6262 L556-731-91-842-J   Justine Rushlo  0
8291-7381-8291-7489 G803-389-53-934-J   Maria Sheldon   0
7828-8391-7737-0822 K991-892-02-578-O   Keenan Ross 1
7834-5445-7823-7843 F894-895-10-215-N   Francesca Leonie    0
8383-9745-1230-4820 M352-811-49-765-N   Nadia Elyse 0
3129-3648-3589-0848 S891-915-48-653-E   Gordon Cade 1
0583-7290-7492-8375 Z538-482-61-543-M   Hanna Fay   0
```

Run Configurations

Create, manage, and run configurations

Name: Hadoop_demo

File: Hadoop/HadoopDemo.scl

Working directory: /user/yava/demo/

Engines

● Map Reduce 2
○ Spark
○ Spark Stream
○ Storm
○ Tez

Map once, deploy anywhere

Name	Size	Modified
female_personal_info_encrypted		11/17/2016 07:59:00
male_personal_info_encrypted		11/17/2016 07:59:29
chiefs.txt	2 KB	11/16/2016 15:50:55
personalInformation	1 KB	11/16/2016 15:48:21
personalInformation2	1 KB	11/21/2016 10:20:54
purchases	1 KB	11/16/2016 15:23:50

Data Viewer

File: /user/yava/demo/male_personal_info_encrypted/hgrid247-00000

```
***************0111  78|?tU=bnRj0_aq:I    Cody Blagg
***************0848  MNRMsa?^'Q%jb&PJ>    Gordon Cade
***************8291  b"mJc45uqv*Z=(YH§    Jacob Blagg
***************0822  >Qu;mQKcY|M-ck^1H    Keenan Ross
```

Data Viewer

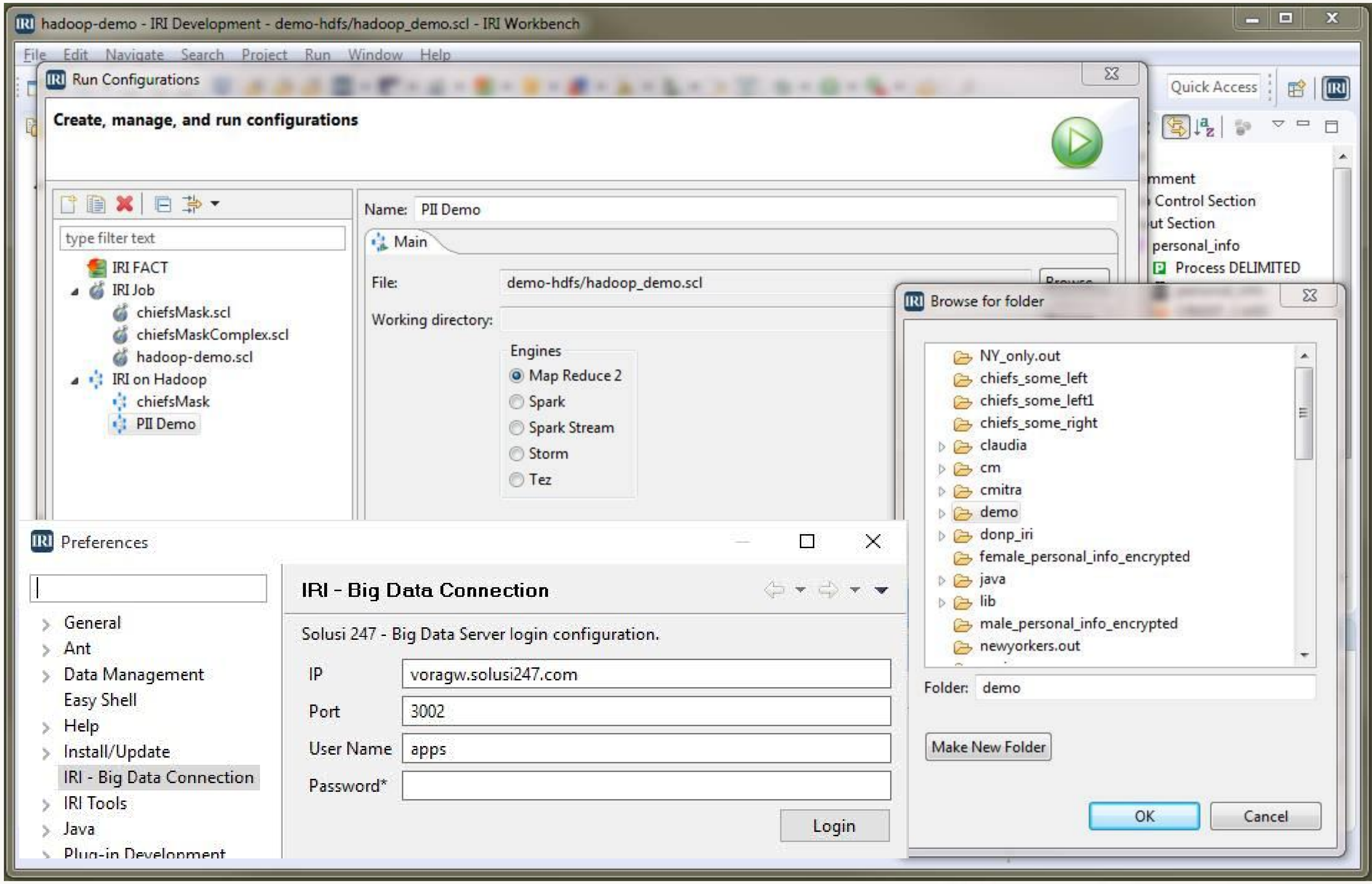
File: /user/yava/demo/female_personal_info_encrypted/hgrid247-00000

```
***************7843,a^oTa]c#S]?xGH/yG,Francesca Leonie
***************8375,|0#?9#CS'er0e'`&F,Hanna Fay
***************8128,QWO>xIW51rURr2M-J,Jessica Steffani
***************6262,jVR Y:1]M7zA07p1C,Justine Rushlo
***************7489,] u,E4M2K@YtzXvzC,Maria Sheldon
```

Preparing a run configuration for ***Hadoop*** *...*

Once our gateway is open, we can tell any job to run in Hadoop.

Here, we specify MR2 as the engine, and our working directory in HDFS.



hadoop-demo - IRI Development - demo-hdfs/hadoop_demo.scl - IRI Workbench

Run Configurations

Create, manage, and run configurations

Name: PII Demo

File: demo-hdfs/hadoop_demo.scl

Working directory:

Engines

- ● Map Reduce 2
- ○ Spark
- ○ Spark Stream
- ○ Storm
- ○ Tez

Preferences

IRI - Big Data Connection

Solusi 247 - Big Data Server login configuration.

IP: voragw.solusi247.com

Port: 3002

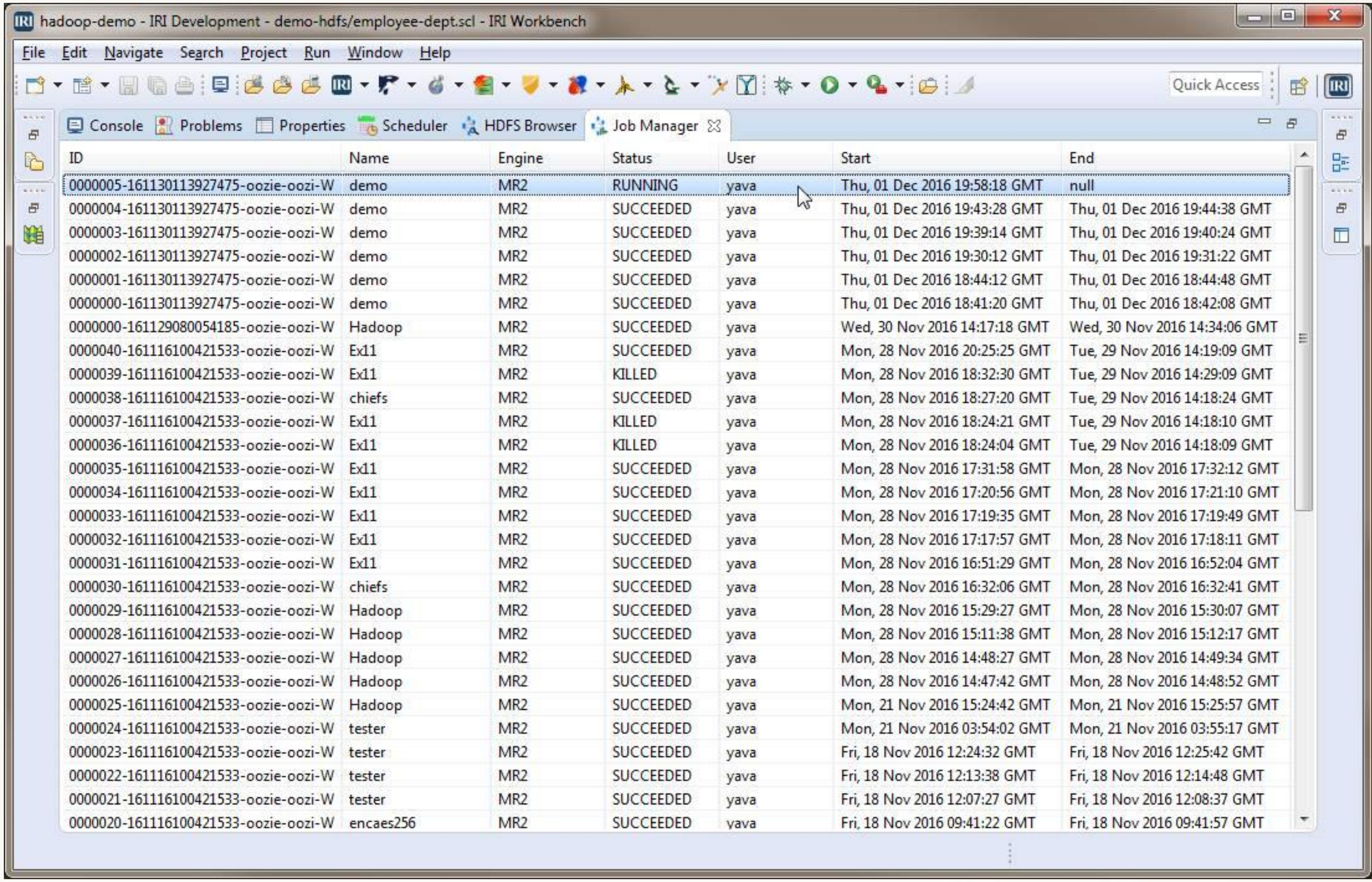
User Name: apps

Password*:

Browse for folder

Folder: demo

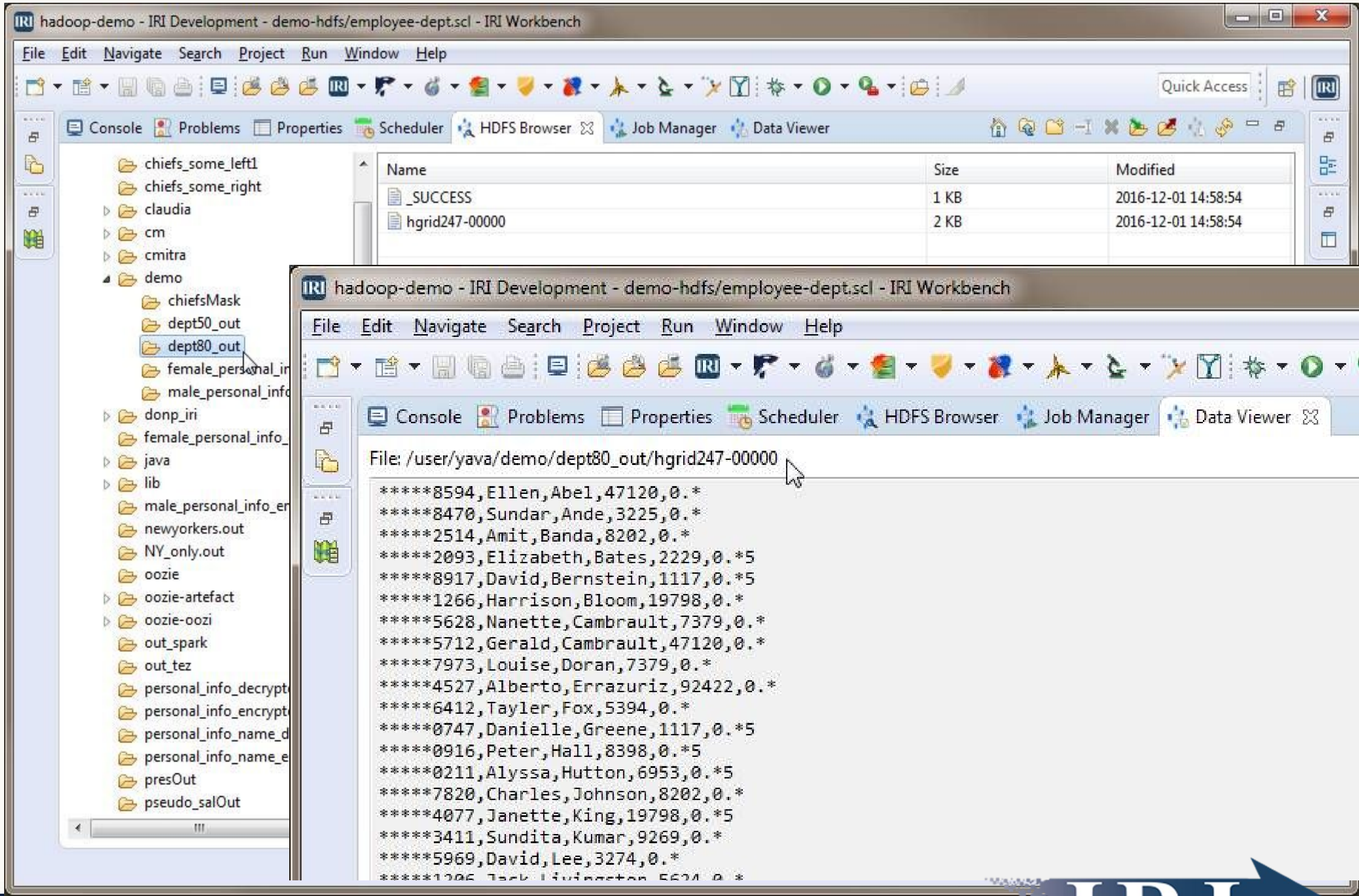
The Job Manager view shows our Hadoop job running, plus the status of other jobs.



ID	Name	Engine	Status	User	Start	End
0000005-161130113927475-oozie-oozi-W	demo	MR2	RUNNING	yava	Thu, 01 Dec 2016 19:58:18 GMT	null
0000004-161130113927475-oozie-oozi-W	demo	MR2	SUCCEEDED	yava	Thu, 01 Dec 2016 19:43:28 GMT	Thu, 01 Dec 2016 19:44:38 GMT
0000003-161130113927475-oozie-oozi-W	demo	MR2	SUCCEEDED	yava	Thu, 01 Dec 2016 19:39:14 GMT	Thu, 01 Dec 2016 19:40:24 GMT
0000002-161130113927475-oozie-oozi-W	demo	MR2	SUCCEEDED	yava	Thu, 01 Dec 2016 19:30:12 GMT	Thu, 01 Dec 2016 19:31:22 GMT
0000001-161130113927475-oozie-oozi-W	demo	MR2	SUCCEEDED	yava	Thu, 01 Dec 2016 18:44:12 GMT	Thu, 01 Dec 2016 18:44:48 GMT
0000000-161130113927475-oozie-oozi-W	demo	MR2	SUCCEEDED	yava	Thu, 01 Dec 2016 18:41:20 GMT	Thu, 01 Dec 2016 18:42:08 GMT
0000000-161129080054185-oozie-oozi-W	Hadoop	MR2	SUCCEEDED	yava	Wed, 30 Nov 2016 14:17:18 GMT	Wed, 30 Nov 2016 14:34:06 GMT
0000040-161116100421533-oozie-oozi-W	Ex11	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 20:25:25 GMT	Tue, 29 Nov 2016 14:19:09 GMT
0000039-161116100421533-oozie-oozi-W	Ex11	MR2	KILLED	yava	Mon, 28 Nov 2016 18:32:30 GMT	Tue, 29 Nov 2016 14:29:09 GMT
0000038-161116100421533-oozie-oozi-W	chiefs	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 18:27:20 GMT	Tue, 29 Nov 2016 14:18:24 GMT
0000037-161116100421533-oozie-oozi-W	Ex11	MR2	KILLED	yava	Mon, 28 Nov 2016 18:24:21 GMT	Tue, 29 Nov 2016 14:18:10 GMT
0000036-161116100421533-oozie-oozi-W	Ex11	MR2	KILLED	yava	Mon, 28 Nov 2016 18:24:04 GMT	Tue, 29 Nov 2016 14:18:09 GMT
0000035-161116100421533-oozie-oozi-W	Ex11	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 17:31:58 GMT	Mon, 28 Nov 2016 17:32:12 GMT
0000034-161116100421533-oozie-oozi-W	Ex11	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 17:20:56 GMT	Mon, 28 Nov 2016 17:21:10 GMT
0000033-161116100421533-oozie-oozi-W	Ex11	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 17:19:35 GMT	Mon, 28 Nov 2016 17:19:49 GMT
0000032-161116100421533-oozie-oozi-W	Ex11	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 17:17:57 GMT	Mon, 28 Nov 2016 17:18:11 GMT
0000031-161116100421533-oozie-oozi-W	Ex11	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 16:51:29 GMT	Mon, 28 Nov 2016 16:52:04 GMT
0000030-161116100421533-oozie-oozi-W	chiefs	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 16:32:06 GMT	Mon, 28 Nov 2016 16:32:41 GMT
0000029-161116100421533-oozie-oozi-W	Hadoop	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 15:29:27 GMT	Mon, 28 Nov 2016 15:30:07 GMT
0000028-161116100421533-oozie-oozi-W	Hadoop	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 15:11:38 GMT	Mon, 28 Nov 2016 15:12:17 GMT
0000027-161116100421533-oozie-oozi-W	Hadoop	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 14:48:27 GMT	Mon, 28 Nov 2016 14:49:34 GMT
0000026-161116100421533-oozie-oozi-W	Hadoop	MR2	SUCCEEDED	yava	Mon, 28 Nov 2016 14:47:42 GMT	Mon, 28 Nov 2016 14:48:52 GMT
0000025-161116100421533-oozie-oozi-W	Hadoop	MR2	SUCCEEDED	yava	Mon, 21 Nov 2016 15:24:42 GMT	Mon, 21 Nov 2016 15:25:57 GMT
0000024-161116100421533-oozie-oozi-W	tester	MR2	SUCCEEDED	yava	Mon, 21 Nov 2016 03:54:02 GMT	Mon, 21 Nov 2016 03:55:17 GMT
0000023-161116100421533-oozie-oozi-W	tester	MR2	SUCCEEDED	yava	Fri, 18 Nov 2016 12:24:32 GMT	Fri, 18 Nov 2016 12:25:42 GMT
0000022-161116100421533-oozie-oozi-W	tester	MR2	SUCCEEDED	yava	Fri, 18 Nov 2016 12:13:38 GMT	Fri, 18 Nov 2016 12:14:48 GMT
0000021-161116100421533-oozie-oozi-W	tester	MR2	SUCCEEDED	yava	Fri, 18 Nov 2016 12:07:27 GMT	Fri, 18 Nov 2016 12:08:37 GMT
0000020-161116100421533-oozie-oozi-W	encaes256	MR2	SUCCEEDED	yava	Fri, 18 Nov 2016 09:41:22 GMT	Fri, 18 Nov 2016 09:41:57 GMT

The HDFS Browser and Data Viewer show the target file and its contents ..

You can also use the viewer window to manage all of your input and output data directly in HDFS..


hadoop-demo - IRI Development - demo-hdfs/employee-dept.scl - IRI Workbench
File Edit Navigate Search Project Run Window Help
Quick Access
Console Problems Properties Scheduler HDFS Browser Job Manager Data Viewer

```
chiefs_some_left1
chiefs_some_right
claudia
cm
cmitra
demo
    chiefsMask
    dept50_out
    dept80_out
    female_personal_ir
    male_personal_info
donp_iri
female_personal_info_
java
lib
male_personal_info_er
newyorkers.out
NY_only.out
oozie
oozie-artefact
oozie-oozi
out_spark
out_tez
personal_info_decrypt
personal_info_encrypt
personal_info_name_d
personal_info_name_e
presOut
pseudo_salOut
```

Name	Size	Modified
_SUCCESS	1 KB	2016-12-01 14:58:54
hgrid247-00000	2 KB	2016-12-01 14:58:54

hadoop-demo - IRI Development - demo-hdfs/employee-dept.scl - IRI Workbench
File Edit Navigate Search Project Run Window Help
Console Problems Properties Scheduler HDFS Browser Job Manager Data Viewer

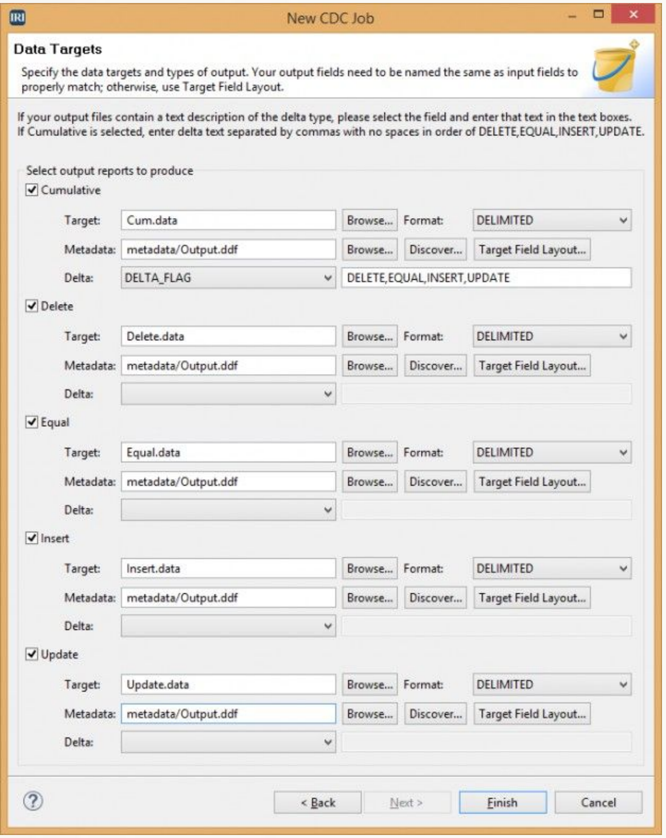
File: /user/yava/demo/dept80_out/hgrid247-00000

```
*****8594,Ellen,Abel,47120,0.*
*****8470,Sundar,Ande,3225,0.*
*****2514,Amit,Banda,8202,0.*
*****2093,Elizabeth,Bates,2229,0.*5
*****8917,David,Bernstein,1117,0.*5
*****1266,Harrison,Bloom,19798,0.*
*****5628,Nanette,Cambrault,7379,0.*
*****5712,Gerald,Cambrault,47120,0.*
*****7973,Louise,Doran,7379,0.*
*****4527,Alberto,Errazuriz,92422,0.*
*****6412,Tayler,Fox,5394,0.*
*****0747,Danielle,Greene,1117,0.*5
*****0916,Peter,Hall,8398,0.*5
*****0211,Alyssa,Hutton,6953,0.*5
*****7820,Charles,Johnson,8202,0.*
*****4077,Janette,King,19798,0.*5
*****3411,Sundita,Kumar,9269,0.*
*****5969,David,Lee,3274,0.*
*****1206,Jack,Livingston,5624,0.*
```

IRI The CoSort Company

Global Big Data Conference

DISCOVER **INTEGRATE** MIGRATE GOVERN ANALYZE



New CDC Job

Data Targets

Specify the data targets and types of output. Your output fields need to be named the same as input fields to properly match; otherwise, use Target Field Layout.

If your output files contain a text description of the delta type, please select the field and enter that text in the text boxes. If Cumulative is selected, enter delta text separated by commas with no spaces in order of DELETE,EQUAL,INSERT,UPDATE.

Select output reports to produce

☑ Cumulative

Target:	Cum.data	Browse...	Format:	DELIMITED
Metadata:	metadata/Output.ddf	Browse...	Discover...	Target Field Layout...
Delta:	DELTA_FLAG	DELETE,EQUAL,INSERT,UPDATE		

☑ Delete

Target:	Delete.data	Browse...	Format:	DELIMITED
Metadata:	metadata/Output.ddf	Browse...	Discover...	Target Field Layout...
Delta:				

☑ Equal

Target:	Equal.data	Browse...	Format:	DELIMITED
Metadata:	metadata/Output.ddf	Browse...	Discover...	Target Field Layout...
Delta:				

☑ Insert

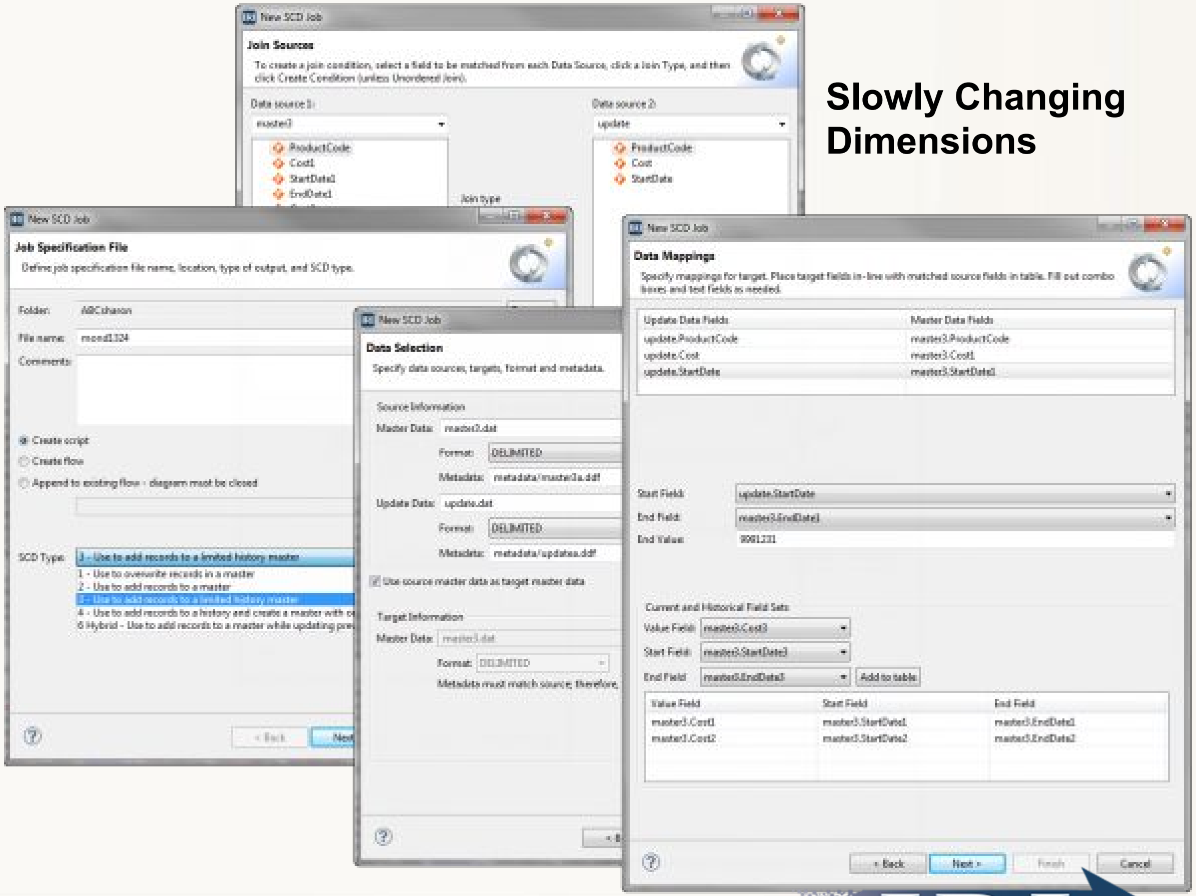
Target:	Insert.data	Browse...	Format:	DELIMITED
Metadata:	metadata/Output.ddf	Browse...	Discover...	Target Field Layout...
Delta:				

☑ Update

Target:	Update.data	Browse...	Format:	DELIMITED
Metadata:	metadata/Output.ddf	Browse...	Discover...	Target Field Layout...
Delta:				

Change Data Capture

Wizards for ...



New SCD Job

Join Sources

To create a join condition, select a field to be matched from each Data Source, click a Join Type, and then click Create Condition (unless Unordered Join).

Data source 1: master3

- ProductCode
- Cost1
- StartDate1
- EndDate1

Data source 2: update

- ProductCode
- Cost
- StartDate

Join type

New SCD Job

Job Specification File

Define job specification file name, location, type of output, and SCD type.

Folder: ABC sharon

File name: mood1324

Comments:

● Create script

○ Create flow

○ Append to existing flow - diagram must be closed

SCD Type: 3 - Use to add records to a limited history master

1 - Use to overwrite records in a master
2 - Use to add records to a master
3 - Use to add records to a limited history master
4 - Use to add records to a history and create a master with o
6 Hybrid - Use to add records to a master while updating pre

New SCD Job

Data Selection

Specify data sources, targets, format and metadata.

Source Information

Master Data: master3.dat

Format: DELIMITED

Metadata: metadata/master3a.ddf

Update Data: update.dat

Format: DELIMITED

Metadata: metadata/updatea.ddf

☑ Use source master data as target master data

Target Information

Master Data: master3.dat

Format: DELIMITED

Metadata must match source; therefore,

New SCD Job

Data Mappings

Specify mappings for target. Place target fields in-line with matched source fields in table. Fill out combo boxes and text fields as needed.

Update Data Fields	Master Data Fields
update.ProductCode	master3.ProductCode
update.Cost	master3.Cost1
update.StartDate	master3.StartDate1

Start Field: update.StartDate

End Field: master3.EndDate1

End Value: 99991231

Current and Historical Field Sets

Value Field: master3.Cost3

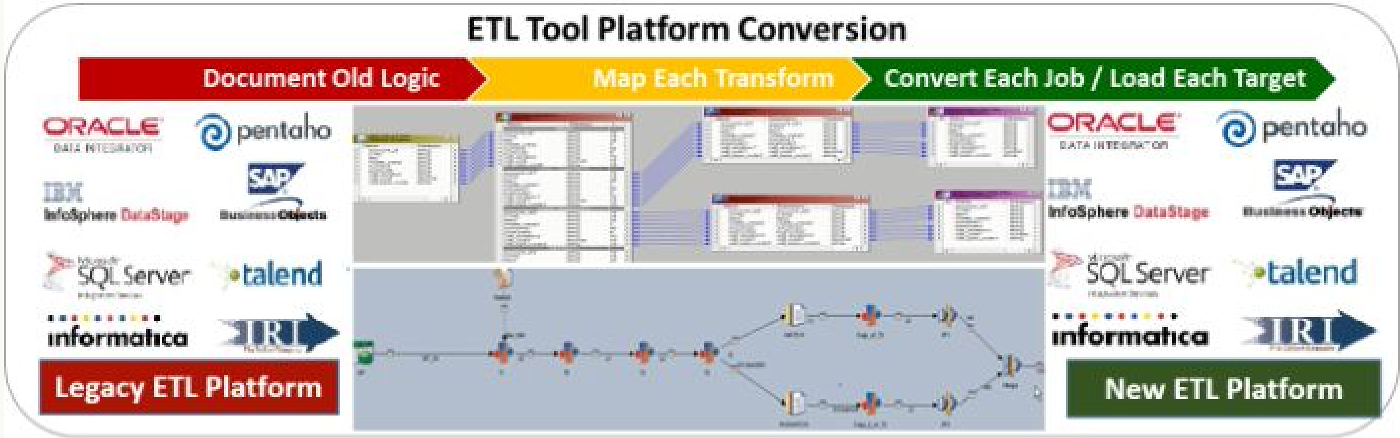
Start Field: master3.StartDate3

End Field: master3.EndDate3

Value Field	Start Field	End Field
master3.Cost1	master3.StartDate1	master3.EndDate1
master3.Cost2	master3.StartDate2	master3.EndDate2

Slowly Changing Dimensions

*With AnalytiX DS, ETL tool and SQL users can **convert** their existing data integration jobs to faster, simpler, and **far less expensive** Voracity workflows.*

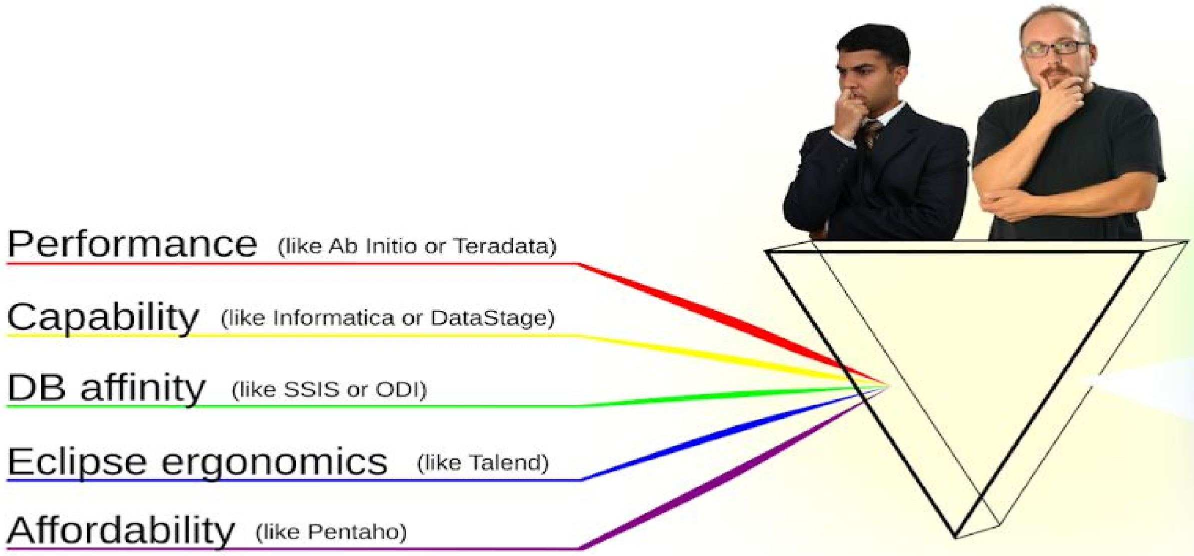


ETL Tool Platform Conversion

Document Old Logic → Map Each Transform → Convert Each Job / Load Each Target

Legacy ETL Platform

New ETL Platform



Performance (like Ab Initio or Teradata)

Capability (like Informatica or DataStage)

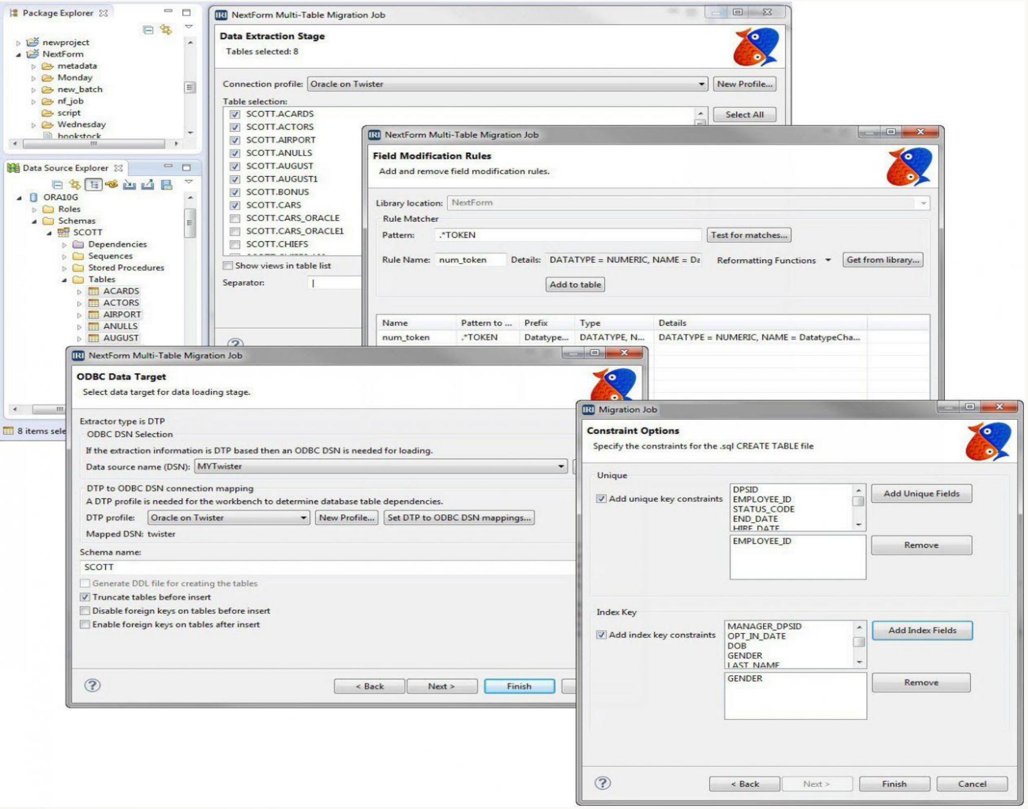
DB affinity (like SSIS or ODI)

Eclipse ergonomics (like Talend)

Affordability (like Pentaho)

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

•



NextForm Multi-Table Migration Job

Data Extraction Stage

Tables selected: 8

Connection profile: Oracle on Twister

Table selection:

☑ SCOTT.ACARDS
☑ SCOTT.ACTORS
☑ SCOTT.AIRPORT
☑ SCOTT.ANULLS
☑ SCOTT.AUGUST
☑ SCOTT.AUGUST1
☑ SCOTT.BONUS
☑ SCOTT.CARS
□ SCOTT.CARS_ORACLE
□ SCOTT.CARS_ORACLE1
□ SCOTT.CHIEFS

□ Show views in table list

Separator: |

NextForm Multi-Table Migration Job

Field Modification Rules

Add and remove field modification rules.

Library location: NextForm

Rule Matcher

Pattern: .*TOKEN

Rule Name: num_token Details: DATATYPE = NUMERIC, NAME = Da

Name	Pattern to ...	Prefix	Type	Details
num_token	.*TOKEN	Datatype...	DATATYPE, N...	DATATYPE = NUMERIC, NAME = DatatypeCha...

NextForm Multi-Table Migration Job

ODBC Data Target

Select data target for data loading stage.

Extractor type is DTP

ODBC DSN Selection

If the extraction information is DTP based then an ODBC DSN is needed for loading.

Data source name (DSN): MYTwister

DTP to ODBC DSN connection mapping

A DTP profile is needed for the workbench to determine database table dependencies.

DTP profile: Oracle on Twister

Mapped DSN: twister

Schema name: SCOTT

□ Generate DDL file for creating the tables
☑ Truncate tables before insert
□ Disable foreign keys on tables before insert
□ Enable foreign keys on tables after insert

Migration Job

Constraint Options

Specify the constraints for the .sql CREATE TABLE file

Unique

☑ Add unique key constraints

DPSID
EMPLOYEE_ID
STATUS_CODE
END_DATE
HIRE_DATE

EMPLOYEE_ID

Index Key

☑ Add index key constraints

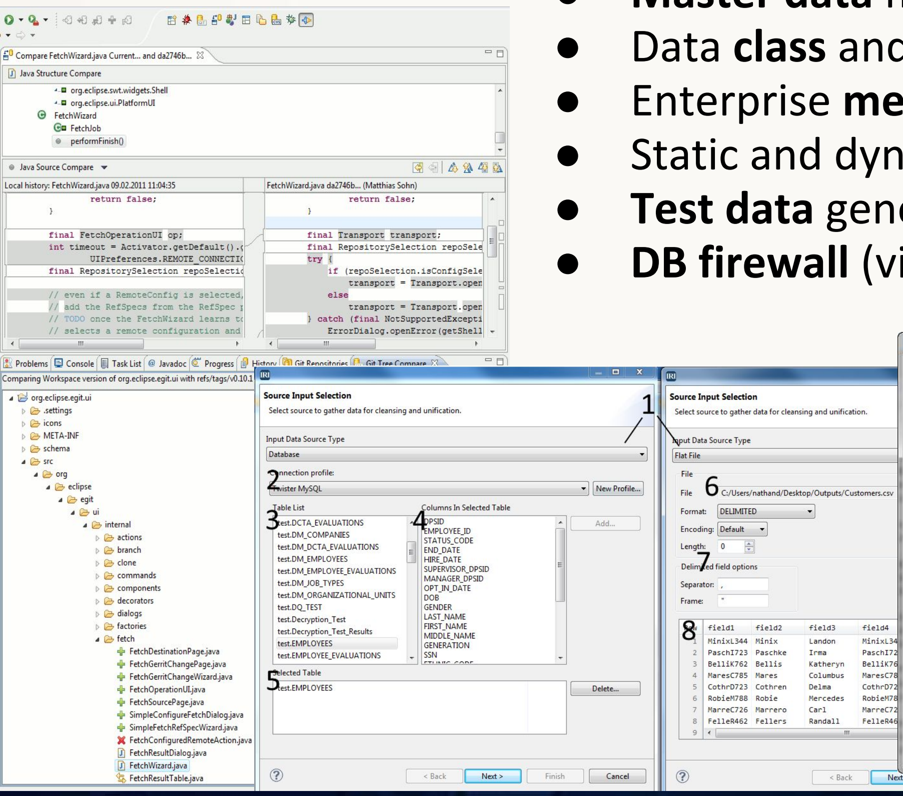
MANAGER_DPSID
OPT_IN_DATE
DOB
GENDER
LAST_NAME

GENDER

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

Voracity's data governance and information stewardship features include:

- **Master data** management
- Data **class** and **rule libraries**
- Enterprise **metadata management**
- Static and dynamic **data masking**
- **Test data** generation & management
- **DB firewall** (via IRI Chakra Max)



Compare FetchWizard.java Current... and da2746b...

Java Structure Compare

- org.eclipse.swt.widgets.Shell
- org.eclipse.ui.PlatformUI
- FetchWizard
- FetchJob
- performFinish()

Java Source Compare

Local history: FetchWizard.java 09.02.2011 11:04:35

```
            return false;
        }

        final FetchOperationUI op;
        int timeout = Activator.getDefault().g
                UIPreferences.REMOTE_CONNECTIC
        final RepositorySelection repoSelectic

        // even if a RemoteConfig is selected,
        // add the RefSpecs from the RefSpec p
        // TODO once the FetchWizard learns tc
        // selects a remote configuration and
```

FetchWizard.java da2746b... (Matthias Sohn)

```
            return false;
        }

        final Transport transport;
        final RepositorySelection repoSele
        try {
            if (repoSelection.isConfigSele
                transport = Transport.open
            else
                transport = Transport.open
        } catch (final NotSupportedExcepti
            ErrorDialog.openError(getShell
```

Comparing Workspace version of org.eclipse.egit.ui with refs/tags/v0.10.1

```
org.eclipse.egit.ui
  .settings
  icons
  META-INF
  schema
  src
    org
      eclipse
        egit
          ui
            internal
              actions
              branch
              clone
              commands
              components
              decorators
              dialogs
              factories
              fetch
                FetchDestinationPage.java
                FetchGerritChangePage.java
                FetchGerritChangeWizard.java
                FetchOperationUI.java
                FetchSourcePage.java
                SimpleConfigureFetchDialog.java
                SimpleFetchRefSpecWizard.java
                FetchConfiguredRemoteAction.java
                FetchResultDialog.java
                FetchWizard.java
                FetchResultTable.java
```

Source Input Selection

Select source to gather data for cleansing and unification.

1 Input Data Source Type: Database

2 Connection profile: Twister MySQL

3 Table List:
test.DCTA_EVALUATIONS
test.DM_COMPANIES
test.DM_DCTA_EVALUATIONS
test.DM_EMPLOYEES
test.DM_EMPLOYEE_EVALUATIONS
test.DM_JOB_TYPES
test.DM_ORGANIZATIONAL_UNITS
test.DQ_TEST
test.Decryption_Test
test.Decryption_Test_Results
test.EMPLOYEES
test.EMPLOYEE_EVALUATIONS

4 Columns In Selected Table:
DPSID
EMPLOYEE_ID
STATUS_CODE
END_DATE
HIRE_DATE
SUPERVISOR_DPSID
MANAGER_DPSID
OPT_IN_DATE
DOB
GENDER
LAST_NAME
FIRST_NAME
MIDDLE_NAME
GENERATION
SSN

5 Selected Table: test.EMPLOYEES

Source Input Selection

Select source to gather data for cleansing and unification.

Input Data Source Type: Flat File

6 File: C:/Users/nathand/Desktop/Outputs/Customers.csv

Format: DELIMITED

Encoding: Default

Length: 0

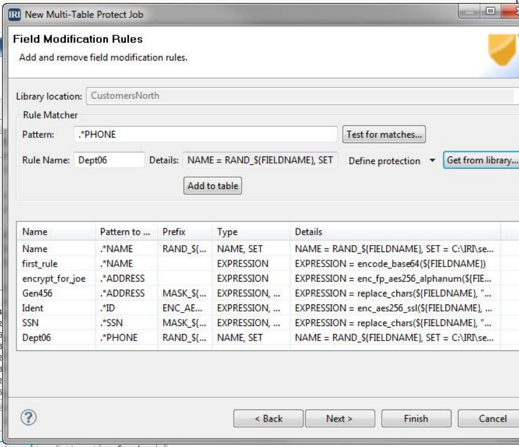
7 Delimited field options

Separator: ,

Frame: "

8

Row	field1	field2	field3	field4
1	MinixL344	Minix	Landon	MinixL34
2	PaschI723	Paschke	Irma	PaschI72
3	BelliK762	Bellis	Katheryn	BelliK76
4	MaresC785	Mares	Columbus	MaresC78
5	CothrD723	Cothren	Delma	CothrD72
6	RobieM788	Robie	Mercedes	RobieM78
7	MarreC726	Marrero	Carl	MarreC72
8	FelleR462	Fellers	Randall	FelleR46
9				



New Multi-Table Protect Job

Field Modification Rules

Add and remove field modification rules.

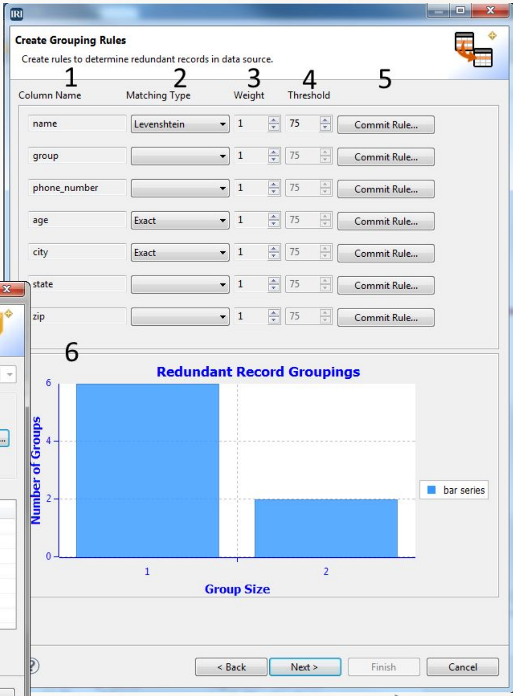
Library location: CustomersNorth

Rule Matcher

Pattern: .*PHONE

Rule Name: Dept06 Details: NAME = RAND_${FIELDNAME}, SET Define protection

Name	Pattern to ...	Prefix	Type	Details
Name	.*NAME	RAND_${...	NAME, SET	NAME = RAND_${FIELDNAME}, SET = C:\IRI\se...
first_rule	.*NAME		EXPRESSION	EXPRESSION = encode_base64(${FIELDNAME})
encrypt_for_joe	.*ADDRESS		EXPRESSION	EXPRESSION = enc_fp_aes256_alphanum(${FIE...
Gen456	.*ADDRESS	MASK_${...	EXPRESSION, ...	EXPRESSION = replace_chars(${FIELDNAME}, "...
Ident	.*ID	ENC_AE...	EXPRESSION, ...	EXPRESSION = enc_aes256_ssl(${FIELDNAME}, ...
SSN	.*SSN	MASK_${...	EXPRESSION, ...	EXPRESSION = replace_chars(${FIELDNAME}, "...
Dept06	.*PHONE	RAND_${...	NAME, SET	NAME = RAND_${FIELDNAME}, SET = C:\IRI\se...



Create Grouping Rules

Create rules to determine redundant records in data source.

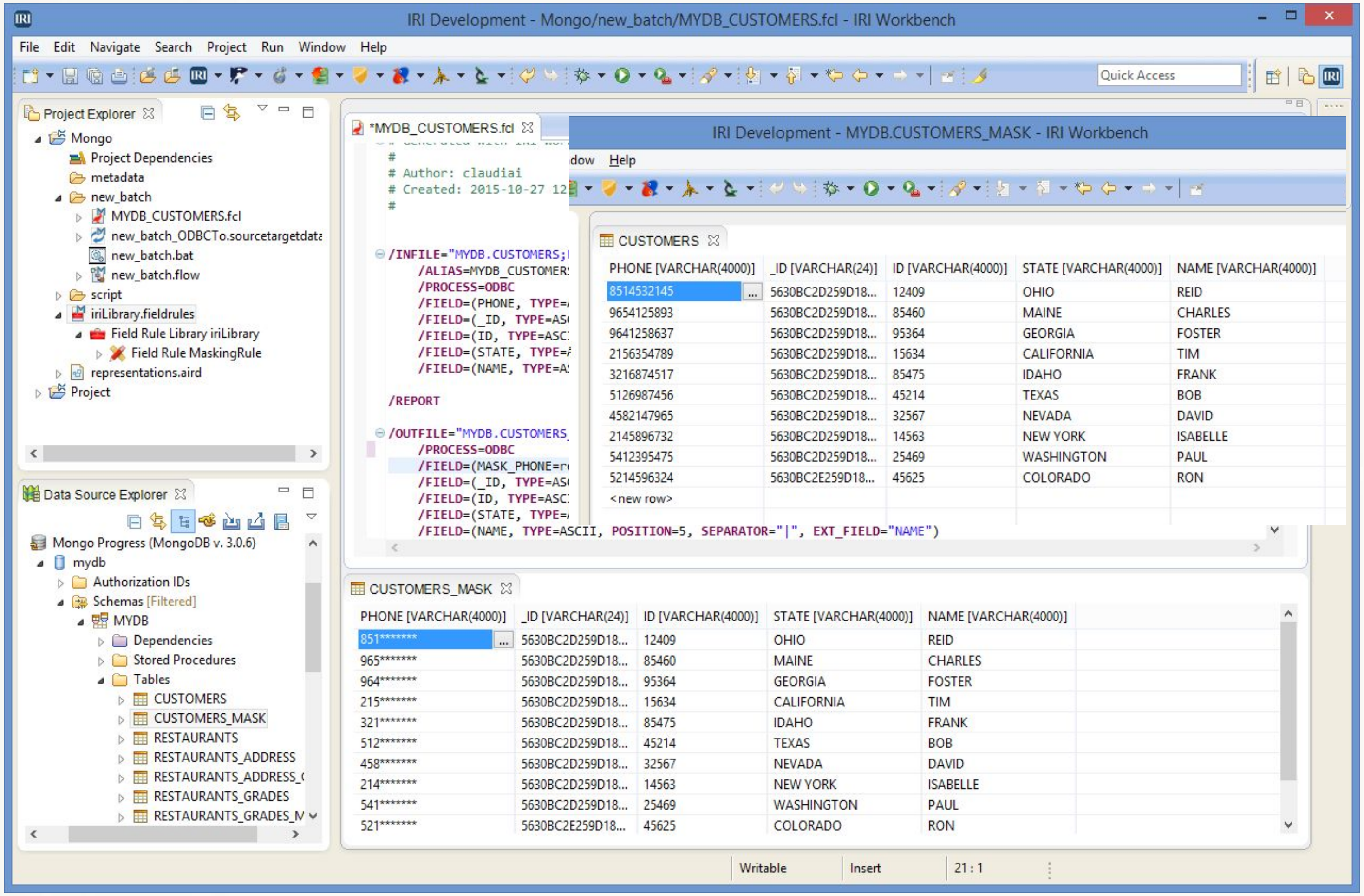
1 Column Name	2 Matching Type	3 Weight	4 Threshold	5
name	Levenshtein	1	75	Commit Rule...
group		1	75	Commit Rule...
phone_number		1	75	Commit Rule...
age	Exact	1	75	Commit Rule...
city	Exact	1	75	Commit Rule...
state		1	75	Commit Rule...
zip		1	75	Commit Rule...

6

Masking Features

- Connect and interact with **multiple sources** and targets, on-prem or cloud
- **Discover and classify** data in DB, flat-file, and dark-data (document) sources
- Mask **static or streaming** inputs, NoSQL DBs, and files in LUW, HDFS and S3
- Select from **12 masking categories** (e.g., encrypt, hash, pseudonymize, redact)
- **Address multiple** protections, targets and recipients all in one job, one I/O
- Apply consistent, cross-table masking rules for **referential integrity**
- Support **conditional security**, based on patterns, values, or ranges
- Specify target protections and formats in **Eclipse or portable** job scripts
- Integrate with **DB apps** via ODBC. Use .NET and Java SDK for dynamic masking
- Retain data **realism via FPE** and pseudonymization for testing, outsourcing
- **Mask during** big data ETL, migration, sub-setting, and BI/analytic jobs
- Log job and system runtime detail to XML audit files to **verify compliance**

MongoDB Masking



```
#
# Author: claudiai
# Created: 2015-10-27 12

/INFILE="MYDB.CUSTOMERS;
    /ALIAS=MYDB_CUSTOMER
    /PROCESS=ODBC
    /FIELD=(PHONE, TYPE=
    /FIELD=(_ID, TYPE=AS
    /FIELD=(ID, TYPE=ASC
    /FIELD=(STATE, TYPE=
    /FIELD=(NAME, TYPE=A

/REPORT

/OUTFILE="MYDB.CUSTOMERS_
    /PROCESS=ODBC
    /FIELD=(MASK_PHONE=r
    /FIELD=(_ID, TYPE=AS
    /FIELD=(ID, TYPE=ASC
    /FIELD=(STATE, TYPE=
    /FIELD=(NAME, TYPE=ASCII, POSITION=5, SEPARATOR="|", EXT_FIELD="NAME")
```

CUSTOMERS

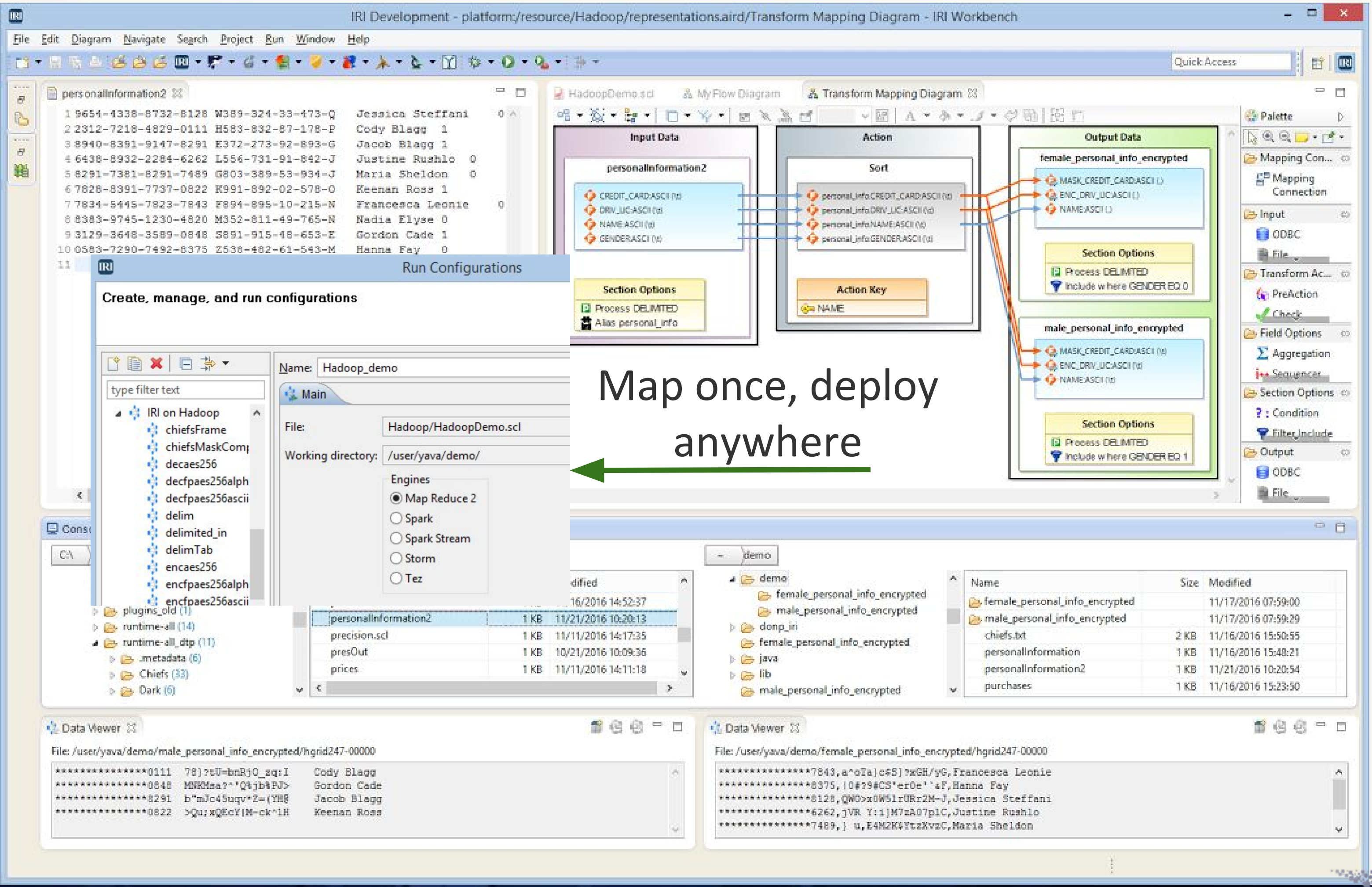
PHONE [VARCHAR(4000)]	_ID [VARCHAR(24)]	ID [VARCHAR(4000)]	STATE [VARCHAR(4000)]	NAME [VARCHAR(4000)]
8514532145	5630BC2D259D18...	12409	OHIO	REID
9654125893	5630BC2D259D18...	85460	MAINE	CHARLES
9641258637	5630BC2D259D18...	95364	GEORGIA	FOSTER
2156354789	5630BC2D259D18...	15634	CALIFORNIA	TIM
3216874517	5630BC2D259D18...	85475	IDAHO	FRANK
5126987456	5630BC2D259D18...	45214	TEXAS	BOB
4582147965	5630BC2D259D18...	32567	NEVADA	DAVID
2145896732	5630BC2D259D18...	14563	NEW YORK	ISABELLE
5412395475	5630BC2D259D18...	25469	WASHINGTON	PAUL
5214596324	5630BC2E259D18...	45625	COLORADO	RON
<new row>				

CUSTOMERS_MASK

PHONE [VARCHAR(4000)]	_ID [VARCHAR(24)]	ID [VARCHAR(4000)]	STATE [VARCHAR(4000)]	NAME [VARCHAR(4000)]
851*******	5630BC2D259D18...	12409	OHIO	REID
965*******	5630BC2D259D18...	85460	MAINE	CHARLES
964*******	5630BC2D259D18...	95364	GEORGIA	FOSTER
215*******	5630BC2D259D18...	15634	CALIFORNIA	TIM
321*******	5630BC2D259D18...	85475	IDAHO	FRANK
512*******	5630BC2D259D18...	45214	TEXAS	BOB
458*******	5630BC2D259D18...	32567	NEVADA	DAVID
214*******	5630BC2D259D18...	14563	NEW YORK	ISABELLE
541*******	5630BC2D259D18...	25469	WASHINGTON	PAUL
521*******	5630BC2E259D18...	45625	COLORADO	RON

DISCOVER INTEGRATE MIGRATE **GOVERN** ANALYZE

Masking in Hadoop



```
 1 9654-4338-8732-8128 W389-324-33-473-Q   Jessica Steffani     0
 2 2312-7218-4829-0111 H583-832-87-178-P   Cody Blagg  1
 3 8940-8391-9147-8291 E372-273-92-893-G   Jacob Blagg 1
 4 6438-8932-2284-6262 L556-731-91-842-J   Justine Rushlo  0
 5 8291-7381-8291-7489 G803-389-53-934-J   Maria Sheldon   0
 6 7828-8391-7737-0822 K991-892-02-578-O   Keenan Ross 1
 7 7834-5445-7823-7843 F894-895-10-215-N   Francesca Leonie     0
 8 8383-9745-1230-4820 M352-811-49-765-N   Nadia Elyse 0
 9 3129-3648-3589-0848 S891-915-48-653-E   Gordon Cade 1
10 0583-7290-7492-8375 Z538-482-61-543-M   Hanna Fay   0
11
```

Create, manage, and run configurations

Name: Hadoop_demo

File: Hadoop/HadoopDemo.scl

Working directory: /user/yava/demo/

Engines

● Map Reduce 2
○ Spark
○ Spark Stream
○ Storm
○ Tez

	Size	Modified
		11/16/2016 14:52:37
personalInformation2	1 KB	11/21/2016 10:20:13
precision.scl	1 KB	11/11/2016 14:17:35
presOut	1 KB	10/21/2016 10:09:36
prices	1 KB	11/11/2016 14:11:18

Name	Size	Modified
female_personal_info_encrypted		11/17/2016 07:59:00
male_personal_info_encrypted		11/17/2016 07:59:29
chiefs.txt	2 KB	11/16/2016 15:50:55
personalInformation	1 KB	11/16/2016 15:48:21
personalInformation2	1 KB	11/21/2016 10:20:54
purchases	1 KB	11/16/2016 15:23:50

File: /user/yava/demo/male_personal_info_encrypted/hgrid247-00000

```
***************0111  78}?tU=bnRjO_zq:I    Cody Blagg
***************0848  MNKMsa?^'Q%jb&PJ>    Gordon Cade
***************8291  b"mJc45uqv*Z=(YH§    Jacob Blagg
***************0822  >Qu;xQEcY|M-ck^1H    Keenan Ross
```

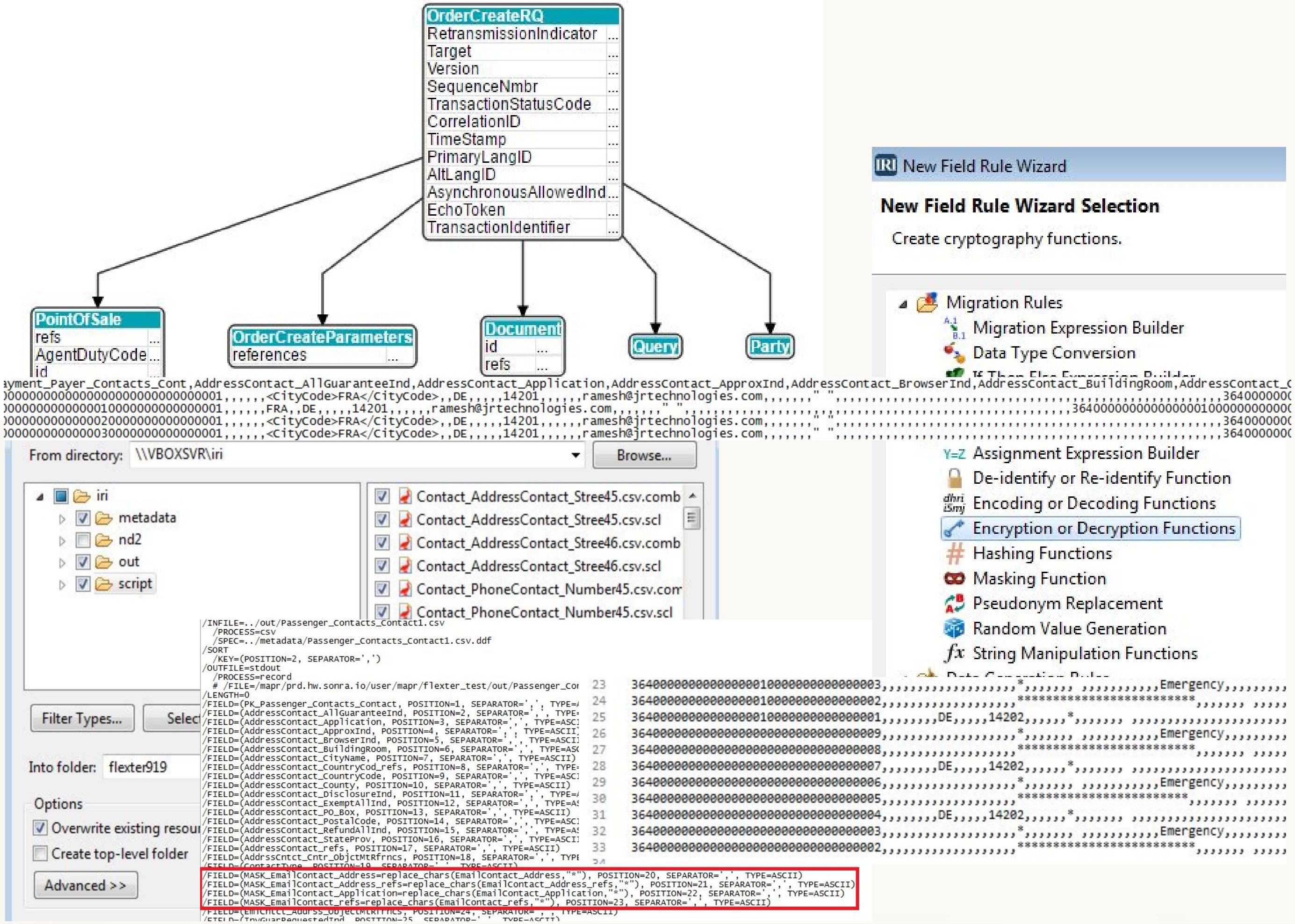
File: /user/yava/demo/female_personal_info_encrypted/hgrid247-00000

```
***************7843,a^oTa]c$S]?xGH/yG,Francesca Leonie
***************8375,|0#?9#CS'er0e'`&F,Hanna Fay
***************8128,QWO>x0W51rURr2M-J,Jessica Steffani
***************6262,jVR Y:1|M7zA07plC,Justine Rushlo
***************7489,} u,E4M2K$YtzXvzC,Maria Sheldon
```

Map once, deploy anywhere

IRI The CoSort Company

Masking Complex XML



```
ayment_Payer_Contacts_Cont,AddressContact_AllGuaranteeInd,AddressContact_Application,AddressContact_ApproxInd,AddressContact_BrowserInd,AddressContact_BuildingRoom,AddressContact_C
)0000000000000000000000000000001,,,,,,<CityCode>FRA</CityCode>,,DE,,,,,14201,,,,,,ramesh@jrtechnologies.com,,,,,,," ",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,364000000
)0000000000000001000000000000001,,,,,,FRA,,DE,,,,,14201,,,,,,ramesh@jrtechnologies.com,,,,,,," ",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,364000000000000000010000000000
)0000000000000002000000000000001,,,,,,<CityCode>FRA</CityCode>,,DE,,,,,14201,,,,,,ramesh@jrtechnologies.com,,,,,,," ",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,364000000
)0000000000000003000000000000001,,,,,,<CityCode>FRA</CityCode>,,DE,,,,,14201,,,,,,ramesh@jrtechnologies.com,,,,,,," ",,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,364000000
```

New Field Rule Wizard
New Field Rule Wizard Selection
Create cryptography functions.
Migration Rules
Migration Expression Builder
Data Type Conversion
If Then Else Expression Builder
Assignment Expression Builder
De-identify or Re-identify Function
Encoding or Decoding Functions
Encryption or Decryption Functions
Hashing Functions
Masking Function
Pseudonym Replacement
Random Value Generation
String Manipulation Functions
Data Generation Rules
From directory: \\VBOXSVR\iri
Browse...
iri
metadata
nd2
out
script
Filter Types...
Select
Into folder: flexter919
Options
Advanced >>

☑ Contact_AddressContact_Stree45.csv.comb
☑ Contact_AddressContact_Stree45.csv.scl
☑ Contact_AddressContact_Stree46.csv.comb
☑ Contact_AddressContact_Stree46.csv.scl
☑ Contact_PhoneContact_Number45.csv.com
☑ Contact_PhoneContact_Number45.csv.scl

☑ Overwrite existing resou
☐ Create top-level folder

```
/INFILE=../out/Passenger_Contacts_Contact1.csv
  /PROCESS=csv
  /SPEC=../metadata/Passenger_Contacts_Contact1.csv.ddf
/SORT
  /KEY=(POSITION=2, SEPARATOR=',')
/OUTFILE=stdout
  /PROCESS=record
  # /FILE=/mapr/prd.hw.sonra.io/user/mapr/flexter_test/out/Passenger_Co
/LENGTH=0
/FIELD=(PK_Passenger_Contacts_Contact, POSITION=1, SEPARATOR=',', TYPE=
/FIELD=(AddressContact_AllGuaranteeInd, POSITION=2, SEPARATOR=',', TYPE=
/FIELD=(AddressContact_Application, POSITION=3, SEPARATOR=',', TYPE=ASCI
/FIELD=(AddressContact_ApproxInd, POSITION=4, SEPARATOR=',', TYPE=ASCII)
/FIELD=(AddressContact_BrowserInd, POSITION=5, SEPARATOR=',', TYPE=ASCI
/FIELD=(AddressContact_BuildingRoom, POSITION=6, SEPARATOR=',', TYPE=AS
/FIELD=(AddressContact_CityName, POSITION=7, SEPARATOR=',', TYPE=ASCII)
/FIELD=(AddressContact_CountryCod_refs, POSITION=8, SEPARATOR=',', TYPE=
/FIELD=(AddressContact_CountryCode, POSITION=9, SEPARATOR=',', TYPE=ASCI
/FIELD=(AddressContact_County, POSITION=10, SEPARATOR=',', TYPE=ASCII)
/FIELD=(AddressContact_DisclosureInd, POSITION=11, SEPARATOR=',', TYPE=
/FIELD=(AddressContact_ExemptAllInd, POSITION=12, SEPARATOR=',', TYPE=AS
/FIELD=(AddressContact_PO_Box, POSITION=13, SEPARATOR=',', TYPE=ASCII)
/FIELD=(AddressContact_PostalCode, POSITION=14, SEPARATOR=',', TYPE=ASCI
/FIELD=(AddressContact_RefundAllInd, POSITION=15, SEPARATOR=',', TYPE=AS
/FIELD=(AddressContact_StateProv, POSITION=16, SEPARATOR=',', TYPE=ASCI
/FIELD=(AddressContact_refs, POSITION=17, SEPARATOR=',', TYPE=ASCII)
/FIELD=(AddrssCntct_Cntr_ObjctMtRfrncs, POSITION=18, SEPARATOR=',', TYPE
/FIELD=(ContactType, POSITION=19, SEPARATOR=',', TYPE=ASCII)
/FIELD=(MASK_EmailContact_Address=replace_chars(EmailContact_Address,"*"), POSITION=20, SEPARATOR=',', TYPE=ASCII)
/FIELD=(MASK_EmailContact_Address_refs=replace_chars(EmailContact_Address_refs,"*"), POSITION=21, SEPARATOR=',', TYPE=ASCII)
/FIELD=(MASK_EmailContact_Application=replace_chars(EmailContact_Application,"*"), POSITION=22, SEPARATOR=',', TYPE=ASCII)
/FIELD=(MASK_EmailContact_refs=replace_chars(EmailContact_refs,"*"), POSITION=23, SEPARATOR=',', TYPE=ASCII)
/FIELD=(EmlCntct_Addrss_ObjctMtRfrncs, POSITION=24, SEPARATOR=',', TYPE=ASCII)
/FIELD=(InvGuarRequestedInd, POSITION=25, SEPARATOR=',', TYPE=ASCII)
```

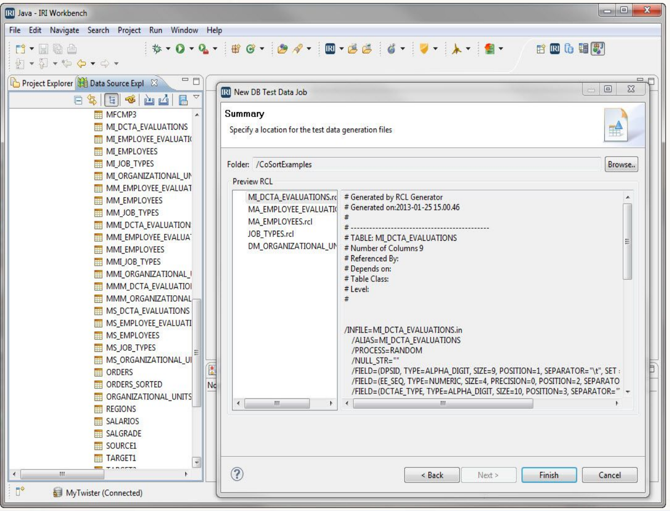
```
3640000000000000001000000000000003,,,,,,,,,,,,,,,,,,,,*,,,,,,, ,,,,,,,,,,Emergency,,,,,,,,,
3640000000000000001000000000000002,,,,,,,,,,,,,,,,,,,,************************,,,,,,, ,,,,,
3640000000000000001000000000000001,,,,,,,,DE,,,,,14202,,,,,,*,,,,,,, ,,,,,,,,,,,,,,,,,,,,,
3640000000000000000000000000000009,,,,,,,,,,,,,,,,,,,,*,,,,,,, ,,,,,,,,,,Emergency,,,,,,,,,
3640000000000000000000000000000008,,,,,,,,,,,,,,,,,,,,************************,,,,,,, ,,,,,
3640000000000000000000000000000007,,,,,,,,DE,,,,,14202,,,,,,*,,,,,,, ,,,,,,,,,,,,,,,,,,,,,
3640000000000000000000000000000006,,,,,,,,,,,,,,,,,,,,*,,,,,,, ,,,,,,,,,,Emergency,,,,,,,,,
3640000000000000000000000000000005,,,,,,,,,,,,,,,,,,,,************************,,,,,,, ,,,,,
3640000000000000000000000000000004,,,,,,,,DE,,,,,14202,,,,,,*,,,,,,, ,,,,,,,,,,,,,,,,,,,,,
3640000000000000000000000000000003,,,,,,,,,,,,,,,,,,,,*,,,,,,, ,,,,,,,,,,Emergency,,,,,,,,,
3640000000000000000000000000000002,,,,,,,,,,,,,,,,,,,,************************,,,,,,, ,,,,,
```

TDM Features

- 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

Synthetic Data for:

- Flat files
- EDW ETL tools
- RDB & NoSQL
- Data lakes
- Mainframe jobs
- SAP, Teradata
- Cloud/SaaS apps



Java - IRI Workbench

MFCMP3
MI_DCTA_EVALUATIONS
MI_EMPLOYEE_EVALUATI
MI_EMPLOYEES
MI_JOB_TYPES
MI_ORGANIZATIONAL_UN
MM_EMPLOYEE_EVALUAT
MM_EMPLOYEES
MM_JOB_TYPES
MMI_DCTA_EVALUATION
MMI_EMPLOYEE_EVALUA
MMI_EMPLOYEES
MMI_JOB_TYPES
MMI_ORGANIZATIONAL_I
MMM_DCTA_EVALUATIOI
MMM_ORGANIZATIONAL
MS_DCTA_EVALUATIONS
MS_EMPLOYEE_EVALUATI
MS_EMPLOYEES
MS_JOB_TYPES
MS_ORGANIZATIONAL_UI
ORDERS
ORDERS_SORTED
ORGANIZATIONAL_UNITS
REGIONS
SALARIOS
SALGRADE
SOURCE1
TARGET1

MyTwister (Connected)

New DB Test Data Job

Summary

Specify a location for the test data generation files

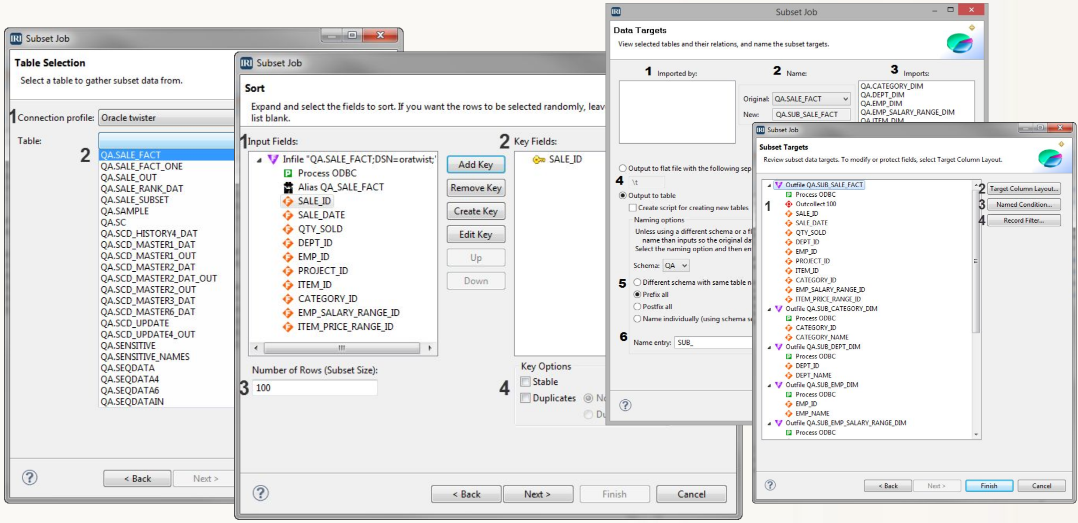
Folder: /CoSortExamples

Preview RCL

MI_DCTA_EVALUATIONS.rc
MA_EMPLOYEE_EVALUATI
MA_EMPLOYEES.rcl
JOB_TYPES.rcl
DM_ORGANIZATIONAL_UN

```
# Generated by RCL Generator
# Generated on:2013-01-25 15.00.46
#
# -----------------------------------------
# TABLE: MI_DCTA_EVALUATIONS
# Number of Columns 9
# Referenced By:
# Depends on:
# Table Class:
# Level:
#


/INFILE=MI_DCTA_EVALUATIONS.in
  /ALIAS=MI_DCTA_EVALUATIONS
  /PROCESS=RANDOM
  /NULL_STR=""
  /FIELD=(DPSID, TYPE=ALPHA_DIGIT, SIZE=9, POSITION=1, SEPARATOR="\t", SET :
  /FIELD=(EE_SEQ, TYPE=NUMERIC, SIZE=4, PRECISION=0, POSITION=2, SEPARATO
  /FIELD=(DCTAE_TYPE, TYPE=ALPHA_DIGIT, SIZE=10, POSITION=3, SEPARATOR="
```

TDM Features



Subset Job

Table Selection

Select a table to gather subset data from.

1 Connection profile: Oracle twister

Table:

2
QA.SALE_FACT
QA.SALE_FACT_ONE
QA.SALE_OUT
QA.SALE_RANK_DAT
QA.SALE_SUBSET
QA.SAMPLE
QA.SC
QA.SCD_HISTORY4_DAT
QA.SCD_MASTER1_DAT
QA.SCD_MASTER1_OUT
QA.SCD_MASTER2_DAT
QA.SCD_MASTER2_DAT_OUT
QA.SCD_MASTER2_OUT
QA.SCD_MASTER3_DAT
QA.SCD_MASTER6_DAT
QA.SCD_UPDATE
QA.SCD_UPDATE4_OUT
QA.SENSITIVE
QA.SENSITIVE_NAMES
QA.SEQDATA
QA.SEQDATA4
QA.SEQDATA6
QA.SEQDATAIN

Subset Job

Sort

Expand and select the fields to sort. If you want the rows to be selected randomly, leav list blank.

1 Input Fields:

Infile "QA.SALE_FACT;DSN=oratwist;'
Process ODBC
Alias QA_SALE_FACT
SALE_ID
SALE_DATE
QTY_SOLD
DEPT_ID
EMP_ID
PROJECT_ID
ITEM_ID
CATEGORY_ID
EMP_SALARY_RANGE_ID
ITEM_PRICE_RANGE_ID

2 Key Fields:

SALE_ID

Number of Rows (Subset Size):

3 100

4 Key Options
□ Stable
□ Duplicates ● No
○ Du

Subset Job

Data Targets

View selected tables and their relations, and name the subset targets.

1 Imported by:

2 Name:
Original: QA.SALE_FACT
New: QA.SUB_SALE_FACT

3 Imports:
QA.CATEGORY_DIM
QA.DEPT_DIM
QA.EMP_DIM
QA.EMP_SALARY_RANGE_DIM
QA.ITEM_DIM

○ Output to flat file with the following sep
4 \t
● Output to table
□ Create script for creating new tables

Naming options
Unless using a different schema or a fi name than inputs so the original da Select the naming option and then en

Schema: QA

5 ○ Different schema with same table n
● Prefix all
○ Postfix all
○ Name individually (using schema se

6 Name entry: SUB_

Subset Job

Subset Targets

Review subset data targets. To modify or protect fields, select Target Column Layout.

1
Outfile QA.SUB_SALE_FACT
Process ODBC
Outcollect 100
SALE_ID
SALE_DATE
QTY_SOLD
DEPT_ID
EMP_ID
PROJECT_ID
ITEM_ID
CATEGORY_ID
EMP_SALARY_RANGE_ID
ITEM_PRICE_RANGE_ID
Outfile QA.SUB_CATEGORY_DIM
Process ODBC
CATEGORY_ID
CATEGORY_NAME
Outfile QA.SUB_DEPT_DIM
Process ODBC
DEPT_ID
DEPT_NAME
Outfile QA.SUB_EMP_DIM
Process ODBC
EMP_ID
EMP_NAME
Outfile QA.SUB_EMP_SALARY_RANGE_DIM
Process ODBC

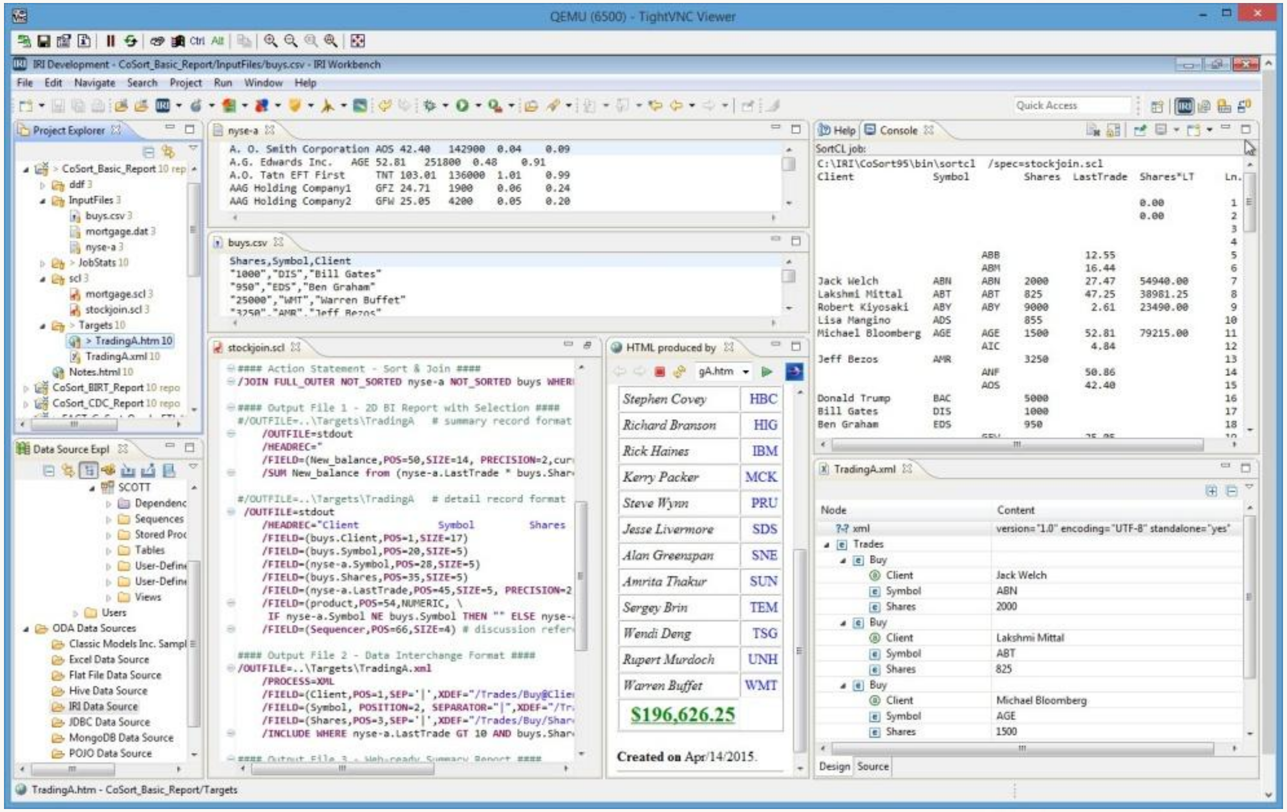
2 Target Column Layout...
3 Named Condition...
4 Record Filter...

Both test data generation/population and DB subsetting wizards with built-in data masking are included in Voracity to facilitate DB and EDW prototyping. Either way, the test data is realistic, referentially-correct, and privacy-law compliant.

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

Voracity Analytic Option 1: Embedded BI

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



nyse-a

```
A. O. Smith Corporation AOS 42.40   142900  0.04    0.09
A.G. Edwards Inc.   AGE 52.81   251800  0.48    0.91
A.O. Tatn EFT First     TNT 103.01  136000  1.01    0.99
AAG Holding Company1    GFZ 24.71   1900    0.06    0.24
AAG Holding Company2    GFW 25.05   4200    0.05    0.20
```

buys.csv

```
Shares,Symbol,Client
"1000","DIS","Bill Gates"
"950","EDS","Ben Graham"
"25000","WMT","Warren Buffet"
"3250","AMR","Jeff Bezos"
```

stockjoin.scl

```
#### Action Statement - Sort & Join ####
/JOIN FULL_OUTER NOT_SORTED nyse-a NOT_SORTED buys WHER

#### Output File 1 - 2D BI Report with Selection ####
#/OUTFILE=..\Targets\TradingA   # summary record format
    /OUTFILE=stdout
    /HEADREC="
    /FIELD=(New_balance,POS=50,SIZE=14, PRECISION=2,cur
    /SUM New_balance from (nyse-a.LastTrade * buys.Shar

#/OUTFILE=..\Targets\TradingA   # detail record format
 /OUTFILE=stdout
    /HEADREC="Client                Symbol          Shares
    /FIELD=(buys.Client,POS=1,SIZE=17)
    /FIELD=(buys.Symbol,POS=20,SIZE=5)
    /FIELD=(nyse-a.Symbol,POS=28,SIZE=5)
    /FIELD=(buys.Shares,POS=35,SIZE=5)
    /FIELD=(nyse-a.LastTrade,POS=45,SIZE=5, PRECISION=2
    /FIELD=(product,POS=54,NUMERIC, \
     IF nyse-a.Symbol NE buys.Symbol THEN "" ELSE nyse-
    /FIELD=(Sequencer,POS=66,SIZE=4) # discussion refer

#### Output File 2 - Data Interchange Format ####
/OUTFILE=..\Targets\TradingA.xml
    /PROCESS=XML
    /FIELD=(Client,POS=1,SEP='|',XDEF="/Trades/Buy@Clie
    /FIELD=(Symbol, POSITION=2, SEPARATOR="|",XDEF="/Tr
    /FIELD=(Shares,POS=3,SEP='|',XDEF="/Trades/Buy/Shar
    /INCLUDE WHERE nyse-a.LastTrade GT 10 AND buys.Shar

#### Output File 3 - Web-ready Summary Report ####
```

HTML produced by

Stephen Covey	HBC
Richard Branson	HIG
Rick Haines	IBM
Kerry Packer	MCK
Steve Wynn	PRU
Jesse Livermore	SDS
Alan Greenspan	SNE
Amrita Thakur	SUN
Sergey Brin	TEM
Wendi Deng	TSG
Rupert Murdoch	UNH
Warren Buffet	WMT

$196,626.25

Created on Apr/14/2015.

Console

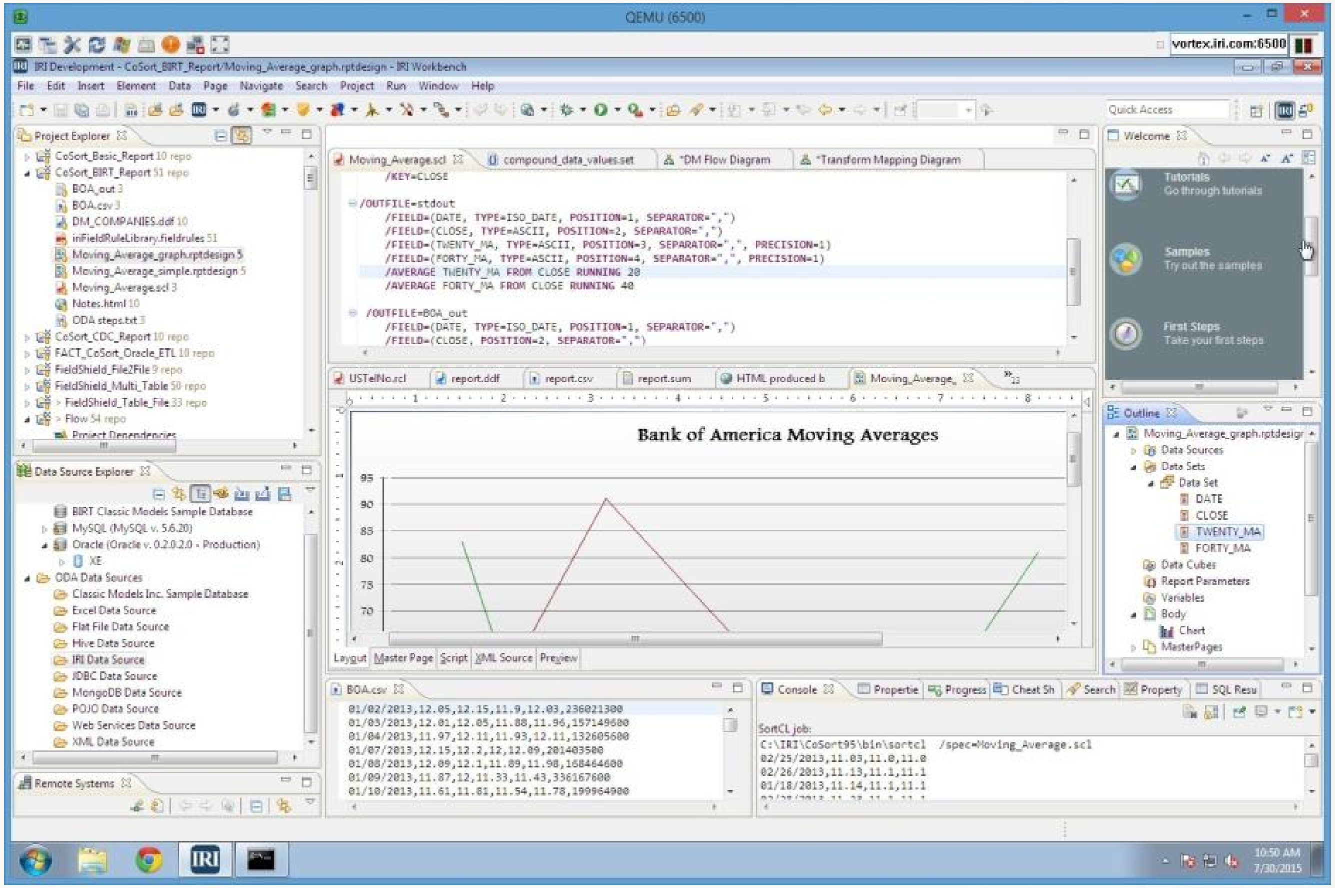
```
SortCL job:
C:\IRI\CoSort95\bin\sortcl  /spec=stockjoin.scl
Client              Symbol          Shares  LastTrade  Shares*LT      Ln.

                                                       0.00             1
                                                       0.00             2
                                                                        3
                                                                        4
                          ABB                12.55                      5
                          ABM                16.44                      6
Jack Welch          ABN   ABN       2000     27.47     54940.00         7
Lakshmi Mittal      ABT   ABT       825      47.25     38981.25         8
Robert Kiyosaki     ABY   ABY       9000      2.61     23490.00         9
Lisa Mangino        ADS             855                                10
Michael Bloomberg   AGE   AGE       1500     52.81     79215.00        11
                          AIC                 4.84                     12
Jeff Bezos          AMR             3250                               13
                          ANF                50.86                     14
                          AOS                42.40                     15
Donald Trump        BAC             5000                               16
Bill Gates          DIS             1000                               17
Ben Graham          EDS             950                                18
```

TradingA.xml

Node	Content
?-? xml	version="1.0" encoding="UTF-8" standalone="yes"
Trades	
Buy	
Client	Jack Welch
Symbol	ABN
Shares	2000
Buy	
Client	Lakshmi Mittal
Symbol	ABT
Shares	825
Buy	
Client	Michael Bloomberg
Symbol	AGE
Shares	1500

Voracity Analytic Option 2: BIRT Integration

Prepare and present data simultaneously from an “IRI Data Source” in BIRT


QEMU (6500)
vortex.iri.com:6500
IRI Development - CoSort_BIRT_Report/Moving_Average_graph.rptdesign - IRI Workbench
File Edit Insert Element Data Page Navigate Search Project Run Window Help
Project Explorer
Data Source Explorer
Remote Systems
Moving_Average.scl
compound_data_values.set
*DM Flow Diagram
*Transform Mapping Diagram
Bank of America Moving Averages
95
90
85
80
75
70
Layout Master Page Script XML Source Preview
Welcome
Outline
BOA.csv
Console
SortCL job:
10:50 AM
7/30/2015

```
    /KEY=CLOSE

/OUTFILE=stdout
    /FIELD=(DATE, TYPE=ISO_DATE, POSITION=1, SEPARATOR=",")
    /FIELD=(CLOSE, TYPE=ASCII, POSITION=2, SEPARATOR=",")
    /FIELD=(TWENTY_MA, TYPE=ASCII, POSITION=3, SEPARATOR=",", PRECISION=1)
    /FIELD=(FORTY_MA, TYPE=ASCII, POSITION=4, SEPARATOR=",", PRECISION=1)
    /AVERAGE TWENTY_MA FROM CLOSE RUNNING 20
    /AVERAGE FORTY_MA FROM CLOSE RUNNING 40

 /OUTFILE=BOA_out
    /FIELD=(DATE, TYPE=ISO_DATE, POSITION=1, SEPARATOR=",")
    /FIELD=(CLOSE, POSITION=2, SEPARATOR=",")
```

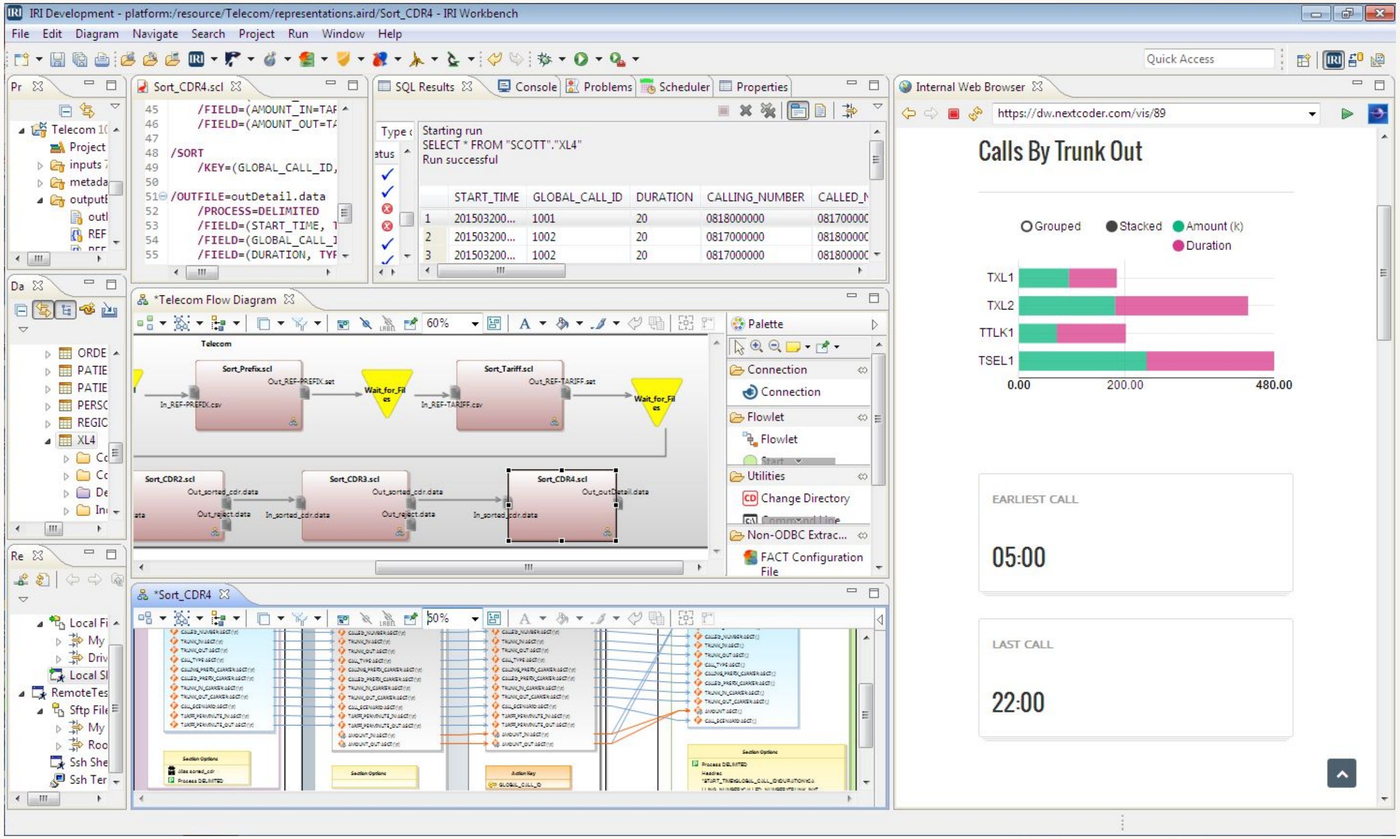
```
01/02/2013,12.05,12.15,11.9,12.03,236021300
01/03/2013,12.01,12.05,11.88,11.96,157149600
01/04/2013,11.97,12.11,11.93,12.11,132605600
01/07/2013,12.15,12.2,12,12.09,201403500
01/08/2013,12.09,12.1,11.89,11.98,168464600
01/09/2013,11.87,12,11.33,11.43,336167600
01/10/2013,11.61,11.81,11.54,11.78,199964900
```

```
C:\IRI\CoSort95\bin\sortcl  /spec=Moving_Average.scl
02/25/2013,11.03,11.0,11.0
02/26/2013,11.13,11.1,11.1
01/18/2013,11.14,11.1,11.1
```

Voracity Analytic Option 3:

Cloud Dashboard

Leverage drill-down, browser-based dashboard applications like this one from NextCoder



```
    /FIELD=(AMOUNT_IN=TAF
    /FIELD=(AMOUNT_OUT=TA

/SORT
    /KEY=(GLOBAL_CALL_ID,

/OUTFILE=outDetail.data
    /PROCESS=DELIMITED
    /FIELD=(START_TIME, T
    /FIELD=(GLOBAL_CALL_I
    /FIELD=(DURATION, TYF
```

```
Starting run
SELECT * FROM "SCOTT"."XL4"
Run successful
```

	START_TIME	GLOBAL_CALL_ID	DURATION	CALLING_NUMBER	CALLED_N
1	201503200...	1001	20	0818000000	081700000
2	201503200...	1002	20	0817000000	081800000
3	201503200...	1002	20	0817000000	081800000

https://dw.nextcoder.com/vis/89

Calls By Trunk Out

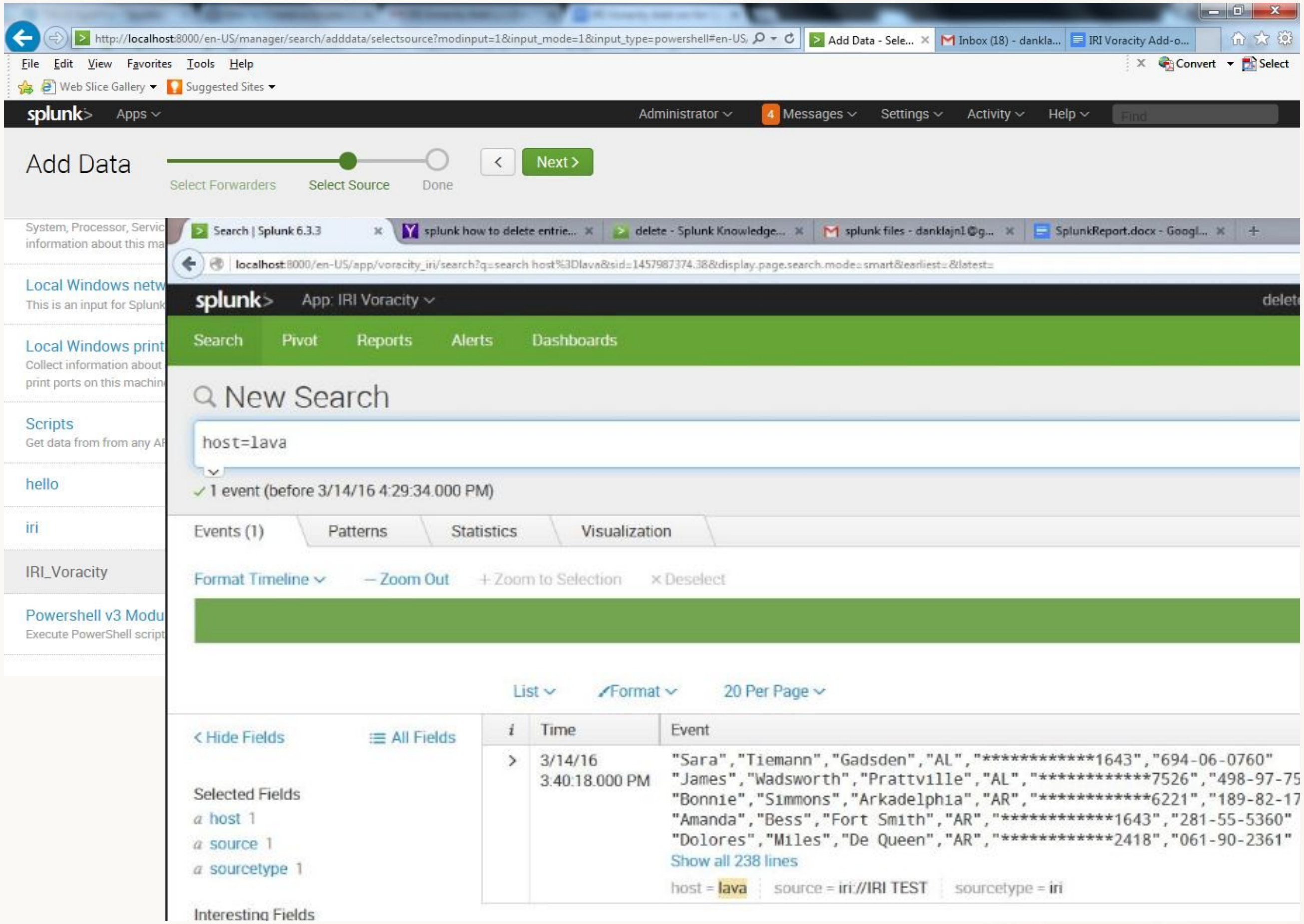
EARLIEST CALL

05:00

LAST CALL

22:00

Voracity Analytic Option 4: Splunk Add-On



http://localhost:8000/en-US/manager/search/adddata/selectsource?modinput=1&input_mode=1&input_type=powershell#en-US
Add Data - Sele...
Inbox (18) - dankla...
IRI Voracity Add-o...
File Edit View Favorites Tools Help
Convert Select
Web Slice Gallery Suggested Sites
splunk> Apps
Administrator 4 Messages Settings Activity Help Find
Add Data
Select Forwarders Select Source Done
< Next >
System, Processor, Servic information about this ma
Local Windows netw This is an input for Splunk
Local Windows print Collect information about print ports on this machin
Scripts Get data from from any AP
hello
iri
IRI_Voracity
Powershell v3 Modu Execute PowerShell script
Search | Splunk 6.3.3
splunk how to delete entrie...
delete - Splunk Knowledge...
splunk files - danklajn1@g...
SplunkReport.docx - Googl...
localhost:8000/en-US/app/voracity_iri/search?q=search host%3Dlava&sid=1457987374.38&display.page.search.mode=smart&earliest=&latest=
splunk> App: IRI Voracity
delete
Search Pivot Reports Alerts Dashboards
New Search

```
host=lava
```

✓ 1 event (before 3/14/16 4:29:34.000 PM)

Events (1) Patterns Statistics Visualization
Format Timeline — Zoom Out + Zoom to Selection × Deselect
List Format 20 Per Page
< Hide Fields All Fields
Selected Fields
a host 1
a source 1
a sourcetype 1
Interesting Fields

i	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","Simmons","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 host = lava source = iri://IRI TEST sourcetype = iri

Prepare data you need to index ad hoc, with a Voracity job launched from Splunk

Voracity Analytic Option 5: Data Blending

Prepare CSV, XML or table subsets to reduce time-to-display 2-20X, along with data quality, privacy, and storage

SAP Business Objects

IBM COGNOS

MicroStrategy

ORACLE BUSINESS INTELLIGENCE

Power BI

iDashboards

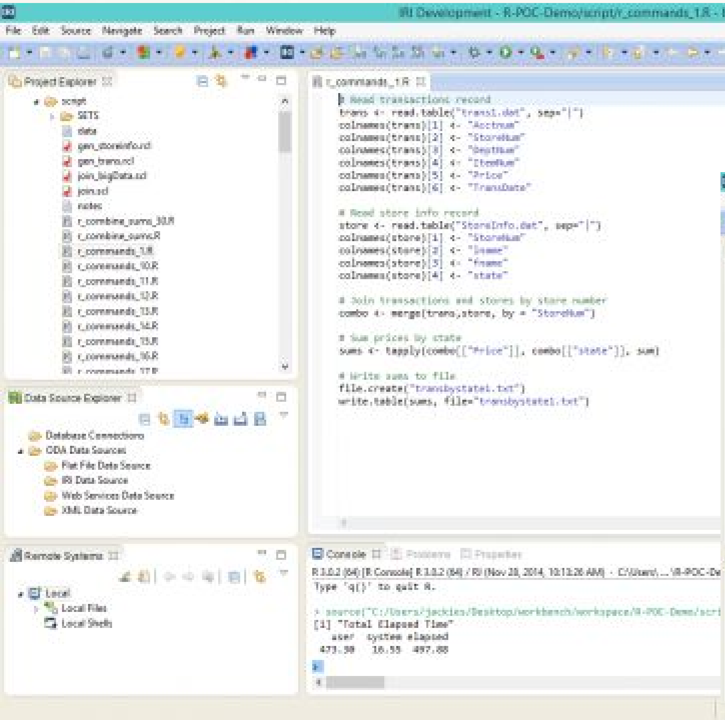
QlikView

TIBCO Spotfire

tableau SOFTWARE

R

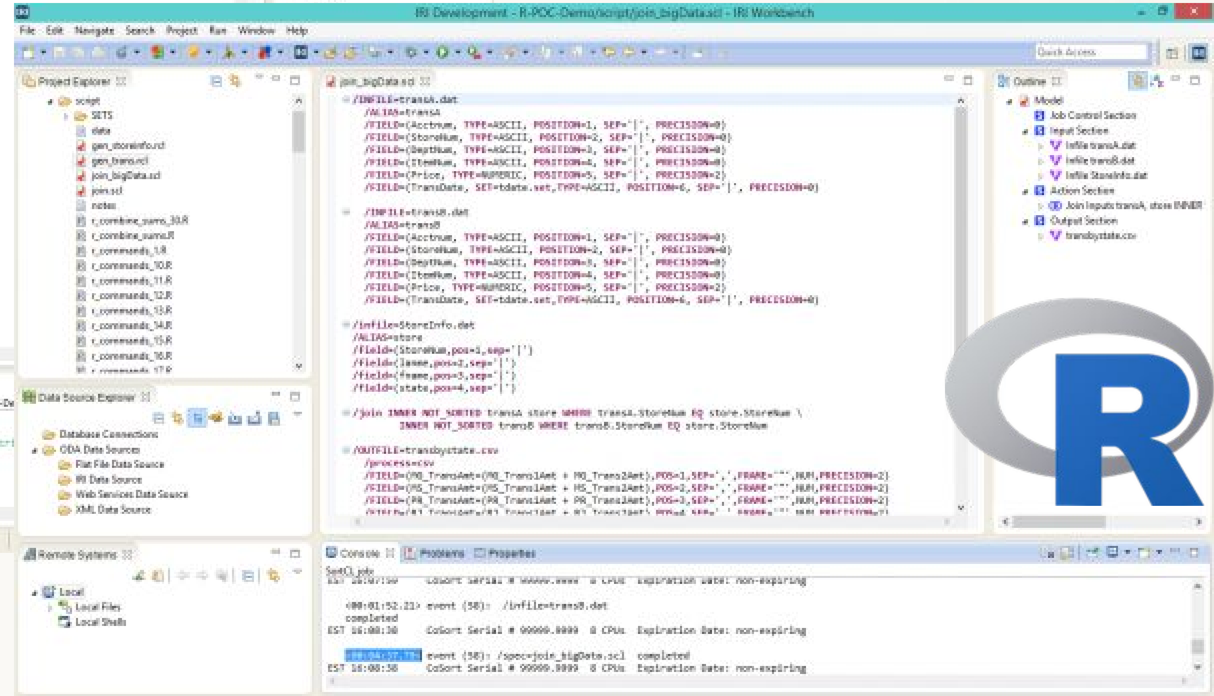
Data Preparation for R ...



```
# Read transactions record
trans <- read.table("trans1.dat", sep="|")
colnames(trans)[1] <- "Acctnum"
colnames(trans)[2] <- "StoreNum"
colnames(trans)[3] <- "DeptNum"
colnames(trans)[4] <- "ItemNum"
colnames(trans)[5] <- "Price"
colnames(trans)[6] <- "TransDate"

# Read store info record
store <- read.table("StoreInfo.dat", sep="|")
colnames(store)[1] <- "StoreNum"
colnames(store)[2] <- "lname"
colnames(store)[3] <- "fname"
colnames(store)[4] <- "state"

# Join transactions and stores by store number
combo <- merge(trans,store, by = "StoreNum")

# Sum prices by state
sums <- tapply(combo[["Price"]], combo[["state"]], sum)

# Write sums to file
file.create("transbystate1.txt")
write.table(sums, file="transbystate1.txt")
```

```
R 3.0.2 (64) [R Console] R 3.0.2 (64) / RJ (Nov 20, 2014, 10:13:26 AM) - C:\Users\...\R-POC-De
Type 'q()' to quit R.

> source("C:/Users/jackies/Desktop/workbench/workspace/R-POC-Demo/scri
[1] "Total Elapsed Time"
   user  system elapsed
 473.30   16.55  497.88
>
```

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

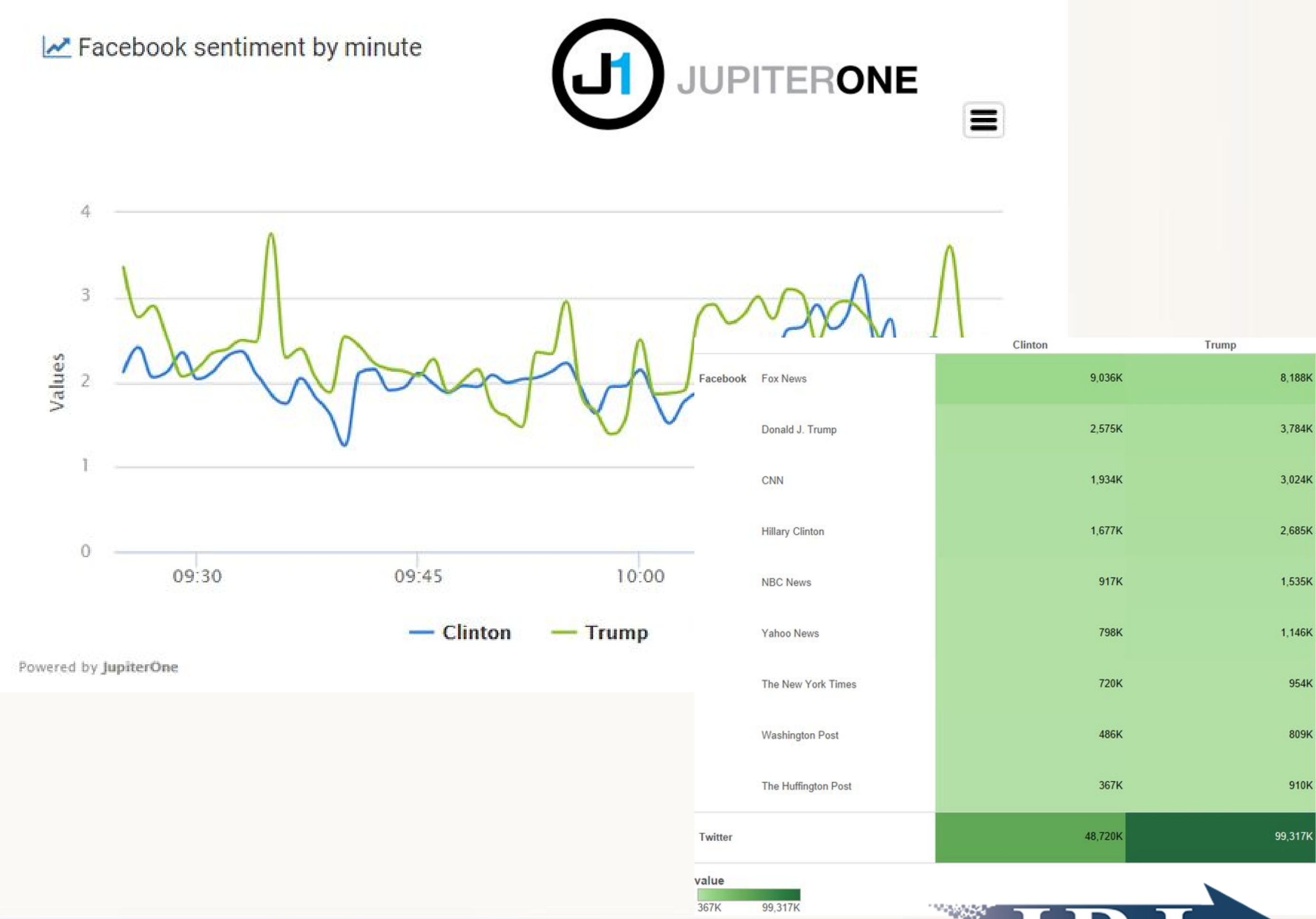


```
/INFILE=transA.dat
   /ALIAS=transA
   /FIELD=(Acctnum, TYPE=ASCII, POSITION=1, SEP='|', PRECISION=0)
   /FIELD=(StoreNum, TYPE=ASCII, POSITION=2, SEP='|', PRECISION=0)
   /FIELD=(DeptNum, TYPE=ASCII, POSITION=3, SEP='|', PRECISION=0)
   /FIELD=(ItemNum, TYPE=ASCII, POSITION=4, SEP='|', PRECISION=0)
   /FIELD=(Price, TYPE=NUMERIC, POSITION=5, SEP='|', PRECISION=2)
   /FIELD=(TransDate, SET=tdate.set,TYPE=ASCII, POSITION=6, SEP='|', PRECISION=0)

   /INFILE=transB.dat
   /ALIAS=transB
   /FIELD=(Acctnum, TYPE=ASCII, POSITION=1, SEP='|', PRECISION=0)
   /FIELD=(StoreNum, TYPE=ASCII, POSITION=2, SEP='|', PRECISION=0)
   /FIELD=(DeptNum, TYPE=ASCII, POSITION=3, SEP='|', PRECISION=0)
   /FIELD=(ItemNum, TYPE=ASCII, POSITION=4, SEP='|', PRECISION=0)
   /FIELD=(Price, TYPE=NUMERIC, POSITION=5, SEP='|', PRECISION=2)
   /FIELD=(TransDate, SET=tdate.set,TYPE=ASCII, POSITION=6, SEP='|', PRECISION=0)

/infile=StoreInfo.dat
/ALIAS=store
/Field=(StoreNum,pos=1,sep='|')
/Field=(lname,pos=2,sep='|')
/field=(fname,pos=3,sep='|')
/field=(state,pos=4,sep='|')

/join INNER NOT_SORTED transA store WHERE transA.StoreNum EQ store.StoreNum \
        INNER NOT_SORTED transB WHERE transB.StoreNum EQ store.StoreNum

/OUTFILE=transbystate.csv
   /process=csv
   /FIELD=(MO_TransAmt=(MO_Trans1Amt + MO_Trans2Amt),POS=1,SEP=',',FRAME='"',NUM,PRECISION=2)
   /FIELD=(MS_TransAmt=(MS_Trans1Amt + MS_Trans2Amt),POS=2,SEP=',',FRAME='"',NUM,PRECISION=2)
   /FIELD=(PR_TransAmt=(PR_Trans1Amt + PR_Trans2Amt),POS=3,SEP=',',FRAME='"',NUM,PRECISION=2)
```

```
SortCL job
EST 16:07:50      CoSort Serial # 99999.9999  8 CPUs  Expiration Date: non-expiring

   <00:01:52.21> event (58):  /infile=transB.dat
   completed
EST 16:08:30      CoSort Serial # 99999.9999  8 CPUs  Expiration Date: non-expiring

   <00:04:57.73> event (58): /spec=join_bigData.scl  completed
EST 16:08:30      CoSort Serial # 99999.9999  8 CPUs  Expiration Date: non-expiring
```

… 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 6: Big SM Streams



		Clinton	Trump
Facebook	Fox News	9,036K	8,188K
	Donald J. Trump	2,575K	3,784K
	CNN	1,934K	3,024K
	Hillary Clinton	1,677K	2,685K
	NBC News	917K	1,535K
	Yahoo News	798K	1,146K
	The New York Times	720K	954K
	Washington Post	486K	809K
	The Huffington Post	367K	910K
Twitter		48,720K	99,317K

Leverage advanced text and social media analytic engines with NLP and Kafka support

DISCOVER INTEGRATE MIGRATE GOVERN ANALYZE

Data Curation

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. Select, fuzzy-search, and merge reference data into master tables and values.

Process & Provide

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

Protect & Audit

De-ID data at the field level as you acquire, transform, report, or franchise. Encrypt, hash, pseudonymize, redact, tokenize, etc.

Express & Predict

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

Convert & Replicate

Migrate legacy databases, or files and data types -- or specify new target record layouts -- in copies, or subsets, of data in any format or schema.

Publish & Share

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

Voracity Uses

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, weather, 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 Uses

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.

Voracity Uses

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 patient information from family, provider, demographic, diagnostic and treatment data silos.

Voracity Uses

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 sensor 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 or BIRT-connected Integeo geospatial reports to warn drivers.

Optimize Fleet Performance

Use IoT analytic 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.

Voracity Uses

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.

Throttling & Enforcement

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

Voracity Uses

Reliance Communications (RC) is broadband and telco company in india with 110M subscribers. To meet daily SLAs in billing and analytics for wireless (mobile) and global (landline) segments, RC must process and report on hundreds of millions of call detail records (CDRs) every day.

RC uses 64-bit Solaris servers and Oracle. The CDRs come from binary switch data mediated into flat files that the CoSort engine in Voracity transforms *before* DataStage ETL & BOBJ reports.

"Prior pilots failed from slow and inaccurate results, and SLAs were missed as call volume grew. After Voracity jobs transformed flat files in the 60GB range, the processing bottleneck disappeared, and our analytic results were always accurate."

RELIANCE Communications
Anil Dhirubhai Ambani Group

Voracity Uses

DataBase Technologies (DBT) in Parsippany, NJ builds and maintains VLDB CRMs for ADP, Verizon, Merrill Lynch, Seagrams, and Universal Studios.

DBT integrates 350M transaction records per day, joining them to files up to 100M rows each, and accumulating the data over time for analysis. Their first 350GB dataset took over two days to load, so it had to be pre-sorted.

"It’s fun to watch the system performance monitor and see all those processors working in the high 90 percentages and the disks utilizing the fast data rates you pay for."

Voracity filter, sort, and join operations, were 10x faster than those in MS SQL Server ….
9.5 minutes versus 98 @350GB.

DATABASE TECHNOLOGIES

Learn and Share

IRI.com IRI blog

IRI Voracity Total Data Management

IRI Voracity Data Management Group on LinkedIn