

- 資料整合與分析
- 資料與 ETL 工具移植
- 資料探索與管理
- 建置於 CoSort®, Hadoop™ & Eclipse™

巨量資料處理



產品簡介

IRI

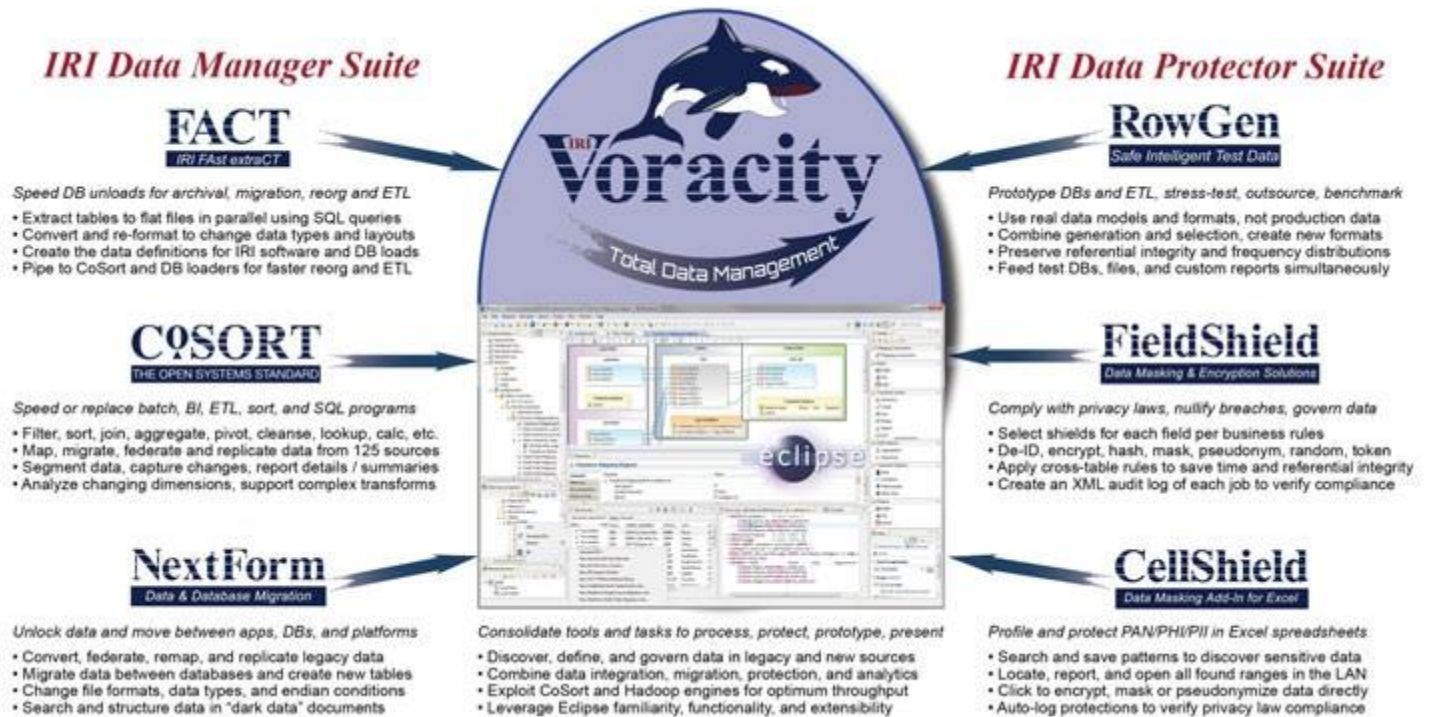
The CoSort Company

什麼是 Voracity?

IRI Voracity® 是一套功能強大的企業級資料管理以及操作平台,

針對下列人員	工作內容
資料倉儲建構師	設計、預覽、以及運行 ETL 工作或者加速與移植其他工具
資料科學家	探索、定義、與預備各種性質不同的資料以進行分析
資料管家	淨化、統一、遮罩、測試、追蹤、與分享資料
資料庫管理師	加速資料卸載、負載、查詢與重整或者變動資料的分析報表
資訊長	從傳統 ETL 工具進行自動移植以及廠商整合

Voracity 平台，結合與加速: 資料探索、整合、移植、管理、分析以及展集(註 1) ... 管理企業資訊的生命週期. Voracity 是基於市場驗證的資料處理效能 IRI CoSort 引擎或者多個 Hadoop 引擎所開發. Voracity 的前端介面為 IRI Workbench GUI，一個基於單一 Eclipse 整體擴展性能的插入模組.



為什麼需要 Voracity?

速度需求

Voracity 將不同的種類的巨量資料基於速度及正確性要求的挑戰下轉換成可被操作的有用訊息。它靈活運用了市場效能驗證的 CoSort 引擎 – 或者是無縫整合了 Hadoop 的 MapReduce、Spark、Storm 與 Tez 等四種 Hadoop 平台選項 – 將資料來源在本地端或者在雲端的資料進行搬運、操作、遮罩與資料淨化並且可整合已結構化、半結構化與未結構化的資料，CoSort 優越的演算與工作整合運算效能可 6 倍快速於 megavendor 的 ETL 工具套件。

節省成本

與傳統 ETL 套件不同，Voracity 可有效降低運作成本、複雜度與風險曲線並且無需關聯 Apache 專案可與 BI 分離，資料聯邦(註 2) 品質，遮罩以及 MDM 平台. 所有工作的設計與運行都是在單一虛擬管理平台上一並且有時候僅在一個輸出入路徑中通過 --Voracity 提供商業所需的結果以及 IT 所需要的整合性能，提供企業財務長一種不以 CPU core 數, 使用者人數, 作業數量或資料表行數計價的可負擔的採購模式。

簡單易用

Voracity 在 Eclipse 當中對於作業程序設計與佈局，在操作上不但直覺且人性化—加上可追蹤，自我解釋說明資料 – 易於操作與維護. 與 AnalytiX DS 對於傳統 ETL 作業的自動化轉換程序相容，與 BIRT 的整合，嵌入式商業智慧，data franchising 資料處理作業以及自助服務, shard-powered 巨量資料分析 (未定), 具備這些功能意味著資訊科技與商業智慧使用者能夠經由一套經過市場證明的資料管理基礎建設的協助進而加速對數位商業的洞察力。

資料探索

Voracity 的資料探索大部分是以樣式與模糊邏輯搜尋對資料進行識別及分類，統計報表與元數據建立開始進行，使用者友好介面精靈中有：

- 資料庫分析與建立實體關聯圖
- 資料萃取中搜尋暗數據(Dark data)
- 不具關聯性的資料分析報表, 搜尋與分類
- 結構性資料的元數據探索與定義



資料整合

Voracity 能快速地移動與操作不同種類的資料，它使用 IRI 產品與 Analytix DS 所共享的簡單元數據。藉由眾多優越的技術加持 Voracity 能加速或取代傳統的 ETL 工具：

- 巨量資料倉儲 ETL 或 ELT 作業
- 資料品質與標準化
- 超級巨量資料庫卸載, 外部資料轉型, 結構重整或載入
- 漸變資料維度(SCD)與異動資料擷取(CDC)

資料移植

Voracity 提供資料移植與重新格式化選項以轉換與重用包含儲存在傳統與現代資料庫的資料索引檔、循序檔以及暗數據文件：

- 重匹配資料型別, 紀錄佈局, 檔案格式與位元組順序
- 在資料庫之間移動欄位值, 佈局與關連 (限制)
- 從一個或多個資料來源複製資料到一個或多個資料表, 檔案或報表目標
- 藉由 ad hoc view 的方式從不同的資料源產生聯合或虛擬化資料



資料管理

Voracity 支援資料管理與訊息管家，專案或者將那些活動加到它的其他資料管理活動中，包含下列的管理能力：

- 主數據管理 - unify 及 bucket 資料表與檔案值
- 企業元數據管理 - 追蹤, 分享以及元數據資產加密
- 靜態與動態資料遮罩 - 符合資料隱私條例
- 測試資料合成 - 建立 DB/ETL 運作與壓力測試的應用程式

分析

在同樣的 Eclipse IDE 介面中，Voracity 從作為進行描述、分析、預測或常規的分析的公開與私有資料中準備、呈現與提供包羅萬象的客製化洞察：

- 嵌入式商業智慧- 以 calc/stat(計算/統計) 功能定製細節與總結報告
- BIRT 整合 - 在報表時段(reporting time)將結果與元數據至 BIRT
- 資料預備- 導入結果以加速可視化的時間
- 自我服務商業智慧 - shard-powered 與立即可視化分析

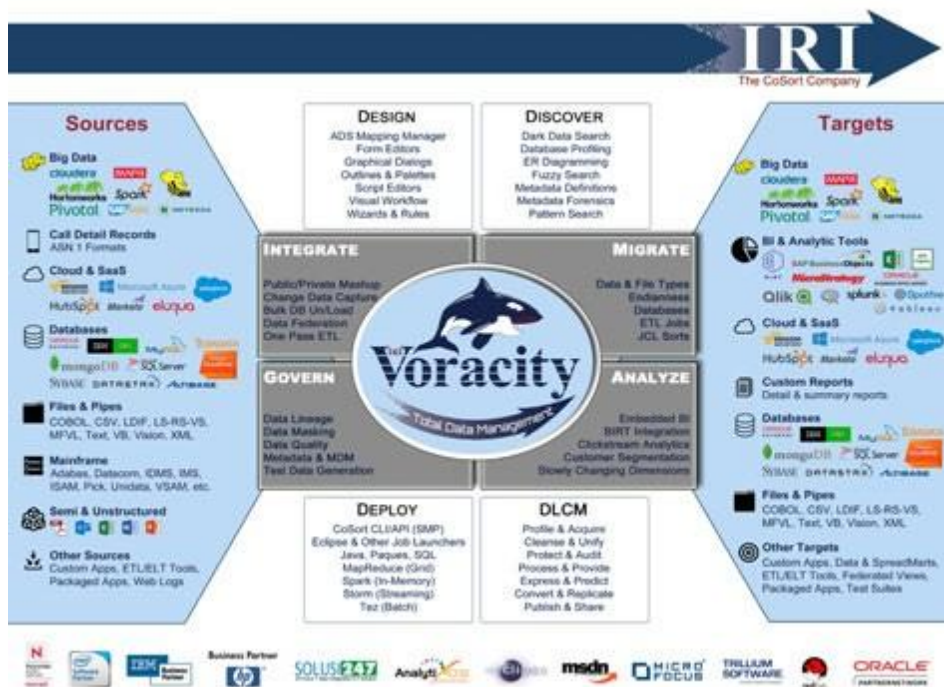
產生不限數量的資料庫、檔案與回報目標 - 立即以多重傳統與現代的資料格式產生實際線上或實務可用的測試資料。將多種工作合併在相同的輸出入通道以節省作業設計及執行時間。



資料生命週期管理

在 Voracity 中使用這些必要的展集功能：

- **Profile & Acquire** 當連結到不同的資料來源時，探索、分類與萃取資料及元數據。定義定制資料結構、遮罩資料格式與建立測試資料。
- **Cleanse & Unify** 在多重來源中過濾、豐富、擦洗以及標準化資料。選擇、模糊搜尋與合併參考資料到主資料表與其值。
- **Protect & Audit** 在欄位級別進行過濾中去除識別數據、轉型、報表、或不去處理，編碼、解碼、雜湊、假名、編輯、記號化 等等以及 XML 中的日誌。
- **Process & Provide** 在相同的作業與 CoSort 或 Hadoop 的輸出入通道中整合，移植，管理與分析資料，可視化與以任何格式導入測試或實際資料目標
- **Express & Predict** 在細節與總結及趨勢報表中的聚合，交叉計算與資料格式化或自動派送結果到你的分析工具或 BIRT 可視化處理。
- **Convert & Replicate** 移植傳統資料庫，檔案與資料型別或指定的新目標紀錄設計。以任何格式或 schema 複製或子集資料。
- **Publish & Share** 資料聯邦、儲存或立即移動到多重目標，如同 GIT 一樣連結到它們及他們在安全儲存庫的元數據以進行異動追蹤與版本控制。



互動設計功能

使用以下 Voracity 的功能組合：

- 符合目的(fit-for-purpose)的作業建立精靈
- 求助啓用的圖形對話與型式編輯器
- 提供動態大綱的語法感知批次編輯器
- 先進的可視工作流程與對應圖樣
- 電子表格樣式、免編碼 ETL 的 AnalytixDS 對應管理

自動化 ETL 工具移植

藉由 IRI 與 Analytix DS 的結盟，Voracity 能以更便宜、更快速與更簡單的方式將傳統 ETL 工具與 SQL 建構工作自動化地將進行移植的作業工作流程。Hadoop 用戶能夠全部或局部的靈活運用 Voracity 的功能。並且藉由 IRI 作為元數據整合的“Gulfstream”應用程式介面，獨立開發商與解決方案開發商得以使用或建置的方式無縫整合 Voracity 的工作流程。

註 1：將已經存在的物品、內容、事件等等，以特殊的方式組合後讓「意義再呈現」。不論是美術館的展覽、精選集唱片、黃頁電話簿或網路懶人包，全部都是展集。透過這種方式，只要讀者覺得有意義、有好處，展集本身也是一種有用的內容，也可以拿來「內容行銷」。

註 2：由 IBM 提出的，目前以該公司的 DB2 Information Integrator (DB2 II) 為支援此概念的主要產品，其主要的目標就是使應用程式能夠存取與整合不同的資料，無論分散式的、主從的系統，結構化的還是非結構化的資料，公共的、私有的等差異，透過中介工具的轉換，讓使用者感覺不到資料的差異，就好像只存取單一資料庫而已。