

COSORT

THE OPEN SYSTEMS STANDARD



NST
New System Technology

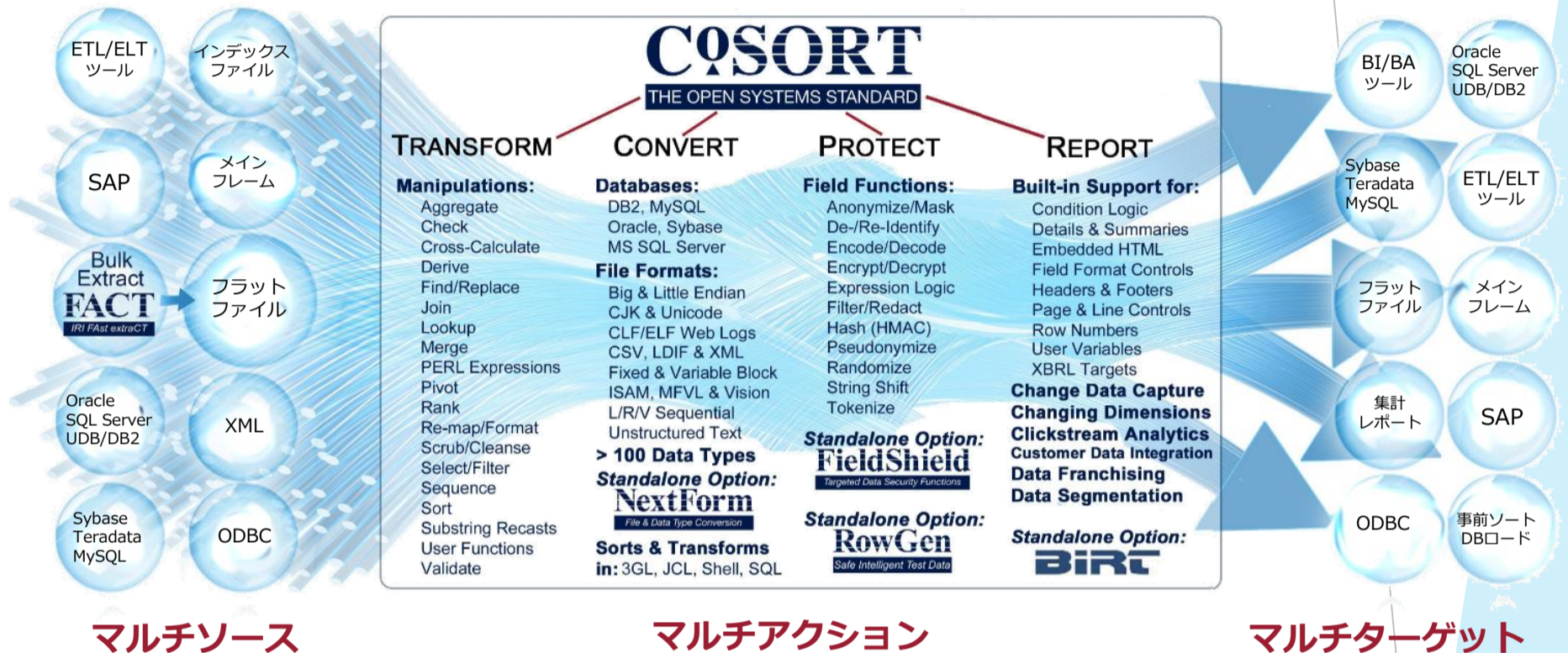
オープンシステム対応ソートツール

選択、結合、ソート、マージ、集計、タイプ変換および再フォーマットは、データを情報化する基本的な要素です。

たとえばソートの基本的な動作は、処理時間、ディスクスペース、メモリおよびコンピューターのその他のリソースを大量に消費します。

データサイズが2倍になると、ハードウェアの必要量は3倍にもなります。

CoSORTは、データウェアハウスの構築や大量のデータ処理における様々なニーズを最小限の資源で実施することができるソートおよびレポート作成のための製品です。



CoSORTの主な特徴

1. 超高速ソートおよび並列変換

CoSORTでは以下のような機能を同時に実行できます。

Aggregate	Hash	Randomize	Convert	Join
Map	Cross-Calculate	Integrate	Report	De-identify
Lookup Sequence	Encrypt	Mask	Sort	
Filter	Merge	Validate		

大規模かつマルチフォーマットの入力を一回のパスで処理します。またPerl互換の正規表現および、個人情報の保護, データクレンジング, 他の複合変換を行う為のカスタムフィールド機能をサポートしています。

2. 任意の入力源、データ型、データベースおよびフラットファイル形式をサポート

CoSORTはEBCDIC、10進数、IPアドレス、Unicode、COBOL、Varchar、Shift JIS、EUC、タイムスタンプを含む、100以上のデータ型を照合、変換します。

また、CSV、ISAM、LDIF、Vision、XML、いくつかのシーケンシャルファイル形式間でファイルを変換することが可能です。

3. 高速データベースおよびデータウェアハウスの操作

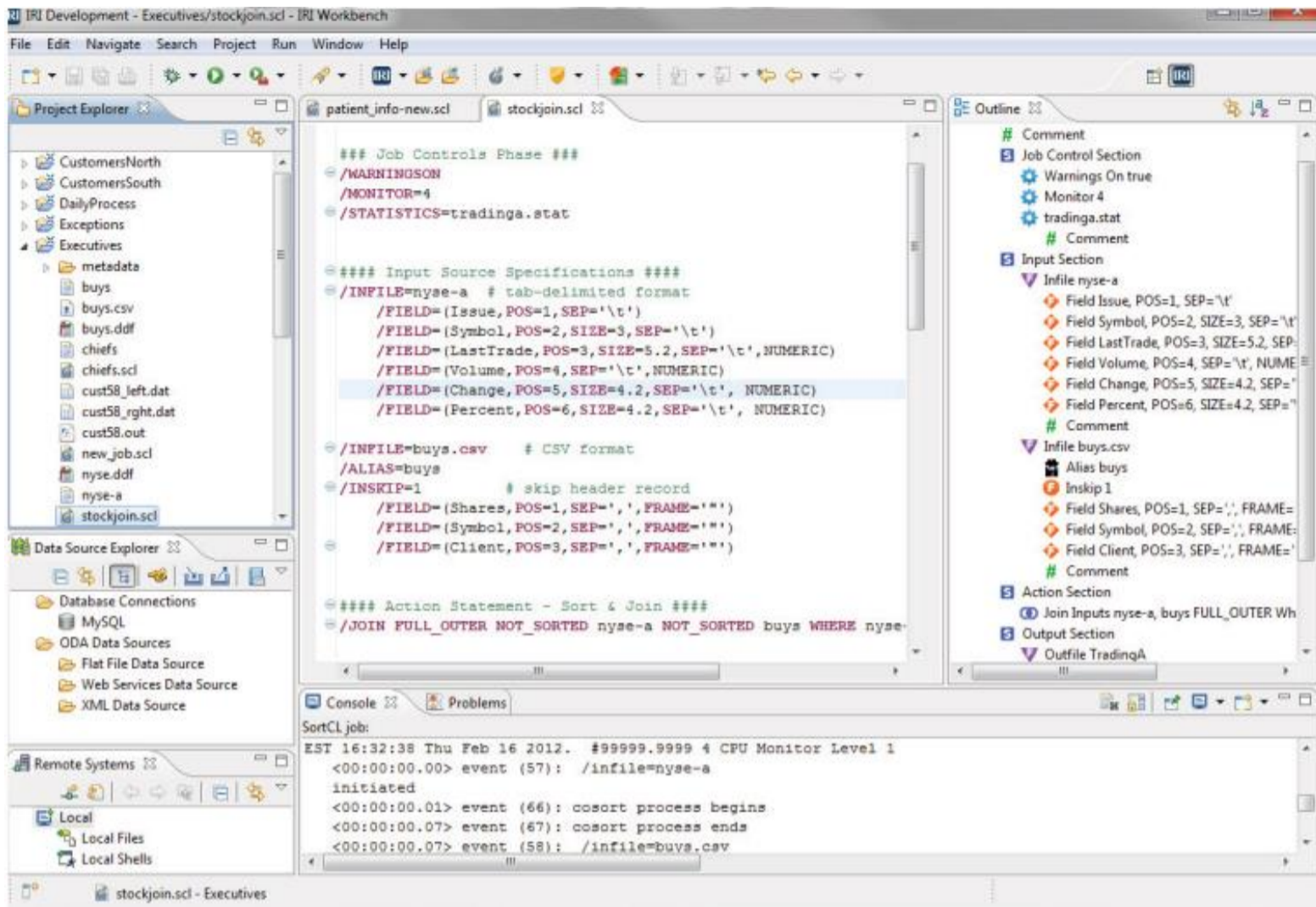
CoSORTはDB2、Oracle、Informatica、DataStageより高速にselect,join,order-by group-byを実行することが可能です。

4. データセキュリティ

CoSORTはフィールドレベルで機密データを暗号化、復号化、フィルター、マスク、ランダム化することが可能です。又、テーブルおよびファイルの機密データを保管、監査証跡の作成も容易に行うことができます。

グラフィカルな統合開発環境

GUI環境により、スムーズでわかりやすい操作が可能。



事前処理による効率化

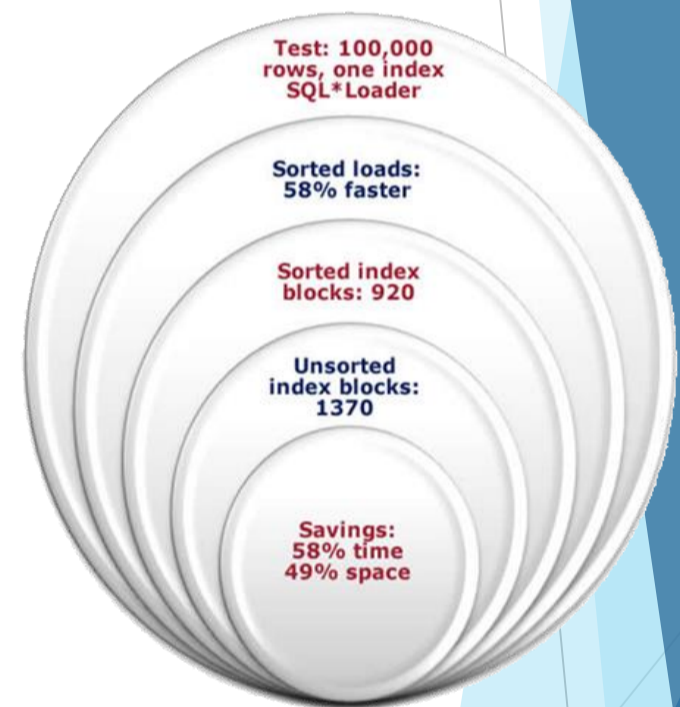
課題：未ソートのデータロードは非効率

インデックス作成:

- 1) より多くの内部処理が必要 (多数のブロック分割)
- 2) より多くの一次領域が必要、ソート済みの2倍の処理時間

解決策：CoSORTによる事前ソート

- 1) FACTでテーブルをフラットファイルへアンロード
- 2) CoSORTで最長のインデックスを指定してソート
- 3) SQL*Loader, bcpなどでDIRECT=TRUEを指定してロード
- 4) ソート済みの為、インデックス作成が高速 (ロード完了後にインデックスを作成する場合、NOSORTオプションを指定)

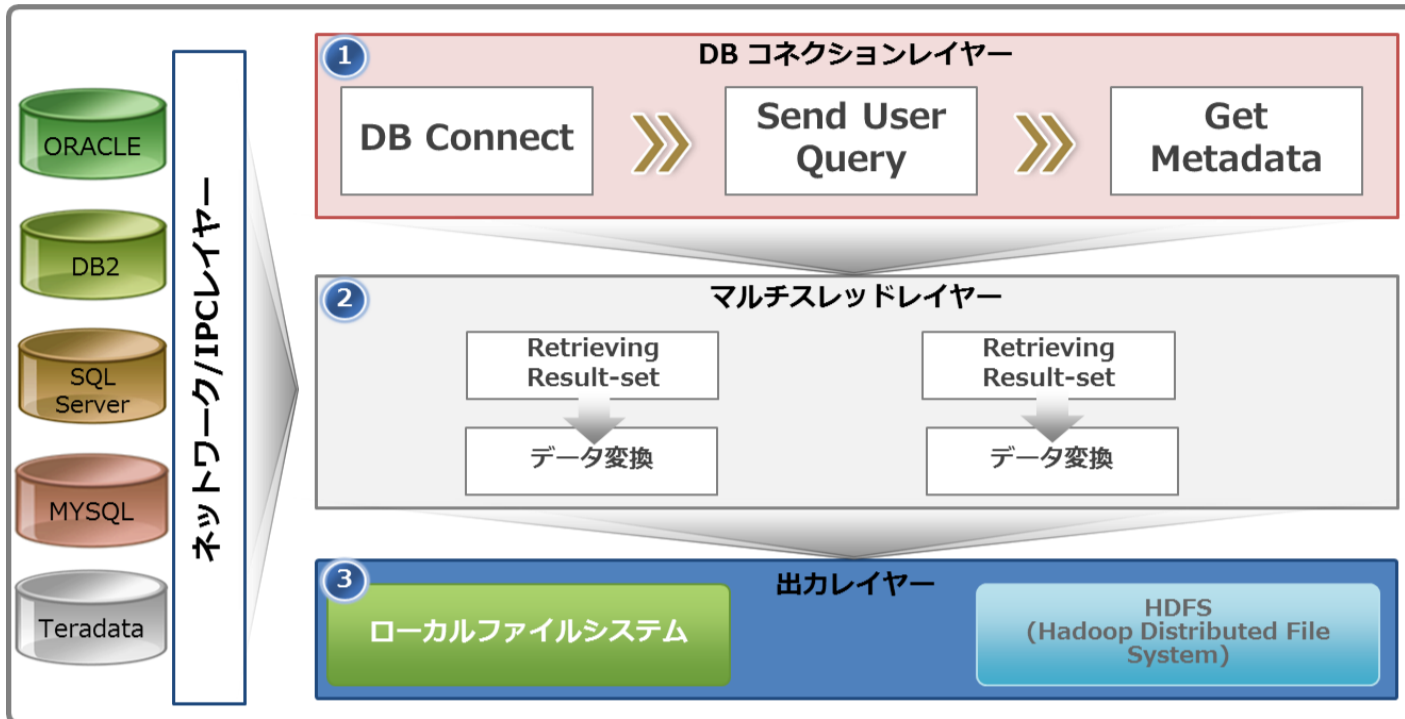


IT部門における課題とCoSORTによる解決策

- ETL利用範囲の拡大、ハードウェアのアップグレード、他のツールのバージョンアップに伴う費用増大
⇒CoSORTを置き換えツールとして、あるいは一部の代替ソリューションとして利用することにより、費用逓減が可能。
- 大量データの為、ソート、マージ、集計、ロードに要する時間が遅い
⇒大量データを扱うバッチ処理における大幅時間短縮。また、DWHにおいては事前に重複データ削除、グルーピング、ソート行うことにより、大幅にロード時間を短縮、削減します。
- 第3代言語のプログラミングやETLツールのコーディング、保守が非常に困難になっている
⇒CoSORTの「sortCL」を使用することにより、ジョブスクリプトと入出力パスの指定だけで、データ変換、ファイル変換、データ保護、そしてレポート出力が可能です。

FACT:Fast extraCT

OracleやDB2などから高速データ抽出



抽出
プロセス

- SQL SELECT文または、RDBMSのAPIを使用
- .DDFメタデータファイルでソート、暗号化、再フォーマット、テストなど
- RDBMSロードユーティリティ用のメタデータを作成
- 大規模なファイルから抽出されたデータはローカルファイル（又はパイプ）や Hadoop HDFSへアンロード

◎適用範囲と特徴

- 通信、金融、小売、医療、運輸、政府、ドラッグストア、その他大量のトランザクション（ファクト）テーブルをもつ企業
- CoSORTとFACTを組み合わせると、大規模なETL、データ再編成、変更部分のレポート、アーカイブ、データ移行、データ準備といった業務におけるソート、ジョイン、集計、計算、変換、クレンジング、再マッピングに最適
- Oracle用の高速アンロードツール市場では、FACTはベンチマークで BMCやCAのツールより高速
- FACTは、CoSort, RowGen, FieldShield, SQL*Loader用のデータレイアウトを出力
- データ定義は、MIMB (Meta Integration® Model Bridge) で多くのサードパーティアプリケーションに変換可能

サポートするデータ（DBMS）および動作環境

サポートするデータおよびファイル

- ASCII、EBCDIC、COBOL、C(バイナリー)形式
- ヨーロッパ、ISO、日本、アメリカのタイムスタンプ
- IPアドレス、数値
- Bulk RDBMS Unloads(FACT経由)
- DBテーブルおよびEXCELスプレッドシート
- ACUCOBOL-GTインデックスファイル
- IBM Unblocked Variable Record Format
- LDIF(LDAP)、Microsoft CSV、Flat XML
- Micro Focus 可変長 およびI-SAMファイル
- シーケンシャルフラットファイル
- Fixed Block File Format
- VSAM(Clerity Mainframe Re-hosting経由)
- W3C Common & Extended Log(Web)
- Unicode、ネイティブマルチバイト文字セット

動作環境

- UNIX (AIX, HP-UX, Solaris, Tru64 & more)
- Linux on x86, Itanium, IBM s/p/i/z, FreeBSD
- Microsoft Windows® (XP, Vista, 7, 8, 2000-2016 Server)

開発元

