

■製品仕様		
●DataStage (サーバ版)		
稼働OS	DataStage サーバ	Windows® Server 2003、Windows® Server 2008、AIX、HP-UX、Solaris、Red Hat Enterprise Linux
	DataStage クライアント	Windows® XP、Windows® Server 2003
対応データベース/ファイルなど	HiRDB、Oracle、Oracle Express、DB2、Teradata、Informix、Sybase、Sybase IQ、Microsoft SQL Server、RedBrick、ODBC、OLE DB、シーケンシャルファイル (固定長/csv/複合ファイル)、XMLファイル、ftp、日立ディスクアレイサブシステム ほか	
	Web Service Client Pack適用時	Webサービス
	Message Adaptors適用時	IBM MQ、SeeBeyond
	Java Pack適用時	Javaプログラム
	PACK for SAP R/3適用時	SAP® R/3®
文字コード変換	メインフレーム系	EBCDIC、EBCDIK、IBM漢字、日立KEIS、富士通JEF、NEC JIPS ほか
	オープン系	シフトJIS、EUC、Unicode ほか
※サポートDBは、DataStage Serverの稼働OSによって異なります。弊社営業までお問合せください。		
●DataStage (パラレル版)		
稼働OS	DataStage サーバ	AIX、HP-UX、Solaris、Red Hat Enterprise Linux
	DataStage クライアント	Windows® XP、Windows® Server 2003
対応データベース/ファイルなど	HiRDB、Oracle、DB2、Teradata、Informix、Sybase、Sybase IQ、RedBrick、シーケンシャルファイル (固定長/csv)、XMLファイル、ftp、日立ディスクアレイサブシステム ほか	
	Web Service Client Pack適用時	Webサービス
	Message Adaptors適用時	IBM MQ、SeeBeyond
	Java Pack適用時	Javaプログラム
	PACK for SAP R/3適用時	SAP® R/3®
文字コード変換	メインフレーム系	EBCDIC、EBCDIK、IBM漢字、日立KEIS、富士通JEF、NEC JIPS ほか
	オープン系	シフトJIS、EUC、Unicode ほか
※サポートDBは、DataStage Serverの稼働OSによって異なります。弊社営業までお問合せください。		
●MetaStage, MetaBrokers		
稼働OS (MetaStageクライアント)		Windows® XP、Windows® 2000
メタデータ格納データベース		Oracle、DB2、Sybase、Microsoft SQL Server
メタデータ連携ツール MetaBrokers	基本セット	DataStage、ODBC ほか
	モデリングツール	ERWin、ER/Studio、Oracle/Designer、Sybase/PowerDesigner ほか
	分析ツール	Business Objects、Impromptu、Microstrategy、Hyperion Essbase ほか
※日立での動作確認済みDB、ツールについては、弊社営業までお問合せください。		
●QualityStage (サーバ版)		
稼働OS	DataStage サーバ	Windows® Server 2003、Windows® 2000、AIX、HP-UX、Solaris、Red Hat Enterprise Linux
	DataStage クライアント	Windows® XP、Windows® 2000
●QualityStage (パラレル版)		
稼働OS	DataStage サーバ	AIX、HP-UX、Solaris、Red Hat Enterprise Linux
	DataStage クライアント	Windows® XP、Windows® 2000

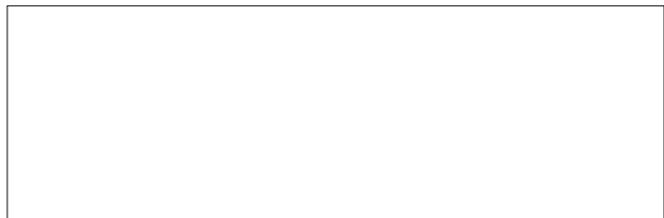
・AIX、DB2、MQSeries、Informix、Impromptu、RedBrick、OS/390は、米国およびその他の国における International Business Machines Corporationの商標です。
 ・Business Objectsは、SAP Franceの商品名称です。
 ・DataStage、QualityStage、MetaBrokerおよびMetaStageは、IBM Corporationの商標です。
 ・ER/Studioは、米国Embarcadero Technologies, Inc.の登録商標です。
 ・ERwinは、Computer Associates International, Inc.の登録商標です。
 ・Oracle、Essbase、Hyperion、Javaは、Oracle Corporation及びその子会社、関連会社の米国及びその他の国における登録商標または商標です。
 ・Linuxは、Linus Torvalds氏の日本およびその他の国における登録商標または商標です。
 ・Microsoft、Windows、Windows NT、Windows Serverは、米国Microsoft Corporationの米国およびその他の国における登録商標または商標です。
 ・Microsoft SQL Serverは、米国Microsoft Corp.の商品名称です。
 ・ODBCは、米国Microsoft Corp.が提唱するデータベースアクセス機構です。

・Red Hat は、米国およびその他の国でRed Hat, Inc.の登録商標若しくは商標です。
 ・SAP、R/3、ABAPは、SAP AGのドイツ及びその他の国における登録商標または商標です。
 ・Solaris、SeeBeyondは、米国Sun Microsystems, Inc.の米国およびその他の国における登録商標または登録商標です。
 ・SOAP (Simple Object Access Protocol) は、分散ネットワーク環境においてXMLベースの情報を交換するための通信プロトコルの名称です。
 ・Sybase、Sybaseのロゴは、米国法人Sybase, Inc.の登録商標です。
 ・Sybase IQは、米国法人Sybase, Inc.の商標です。
 ・Teradataは、Teradata Corporationの登録商標です。
 ・UNIXは、The Open Groupの米国ならびに他の国における登録商標です。
 ・その他記載の会社名、製品名は、それぞれの会社の商標もしくは登録商標です。

- カタログに記載の仕様は、製品の改良などのため予告なく変更することがあります。
- 製品の色は印刷されたものですので、実際の製品の色調と異なる場合があります。
- 本製品を輸出される場合には、外国為替及び外国貿易法の規制ならびに米国の輸出管理規則など外国の輸出関連法規をご確認のうえ、必要な手続きをお取りください。
 なお、ご不明な場合は、弊社担当営業にお問い合わせください。

製品に関する詳細・お問い合わせは下記へ

- 製品情報サイト
<http://www.hitachi.co.jp/datastage/>
- インターネットでのお問い合わせ
<http://www.hitachi.co.jp/soft/ask/>
- 電話でのお問い合わせはHMCC (日立オープンモデルウェア 問い合わせセンター)へ
 0120-55-0504 利用時間 9:00~12:00、13:00~17:00 (土・日・祝日・弊社休日を除く)



DataStage

信頼ある意思決定を、
確実に支える情報のために。

企業内に散在する情報を、 戦力として活かすために。 より速く、確実なデータ統合を実現する DataStageにおまかせください。

膨大な情報を、ビジネスに価値ある資産として活用できるか。CRM、SCM、SFAといった新たなビジネスモデルへの変革や企業統合など、経営環境のめまぐるしい変化の中で、ビジネス戦略を成功へ導く鍵はそこにあります。今、企業に必要とされるのは、企業内のシステムに分散する情報を統合できる情報基盤です。日立のDataStageシリーズでは、複数の異なるデータソースからデータを統合。そして新たに登場したQualityStageのデータクレンジング機能により、データ品質を確保することで真に価値ある情報として活用できます。パラレル機能や、メタデータの一元管理といった高度な管理機能も装備。さあビジネスの進化に向けて今こそ、データ統合を。

CRMを成功させる
データウェアハウス構築を。

SCM/SFAなどの
企業内データの活用を。

企業統合に対応した
多量データの統合を。

(データ分析)

データ品質を保ちながら
分析要件に応じた
DWHの構築を
支援します。

(データ活用)

複数システムに蓄積する
業務データを
情報として迅速に
提供します。

(データ統合)

大容量データの
高速処理を行う
パラレル機構を
提供します。

DataStage

情報資産管理基盤

戦略的情報活用を支援するDataStage。 その3つのアドバンテージ。

Speed

スピーディーなシステム構築・拡張

データ統合アプリケーションの開発時間を短縮。
再利用性を促進することで、開発生産性を向上。
既存処理の把握を容易にし、システム拡張時の効率を向上。

Cost

システム開発・保守コストの削減

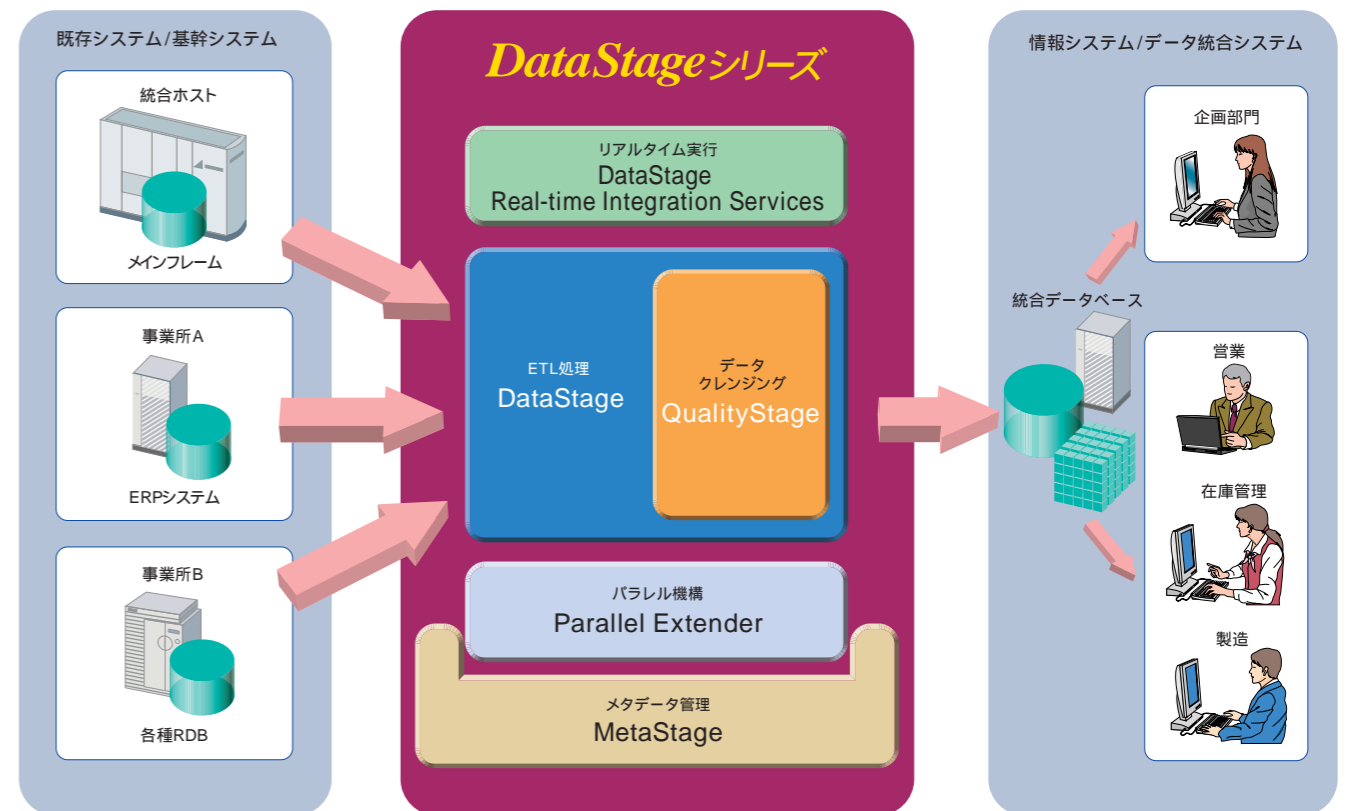
統合データベース、データウェアハウスを開発・維持するための工数を削減。
高度な技術を必要としない、容易な開発環境。
手離れの良いシステム構築で、開発者以外でも保守が可能。

Quality

データ品質の確保

複数のプログラムにわたるメタデータを一元管理し、データの属性を的確に把握。
重複情報の削除やデータ補正・データ品質の劣化を防止するため、データクレンジング処理の自動化を実現。

データを情報資産へと導くDataStageの各機能



ETL : Extraction Transformation and Loading
CRM : Customer Relationship Management
ERP : Enterprise Resource Planning
SCM : Supply Chain Management
SFA : Sales Force Automation

顧客中心型マーケティングに重要なデータ品質管理を実現。

システム上に、正確かつ安全な情報を確実に行き渡らせるため、大規模データの統合で発生するデータの 不整合や重複問題を効果的に解消。世帯名寄せや企業名寄せなどのデータクレンジングを効率的に実現します。



データの重複を容易に排除

最高のデータ品質を確保するデータのリエンジニアリングの実現に向けて、データクレンジング、標準化、補正、マッチング、重複削除などの処理を自動化。レコード間の関連性が識別でき、レコードの重複を排除できます。複数のシステムにまたがる情報を一元化し、データの同一性を判断することで、システム間のデータの重複を削除し、データの少量化が実現できます。

データの問題発見と補正を実現

日々の企業活動で追加される情報における誤入力などに配慮し、マスターとなるデータに対して重複関係の発見やデータの補正を行います。蓄積された情報も分析・調査によりビジネスデータとしての価値の評価が行えます。たとえば、郵便番号「999-9999」が不正情報である、という意味を理解し補正に役立ちます。

充実した機能により、高度なデータクレンジングを実現

確率性レコードリンク設定により高いマッチング率を実現し、正確で完全度の高いデータを生成。パターンベースのマッチングに比べ、クレンジング品質に優れます。各種ルールに従い柔軟性に富んだ標準化機能を提供。複雑なフリーフォーマットフィールドを扱うことが可能です。データ規模に応じスケラブルに複数のCPUを使用し、ハイパフォーマンスを実現するパラレル機能を提供します。

〈分析・調査〉

ソースデータを検証し品質や内容を把握することで、データリエンジニアリングフェーズで利用するデータレベルでのビジネスルールを決定できます。GUIでの簡単な設定により、パターン解析を行うパターン分析、フィールド値の頻度分布を示す単語頻度分析や、割り当てられた分析タイプを示す単語分類分析、といった各レポートを生成します。

〈マッチング〉

1つ以上のファイルの重複するデータを識別し、関連性を識別できるIDを作成。既存データを外部の新しい情報で補足することも可能です。マッチングはブロッキングとマッチングの2段階で行います。マッチングでは、1:1マッチング、多対1マッチング、重複排除を行います。

〈標準化〉

次プロセスのマッチングを効果的に行うため、入力データを標準化し、各データタイプを内容や形式からタイプごとにまとめ、内部的な一貫性を生成します。各要素を正しく解析・識別し、適切なタイプに分けるため、特定データタイプの規則を満たすように設計されたルールセットを使用します。

〈サバイバースhip〉

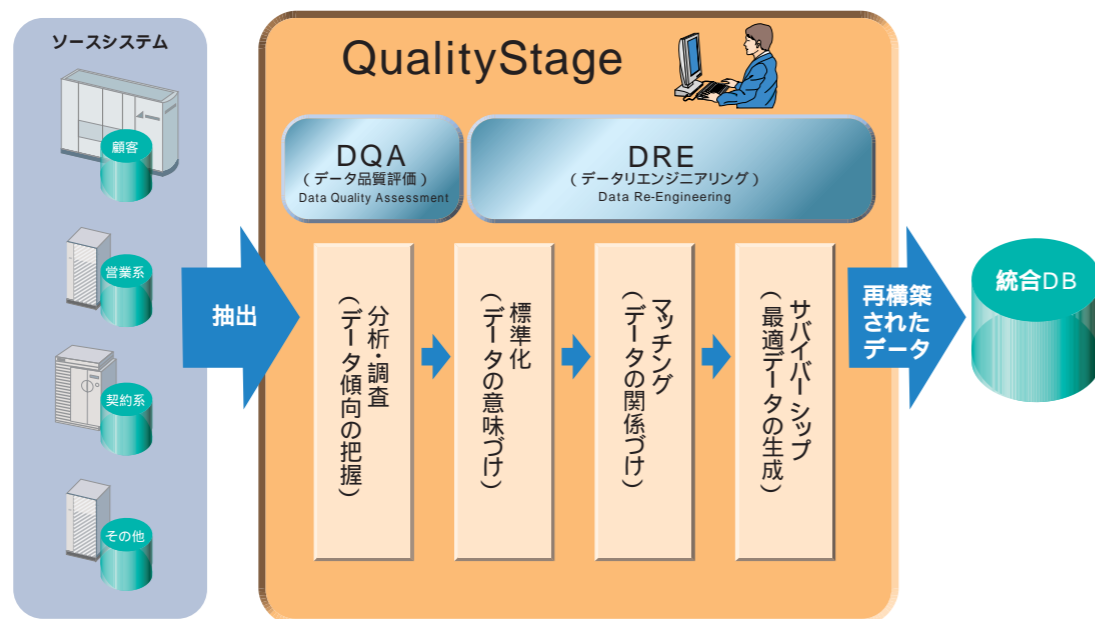
重複したレコードから最適な唯一のレコードを生成するためのルールを指定。指定したルールを利用し、データクレンジング後の残すべきレコードの選択を自動化できます。

4つのプロセスで段階的にデータクレンジング

QualityStageは質の高いデータを得るため、データ品質を改善する手法に基づいて4つのプロセスを提供します。まず「分析・調査」プロセスでは検証機能を活用し、データの傾向を把握。「標準化」プロセス

では標準化機能を活用し、データをクレンジングします。「マッチング」プロセスでは情報の関係を示すIDを作成し、「サバイバースhip」プロセスでは、情報に最適なレコードを生成します。

複数のシステム間のデータを統合し、相互参照のリンク情報を付与。一意のIDを維持することで、複数の重複データから唯一のデータを生成するとともに、既存データからの生成履歴を管理します。



〈データクレンジングの一例〉

統一性のないデータを、設定したルールに基づいて標準化。名寄せ作業が効率的に行えます。

名前	郵便番号	住所	注文数
日立 太郎	123-4567	神奈川県A市B町1丁目3番 Dビル	180,000
日立 たろう	123-	神奈川県A市B町1-3	250,000
日立 太郎本部長	123-0000	A市B町1 Dビル	300,000

全てのレコードをQualityStageの設定したルールに従い、標準化

(例)住所情報の標準化

府県	タイプ	市区	タイプ	町	タイプ	数値1	タイプ	数値2	タイプ	建物
神奈川	県	A	市	B	町	1	丁目	3	番	Dビル
神奈川	県	A	市	B	町	1		3		
		A	市	B	町	1				Dビル

ルールに基づき項目ごとに信頼度を算出

名寄せを行った後、分析ツールにデータを転送し、顧客分析を実行

DataStageとのシームレスな連携

QualityStageで作成したジョブは、DataStageのプラグインからシームレスに連携できます。DataStageのジョブのなかで、QualityStageの提供する機能を1つのステージとして利用できます。DataStageのデータの集約・加工・編集処理とQualityStageのデータクレンジング処理を組み合わせることで、データ品質を確保したデータ処理のジョブ開発が可能となります。



DataStageからの利用例画面

標準化処理設定画面

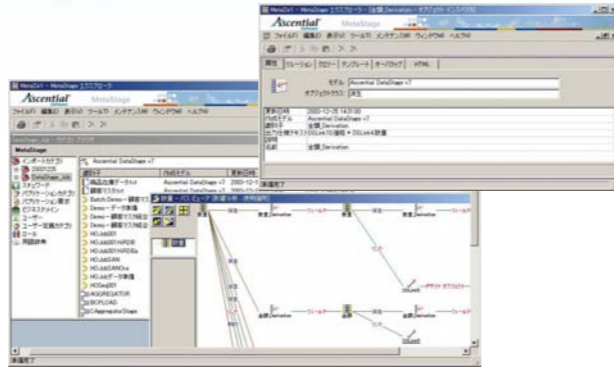
システムの柔軟な拡張に メタデータ管理を実現。

システム保守・拡張時の修正に伴い発生する影響範囲の調査、参照データが、どのシステムに存在しているか把握したい。こうした高度なデータ管理を実現するメタデータ管理のためのコンポーネントを提供し、情報資産の有効活用を促進します。

より高度なメタデータを統合管理

複数のDataStageのプロジェクトからメタデータを取り込み、統合して管理できます。テーブル定義などのテクニカルメタデータはもちろん、データ変換ルールなどのビジネスメタデータも収集。取り込んだメタデータは、バージョン管理され、過去の状況を確認できます。

MetaBrokerを使用して、各種ツールとメタデータ交換ができます。DataStage用、ODBC用などにオプションを追加することで、ERWin、ER/StudioなどのモデリングツールやBusiness Objects、MicroStrategyなどのBIツールとのメタデータ交換も可能です。任意のメタデータを変更した場合、どの変換処理やターゲットテーブルに影響を与えるか、といった影響分析ができます。また、DataStageの稼働記録をプロセスメタデータとして管理できます。



メタデータの影響分析画面

メタデータの内容や分析結果をGUIで視覚的に確認できます。結果は、HTML、XMLなどの形式で出力でき、メタデータの共有や分析レポートとして利用できます。

スケーラブルなパラレル機構により、 大規模データを高速処理。

増え続ける膨大なデータに対応するため、スケーラブルなパラレル機構を利用して、高速処理を実現。従来のDataStageの開発環境を利用してパラレルジョブの開発が行えます。

容易な開発環境

DataStageの使いやすさはそのままに、データフロー定義とパラメータ設定でパラレル対応のジョブを開発。パラレル開発のための高度なスキルを必要とせず、高性能な処理を開発できます。

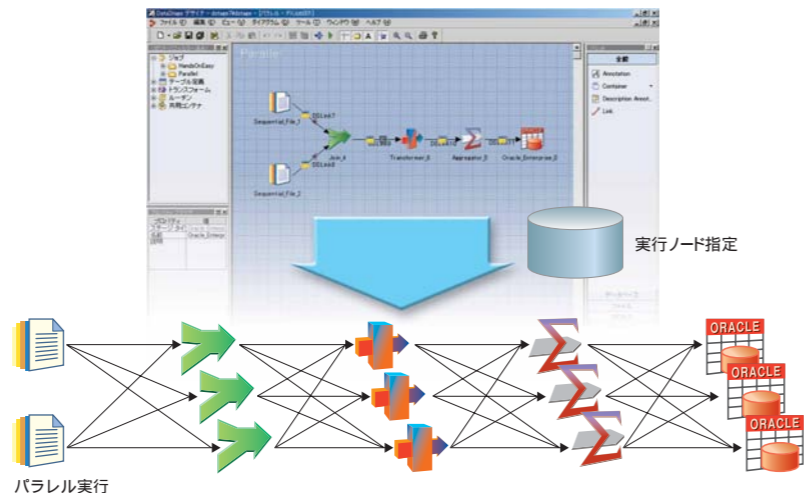
DataStageと同様、各種DBに対応するとともに、データ処理のための豊富な部品群を提供。GUIで作成したジョブを部品化し、再利用性を高める共有コンテナも利用できます。

ハイスケーラビリティを実現

パラレルジョブは、SMP環境、MPP環境のいずれにも対応。CPUの増設に対応して高速化するスケーラビリティを実現します。

SMP: Symmetric Multiple Processor
MPP: Massively Parallel Processor

データは自動的にパーティショニングされ、分割データごとに割り当てた処理プロセスでパイプライン実行を実現。大量データを高速処理できます。



柔軟なジョブの設計

コンフィグレーションファイルでの指定により、実行ノード数を制御。ジョブを変更することなく、実行環境を柔軟に変更できます。ジョブフローの途中で、データのパーティショニング方式を変更できます。ハッシュ分割、レンジ分割、ラウンドロビン分割、ランダム分割などをサポート。それぞれの処理内容に適したデータ分割を柔軟に行うことで、さらに性能をアップします。

SAP® R/3®システムとのシームレスな連携により、 ERP基幹データを有効活用。

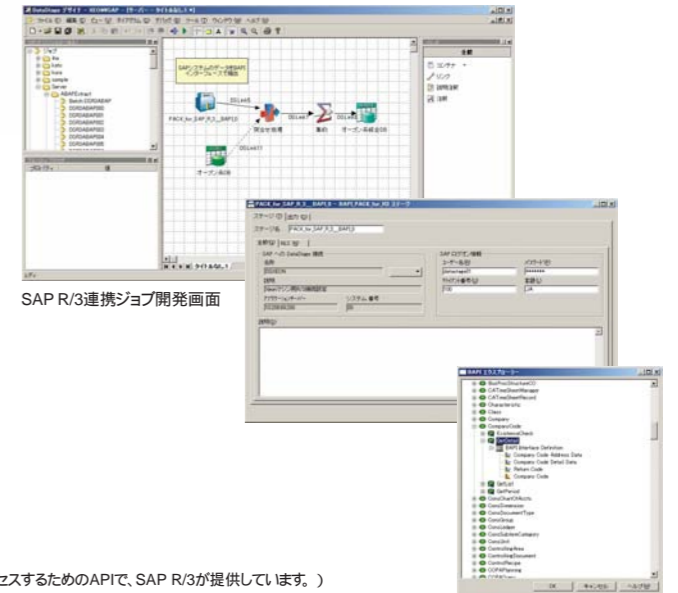
DataStageのビジュアル開発環境に、SAP R/3のデータ抽出や格納機能をアドオン。ERPに存在する基幹データと外部データを組み合わせ、有効に活用することができます。

SAP R/3のデータ処理を容易に開発

DataStageのETL処理上にアドオンし、SAP R/3のデータにアクセスするプラグイン群を提供します。BAPI/IDOC Extract/IDOC Load/ABAP Extractプラグインを用意しています。データ統合・データ連携のシステムを、DataStageの豊富なETL(データ変換)機能も含め、GUIを使用して短時間に構築し、保守性と柔軟性のあるシステムを実現できます。ABAPのコードに比べて高い生産性を確保できます。

BAPIプラグインでは、BAPIを使用してSAP R/3にアクセスし、データの抽出・ロードなどを行います。BAPIは、データとそれを扱うメソッドをオブジェクトとしてカプセル化し、DataStageからの呼び出しは、メソッドの単位となります。また、DataStageの入力リンク・出力リンクのカラムは、BAPIのメソッドの引数(入力パラメータ・出力パラメータ)にマッピングされます。

BAPI: Business Application Programming Interface (SAP R/3のBusiness ObjectにアクセスするためのAPIで、SAP R/3が提供しています。)
IDOC: Intermediate Document



SAP R/3連携ジョブ開発画面

DataStageのジョブをWebサービス化し、 リアルタイム系処理に対応。

DataStageのジョブをリアルタイムなサービスとして提供するためのフレームワークです。情報ビジネスに求められる複雑なデータ変換、リアルタイムなデータ連携、コード体系のデータ標準化などのデータ統合サービスを実現し、DataStageをより便利に活用できます。

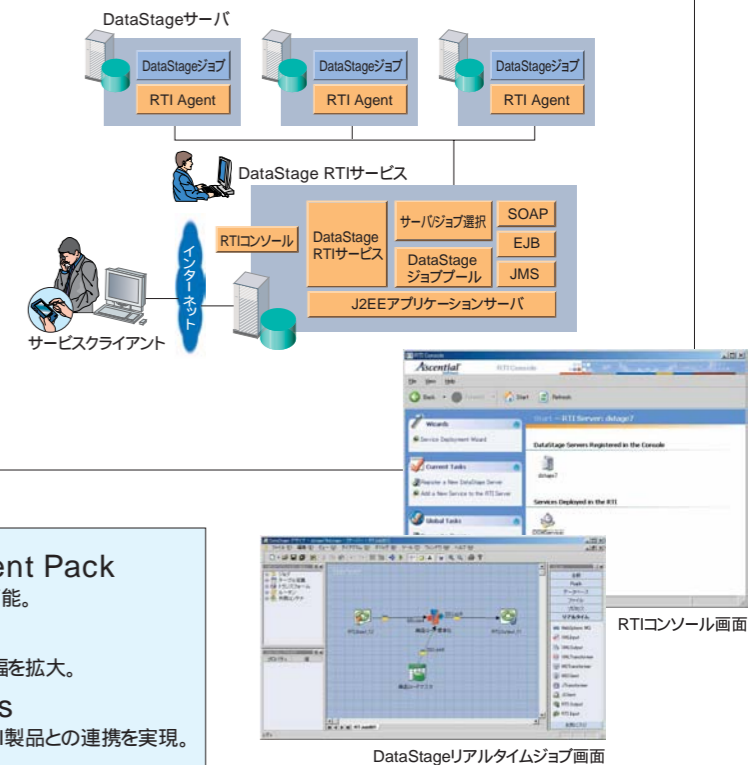
DataStageジョブをリアルタイムサービス化

DataStageのジョブをWebサービス(SOAP)のオペレーション、EJBメソッド、JMSリスナーのいずれかとして提供できます。

SOAP: Simple Object Access Protocol
EJB: Enterprise Java Beans

Webサービスとして提供するジョブの開発をDataStageのビジュアルな開発環境で作成可能。Javaなどに習熟した開発者でなくともWebサービスを提供するシステムを容易に開発できます。

Real-time Integrationのサーバから、複数サーバに存在するDataStageリアルタイムジョブを管理。RTIコンソールで、リアルタイムサービスの設定をGUIから行えます。



その他

DataStage Web Services Client Pack

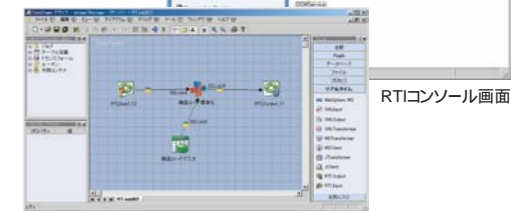
DataStageのジョブからWebサービス呼び出し可能。

DataStage Java PACK

Java™プログラムと連携し、DataStageの活用の幅を拡大。

DataStage Message Adaptors

DataStageのジョブからIBM MQやJMSで他社EAI製品との連携を実現。



DataStageリアルタイムジョブ画面



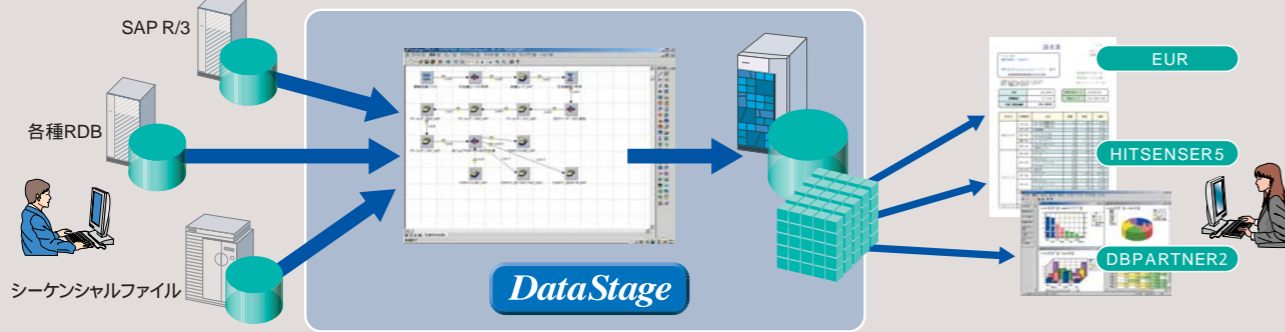
DataStageが活用されるいくつかのビジネスシーンをご紹介します。

さまざまな企業システムにおいてニーズが高まるデータ統合。DataStageはそれぞれのシステムに柔軟に対応し、データ統合を行うことでビジネスメリットを最大限に引き出します。

事例 1 データウェアハウスシステムへの適用

複数のシステムに分散したデータを統合し、効果的な顧客分析や商品分析を実現します。

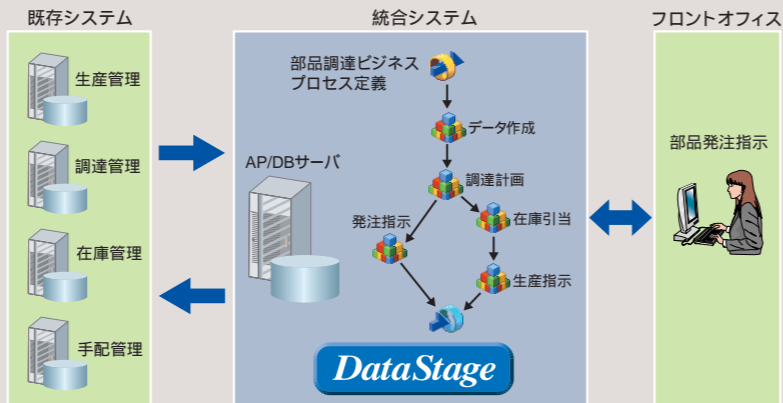
各種システムのデータの集約・統合・編集を行い、それらを提供することで、短期間に経営環境に応じたセントラルウェアハウスやデータマートの構築を支援。日立のHITSENER5などのBIツールによるデータ分析を活用した戦略的な経営を実現します。



事例 2 EAIソリューションへの適用

ビジネスプロセス統合により、生産から手配までのリードタイムを削減します。

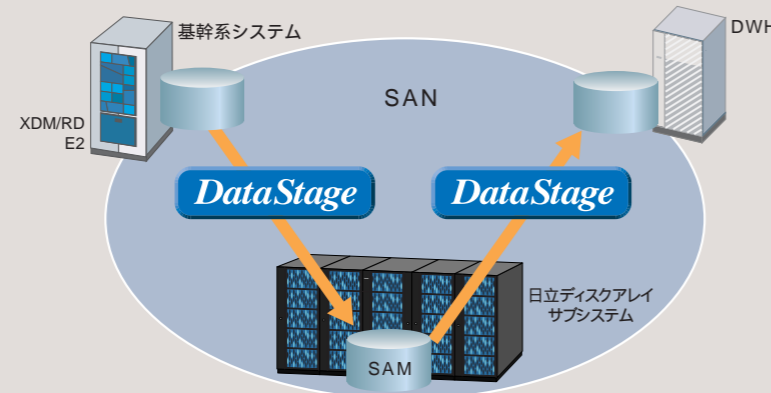
Cosminexus Enterprise Integration Serverとの連携により、ビジネスプロセスレベルでのデータ統合を実現し、各種アプリケーションを連携。お互いの業務結果が情報としてよりリアルかつ正確に反映でき、必要な場面に必要な情報として提供できます。仕様変更時の対応が容易になり、拡張性の高いシステム環境も実現します。



事例 3 SANソリューションへの適用

日立ディスクアレイサブシステム連携によりメインフレームからのデータ統合処理の性能を向上します。

日立ディスクアレイサブシステムと連携することで、ネットワークを経由せずにSAN経由で大量データを高速転送できます。メインフレームからの大量データの反映時間や反映間隔を短縮することで、より的確で鮮度の高いデータを供給し、確実な意思決定を支援します。



DataStageは、日立から。

お客さまにとって最適なシステムを構築し、守り続けるために。

日立では、製品開発元との技術・開発レベルでの協業体制の強化により、サポート体制も充実。

もちろん、システム構築の専門知識や運用に関するお悩みも、信頼の実績と経験豊富なノウハウと技術力を駆使して解決します。

迅速

日立では開発元である、IBM Corporationとともに、DataStage/QualityStageの開発に協力。製品を深く理解した技術者のサポートにより、お客さまの質問やお問い合わせに迅速におこたえします。

充実

日立データベースとの連携により、DataStage接続モジュールを自社開発。開発で培った充実のノウハウによる精度の高い対応が行えます。

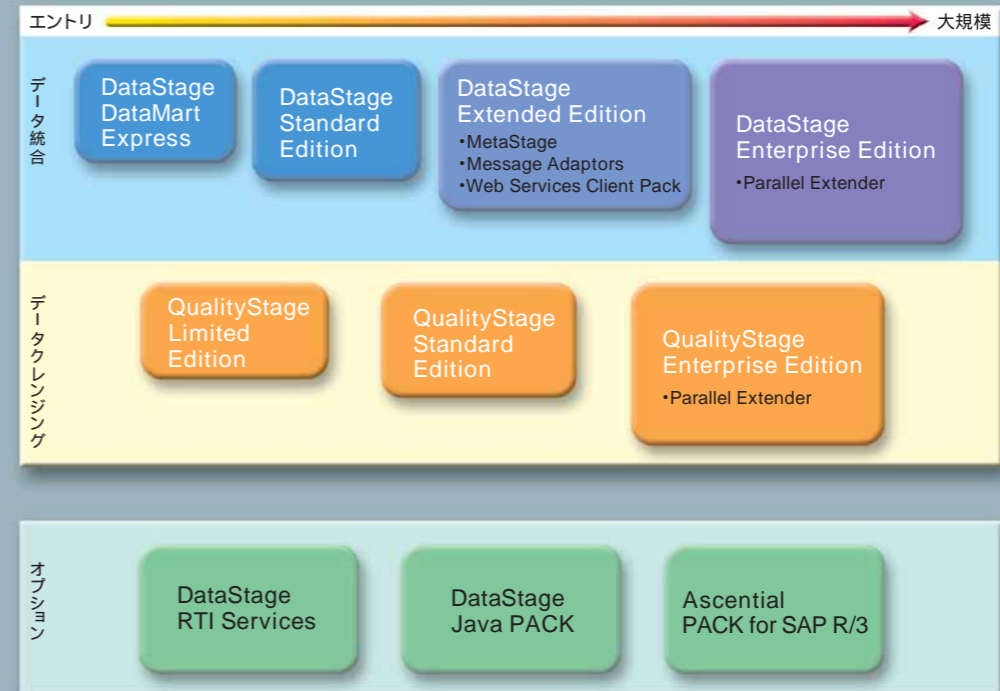
多彩

DataStageをスムーズにお使いいただくためのテクニカルサービス・構築サービスなど多彩なサービスを提供します。



製品体系

プロダクト



サービス商品
サポートサービス
テクニカルサービス