

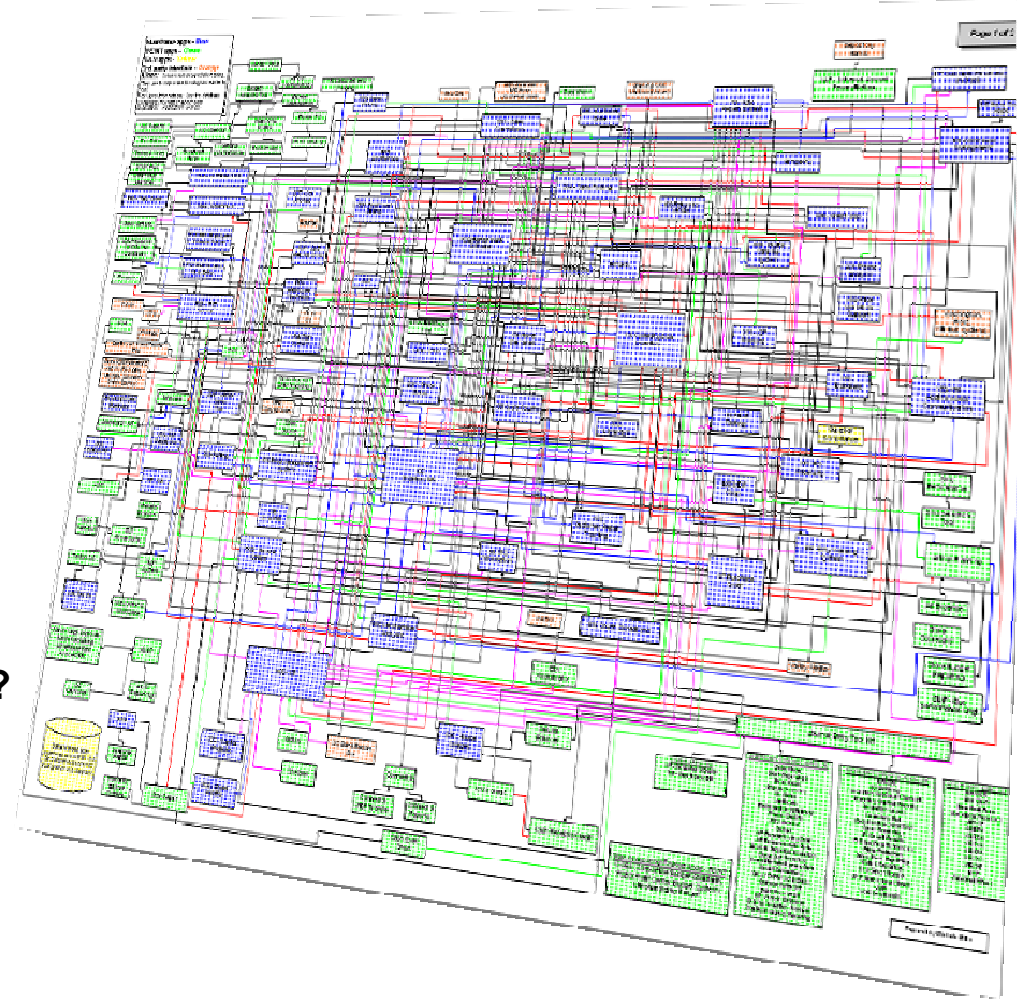
*IBM Metadata Yönetimi ile Verilerinizi
Yakın Takibe Alın*

*A.Okay Akyüz
Profesyonel Hizmetler Direktörü, OSİS*

Trusted Information

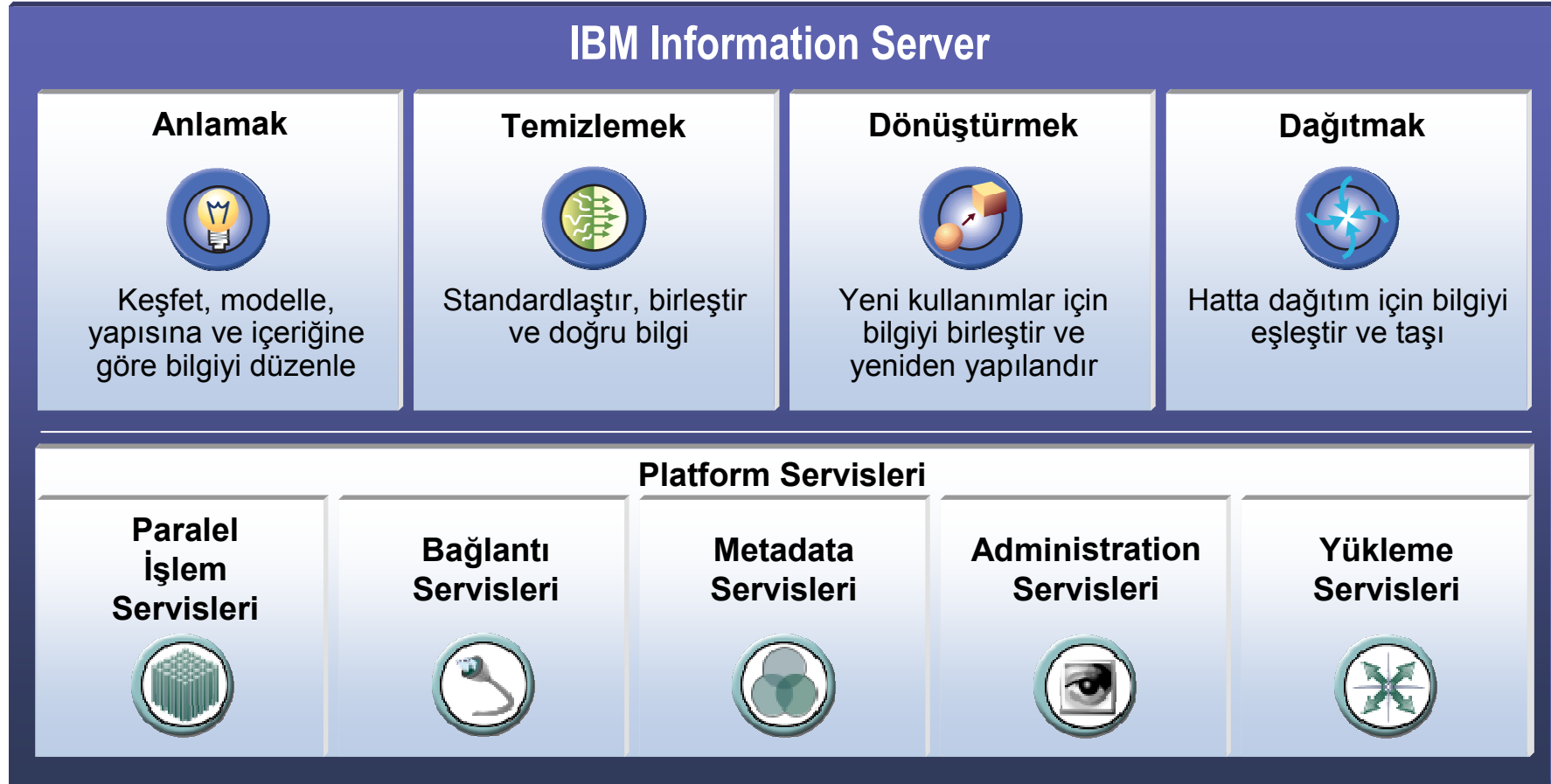
Bilgi...

- Bilgilerim nerede?
- İhtiyacım olduğunda nasıl erişirim?
- Bilgilerimin içeriğinde ne var?
- Güvenilir mi?
- Kalitesini nasıl arttırabilirim?
- İstedğim şekilde nasıl erişirim?
- Bilgim nerden geliyor, nereye gidiyor?
- Nasıl kontrol edebilirim?



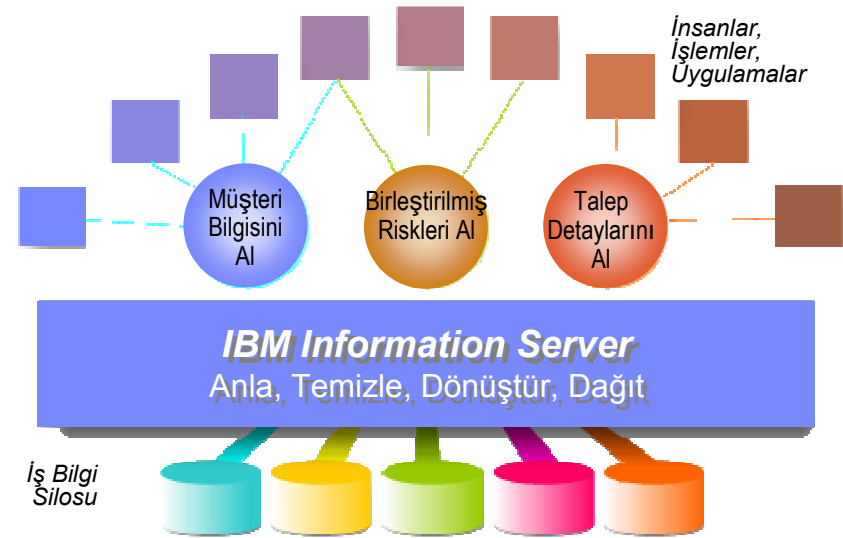
IBM Çözümü: IBM Information Server

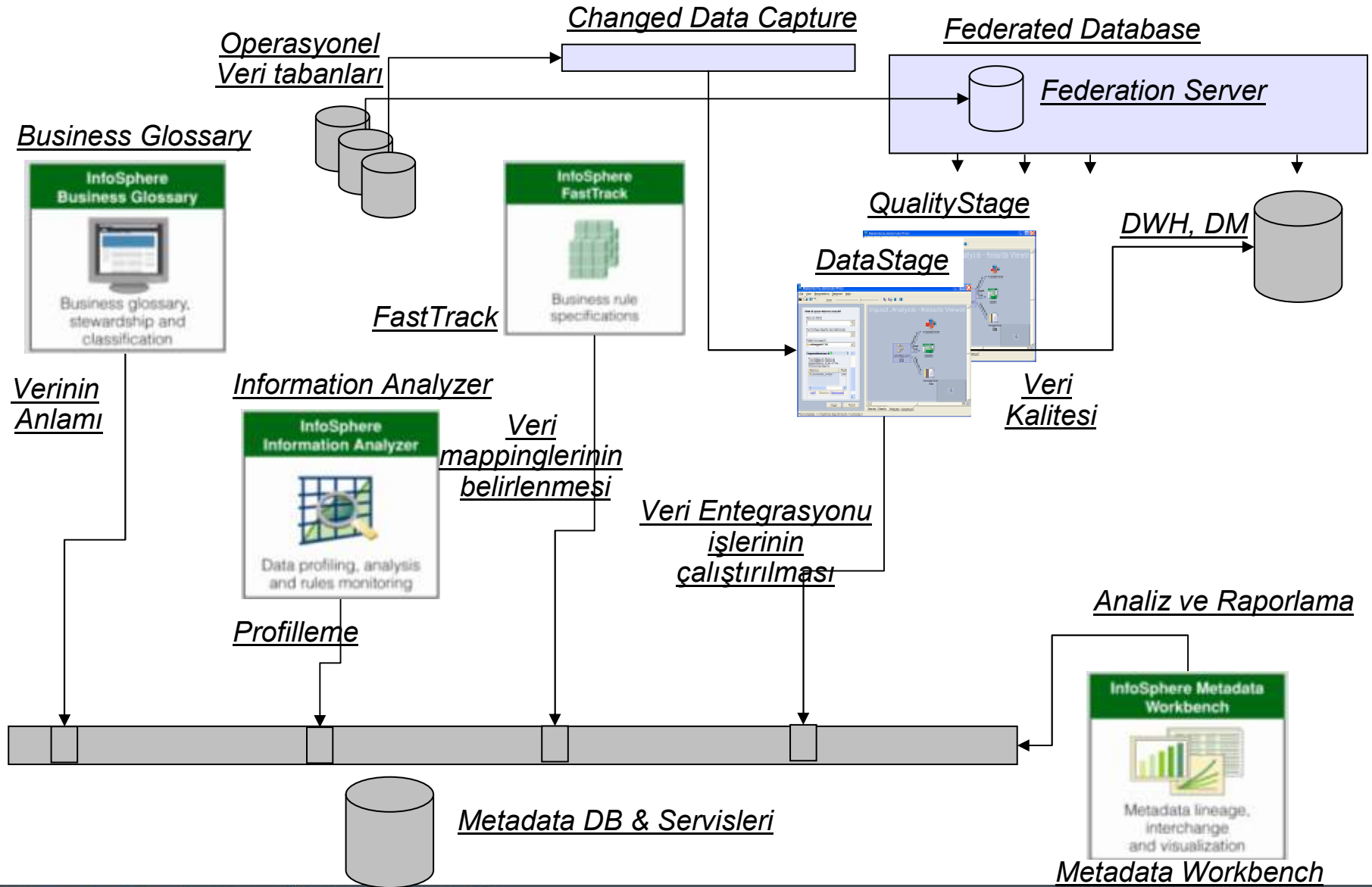
Güvenebileceğiniz Bilgi Sunar



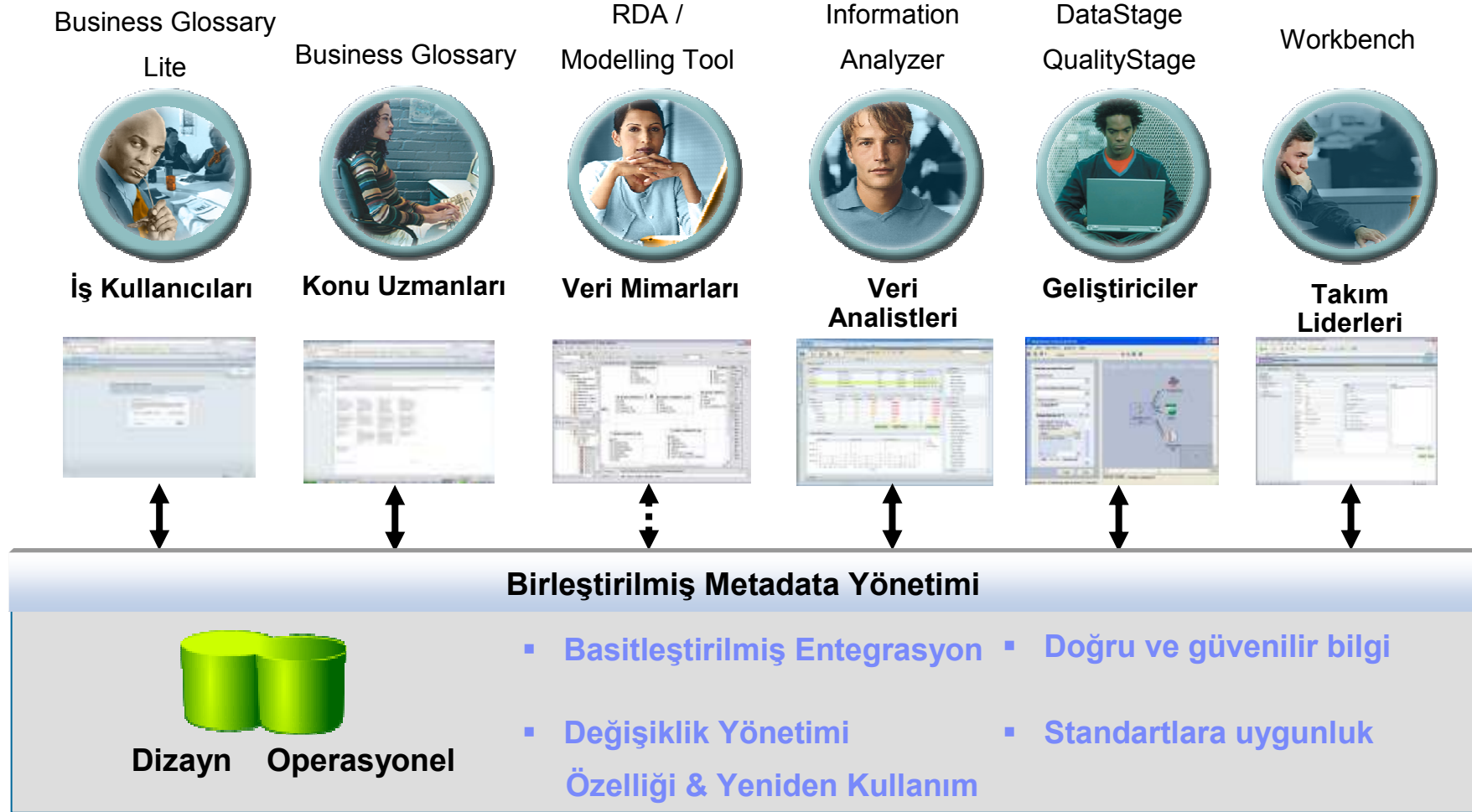
IBM Information Server Bileşenleri

- **Information Analyzer**
 - ▶ Veri Analizi, raporlama
- **DataStage**
 - ▶ Verilerin ETL süreçlerini yönetir
- **QualityStage**
 - ▶ Veri standartlaştırması ve temizliği
- **Business Glossary**
 - ▶ Veri Sahipliği, Metadata Yönetimi
- **Metadata Workbench**
 - ▶ Metadata Arayüzü, Etki Analizleri, Veri Akışı Analizleri
- **Fast Track**
 - ▶ Hızlı ETL geliştirme ve proje yönetimi
- **Change Data Capture**
 - ▶ Değişen veriyi yakalama, realtime datawarehouse
- **Information Services Director**
 - ▶ Web Servis ile kurumsal veri katmanı oluşturulması





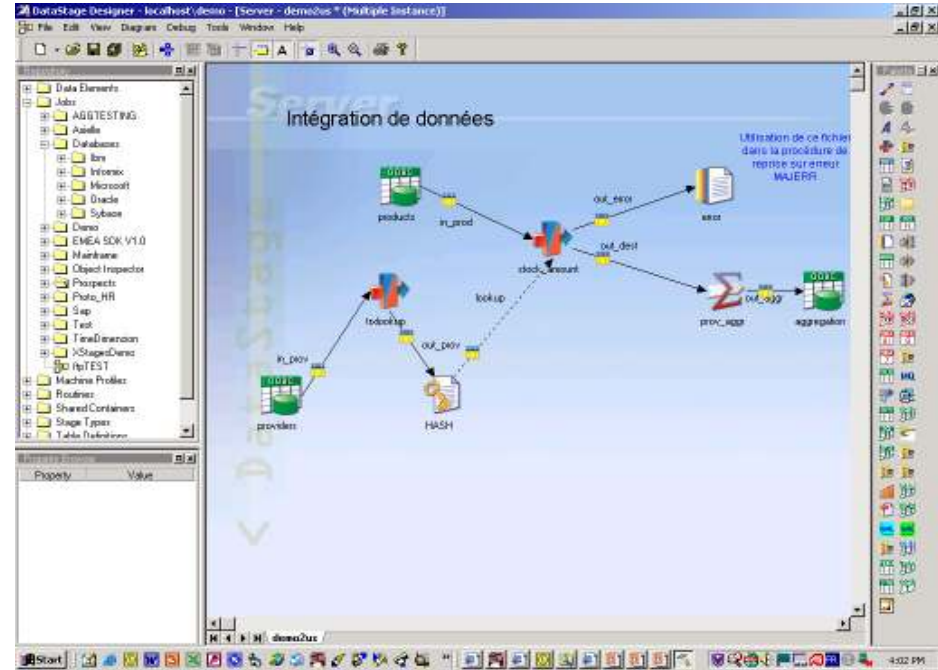
Geliştirilmiş İş Birliği & Birleştirilmiş Metadata ile Rol Tabanlı Uygulamalar Dizisi



DataStage

DataStage Akış Modeli

- Düşündüğünüz Gibi Çalışır
 - ▶ Kod Yazmanıza Gerek Yok
 - ▶ Otomatik Dökümantasyon
 - ▶ Esneklik
 - ▶ Tekrar Kullanılabilen Bileşenler (Reusable Component)
- Güçlü Konnektörler
 - ▶ ODBC, Native, Bulk Loaders
 - ▶ [XML](#), [Mainframe](#), [CDC](#), [MQ Series](#), [Applications](#), [Web Services](#)
- Çoklu Kaynak ve Çoklu Hedefi Bir Kere İşler
 - ▶ Otomatik Bölümlenme (Automatic partitioning)
 - ▶ Bellekte Lookup



Bazı Popüler Adımlar(Stages)



- Geleneksel ETL Kaynak ve Hedefleri:
 - *RDBMS, Sequential File, Data Set*



- Verileri Birleştirme:
 - *Lookup, Joins, Merge*
 - *Aggregator*



- Veri Döküşümü:
 - *Transformer, Remove Duplicates*



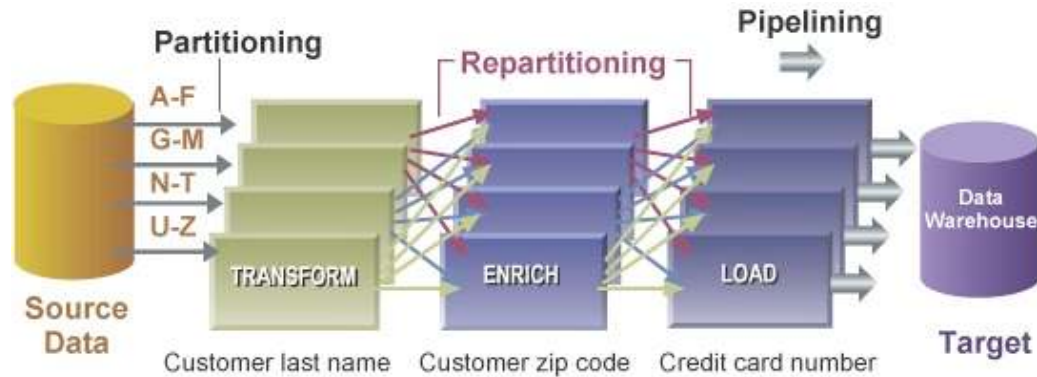
- Yardımcı Fonksiyonlar:
 - *Row Generator, Peek, Sort*

Performans: Paralel Çalıştırma

Mevcut Sistemler



Gerçek zamanlı Enterprise Gereksinimi



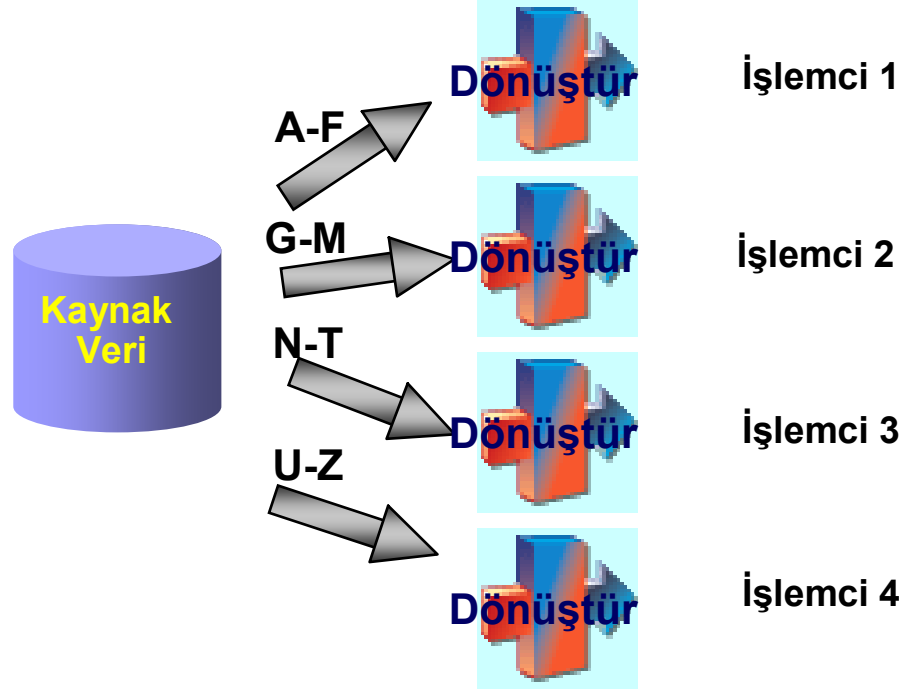
Kritik Problemler

- Servis seviyesi anlaşmaları
- Beklenmeyen İstekler
- Donanım yatırımı optimizasyonu

Alternatif Çözümler

- Manuel Parçalama
- Hand Kod Besleme
- Tek Sıra Darboğaz
- Parçalar arasında diske çıkma
- Donanıma, birleşmiş değişikliklere göre program değişiklikleri
- Hand kod paralel veritabanı ve parçalanmış şemalar

Veri Bölümleme

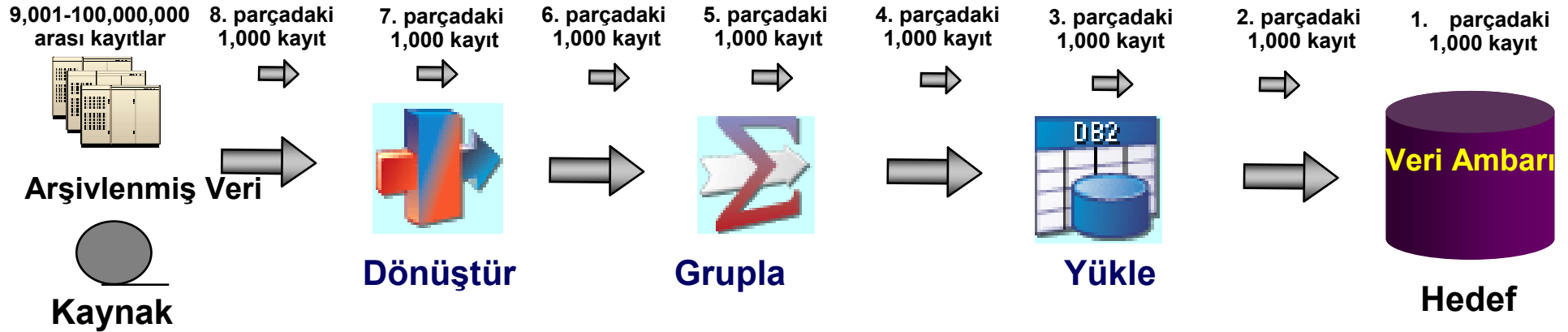


- Büyük veriyi parçalara ayırmak
- Her işlemci üzerinde bir parça çalıştırmak
- 4 işlemci 4 katı hız; 100 işlemci 100 kat daha hızlı
- Bu tamamen paralel database lerin çalışmasıdır!

Ardışık Akan Veri

- *Dönüştürür, temizler ve aynı işlemci üzerinde gerçek zamanlı olarak yükleme işlemlerini çalıştırır.*
- *Akış vasıtasıyla kayıtlar ileri taşınır.*

Operasyonel Veri

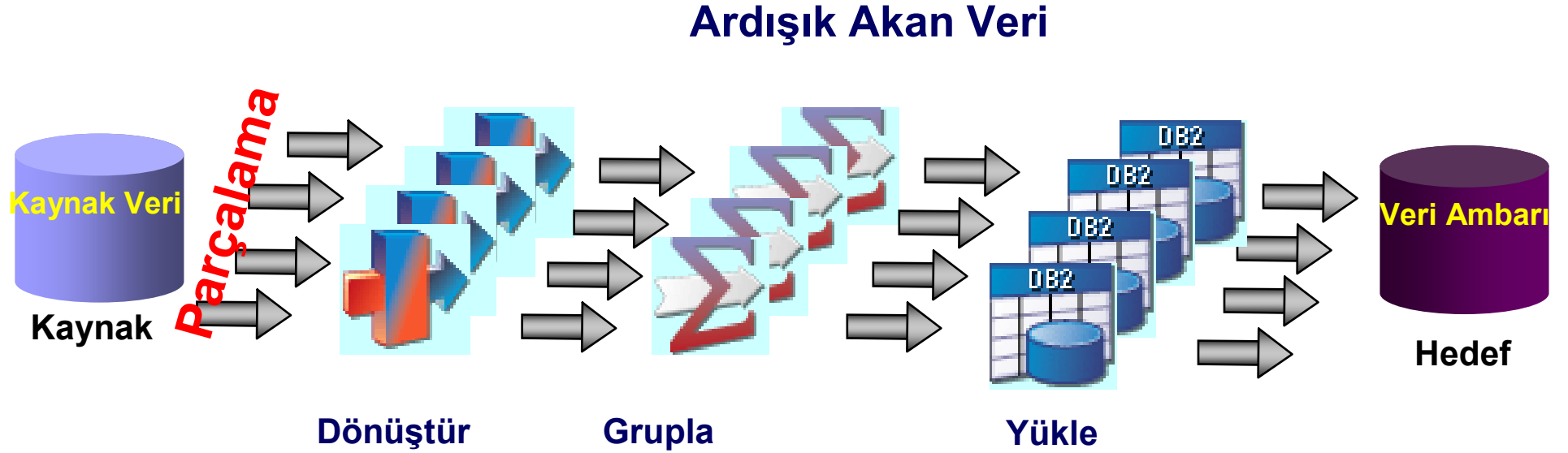


- İşlemler arasındaki diskten okuma ve diske yazmayı ortadan kaldırır
- Bir yazma işlemi hala devam ederken, bir okuma işlemi başlatır.
- Büyük veriler için kritik olan veriyi ara diskte tutma konusunu ortadan kaldırır.
- Bu aynı zamanda işlemcileri meşgul tutar. (This also keeps the processors busy)
- Ölçeklenebilirlikte limitlere sahiptir.

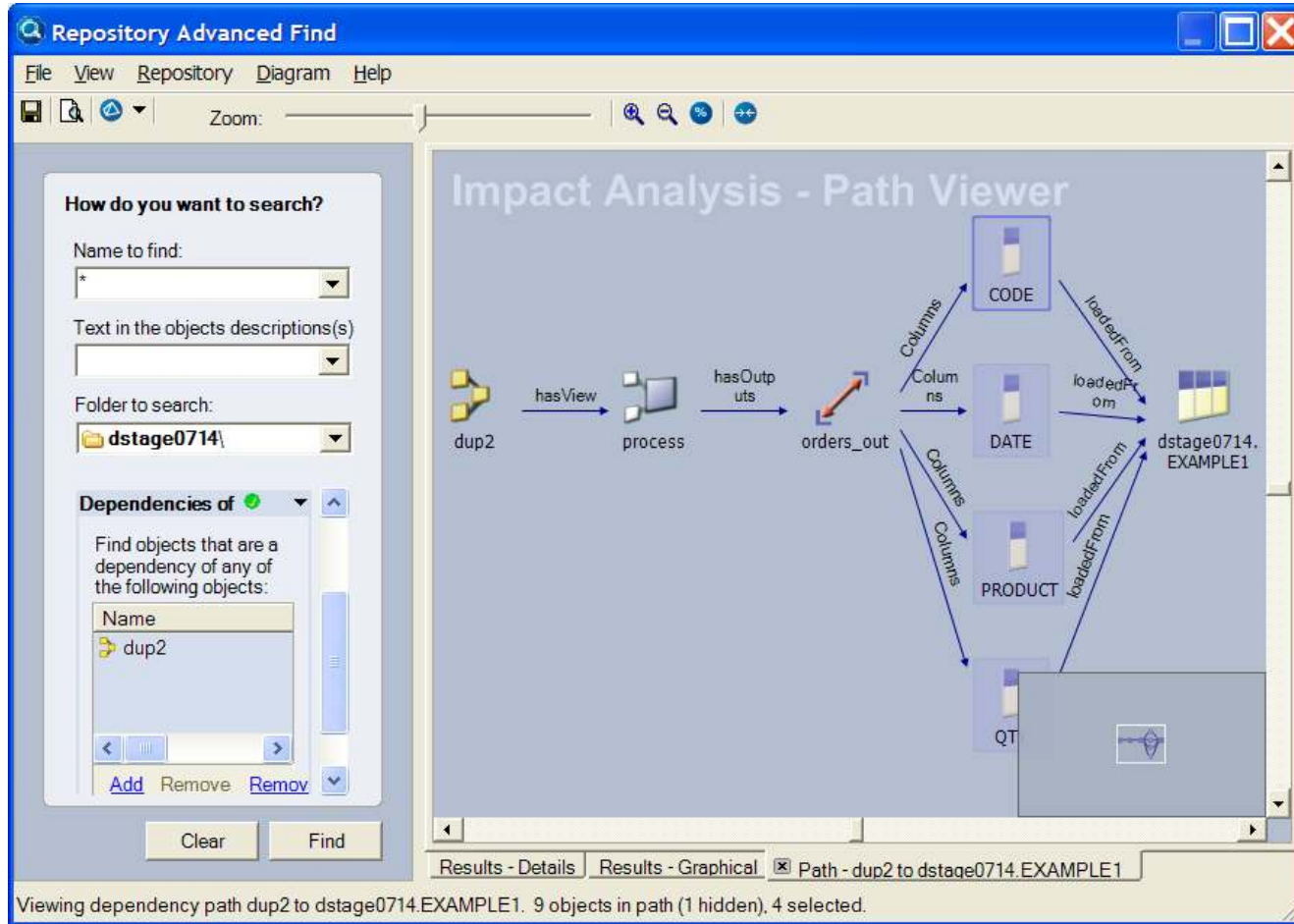
İşlemden işleme kayıtları taşıyan bir taşıma bandı düşünün.

Paralel Veri Akışı

Hep Birlikte Akışa Bakalım



Etki Analizi – Yol Görünümü

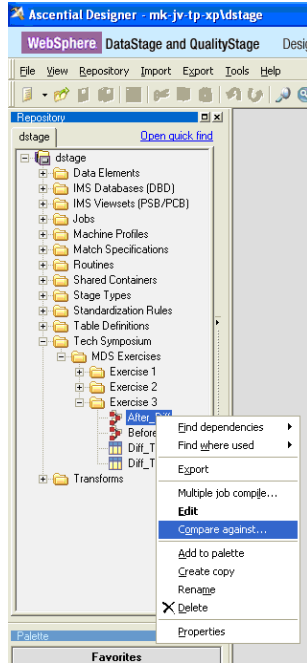


- Sonuç listesinden bir nesnenin seçimine izin verir ve sonra o nesnenin bir akışta nerede ve nasıl kullanıldığını gösterir.
- Nesne editörü bu slayttan çıkmış olabilir.

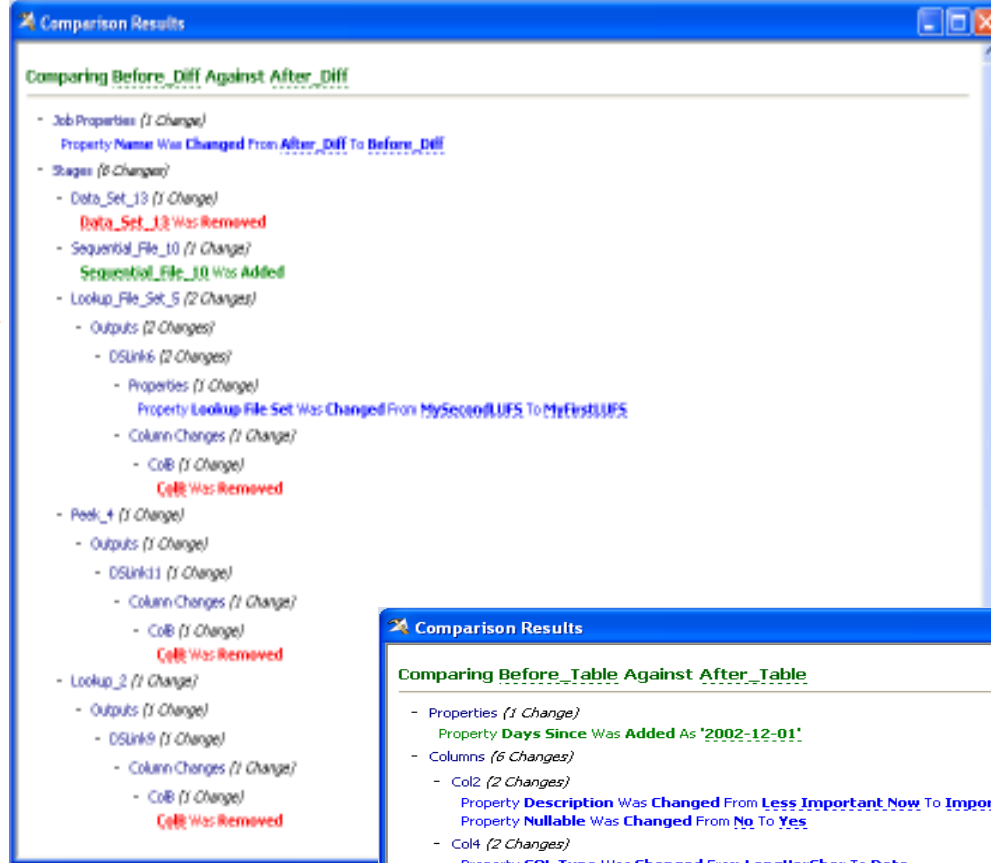
E.g. [dstage014.EXAMPLE1](#) den gelen Job [dup2](#) bir [process](#) adımına sahip, 4 kolonu olan bir [orders_out](#) çıkış linki var

Job, Tablo veya Rutin Farkı

Job, Tablo ve Rutinler için Kullanılabilenler



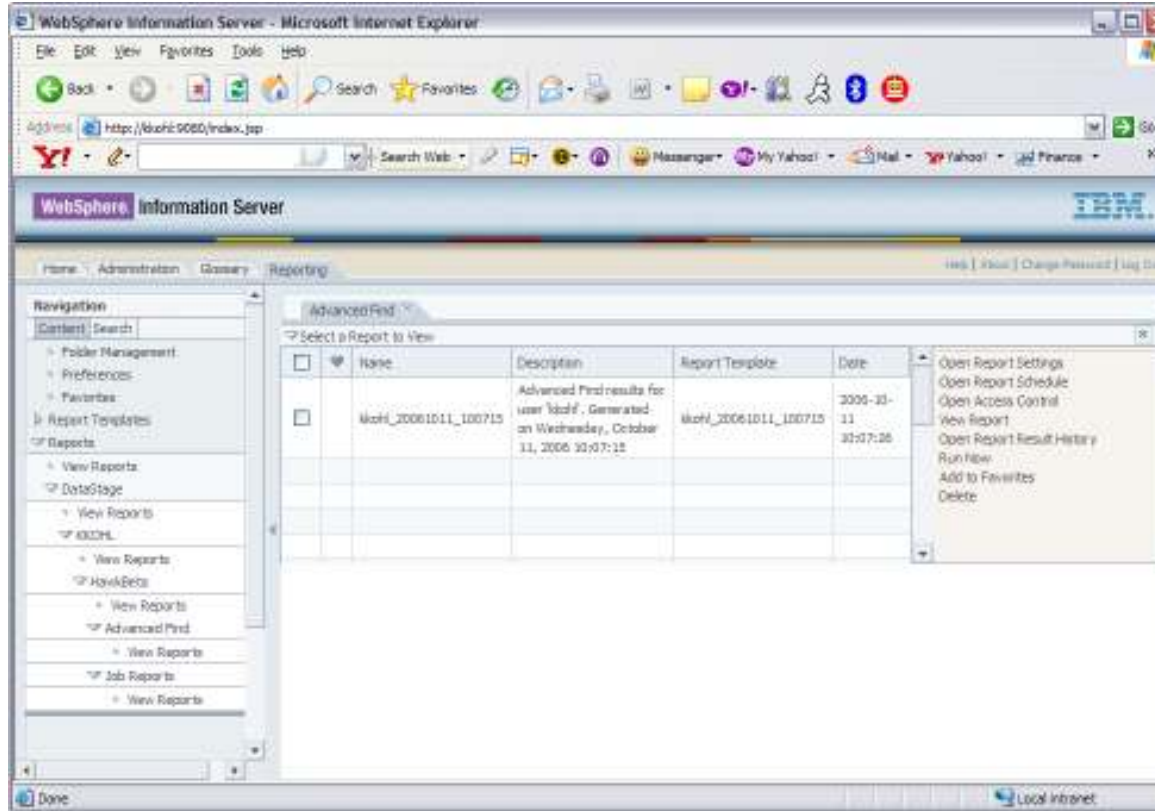
Designer daki ilişkili editörlere geçiş bağı ile metinsel rapor.



Tablolar →



Raporlama Konsolu



- Raporlar DataStage den IBM Information Server Raporlama Konsoluna yayımlanabilir.
- Job Raporları, Geliştirilmiş Arama, Etki Analizleri, etc.

Kaynaktan Hedefe ve Hedeften Kaynağa

Source to Target Report

Report Name : Source and Target by Job
 Date/ Time Executed : 08/17/2006 19:20:38
 Customer Description : Report on Source and Target Details by Job

Target Fields of email

employee_number	A unique number that identifies the employee.
employee_first_name	Employee's name.
store_number	A unique number that identifies the store.
email	Email address of the employee.

Report Name : Source and Target by Database
 Date/ Time Executed : 08/17/2006 19:13:16
 Customer Description : Report on Source and Target Details by Database

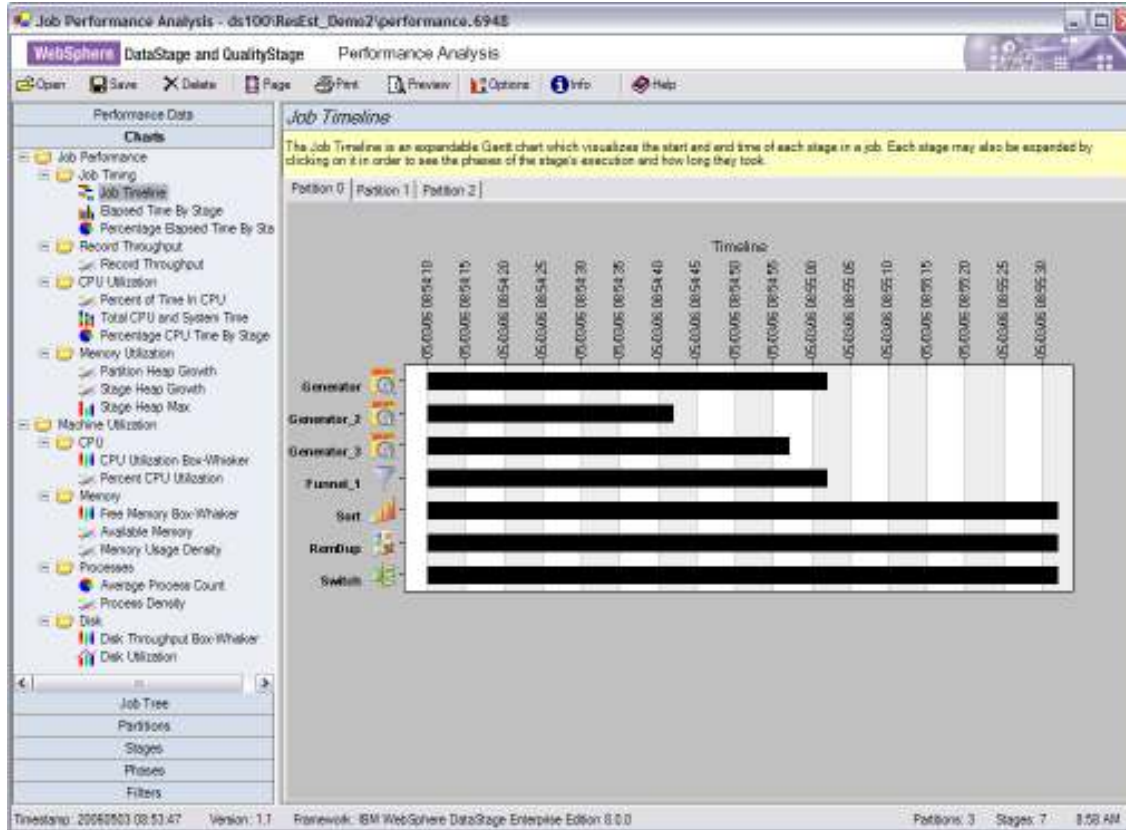
Target Fields of email

Source DB	Source Table	Source Field	Derivation	Target Field	Target Table	Target DB
DSJob_1 Job						
ERwin	EMPLOYEE	email	Derivation1	movie_url	MOVIE	ERwin

Target Fields of employee_number

Source DB	Source Table	Source Field	Derivation	Target Field	Target Table	Target DB
DSJob8 Job						
ERwin	EMPLOYEE	employee_number		movie_number	MOVIE	ERwin
DSJob_1 Job						
ERwin	EMPLOYEE	employee_numberDerivation3		movie_number	MOVIE	ERwin
DSJob_2 Job						
ERwin	EMPLOYEE	employee_numberDerivation2		employee_numberPAYMENT		ERwin

Job Performans Analizi



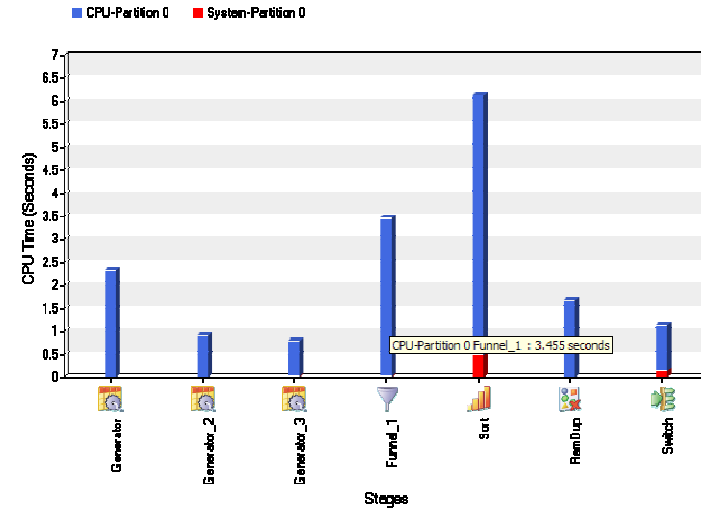
Uygulamaya Bir Bakış:

- Job çalışma davranışlarına daha derin bir kavrama sağlar.
- Çeşitli kategoriler sağlar:
 - Kayıt İşlem Hacmi
 - CPU Kullanımı
 - Job Zamanlaması
 - Job Bellek Kullanımı
 - Fiziksel Makina Kullanımı
- Dizayn taslağı üzerinde yeralan müşteri adımlarının vurgulanmasıyla çalışma karmaşıklığını saklar

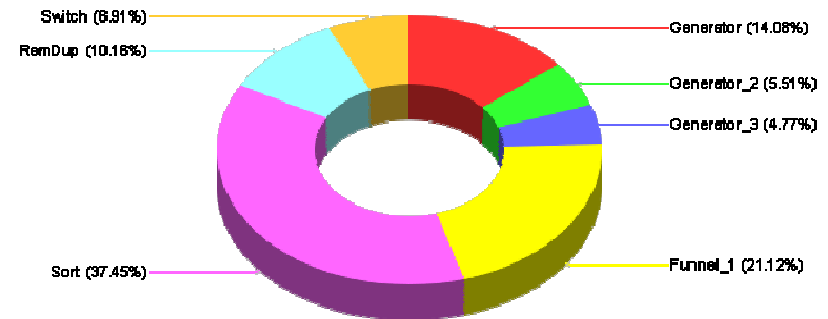
CPU Kullanımı

- Her işlemcinin CPU daki zamanını gösterir.
- Çalışma esnasında farklı noktalarda CPU daki hangi işlemcilerin baskın olduğunu gösterir.
- Yüzde Grafiği Percentage view shows what percentage of the CPU load of the job each stage on the canvas was responsible for.
- Inserted operators and Composite sub-operators automatically get bundled up in these results.
- Gelişmiş kullanıcılar her işlemi ve içerilen adımları yansıtan bu grafikteki değişiklik seçeneklerini görebilir.

Toplam CPU ve Sistem Zamanı

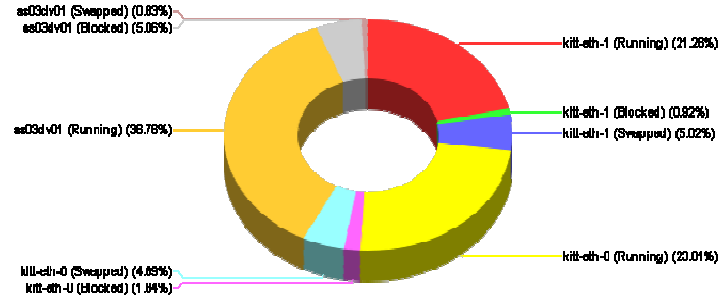


Pasta grafiğinde CPU Yüzdesi

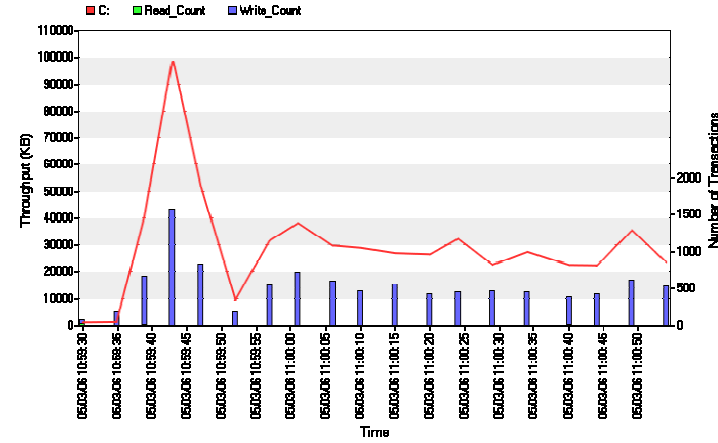


Fiziksel Makina Kullanımı

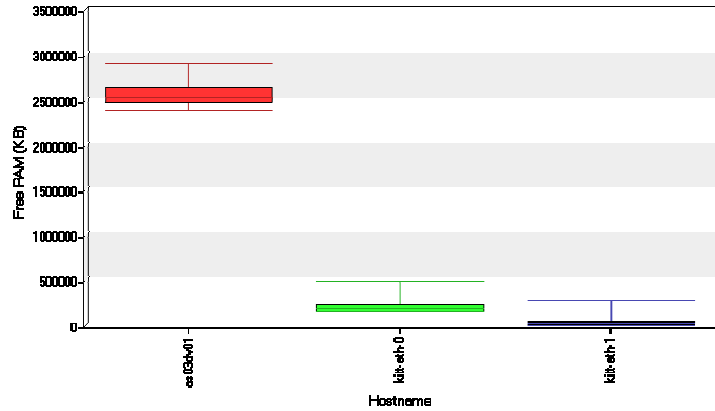
İşlem Dağılım Ortalaması



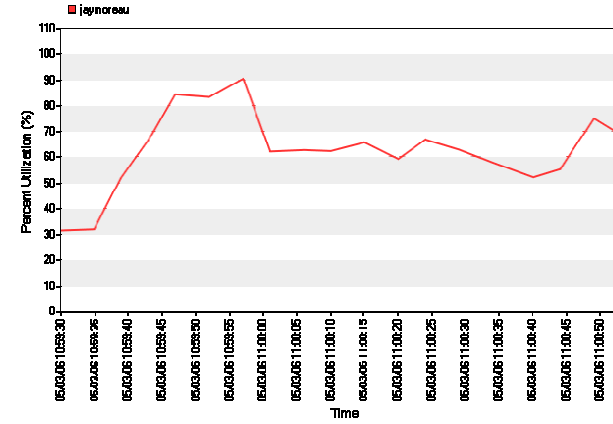
Disk Verimliliği



Boş Bellek Favori Kutusu



CPU Kullanım Yüzdesi



Kaynak Tahmin Uygulamasına Bakış

The screenshot shows the 'Resource Estimation' application window. The main panel is titled 'Partition Overview' and contains a text description: 'The Partition Overview Panel describes the total predicted utilization of each Model given the current selected Input Projection. Each model can be compared by clicking on the tabs for each model. Totals for each stage running on the partition are also displayed, allowing the user to see which stages were responsible for the usage of resources.'

Below the text, there are four summary fields for 'Partition 0':

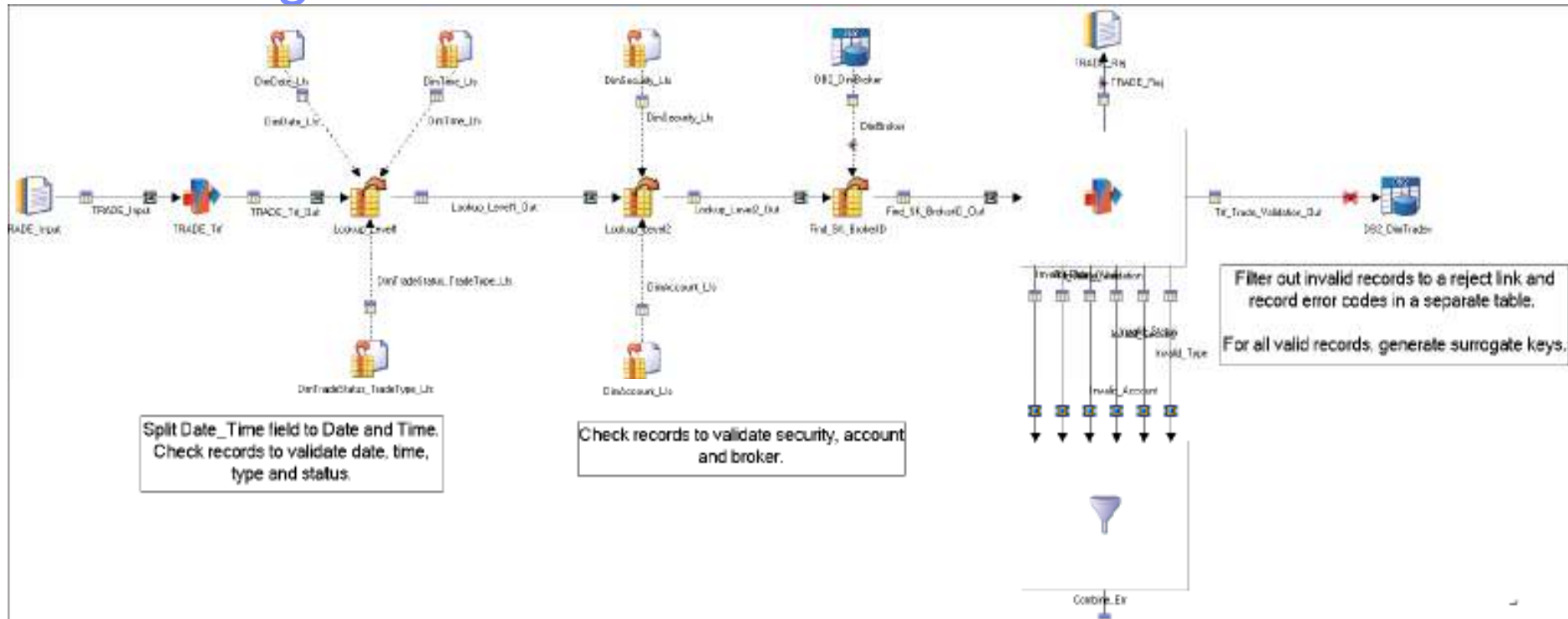
- Input Data (mb): 572.205
- CPU (sec): 29.1846
- Disk (mb): 42.875
- Scratch (mb): 186.378

A 'Partition Utilization' table is also displayed, showing the resource usage for various stages:

Stage	Input Data (mb)	CPU (Sec)	Disk (mb)	Scratch (mb)
Generator_2	190.735	5.487	0	0
Generator	190.735	5.628	0	0
Generator_3	190.735	5.037	0	0
Funnel_1	0	3.50608	0	0
Sort	0	6.54592	0	186.378
RemDup	0	2.17752	0	0
Switch	0	0.803104	0	0
/tmp/dataset1	0	0	21.5	0
/tmp/dataset2	0	0	21.375	0

The interface also includes a 'Job Tree' on the left side with a hierarchy of models and partitions, and a status bar at the bottom showing 'Input Projection: default' and the time '1:55 PM'.

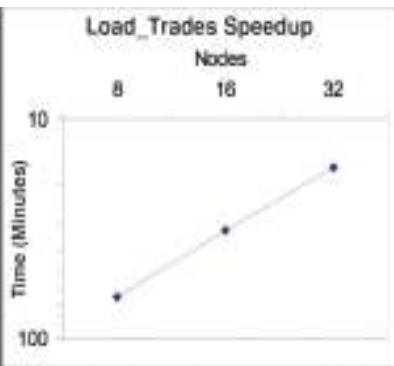
DataStage Performans Örn. Load Trades



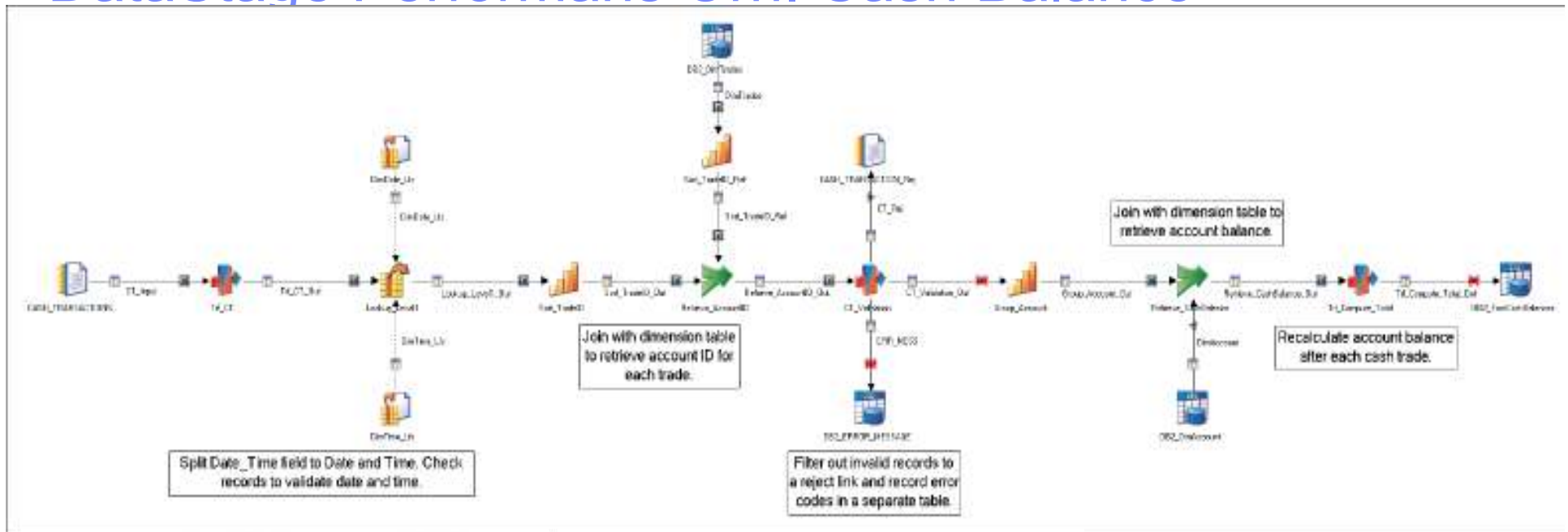
Split Date_Time field to Date and Time.
Check records to validate date, time, type and status.

Check records to validate security, account and broker.

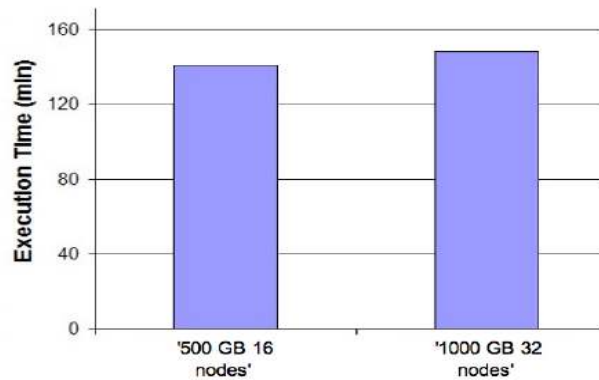
Filter out invalid records to a reject link and record error codes in a separate table.
For all valid records, generate surrogate keys.



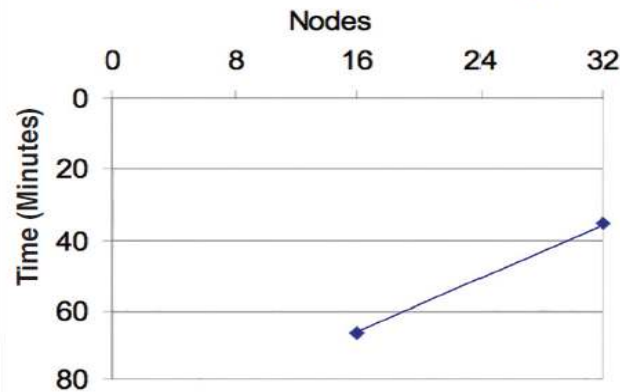
DataStage Performans Örn. Cash Balance



Load_CashBalances Scalability

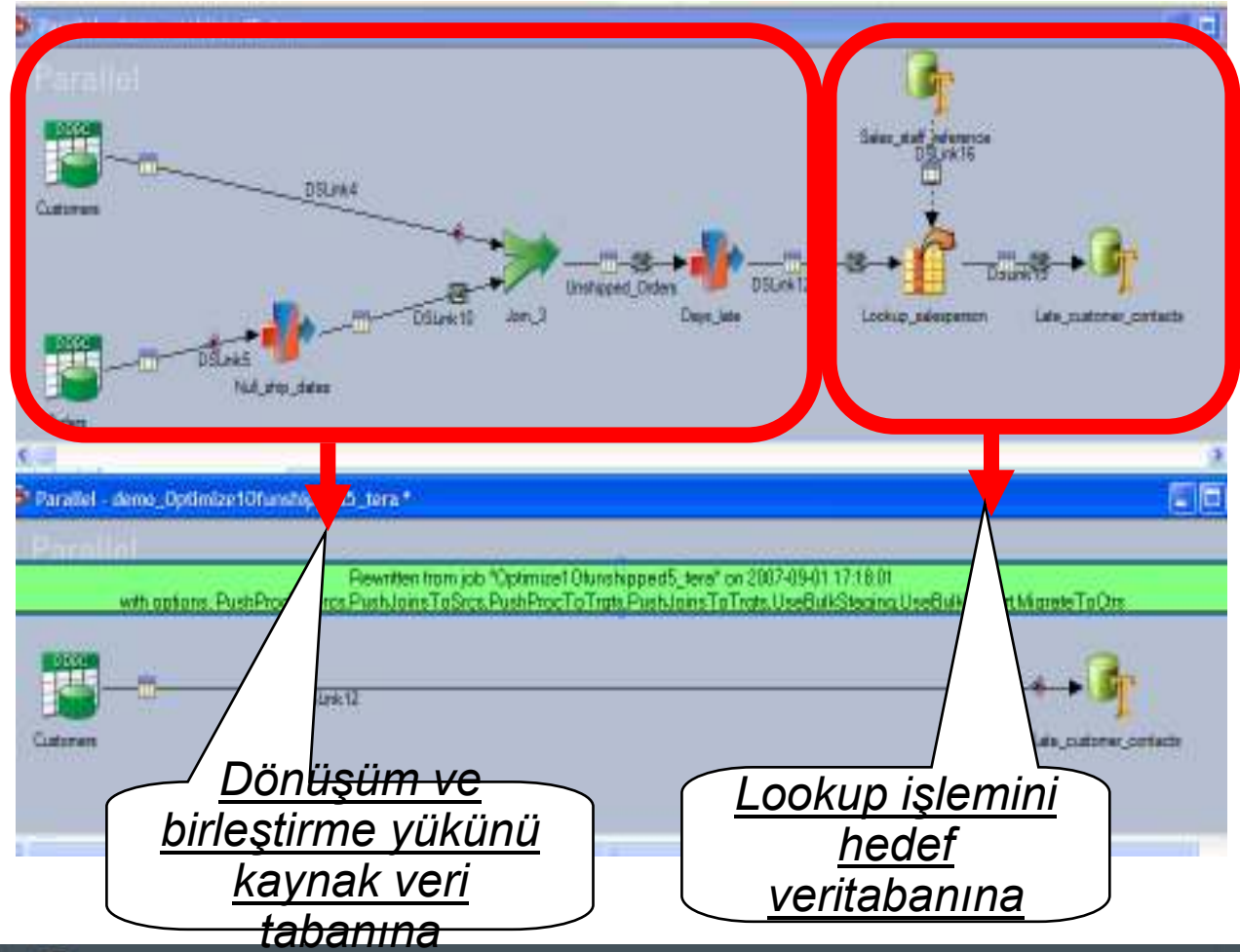


Load_CashBalances Speedup



InfoSphere DataStage Balanced Optimization To Leverage your existing IT infrastructure

- **Minimum I/O**
 - ▶ Kaynakta veriyi işleme
 - ▶ Avoid target extracts
- **Minimize copy & move**
- **Maximize optimization within sources or targets**
 - ▶ İndex ve veritabanı özelindeki özelliklerin kullanılabilmesi
- **Maximize parallelism**
 - ▶ I/O from/to veritabanları
 - ▶ DataStage parallel engine
 - ▶ İşlem yükü veri tabanına gönderilir



Kullanılabilirlik: Metadata Paylaşımı

DataStage, QualityStage & Information Analyzer

- InfoSphere Information Analyzer ile metadata paylaşımı
 - ▶ Her iki uygulama da metadata tablolarını ortak veri havuzunda saklar.
 - ▶ DataStage kullanıcıları Information Analyzer dan metadata tablosunu görebilir .
 - Metadata tanımlarının paylaşılmasına izin verir
 - Her iki uygulamadaki kullanım için veri kaynağından tek metadata yüklemesi sağlar.
 - Paylaşılmış tablolar için IA analiz verisini görmesi için DS kullanıcılarını aktif yapar.

- DS/QS Designer dan IA analiz bilgisine nereden ulaşılır?

- ...bir Tablo Tanımı
- ...bir Adım(stage) Editörü nden bireysel bir kolonun detaylarına baktığımızda EditRow üzerindeki “Analytical Information” sekmesi
- ▶ Tablo Tanımlamadaki “Analytical Information” sekmesi



Bussiness Glossary – Neden İhtiyaç?

- Kurumlardaki uzmanlıklar veriyi ve problemleri anlamakta, sonuçları yorumlamakta kritik rol oynarlar
- Yetersiz uzmanlık ve bilgi birikimi düşük veri kalitesinin başlıca nedenidir, veri kullanılmaz hale dönüşür
- Genellikle çalışanların hafızalarında – nadiren dokümante edilir
- Kurum içerisinde dağılmış haldedir
- İşten ayrılmalarda ve proje değişimlerinde kaybolur
- Dokümante edilmemişse, bilgi ve yorumlama bozulmakta ve zamanla bulanıklaşmaktadır



Uygulama
Geliştirici



Son
Kullanıcılar



İş Analisti



Veri Mimarı



Yazılım Mimarları



IT Yöneticisi



Veri
Yöneticisi

İş Ana Verisi: Business Glossary

- Web tabanlı iş ana verisinin paylaşımı, yönetimi ve yazılması
- IT'nin çalışmalarını iş hedefleri ile eşleştirir
- IT değerlerine iş içeriğinin eklenmesini sağlar
- Sahiplik ve sorumluluk tanımlamalarını yapar

Database = DB2

Schema =
NAACCT

Table =
DLYTRANS

Column =
ACCT_NO

data type =
char(11)



Teknik



İş

Müş. Hesap No

On haneli hesap numarası.
ACCNO ve HESNO diye de adlandırılır L-FIIIVVVV formatında yazılır



Konu Uzmanları



İş Kullanıcıları

Anlamak



Business Glossary

İş terimleri kütüphanesini yaratır ve yönetir. İş terimlerinin ilişkilerini tanımlar, fiziksel dizayn ile iş terimlerini ilişkilendirir

IBM

USA: Finance, Manufacturing, Customer Relationship

Europe: Finance, Manufacturing, Customer Relationship

Data Model: Top-level category, which groups all the elements by type

Asia: Finance, Manufacturing, Customer Relationship

See All Categories

İş Odaklı Görünüm

Yararları: Fikirler, düşünceler ve bilgi birikimlerinin paylaşımı ile iş birliğini yaratır



Muhasebe Bölüm Kodu

SAHIBI: Muhasebe Bölümü

FORMAT: X(7)

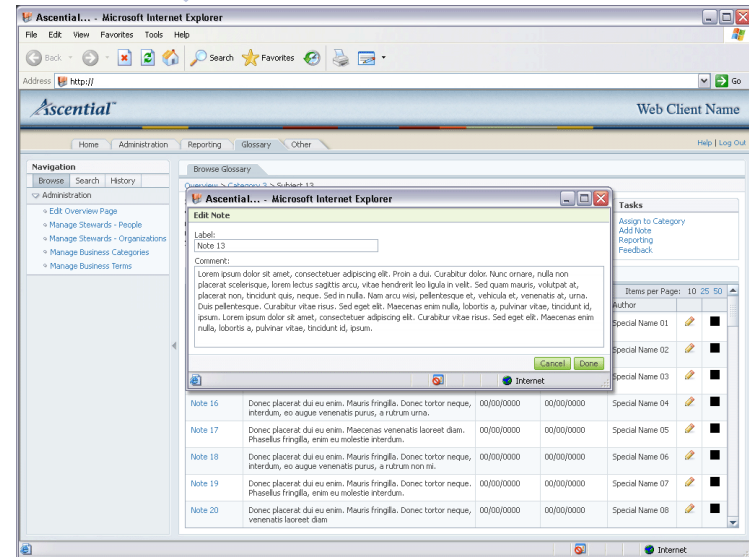
TANIM: Hesabın bağlı olduğu organizasyon bölümünü tanımlayan yedi haneli numara

Standart Tanımı Yap

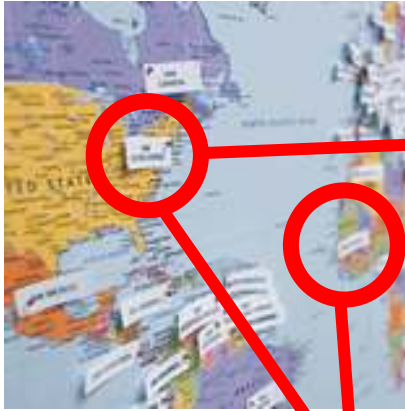


Herhangi bir nesneye not ekle

Alt bölümü tanımlayan Bölüm Kodunun son 2 iki hanesinin genelde boş olduğunu farkettim



Bilgi Varlıklarınızın Tamamına Hakim Olun



Business Glossary

- İşinizi tanımlayan terimler nelerdir?
- Ne anlam ifade ederler?
- Bu terimler birbirleri ile bağlantıları nelerdir?
- Bu terimlerin iş sahipleri kimler?

Birleşik Değer

- Bu sistemdeki veri ne anlama geliyor?
- Bu sistemdeki iş terimlerinin kuralları nedir?

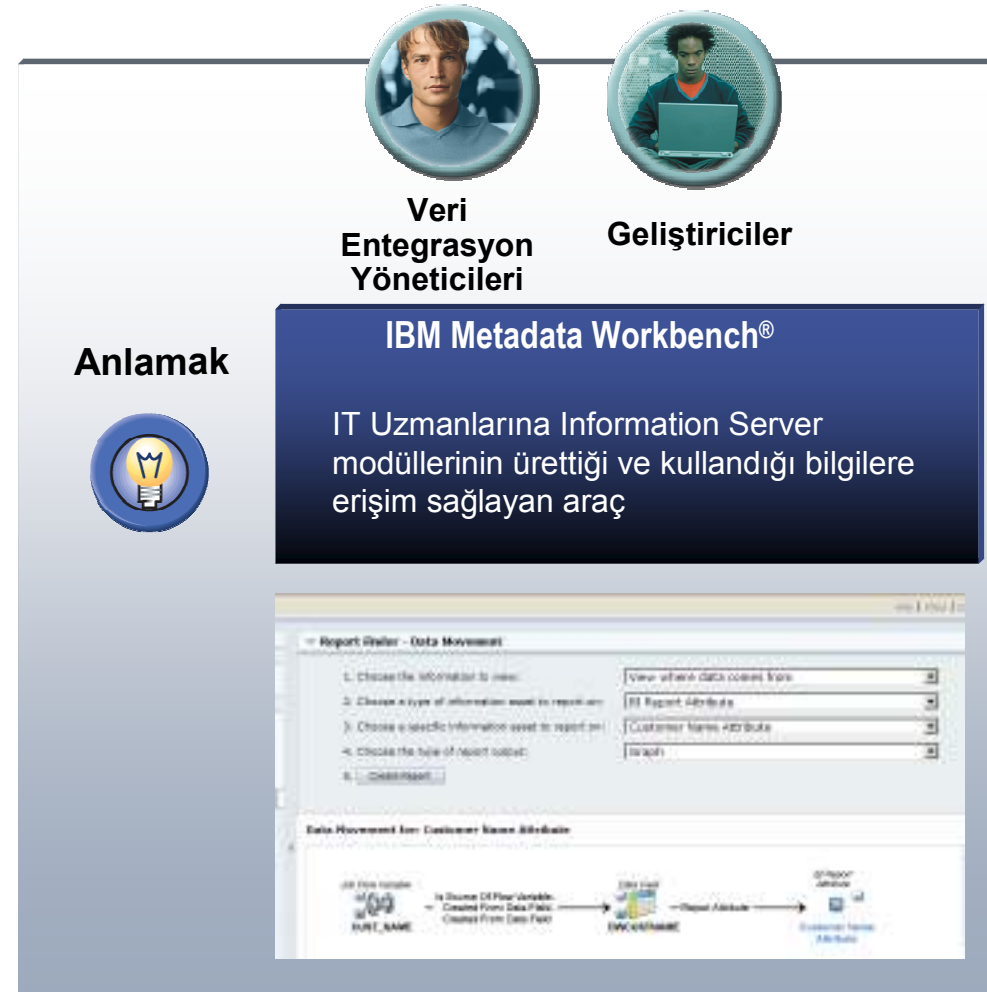


Information Analyzer

- Bu sistemdeki verinin yapısı nasıl?
- Verinin biçimi nedir?
- Veri kalitesi nasıl?
- Bu iki sistem nasıl bağlantılılar?

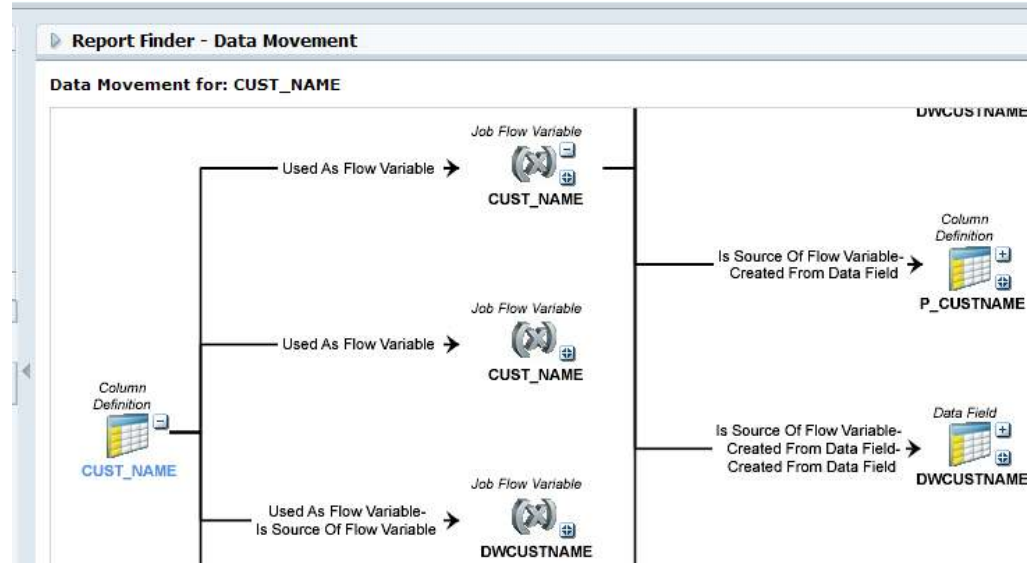
Ana Verinin İzlenmesi: IBM Metadata Workbench

- Information Server modülleri tarafından kullanılan ve üretilmiş bilgi varlıklarına erişim sağlayan web tabanlı uygulama
- Verinin aktarılmasının takibi, iş terimleri, etki ve bağımlılık analizlerine erişimi sağlar
- İş Zekası raporlarının verisinin aktarılmasının takibini yaparak Sarbanes-Oxley ve Basel II gibi kanuni zorunluluklara uyum için temel oluşturur



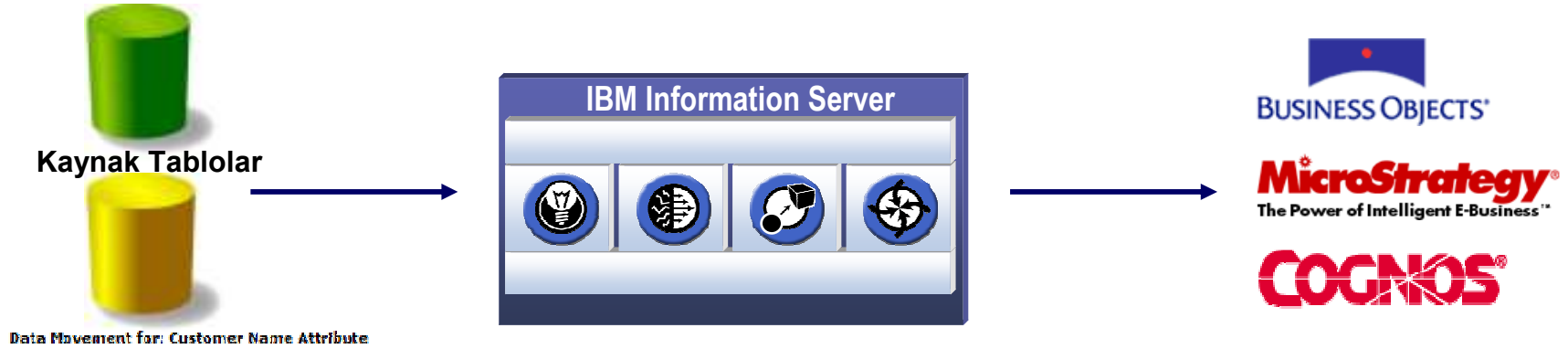
Bu kolonu deęiřtirirsem neler etkilenir?

- Deęiřiklik etkisini grafik veya liste olarak gsterir
- İř Zekası raporlarındaki etkileri de raporlar
- Herhangi bir nesne zerinde etki analizi yapılabilir
- **IT'deki deęiřimler nedeniyle oluřan maliyetleri dřrr**

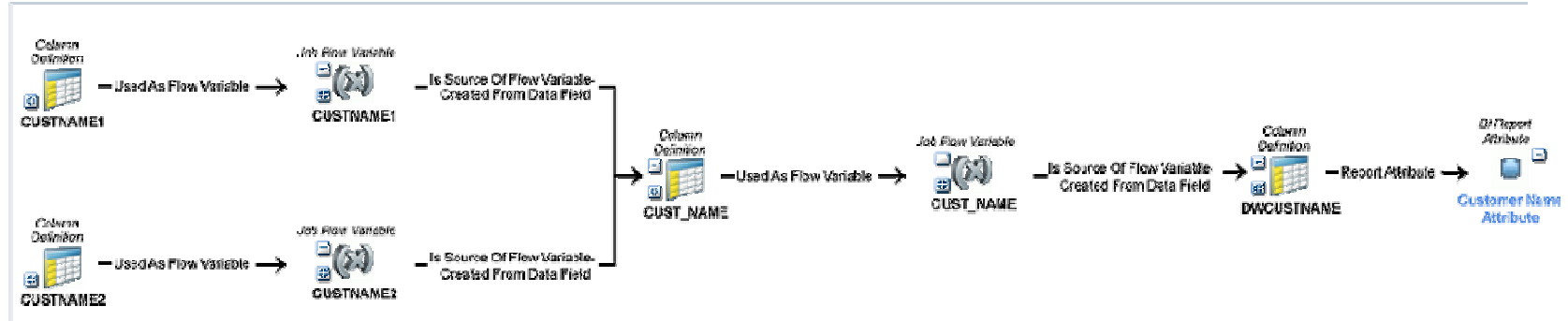


Bu rapordaki bir alanın verisi hangi kaynaktan geliyor?

- İş Zekası Raporlarının Ana Verilerinin alınması ve incelenmesi
- Rapor özniteliklerinin incelenmesi
- Görsel olarak verinin ürünler ve sistemler arasındaki geçişinin takibi
- **İş bilgilerinin anlaşılması ve güvenin artmasını sağlar**

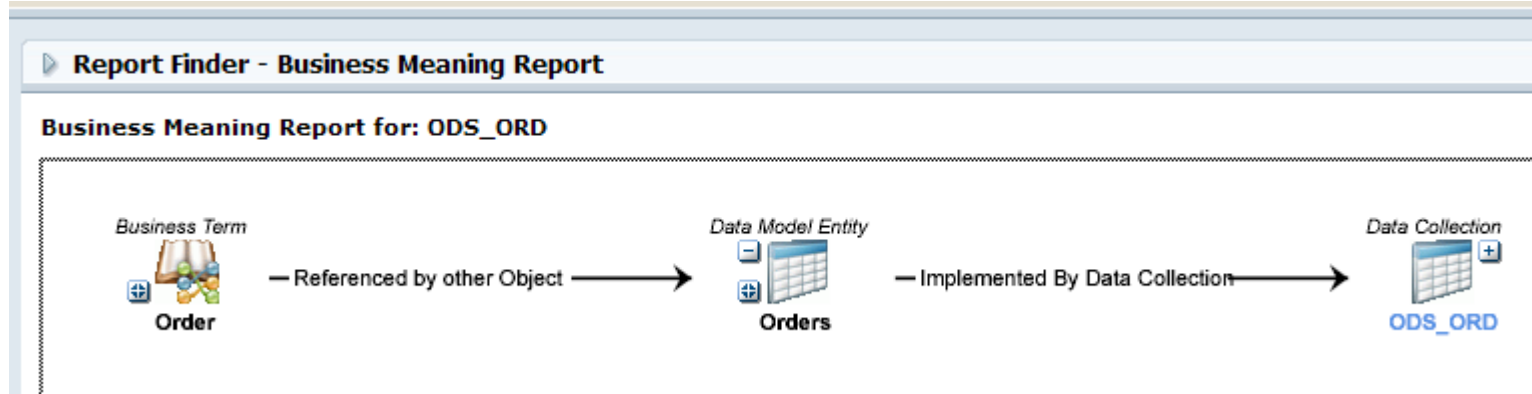


Data Movement for: Customer Name Attribute




Bu kolon ne anlama geliyor?

- İş terimleri, veri modeli nesneleri, rapor alanları arasındaki ilişkileri gösterir
- Bir alanın anlamını anlaşılır kılar
- İş terimlerinin anlamını anlaşılır kılar
- **Veri sahipliğinin etkin biçimde düzenler**

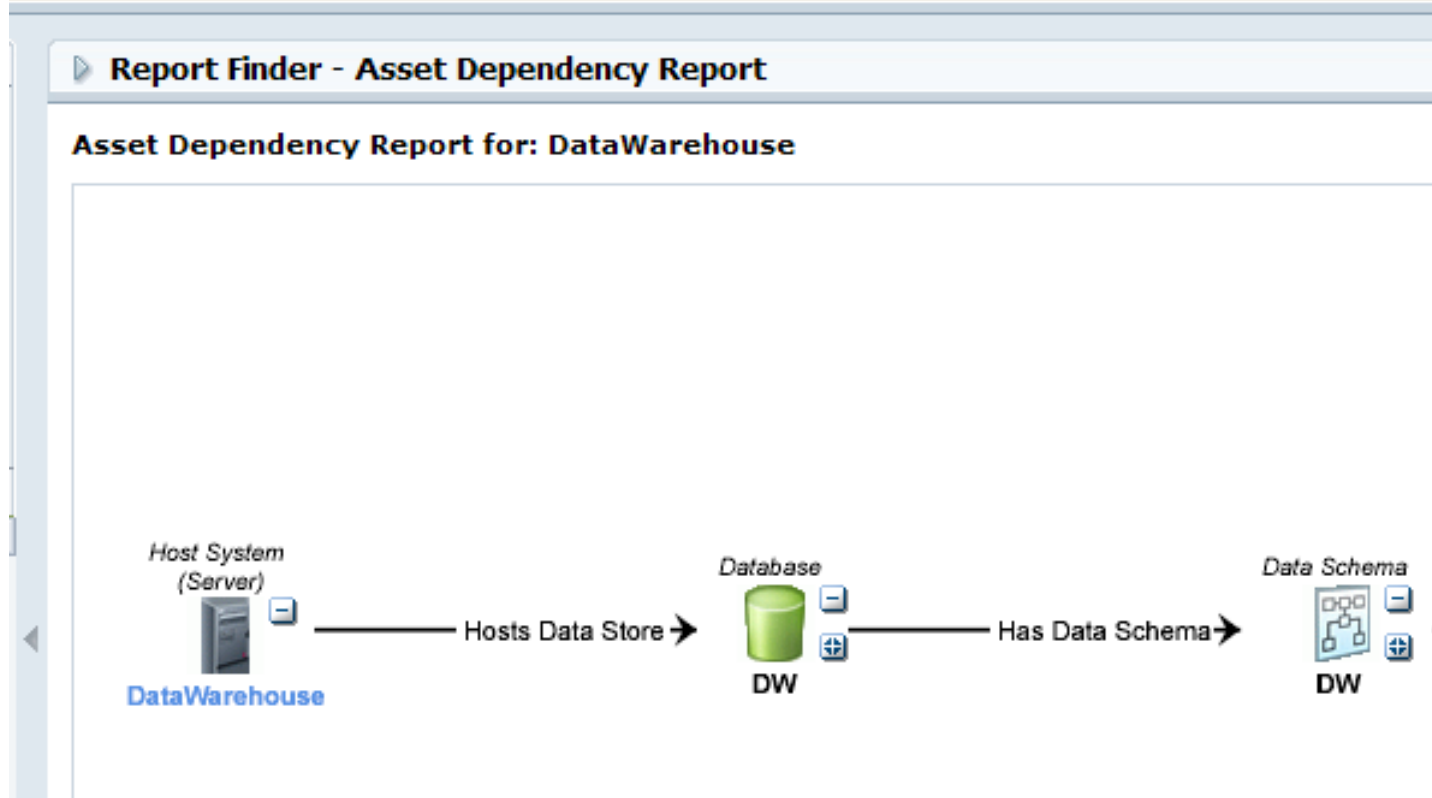


Information Server Varlıklarını İnceleme

Data Collection: DW_CUSTOMER	
Object Details	
Name	DW_CUSTOMER
Short Description	Customer Records and Information
Long Description	Customer Records stored in the Data Warehouse
Alias	None
System Object	False
Object Usage and Assignment	
Business Term	None
Pertains to Business Rule	None
Annotation(s)	None
Data Steward	 SK
Object Creation Details	
Created by User	admin

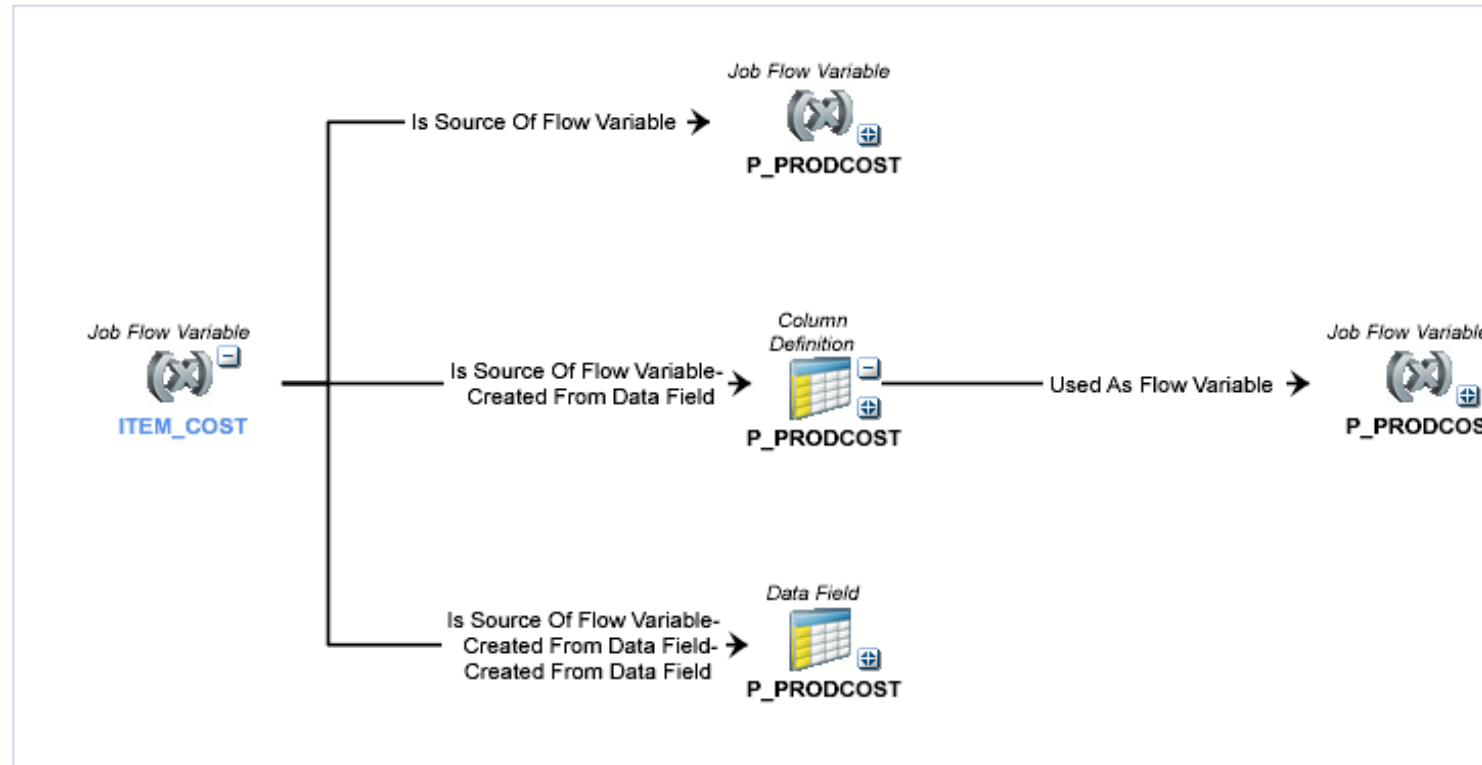
DW_Customer tablosundan aldığım veri ile ilgili bir sorunum var. Bu tablodan kim sorumlu?

Sistem Bağımlılıklarının Analizi



Veri ambarı sunucusunda çalışma yapılacak. Kapalı kalacağı süreyi kiminle koordine etmeli ve haber vermeliyim?

- Metadata Workbench Demosu



Teşekkürler

A.Okay Akyüz

okay.akyuz@osisoftware.com

OSiS
Your Information Builder

