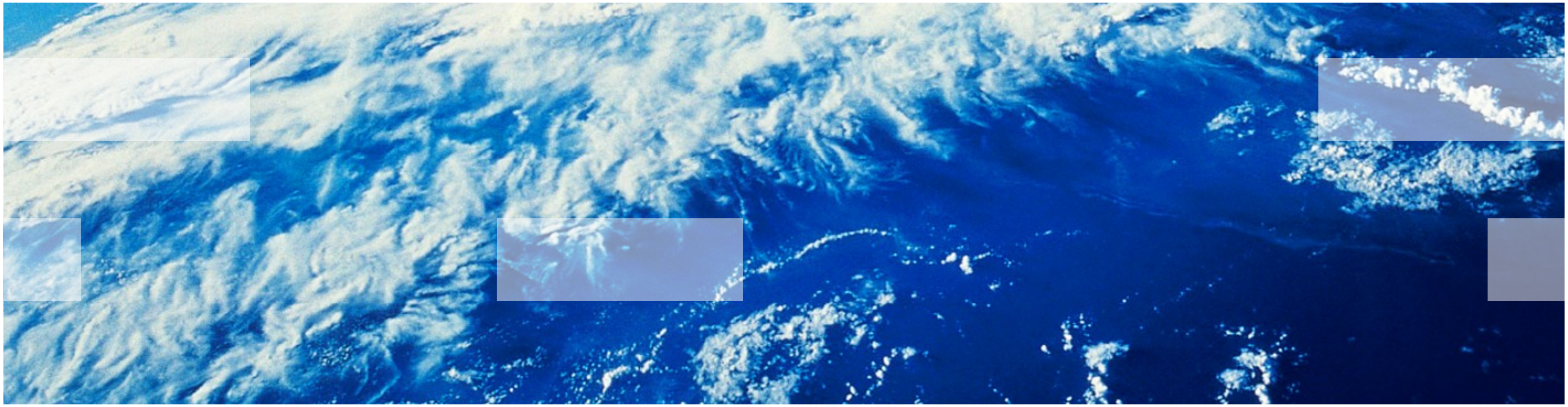
style

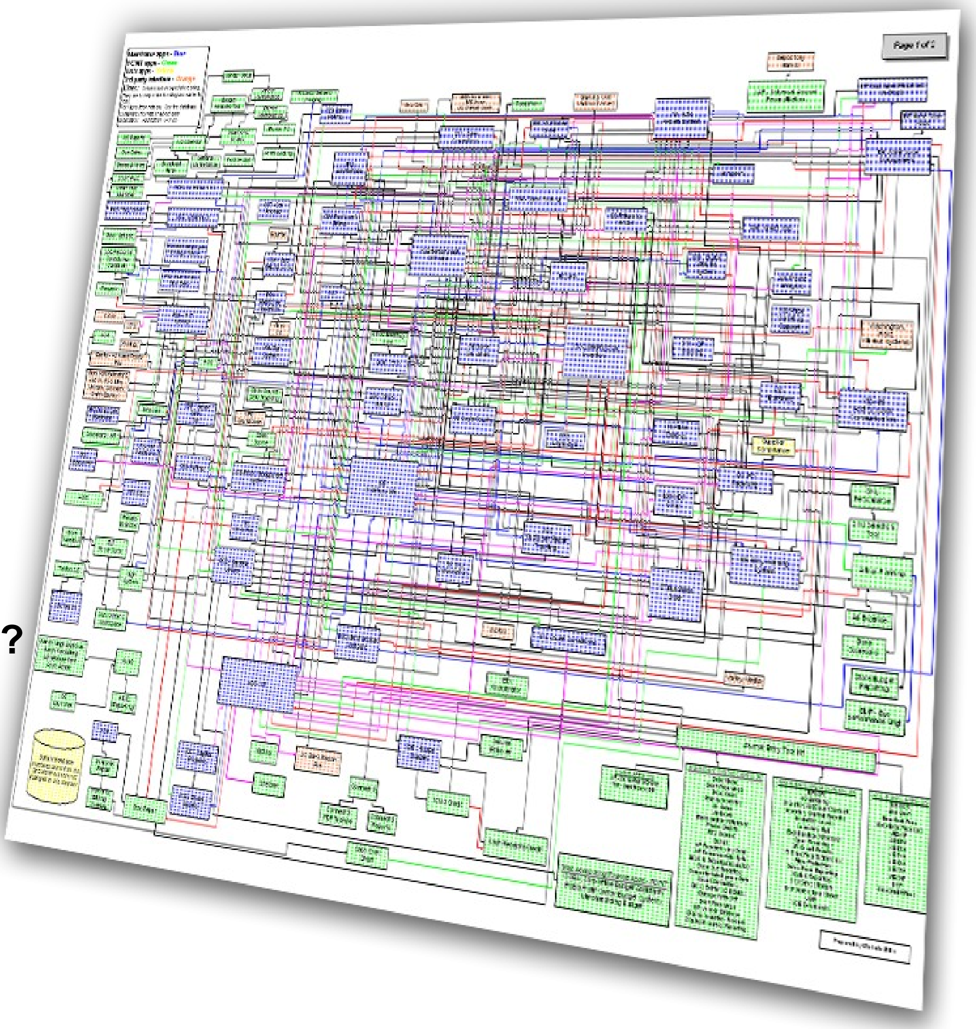
Tevfik Baskaya – Vodafone & Okay Akyüz - OSİS

22/10/2009

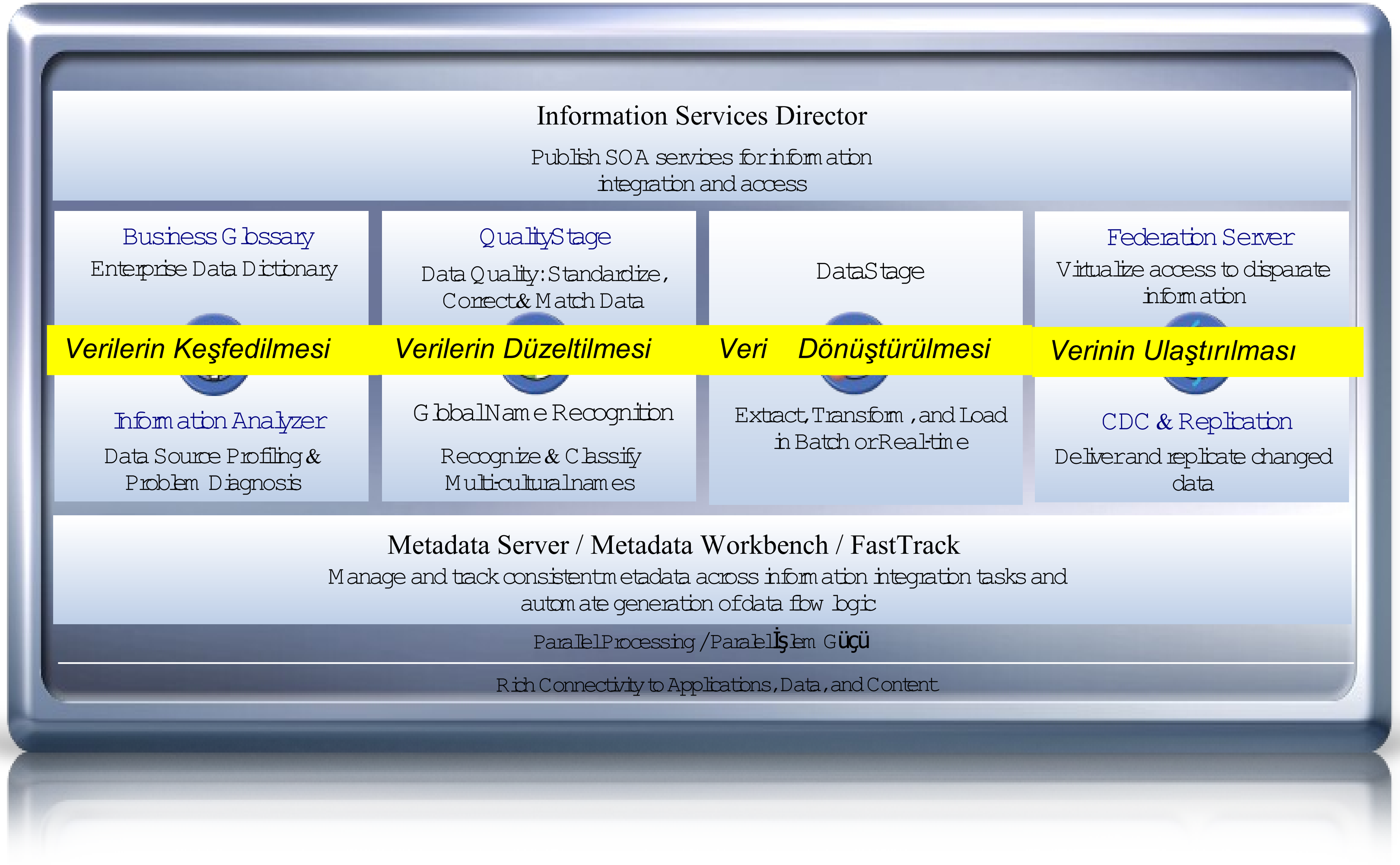


Bilgi...

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



InfoSphere Information Server – Ürünleri



Information Services Director

Publish SOA services for information integration and access

Business Glossary

Enterprise Data Dictionary

QualityStage

Data Quality: Standardize, Correct & Match Data

DataStage

Federation Server

Virtualize access to disparate information

Verilerin Keşfedilmesi

Verilerin Düzeltilmesi

Veri Dönüştürülmesi

Verinin Ulaştırılması

Information Analyzer

Data Source Profiling & Problem Diagnosis

Global Name Recognition

Recognize & Classify Multi-cultural names

Extract, Transform, and Load in Batch or Real-time

CDC & Replication

Deliver and replicate changed data

Metadata Server / Metadata Workbench / FastTrack

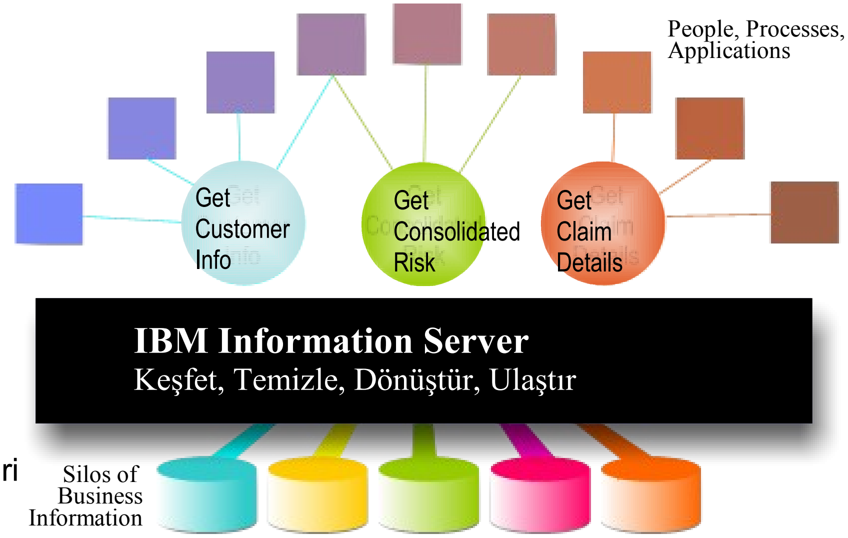
Manage and track consistent metadata across information integration tasks and automate generation of data flow logic

Parallel Processing / Paralel İşlem Gücü

Rich Connectivity to Applications, Data, and Content

IBM Information Server Bileşenleri

- Information Analyzer
  - Veri Analizi, raporlama
- DataStage
  - Verilerin ETL süreçlerini yönetir
- QualityStage
  - Veri standartlaştırması ve temizliği
- Bussiness 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


People, Processes, Applications
Get Customer Info
Get Consolidated Risk
Get Claim Details
IBM Information Server
Keşfet, Temizle, Dönüştür, Ulaştır
Silos of Business Information

IBM Information Server – 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)

Müş. Hesap No
On haneli hesap numarası. ACCNO ve HESNO diye de adlandırılır L-FIIIIVVVV formatında yazılır

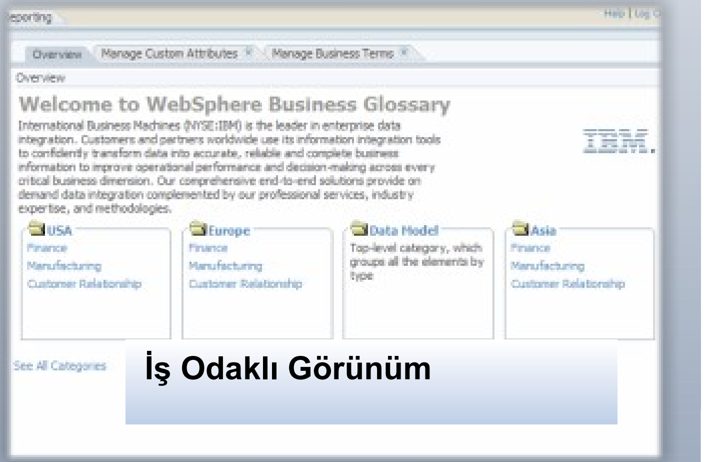
Subject Matter Experts

Business Users

Understand

Business Glossary

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



Welcome to WebSphere Business Glossary

International Business Machines (NYSE:IBM) is the leader in enterprise data integration. Customers and partners worldwide use its information integration tools to confidently transform data into accurate, reliable and complete business information to improve operational performance and decision-making across every critical business dimension. Our comprehensive end-to-end solutions provide on demand data integration complemented by our professional services, industry expertise, and methodologies.

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

İş Odaklı Görünüm

IBM Information Server – Business Glossary

Business 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 hafızalarındadır – nadiren dokümante edilir
- Kurum içerisinde dağılmış haldedir
- İşten ayrılmalarda ve proje değişimlerinde kaybolur
- Dokümante edilmemişse, bilgi ve yorumlama bozulmakta ve zamanla bulanıklaşmaktadır

Developer

Business Analyst

End Users

Software Architect

Data Architect

Data Admin

IT Admin

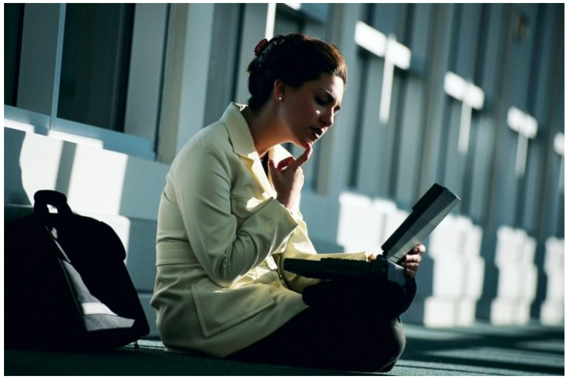
Business Glossary

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

Muhasebe Bölüm Kodu

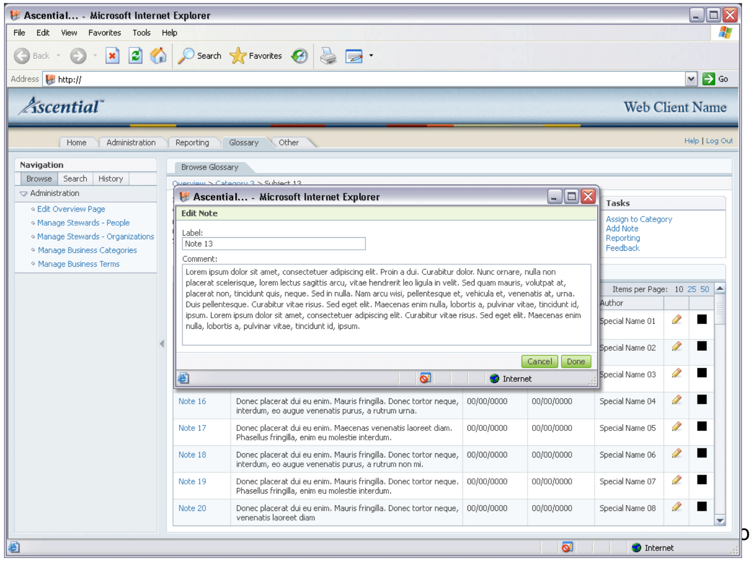
SAHIBI: Muhasebe Bölümü
FORMAT: X(7)
TANIM: Hesabın bağlı olduğu organizasyon bölümünü tanımlayan yedi haneli numara

Standart Tanımı Yap



Herhangi bir nesneye not ekle

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



Ascential... - Microsoft Internet Explorer

Edit Note

Label: Note 13

Comment:
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Proin a dui. Curabitur dolor. Nunc ornare, nulla non placerat scelerisque, lorem lectus sagittis arcu, vitae hendrerit leo ligula in velit. Sed quam mauris, volutpat at, placerat non, tincidunt quis, neque. Sed in nulla. Nam arcu wisi, pellentesque et, vehicula et, venenatis at, urna. Duis pellentesque. Curabitur vitae risus. Sed eget elit. Maecenas enim nulla, lobortis a, pulvinar vitae, tincidunt id, ipsum. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Curabitur vitae risus. Sed eget elit. Maecenas enim nulla, lobortis a, pulvinar vitae, tincidunt id, ipsum.

Cancel Done

				Author		
				Special Name 01	✎	■
				Special Name 02	✎	■
				Special Name 03	✎	■
Note 16	Donec placerat dui eu enim. Mauris fringilla. Donec tortor neque, interdum, eo augue venenatis purus, a rutrum urna.	00/00/0000	00/00/0000	Special Name 04	✎	■
Note 17	Donec placerat dui eu enim. Maecenas venenatis laoreet diam. Phasellus fringilla, enim eu molestie interdum.	00/00/0000	00/00/0000	Special Name 05	✎	■
Note 18	Donec placerat dui eu enim. Mauris fringilla. Donec tortor neque, interdum, eo augue venenatis purus, a rutrum non mi.	00/00/0000	00/00/0000	Special Name 06	✎	■
Note 19	Donec placerat dui eu enim. Mauris fringilla. Donec tortor neque. Phasellus fringilla, enim eu molestie interdum.	00/00/0000	00/00/0000	Special Name 07	✎	■
Note 20	Donec placerat dui eu enim. Mauris fringilla. Donec tortor neque, venenatis laoreet diam	00/00/0000	00/00/0000	Special Name 08	✎	■

IBM Information Server – Business Glossary

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?

IBM Information Server – Business Glossary

- Business Glossary olmayan firmalarda bütün iş sözlüğü bir excel listesinde tutulmaktadır.
- Business Analistlerin ve bütün kullanıcıların iş sözlüğüne erişmeleri saplanıyor
- Herhangi bir metrik'e ihtiyaçımız olursa bunu ortak common metadata sayesinde iş sözlüğün'de bakıyoruz
- Herhangi bir metrik'i arayıp buluruz ve bu metrik ile ilgili tüm detaylar business – glossary'de mevcuttur.
  - Short description
  - Long description
  - Steward
  - Related It assets
  - Etc.

Information Analyzer

- Varolan sisteminizi derinlemesine analiz eder
  - Veritabanları ve dosya bazlı kaynakların içerik, kalite ve yapısal olarak veri odaklı analizini yapar
  - Kolonların, kolonlar arasındaki ve veritabanları arasındaki ilişkilerin detaylı profilini çıkartır
- Veri kalitesini, önceki raporlarla karşılaştırarak sürekli olarak gözlemlenebilmesini sağlar
- Bilginin kurum içinde nerede ve nasıl yönetildiğinin ana veri bilgisini oluşturur
  - Kurum içindeki uzmanlık bilgisinin de eklenmesi ile, yeni projelerin daha hızlı yapılmasına ve doğru bilgi ile yapılmasını sağlar

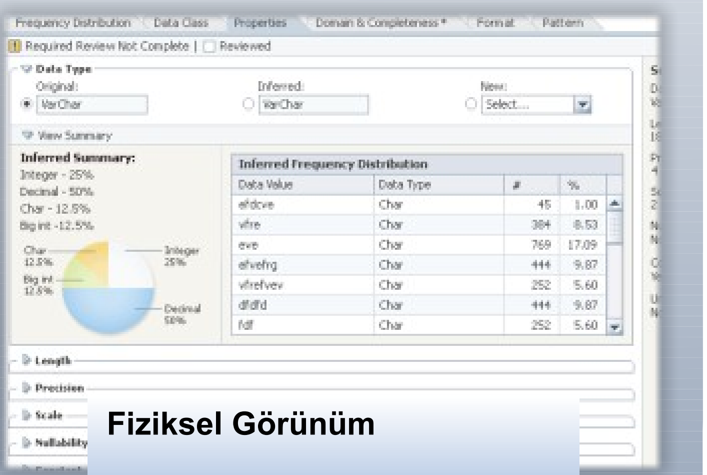
Subject Matter Experts

Data Analysts

Information Analyzer

Understand

Kaynak veri yapılarını inceler, veri katilesini ve bütünlüğünü gözlemler



Required Review Not Complete | □ Reviewed

Data Type

Original: ● VarChar

Inferred: ○ VarChar

New: ○ Select...

View Summary

Inferred Summary:
Integer - 25%
Decimal - 50%
Char - 12.5%
Big int -12.5%

Inferred Frequency Distribution

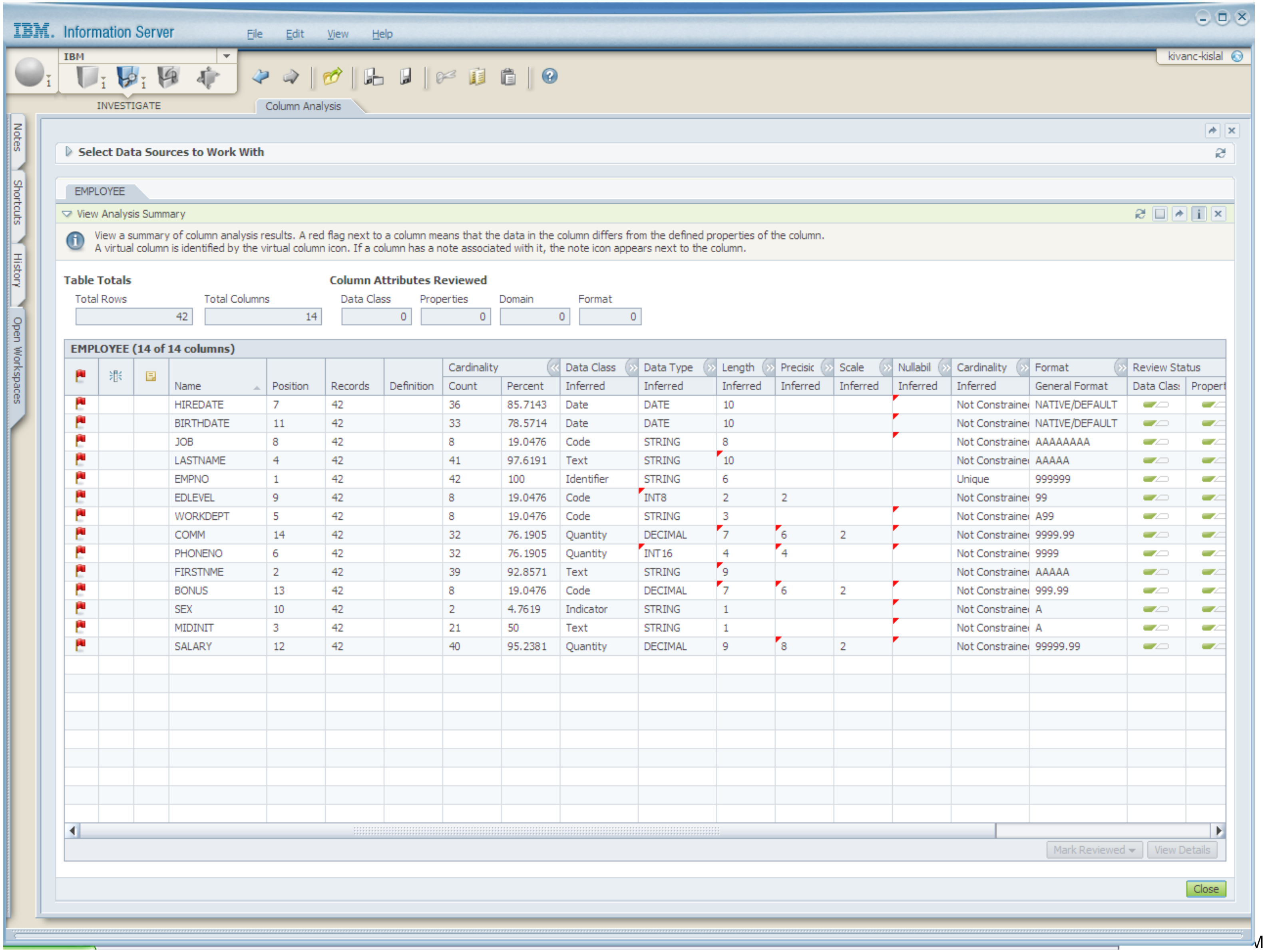
Data Value	Data Type	#	%
efdcve	Char	45	1.00
vfre	Char	384	8.53
eve	Char	769	17.09
efvefrg	Char	444	9.87
vfrefvev	Char	252	5.60
dfdfd	Char	444	9.87
fdf	Char	252	5.60

Fiziksel Görünüm

IBM Information Server – Information Analyzer

- Information Analyzer Fonksiyonel Bakış
  - The Project Paradigm
    - Project Artifacts & Elements
  - Managing Data
    - Platform Subject Administration
    - Metadata Import
  - Profiling & Review
    - Column Analysis
    - Primary Key Analysis
    - Foreign Key Analysis
    - Redundancy Analysis
    - Cross Domain Analysis
    - Baseline Analysis
    - Sharing Results
  - Developing Data Rules & Metrics
    - Data Rules
    - Benchmarks
    - Metrics

Information Analyzer : Kolon Analizi



Select Data Sources to Work With

EMPLOYEE

View Analysis Summary

View a summary of column analysis results. A red flag next to a column means that the data in the column differs from the defined properties of the column.
A virtual column is identified by the virtual column icon. If a column has a note associated with it, the note icon appears next to the column.

Table Totals

Total Rows: 42

Total Columns: 14

Column Attributes Reviewed

Data Class: 0

Properties: 0

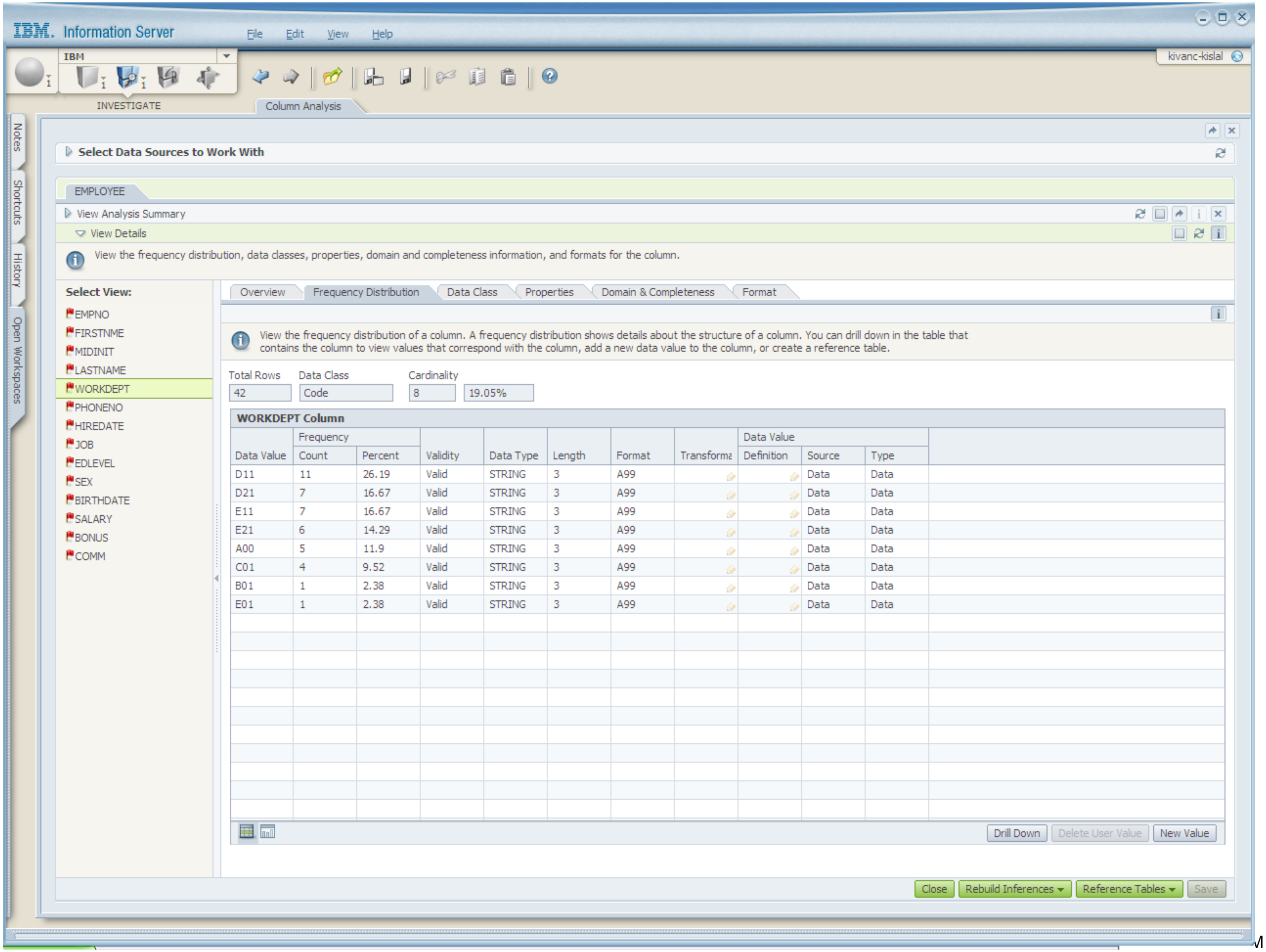
Domain: 0

Format: 0

EMPLOYEE (14 of 14 columns)

🚩							Cardinality		Data Class	Data Type	Length	Precisic	Scale	Nullabil	Cardinality	Format	Review Status	
			Name	Position	Records	Definition	Count	Percent	Inferred	Inferred	Inferred	Inferred	Inferred	Inferred	Inferred	General Format	Data Clas	Propert
🚩			HIREDATE	7	42		36	85.7143	Date	DATE	10				Not Constraine	NATIVE/DEFAULT		
🚩			BIRTHDATE	11	42		33	78.5714	Date	DATE	10				Not Constraine	NATIVE/DEFAULT		
🚩			JOB	8	42		8	19.0476	Code	STRING	8				Not Constraine	AAAAAAAA		
🚩			LASTNAME	4	42		41	97.6191	Text	STRING	10				Not Constraine	AAAAA		
🚩			EMPNO	1	42		42	100	Identifier	STRING	6				Unique	999999		
🚩			EDLEVEL	9	42		8	19.0476	Code	INT8	2	2			Not Constraine	99		
🚩			WORKDEPT	5	42		8	19.0476	Code	STRING	3				Not Constraine	A99		
🚩			COMM	14	42		32	76.1905	Quantity	DECIMAL	7	6	2		Not Constraine	9999.99		
🚩			PHONENO	6	42		32	76.1905	Quantity	INT16	4	4			Not Constraine	9999		
🚩			FIRSTNME	2	42		39	92.8571	Text	STRING	9				Not Constraine	AAAAA		
🚩			BONUS	13	42		8	19.0476	Code	DECIMAL	7	6	2		Not Constraine	999.99		
🚩			SEX	10	42		2	4.7619	Indicator	STRING	1				Not Constraine	A		
🚩			MIDINIT	3	42		21	50	Text	STRING	1				Not Constraine	A		
🚩			SALARY	12	42		40	95.2381	Quantity	DECIMAL	9	8	2		Not Constraine	99999.99		

Information Analyzer : Kolon Analizi



Select Data Sources to Work With

EMPLOYEE

View Analysis Summary

View Details

View the frequency distribution, data classes, properties, domain and completeness information, and formats for the column.

Select View:

- EMPNO
- FIRSTNME
- MIDINIT
- LASTNAME
- WORKDEPT
- PHONENO
- HIREDATE
- JOB
- EDLEVEL
- SEX
- BIRTHDATE
- SALARY
- BONUS
- COMM

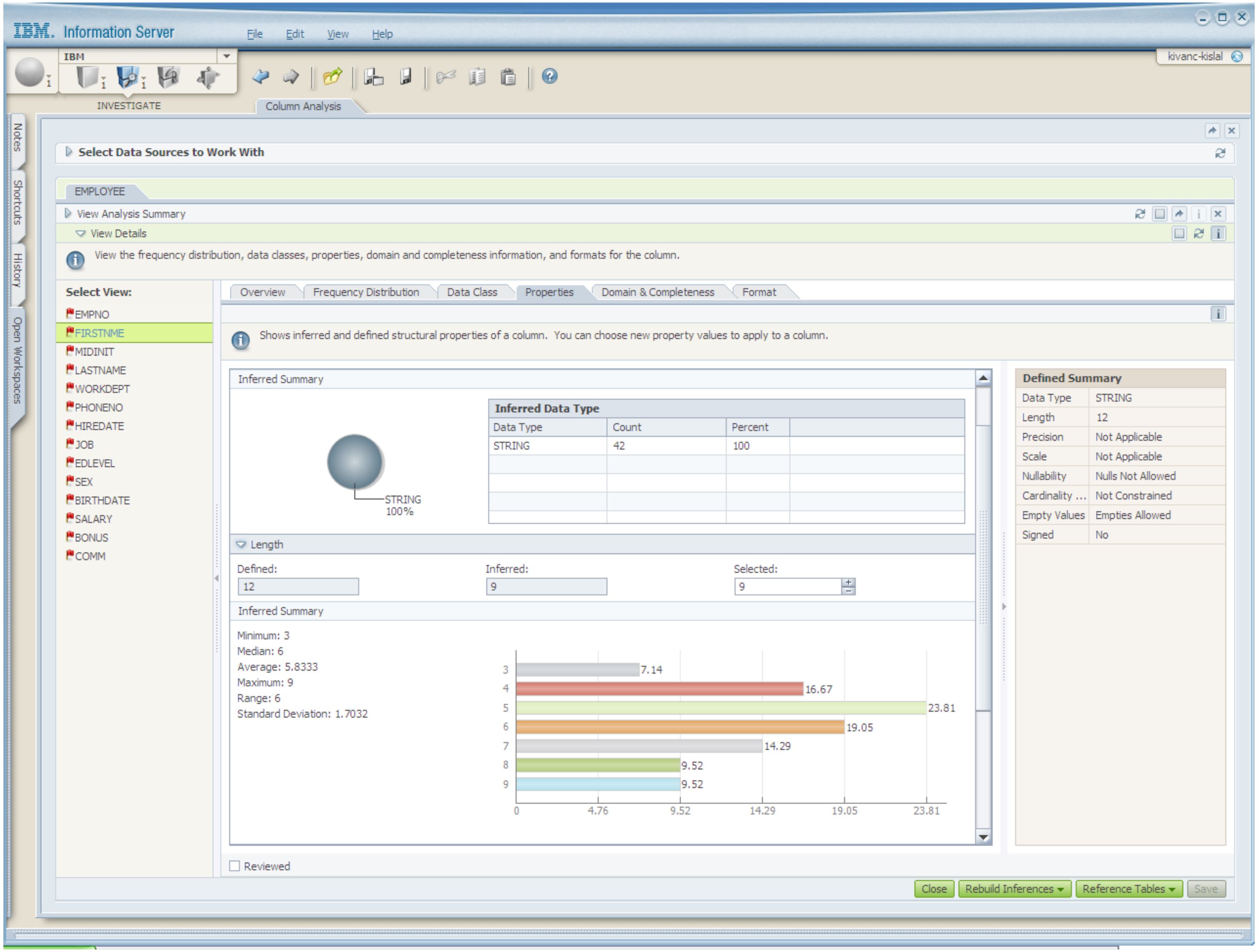
View the frequency distribution of a column. A frequency distribution shows details about the structure of a column. You can drill down in the table that contains the column to view values that correspond with the column, add a new data value to the column, or create a reference table.

Total Rows: 42
Data Class: Code
Cardinality: 8 19.05%

WORKDEPT Column

	Frequency							Data Value		
Data Value	Count	Percent	Validity	Data Type	Length	Format	Transforma	Definition	Source	Type
D11	11	26.19	Valid	STRING	3	A99			Data	Data
D21	7	16.67	Valid	STRING	3	A99			Data	Data
E11	7	16.67	Valid	STRING	3	A99			Data	Data
E21	6	14.29	Valid	STRING	3	A99			Data	Data
A00	5	11.9	Valid	STRING	3	A99			Data	Data
C01	4	9.52	Valid	STRING	3	A99			Data	Data
B01	1	2.38	Valid	STRING	3	A99			Data	Data
E01	1	2.38	Valid	STRING	3	A99			Data	Data

Information Analyzer : Kolon Analizi



View the frequency distribution, data classes, properties, domain and completeness information, and formats for the column.

Select View:

- EMPNO
- FIRSTNME
- MIDINIT
- LASTNAME
- WORKDEPT
- PHONENO
- HIREDATE
- JOB
- EDLEVEL
- SEX
- BIRTHDATE
- SALARY
- BONUS
- COMM

Shows inferred and defined structural properties of a column. You can choose new property values to apply to a column.

Inferred Summary

Inferred Data Type

Data Type	Count	Percent
STRING	42	100

Length

Defined: 12

Inferred: 9

Selected: 9

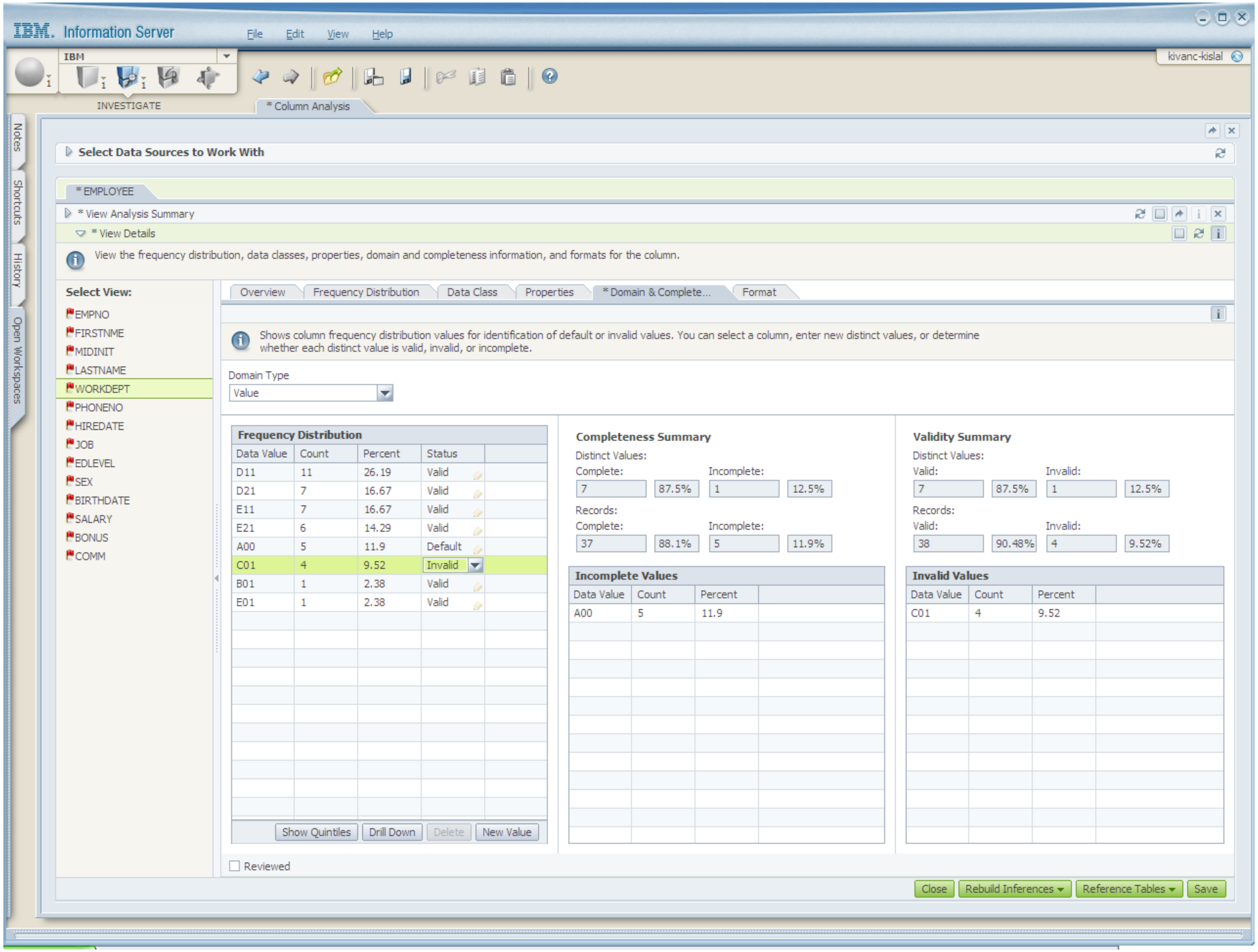
Inferred Summary

Minimum: 3
Median: 6
Average: 5.8333
Maximum: 9
Range: 6
Standard Deviation: 1.7032

Defined Summary	
Data Type	STRING
Length	12
Precision	Not Applicable
Scale	Not Applicable
Nullability	Nulls Not Allowed
Cardinality ...	Not Constrained
Empty Values	Empties Allowed
Signed	No

□ Reviewed

Information Analyzer : Kolon Analizi



Select Data Sources to Work With

* EMPLOYEE

* View Analysis Summary

* View Details

View the frequency distribution, data classes, properties, domain and completeness information, and formats for the column.

Select View:

- EMPNO
- FIRSTNME
- MIDINIT
- LASTNAME
- WORKDEPT
- PHONENO
- HIREDATE
- JOB
- EDLEVEL
- SEX
- BIRTHDATE
- SALARY
- BONUS
- COMM

Shows column frequency distribution values for identification of default or invalid values. You can select a column, enter new distinct values, or determine whether each distinct value is valid, invalid, or incomplete.

Domain Type: Value

Frequency Distribution

Data Value	Count	Percent	Status
D11	11	26.19	Valid
D21	7	16.67	Valid
E11	7	16.67	Valid
E21	6	14.29	Valid
A00	5	11.9	Default
C01	4	9.52	Invalid
B01	1	2.38	Valid
E01	1	2.38	Valid

Completeness Summary

Distinct Values:

Complete:		Incomplete:	
7	87.5%	1	12.5%

Records:

Complete:		Incomplete:	
37	88.1%	5	11.9%

Validity Summary

Distinct Values:

Valid:		Invalid:	
7	87.5%	1	12.5%

Records:

Valid:		Invalid:	
38	90.48%	4	9.52%

Incomplete Values

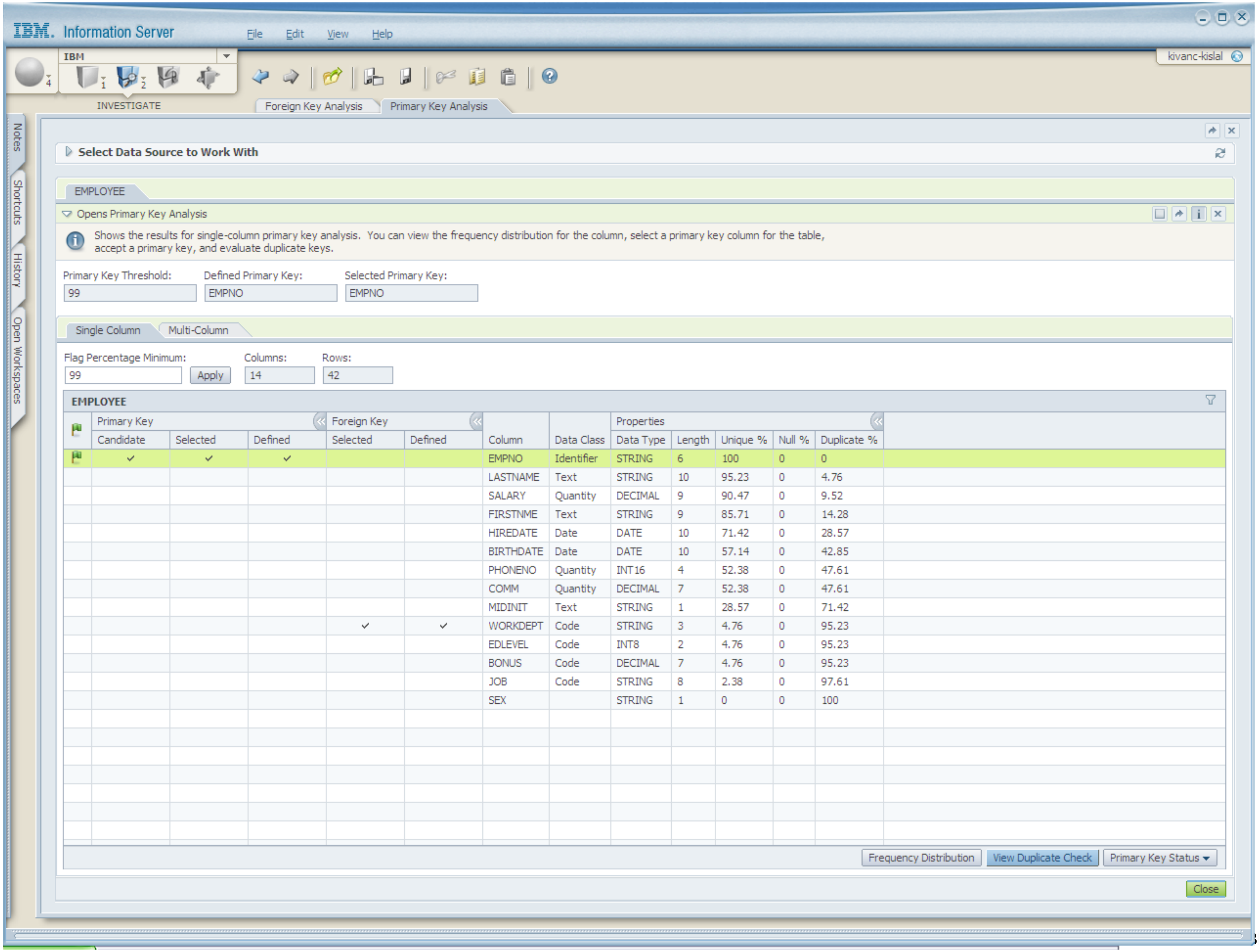
Data Value	Count	Percent
A00	5	11.9

Invalid Values

Data Value	Count	Percent
C01	4	9.52

☐ Reviewed

Information Analyzer : PK, FK Analizi



Select Data Source to Work With

EMPLOYEE

Opens Primary Key Analysis

Shows the results for single-column primary key analysis. You can view the frequency distribution for the column, select a primary key column for the table, accept a primary key, and evaluate duplicate keys.

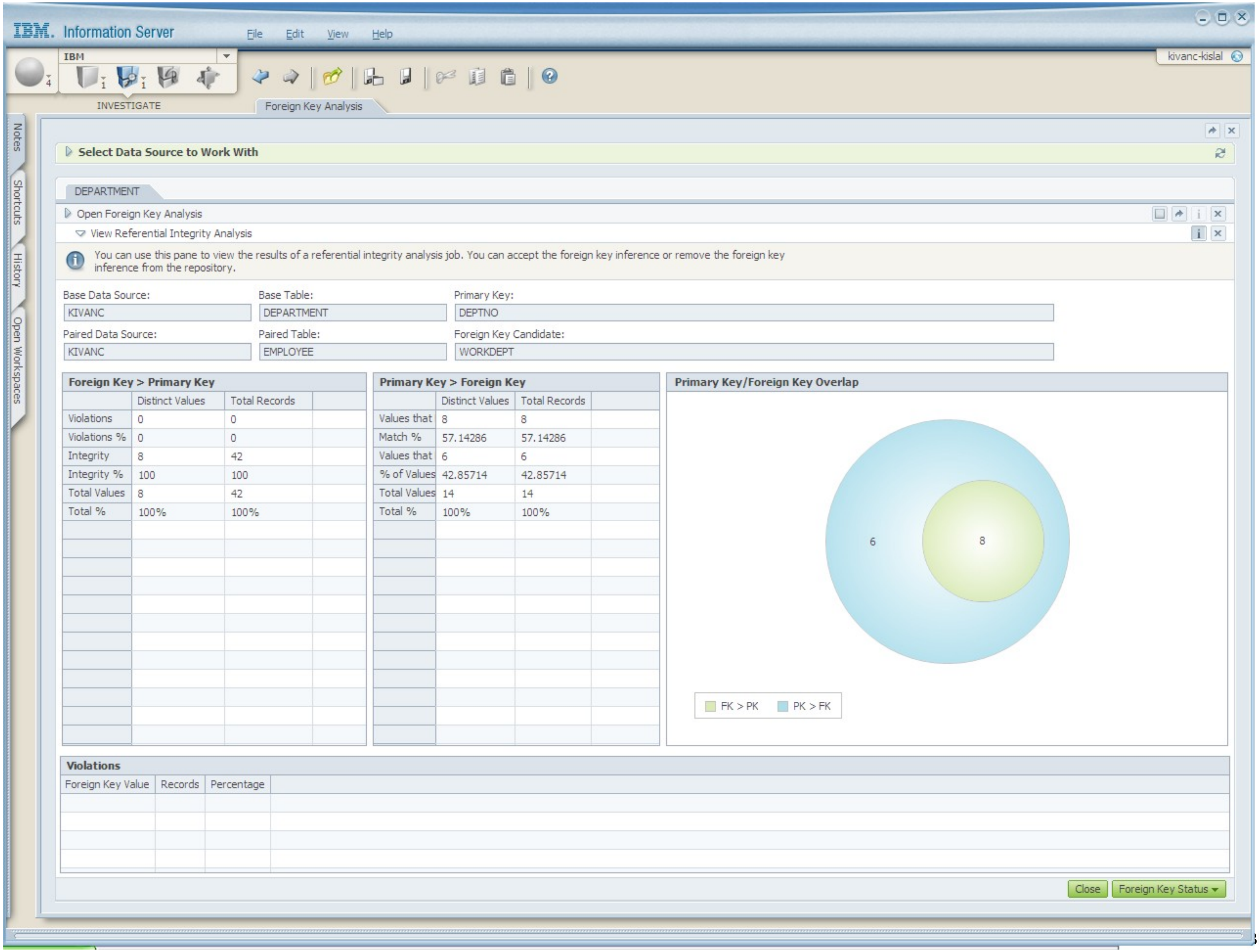
Primary Key Threshold: 99
Defined Primary Key: EMPNO
Selected Primary Key: EMPNO

Flag Percentage Minimum: 99 Apply
Columns: 14
Rows: 42

EMPLOYEE

⚑	Primary Key			Foreign Key				Properties				
	Candidate	Selected	Defined	Selected	Defined	Column	Data Class	Data Type	Length	Unique %	Null %	Duplicate %
⚑	✓	✓	✓			EMPNO	Identifier	STRING	6	100	0	0
						LASTNAME	Text	STRING	10	95.23	0	4.76
						SALARY	Quantity	DECIMAL	9	90.47	0	9.52
						FIRSTNME	Text	STRING	9	85.71	0	14.28
						HIREDATE	Date	DATE	10	71.42	0	28.57
						BIRTHDATE	Date	DATE	10	57.14	0	42.85
						PHONENO	Quantity	INT16	4	52.38	0	47.61
						COMM	Quantity	DECIMAL	7	52.38	0	47.61
						MIDINIT	Text	STRING	1	28.57	0	71.42
				✓	✓	WORKDEPT	Code	STRING	3	4.76	0	95.23
						EDLEVEL	Code	INT8	2	4.76	0	95.23
						BONUS	Code	DECIMAL	7	4.76	0	95.23
						JOB	Code	STRING	8	2.38	0	97.61
						SEX		STRING	1	0	0	100

Information Analyzer : PK, FK Analizi



You can use this pane to view the results of a referential integrity analysis job. You can accept the foreign key inference or remove the foreign key inference from the repository.

Base Data Source: KIVANC	Base Table: DEPARTMENT	Primary Key: DEPTNO
Paired Data Source: KIVANC	Paired Table: EMPLOYEE	Foreign Key Candidate: WORKDEPT

Foreign Key > Primary Key

	Distinct Values	Total Records
Violations	0	0
Violations %	0	0
Integrity	8	42
Integrity %	100	100
Total Values	8	42
Total %	100%	100%

Primary Key > Foreign Key

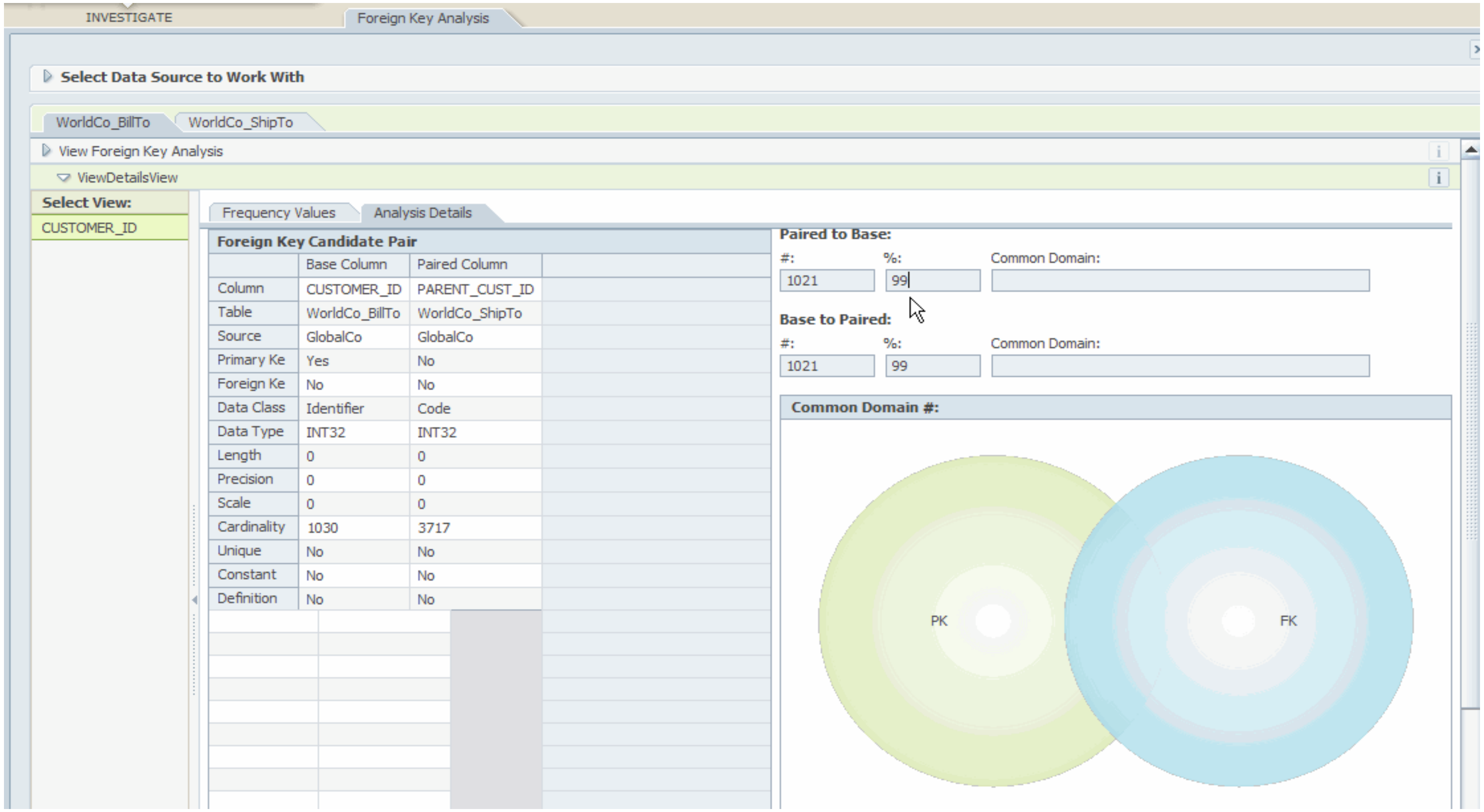
	Distinct Values	Total Records
Values that	8	8
Match %	57.14286	57.14286
Values that	6	6
% of Values	42.85714	42.85714
Total Values	14	14
Total %	100%	100%

Primary Key/Foreign Key Overlap

Violations

Foreign Key Value	Records	Percentage

Tablolar Arası/Sistemler Arası Seviyesinde Bilgilendirme



INVESTIGATE | Foreign Key Analysis

Select Data Source to Work With

WorldCo_BillTo | WorldCo_ShipTo

View Foreign Key Analysis

ViewDetailsView

Select View:

CUSTOMER_ID

Frequency Values | Analysis Details

Foreign Key Candidate Pair

	Base Column	Paired Column
Column	CUSTOMER_ID	PARENT_CUST_ID
Table	WorldCo_BillTo	WorldCo_ShipTo
Source	GlobalCo	GlobalCo
Primary Ke	Yes	No
Foreign Ke	No	No
Data Class	Identifier	Code
Data Type	INT32	INT32
Length	0	0
Precision	0	0
Scale	0	0
Cardinality	1030	3717
Unique	No	No
Constant	No	No
Definition	No	No

Paired to Base:

#: 1021
%: 99
Common Domain:

Base to Paired:

#: 1021
%: 99
Common Domain:

Common Domain #:

İlişkisel Bütünlük

- Foreign Keys
  - Tekil veya Çoğul Alan
  - Referential Integrity
- Cross-Domain
  - Benzerlik
  - Redundancy

QualityStage

- Bütünleşik Veri kalitesi süreci sunar
  - Temiz, standart ve tekil veriye sahip olmanızı sağlar
  - Gerçeğin tek bir versiyonunu oluşturur
  - Global Adres doğrulama

- Veri kalitesi ve eşleştirme algoritmaları için görsel araç
  - Quality Stage ile tam entegre (tek motor, tek veri modeli, tek arayüz)
  - Eşleştirme kurallarını kolay ve kesin olarak ayarlama

- Veri kalitesi mantığını ETL içine kolayca ekler

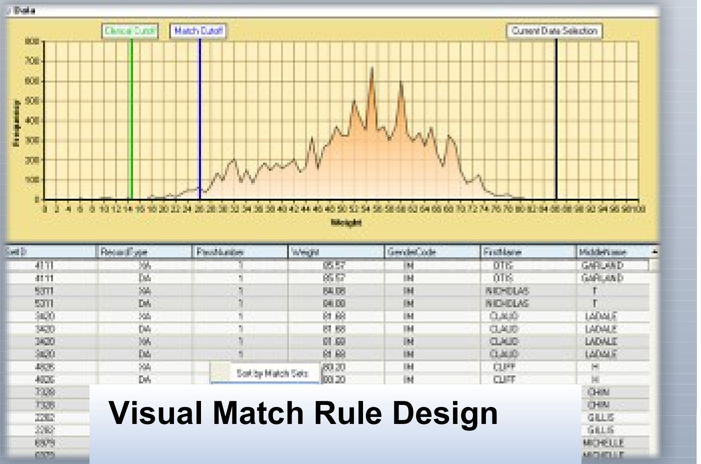
Subject Matter Experts

Data Analysts

Cleanse

QualityStage™

Alanları standartlaştırır ve düzeltir, değişik kaynaklardaki kayıtları eşleştirerek verinin tek bir görüntüsünü oluşturur



SetID	RecordType	PassNumber	Weight	GenderCode	FirstName	MiddleName
4111	XA	1	85.57	M	OTIS	GARLAND
4111	DA	1	85.57	M	OTIS	GARLAND
5311	XA	1	84.08	M	NICHOLAS	T
5311	DA	1	84.08	M	NICHOLAS	T
3420	XA	1	81.68	M	CLAUD	LADALE
3420	DA	1	81.68	M	CLAUD	LADALE
3420	XA	1	81.68	M	CLAUD	LADALE
3420	DA	1	81.68	M	CLAUD	LADALE
4826	XA		80.20	M	CLIFF	H
4826	DA		80.20	M	CLIFF	H
7328						CHIN
7328						CHIN
2282						GILLIS
2282						GILLIS
6979						MICHELLE
6979						MICHELLE

Visual Match Rule Design

QualityStage

Genel Veri Problemleri

- **Verinin standartlaşma eksikliği**
  - Değişik sistemlerde değişik şekillerde ve yapılarda durmakta
- **Alanlardaki farklı veriler**
  - Veri veritabanında yanlış kolonlarda durmakta
- **Bilgi, serbest metin içinde kaybolmakta**
- **Veri miyopluğu**
  - Farklı kaynaklarda duran, ilişkilendirilemeyen ilişkili veriler
- **Veri tekrarı**
  - Standartları farklı kopya bilgiler

```
Kate A. Roberts   416 Columbus Ave #2, Boston, Mass 02116
Catherine Roberts Four sixteen Columbus APT2, Boston, MA 02116
Mrs. K. Roberts   416 Columbus Suite #2, Suffolk County 02116
```

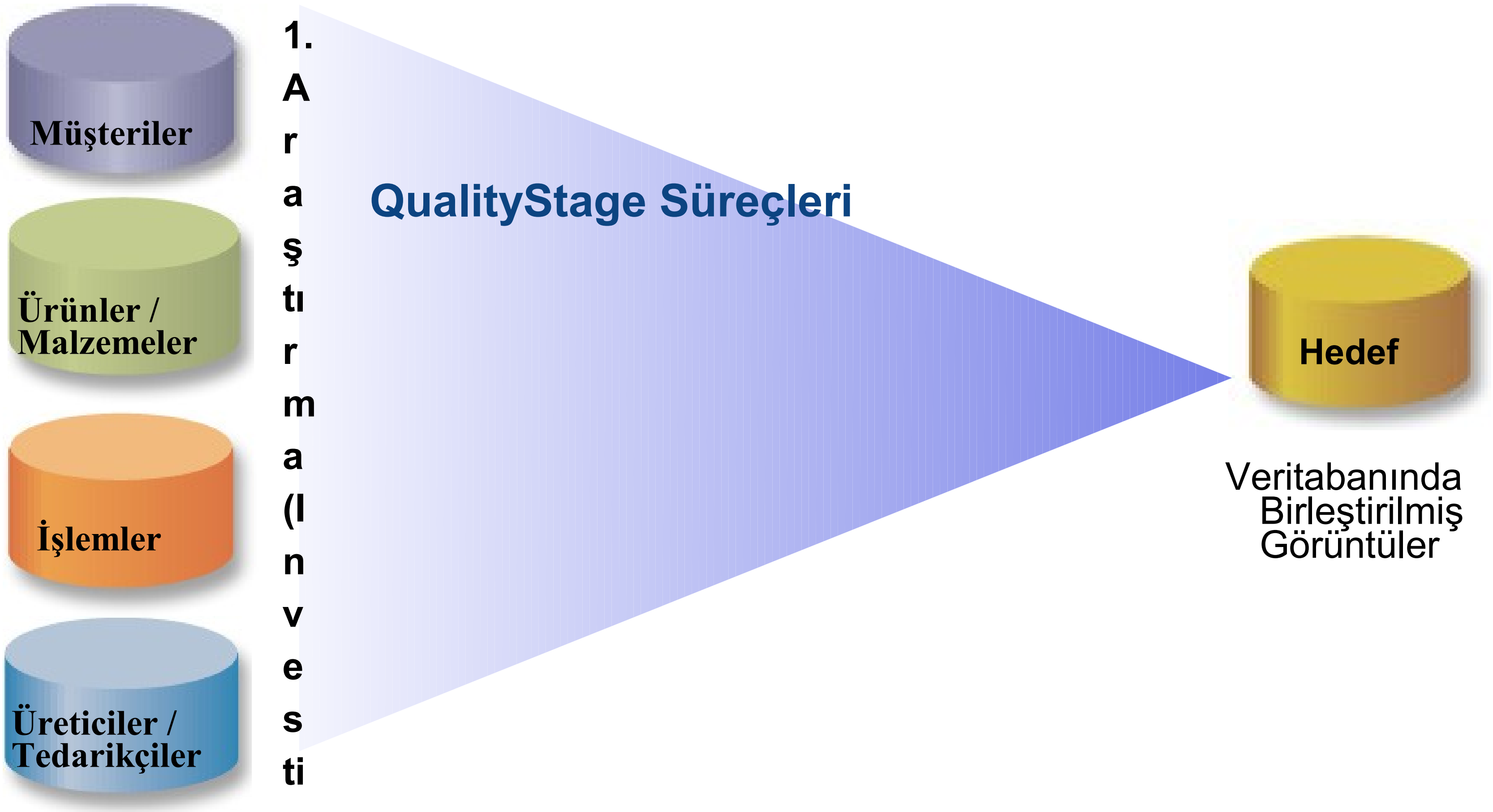
```
Name            Tax ID           Telephone
J Smith DBA Lime Cons.   228-02-1975    6173380300
Williams & Co. C/O Bill  025-37-1888    415-392-2000
1st Natl Provident      34-2671434     3380321
HP 15 State St.    508-466-1200   Orlando
```

```
WING ASSY DRILL 4 HOLE USE 5J868A HEXBOLT 1/4 INCH
WING ASSEMBY, USE 5J868-A HEX BOLT .25" - DRILL FOUR HOLES
USE 4 5J868A BOLTS (HEX .25) - DRILL HOLES FOR EA ON WING ASSEM
RUDER, TAP 6 WHOLES, SECURE W/KL2301 RIVETS (10 CM)
```

```
19-84-103    RS232 Cable 6' M-F CandS
CS-89641     6 ft. Cable Male-F, RS232 #87951
C&SUCH6      Male/Female 25 PIN 6 Foot Cable
```

```
90328574    IBM               187 N.Pk. Str. Salem NH 01456
90328575    I.B.M. Inc.             187 N.Pk. St. Salem NH 01456
90238495    Int. Bus. Machines      187 No. Park St Salem NH 04156
90233479    International Bus. M. 187  Park Ave Salem NH 04156
90233489    Inter-Nation Consults 15 Main Street Andover MA 02341
90345672    I.B. Manufacturing      Park Blvd. Bostno MA  04106
```

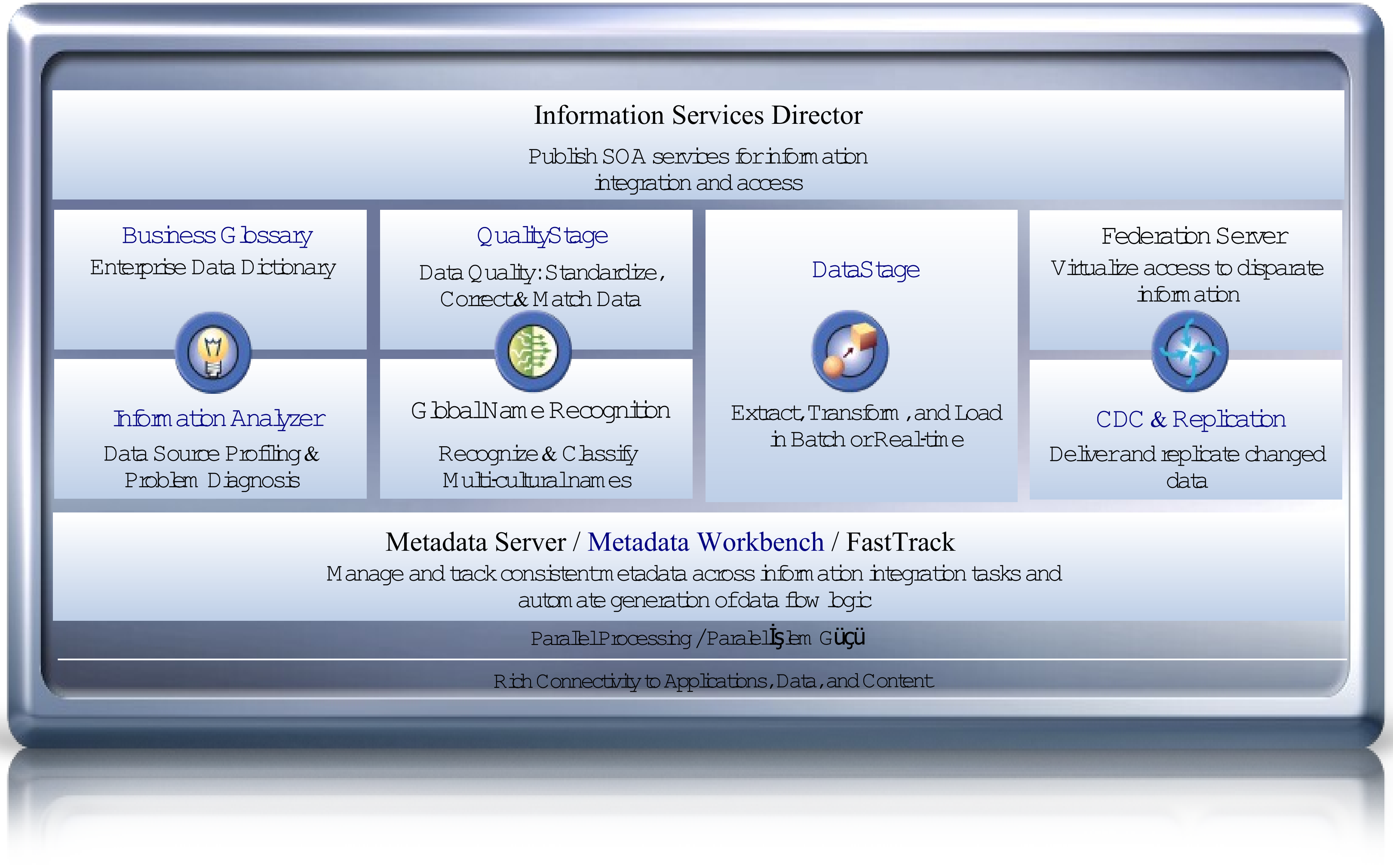
İşinizde Tutarlı, Doğru ve Birleştirilmiş Veriyi Nasıl Elde Edeceksiniz?



1. Araştırma (Investigatio

Veritabanında Birleştirilmiş Görüntüler

InfoSphere Information Server – Ürünleri



Information Services Director

Publish SOA services for information integration and access

Business Glossary

Enterprise Data Dictionary

Information Analyzer

Data Source Profiling & Problem Diagnosis

QualityStage

Data Quality: Standardize, Correct & Match Data

Global Name Recognition

Recognize & Classify Multi-cultural names

DataStage

Extract, Transform, and Load in Batch or Real-time

Federation Server

Virtualize access to disparate information

CDC & Replication

Deliver and replicate changed data

Metadata Server / Metadata Workbench / FastTrack

Manage and track consistent metadata across information integration tasks and automate generation of data flow logic

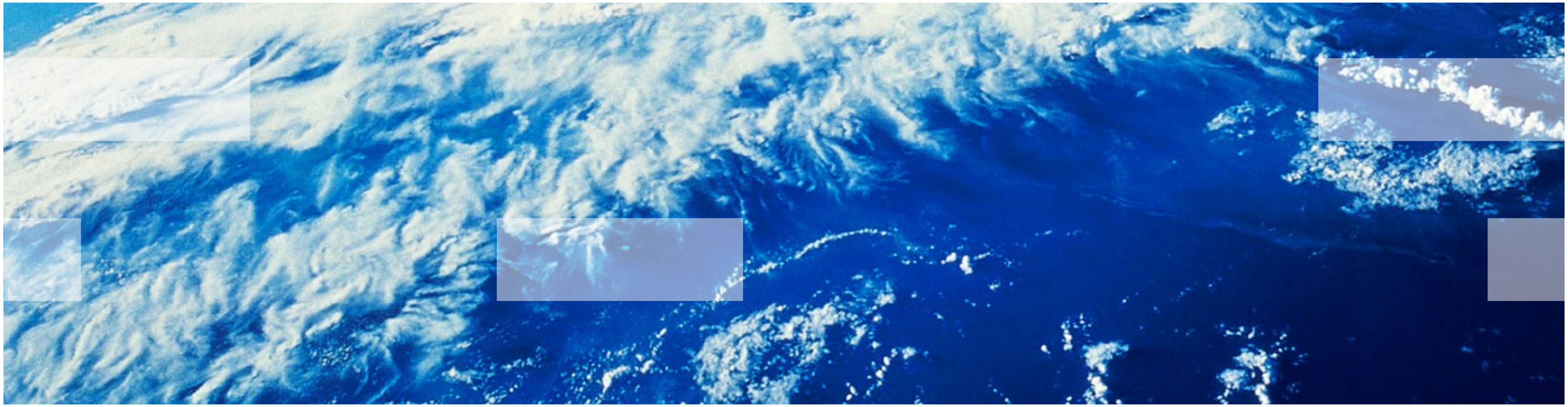
Parallel Processing / Paralel İşlem Gücü

Rich Connectivity to Applications, Data, and Content

Tevfik Başkaya - Vodafone

22/10/2009

IBM yazılım zirvesi '09 style



Vision

With this project, we focused on two main capabilities

1. Data Profiling
2. Automating Data Quality Check with flexible and reportable(open to users) applications.

Data Profiling

- Telecommunication Data
  - Huge Data size and growing fast
  - New data and concepts dynamic business
  - New coming applications in portfolio.

- During journey of information so many applications and systems have integration points. So high risk for inconsistent information.

- By Using Data Profiling tool
  - Identify Anomalies
  - Identify Inconsistencies
  - Taking Corrective Actions – Planning DQ approach.

Why Data Quality Tool

1. Operational Efficiency

1. Maintainability & Manageability

1. Low Impact to Source Systems

1. Performance & Scalability

1. Open Environment – Reports available to business users

Selection Criterias

- Vodafone is again pioneer; writing Data Quality Jobs that checks data in DWH is somehow new in Turkey.

- So we needed a strong tool that is supported by strong Vendor.

- Product must be well-supported against problems.

- We needed a strong solution partner.

- Solution partner should have deep knowledge both in technology and customer’s business.

  -I

Selection Criterias

- Performance
- Easy to develop/modify applications via a GUI
- Interface/Connectivity to other systems
- Integration with other DWH tools
- Vendor/Support
- Solution Integrator
  - I

Project Actions

- Analysis for data to be checked – Identifying DQ Controls
- Thresholds for each data and controls
- Identification corresponding actions (alarms, mail)
- Identification of reporting requirements
- Development for Phase 1
- Pilot run
- Production (Not Yet)

Current status is pilot run completed in IT for Phase 1.

-I

General Requirements

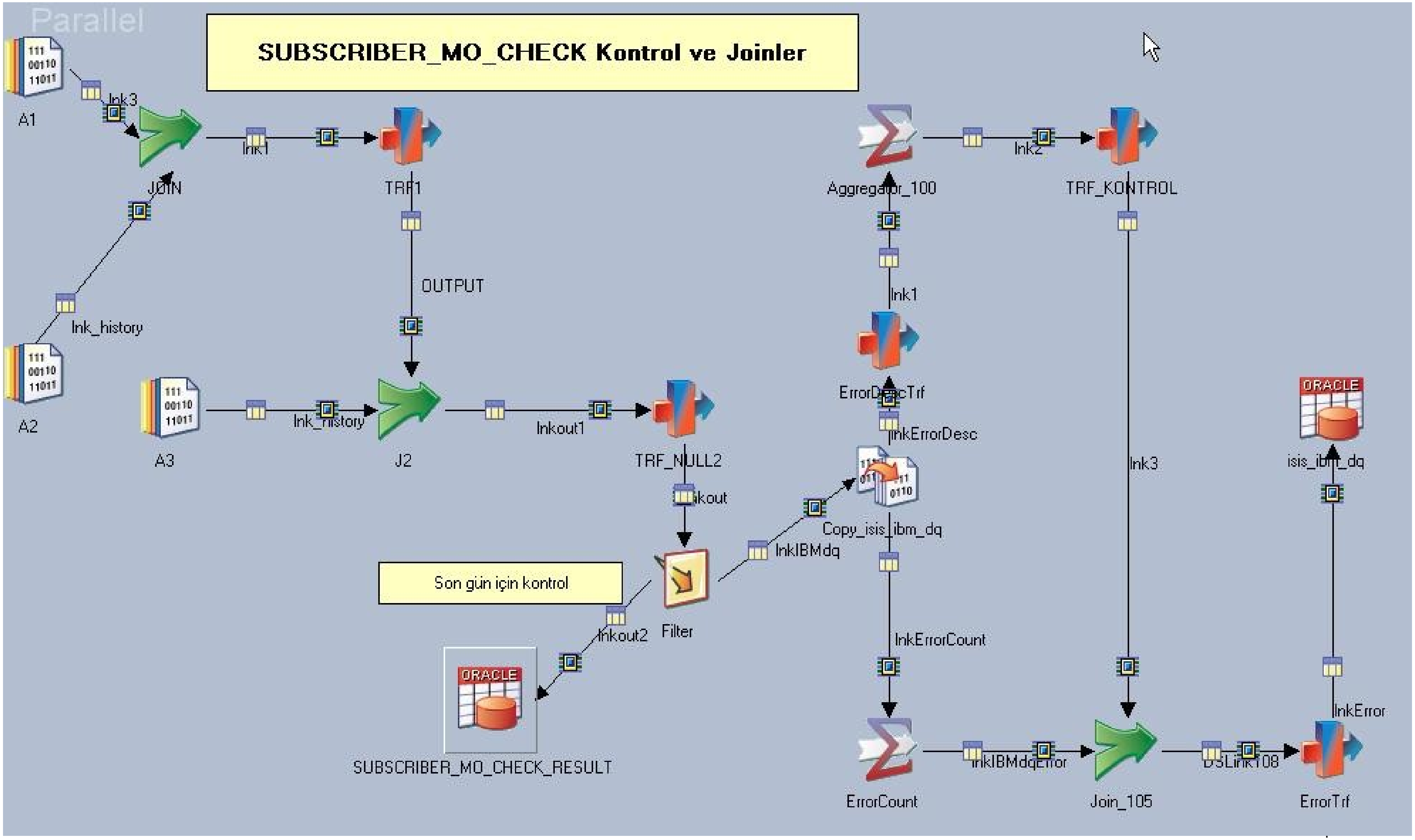
- Flexible
  - Parametric threshold definition for each control
  - Parametric action definition
  - Easy to change thresholds and actions
- Manageable & Reportable
  - Results are stored in Relational Database
  - Results are accessed via reporting interface, open to all users
  - Results are stored historically to show trends , collaborative
- All jobs should use the same framework.
- Opening results & trends to users produce auditable framework.
  - I

Sample Case

Day	Snapshot	Correct
9 Oct	5 Oct	5 Oct
8 Oct	5 Oct	5 Oct
7 Oct	5 Oct	5 Oct
6 Oct	5 Oct	5 Oct
5 Oct	5 Oct	5 Oct

Day	Snapshot	Correct
10 Oct	10 Oct	10 Oct
9 Oct	5 Oct	8 Oct
8 Oct	5 Oct	8 Oct
7 Oct	5 Oct	5 Oct
6 Oct	5 Oct	5 Oct
5 Oct	5 Oct	5 Oct

Örnek uygulama



Bu sunum 22 Ekim 2009 tarihinde İstanbul Swissotel the Bosphorus'da yapılan Yazılım Zirvesi 2009 için hazırlanmıştır.

http://www.ibm.com/software/tr