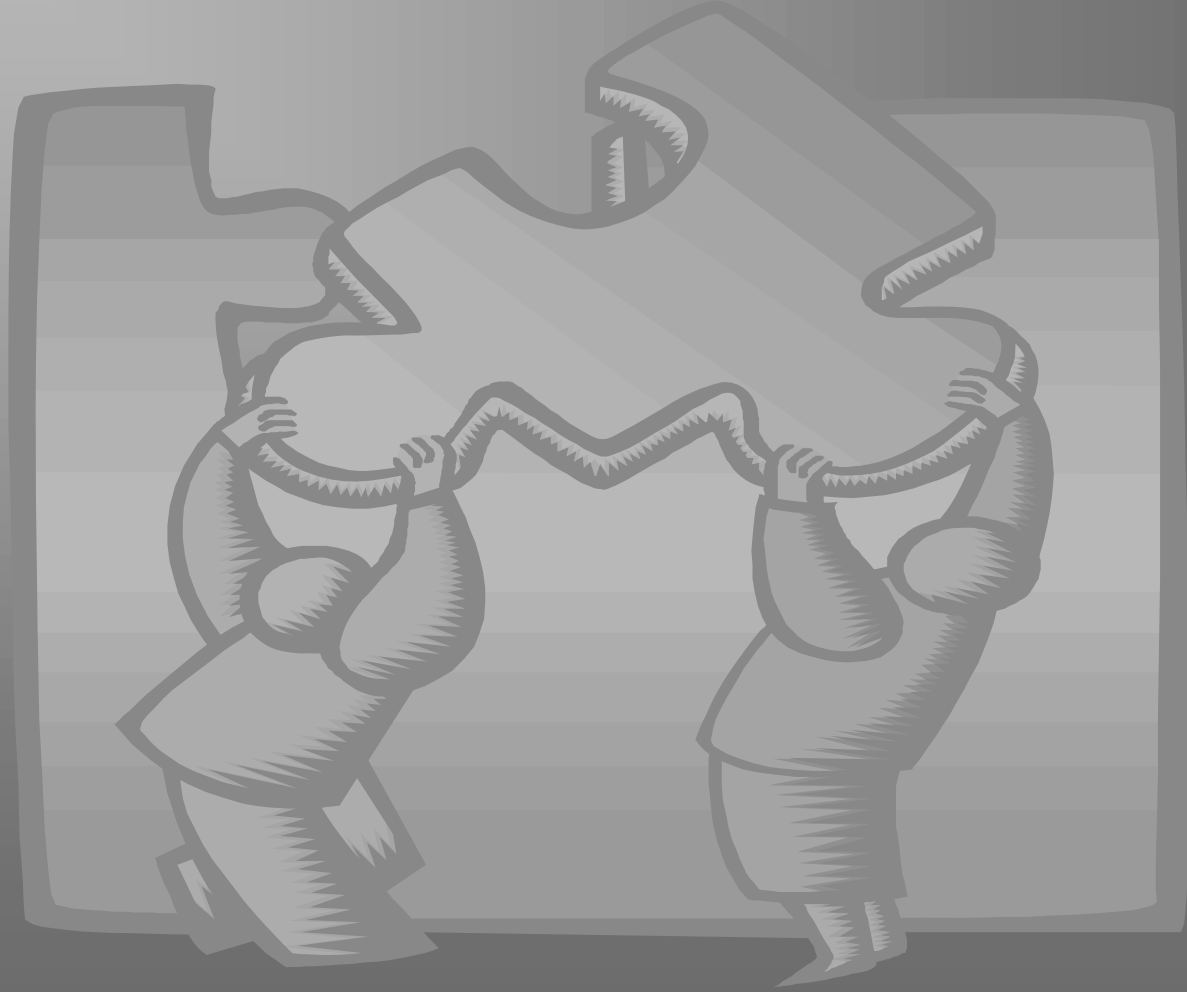




Imperative Suite – A Tool for Data Quality and Integrity Assurance



An *Xactis* Product Overview

Copyright 2005

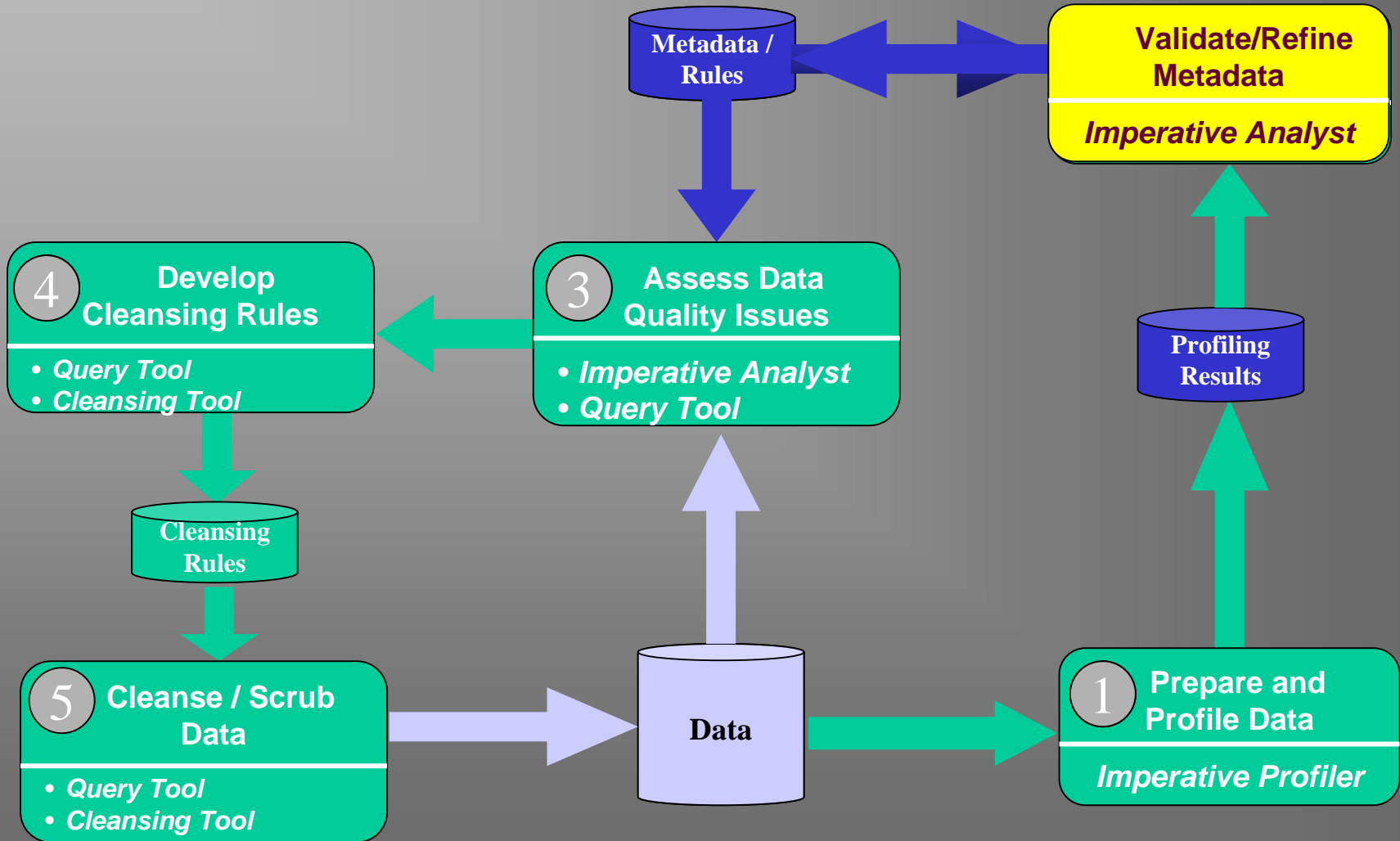
Data Profiling, Metadata Refinement, and Quality Assessment Using Imperative

The foundation of any data quality initiative is *quality metadata*, i.e., complete and accurate specifications of the rules governing the data, based on the requirements of the organization and its systems. Our primary tools for refining data rules and metadata are *Imperative Profiler* and *Imperative Analyst*.

- *Imperative Profiler* analyzes data values to infer possible data types, null rules, patterns/formats, and other such metadata.
- *Imperative Analyst* uses the data profiling results to support a subject matter expert in validating and refining the data rules, and in assessing quality issues relative to the rules.



Overview of Data Quality Assurance Approach





Data Profiling, Metadata Refinement, and Quality Assessment Using Imperative

Order			
Customer			
CustomerID	DOB	Lname	...
0008114	12/02/55	Van Kamp	
0010095	01/12/83	Aurtio	
0000092	06 Jun 1974	Stein	
0814055	04/30/00	Anderson	
<NULL>	10/14/968	Osgood	
4441050	Sep 1 1990	Clarson	
0124078	09/24/55	Ahsana	
0022004	Unknown	Roy	
...			

Data Tables and Files

Imperative Profiler

Infer Data Properties

- Data Type(s)
- Values and Frequencies
- Patterns and Frequencies
- Minimum Values
- Maximum Values
- % of Null Values
- % of Unique Values
- ...

Imperative Analyst

- Refine Rules
- Assess Data Quality



Subject Matter Experts

Data Dictionary

Data Quality Statistics

Rule Coverage Statistics



i now Analysis 1.0 - [Project Attributes]

File Window Help

Project Attributes

Explorer

Project Attributes

System	Entity	Attribute	Min	Max	DType	DomPat	Records	[%]Null	[%]Dist	Profiled
Northwind	[Categories]	[Description]	Breads, crac...	Sweet an...	VARCHAR(58)	No pattern gener...	8	0	100	6/18/2003 ..
Northwind	[Categories]	[CategoryName]	Beverages	Seafood	VARCHAR(14)	Cc(6)	8	0	100	6/18/2003 ..
Northwind	[Categories]	[CategoryID]	1	8	SMALLINT	9	8	0	100	6/18/2003 ..
Northwind	[Categories]	[PostalCode]	01-012	WX3 6FW	VARCHAR(9)	9(5)	91	1.09	95.6	6/18/2003 ..
Northwind	[Categories]	[Address]	1 rue Alsace...	Åkergata...	VARCHAR(46)	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Categories]	[CompanyName]	Alfreds Futt...	Wolski Za...	VARCHAR(36)	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Categories]	[ContactTitle]	Accounting ...	Sales Rep...	VARCHAR(30)	Cc(4)	91	0	13.18	6/18/2003 ..
Northwind	[Categories]	[ContactName]	Zbyszek P...	Zbyszek P...	VARCHAR(23)	Cc(4)WCc(6)	91	0	100	6/18/2003 ..
Northwind	[Categories]	[ContactID]	981-443655	981-443655	VARCHAR(17)	(9(3))W9(3)-9(4)	91	0	100	6/18/2003 ..
Northwind	[Categories]	[City]	Århus	Århus	VARCHAR(15)	Cc(5)	91	24.17	76.92	6/18/2003 ..
Northwind	[Categories]	[Country]	AK	WY	VARCHAR(13)	C(2)	91	65.93	20.87	6/18/2003 ..
Northwind	[Categories]	[Country]	Argentina	Venezuela	VARCHAR(11)	Cc(5)	91	0	23.07	6/18/2003 ..
Northwind	[Customers]	[CustomerID]	ALFKI	WOLZA	CHAR(5)	C(5)	91	0	100	6/18/2003 ..
Northwind	[Customers]	[EmployeeID]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[LastName]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[FirstName]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[Title]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[TitleOfCourtesy]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[BirthDate]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[HireDate]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[Address]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[City]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[Region]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[PostalCode]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[Country]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[HomePhone]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[Extension]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[Notes]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Customers]	[ReportsTo]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Orders]	[OrderID]	10248	11077	SMALLINT	9(5)	2155	0	38.51	6/18/2003 ..

74 items Opened : 11:52:16 Loaded: 11:52:18 AM

User[Mike Neely] Mode[Normal] Logged[11:52:09 AM] Server[localhost]

Start i now... 12:44 PM

Review Attribute Profiling Results via Tabular Display Showing

- System (*Northwind*)
- Entity (*Categories*)
- Attribute (*Postal Code*)
- Min and Max Values
- Dominant Data Type
- Dominant Pattern
- Number of Records Profiled
- Number and % Null
- Date and Time Profiled
- And More...



Data Profiling with Imperative Profiler (3 of 6)

i now Analysis 1.0 - [Project Attributes]

File Window Help

Project Attributes

Explorer

- Project
 - Notes (0)
 - Issues (0)
 - Attributes (76)
 - Attribute profiles (74)
 - Servers (1)
 - Data sources (1)
 - Data stores (1)
 - Entities (8)
 - Entity profiles (8)
 - [Categories].6/1/...
 - Notes (0)
 - Issues (0)
 - Attribute profi...
 - Category...
 - Notes ...
 - Issues ...
 - Unique...
 - Patterns
 - Dataty...
 - Depen...
 - [Category...]
 - [Descriptio...]
 - Data
 - Dependents (0)
 - [Customers].6/18/...
 - [Employees].6/18/...
 - [Order Details].6/...
 - [Orders].6/18/20...
 - [Products].6/18/...
 - [Shippers].6/18/2...
 - [Suppliers].6/18/...
 - Loads (1)
 - Load Entities (8)
 - Runs (1)
 - Systems (1)

Project Attributes

System	Entity	Attribute	Min	Max	DType	DomPat	Records	[%]Null	[%]Dist	Profiled
Northwind	[Categories]	[Description]	Breads, crac...	Sweet an...	VARCHAR(58)	No pattern gener...	8	0	100	6/18/2003 ..
Northwind	[Categories]	[CategoryName]	Beverages	Seafood	VARCHAR(14)	Cc(6)	8	0	100	6/18/2003 ..
Northwind	[Categories]	[CategoryID]	1	8	SMALLINT	Cc(6)	8	0	100	6/18/2003 ..
Northwind	[Customers]	[PostalCode]	01-012	WX3 6FW	VARCHAR(9)	9(5)	91	1.09	95.6	6/18/2003 ..
Northwind	[Customers]	[Address]	1 rue Alsace...	Åkergata...	VARCHAR(46)	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[CompanyName]	A...	Wolski Za...	VARCHAR(36)	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[ContactTitle]	Accou...	Sales Rep...	VA...					
Northwind	[Customers]	[ContactName]	Alejandra C...		VA...					
Northwind	[Customers]	[Phone]	(02) 201 24 67							
Northwind	[Customers]	[Fax]	(02) 201 24 68	981						
Northwind	[Customers]	[City]	Aachen	Århus						
Northwind	[Customers]	[Region]	AK	WY						
Northwind	[Customers]	[Country]	Argentina	Venezuela	VA...					
Northwind	[Customers]	[CustomerID]	ALFKI	WOLZA	CH...					
Northwind	[Employees]	[EmployeeID]	<NULL>	<NULL>	CH...					
Northwind	[Employees]	[LastName]	<NULL>	<NULL>	CH...					
Northwind	[Employees]	[FirstName]	<NULL>	<NULL>	CH...					
Northwind	[Employees]	[Title]	<NULL>	<NULL>	CH...					
Northwind	[Employees]	[TitleOfCourtesy]	<NULL>	<NULL>	CH...					
Northwind	[Employees]	[BirthDate]	<NULL>	<NULL>	CH...					
Northwind	[Employees]	[HireDate]	<NULL>	<NULL>	CH...					
Northwind	[Employees]	[Address]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Employees]	[City]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Employees]	[Region]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Employees]	[PostalCode]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Employees]	[Country]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Employees]	[HomePhone]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Employees]	[Extension]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Employees]	[Notes]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Employees]	[ReportsTo]	<NULL>	<NULL>	CHAR(1)	[NULL]	0	0	0	6/18/2003 ..
Northwind	[Order De	[OrderID]	10248	11077	SMALLINT	9(5)	2155	0	38.51	6/18/2003 ..

74 items Opened : 11:52:16 Loaded: 11:52:18 AM

User[Mike Neely] Mode[Normal] Logged[11:52:09 AM] Server[localhost]

Start i no... 12:44 PM

“Drill Down” to a List of Possible Data Types for “PostalCode”

Data type	#Matching	%Matching
SMALLINT	34	37.03
INTEGER	32	34.85
CHAR(9)	9	9.81
CHAR(7)	8	8.72
CHAR(8)	4	4.36
CHAR(6)	3	3.27
NULL	1	1.09

i now Analysis 1.0 - [Project Attributes]

File Window Help

Project Attributes

Explorer

- Project
 - Notes (0)
 - Issues (0)
 - Attributes (76)
 - Attribute profiles (74)
 - Servers (1)
 - Data sources (1)
 - Data stores (1)
 - Entities (8)
 - Entity profiles (8)
 - [Categories].6/1...
 - Notes (0)
 - Issues (0)
 - Attribute profi...
 - Notes ...
 - Issues ...
 - Unique...
 - Patterns

Project Attributes

System	Entity	Attribute	Min	Max	DType	DomPat	Records	[%]Null	[%]Dist	Profiled
Northwind	[Categories]	[Description]	Breads, crac...	Sweet an...	VARCHAR(58)	No pattern gener...	8	0	100	6/18/2003 ..
Northwind	[Categories]	[CategoryName]	Beverages	Seafood	VARCHAR(14)	Cc(6)	8	0	100	6/18/2003 ..
Northwind	[Categories]	[CategoryID]	1	8	SMALLINT	Cc(6)	8	0	100	6/18/2003 ..
Northwind	[Customers]	[PostalCode]	01-012	WX3 6FW	VARCHAR(9)	9(5)	91	1.09	95.6	6/18/2003 ..
Northwind	[Customers]	[Address]	1 rue Alsace...	Åkergata...	VARCHAR(46)	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[CompanyName]	ands Futt...	Wolski Za...	VARCHAR(36)	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[ContactTitle]		Sales Rep...	VA					003 ..
Northwind	[Customers]	[ContactName]	Ale	Zbyszek P...	VA					003 ..
Northwind	[Customers]	[Phone]	(02) 201 2	3655	VA					003 ..
Northwind	[Customers]	[Fax]	(02) 201 2		VA					003 ..
Northwind	[Customers]	[City]	Aachen		VA					003 ..
Northwind	[Customers]	[Region]	AK		VA					003 ..
Northwind	[Customers]	[Country]	Argentina	Ven	VA					003 ..
Northwind	[Customers]	[CustomerID]	ALFKI	WOLZ	VA					003 ..
Northwind	[Employees]	[EmployeeID]	<NULL>	<NULL>	VA					003 ..
Northwind	[Employees]	[LastName]	<NULL>	<NULL>	VA					003 ..

“Drill Down” to a List of Values and Occurrences for “PostalCode”

[PostalCode]=	Frequency=
01-012	1
01307	1
02389-673	1
02389-890	1
04179	1
04876-786	1
05021	1
05022	1
05023	1
05033	2
05432-043	1
05442-030	1

Show All Records Containing a Given Value

[CustomerID]=	[CompanyName]=	[ContactName]=	[ContactTitle]=	[Address]=	[City]=	[Region]=	[PostalCode]=	[Country]=	[Phone]=
GOURL	Gourmet Lanchone	André Fonseca	Sales Associate	Av. Brasil, 44; Campinas	SP		04876-786	Brazil	(11) 555-948

74 items | Opened : 11:52:16 | Loaded: 11:52:18 AM

User[Mike Neely] | Mode[Normal] | Logged[11:52:09 AM] | Server[localhost]

Start | i no... | 12:44 PM



Data Profiling with Imperative Profiler (5 of 6)

i now Analysis 1.0 - [Project Attributes]

File Window Help

Project Attributes

Explorer

Project Attributes

System	Entity	Attribute	Min	Max	DType	DomPat	Records	[%]Null	[%]Dist	Profiled
Northwind	[Categories]	[Description]	Breads, crac...	Sweet an...	VARCHAR(58)	No pattern gener...	8	0	100	6/18/2003 ..
Northwind	[Categories]	[CategoryName]	Beverages	Seafood	VARCHAR(14)	Cc(6)	8	0	100	6/18/2003 ..
Northwind	[Categories]	[CategoryID]	1	8	SMALLINT	9	8	0	100	6/18/2003 ..
Northwind	[Customers]	[PostalCode]	01-012	WX3 6FW	VARCHAR(9)	9(5)	91	1.09	95.6	6/18/2003 ..
Northwind	[Customers]	[Address]	1 rue Alsace...	Åkergata...	VARCHAR(46)	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[CompanyName]	A. Jensen Fu...	Wolski Za...	VARCHAR(25)	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[ContactName]	Alejandro...	Sales Rep...	VAR...	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[ContactTitle]	Accounting	Sales Rep...	VAR...	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[Phone]	(02) 201 24		VAR...	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[Fax]	(02) 201 24 66		VAR...	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[City]	Aachen		VAR...	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[Region]	AK		VAR...	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[Country]	WY	Venezuel...	VAR...	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[Country]	WOLZA		VAR...	No pattern gener...	91	0	100	6/18/2003 ..
Northwind	[Customers]	[Country]	<NULL>		VAR...	No pattern gener...	91	0	100	6/18/2003 ..

“Drill Down” to a List of Patterns and Occurrences for “PostalCode”

Pattern	Frequency
9(5)	48
9(4)	14
9(5)-9(3)	9
C(2)9W9C(2)	5
C9CW9C9	3
C(2)9(2)W9C(2)	2
C-9(4)	2
C-9(3)W9(2)	2
9(2)-9(3)	1
<NULL>	1

Show All Records Fitting a Given Pattern

[CustomerID]=	[CompanyName]=	[ContactName]=	[ContactTitle]=	[Address]=	[City]=	[Region]=	[PostalCode]=	[Country]=	[Phone]=
RANCH	Rancho grande	Sergio Gutiérrez	Sales Representative	Av. del Libertad	Buenos Aires		1010	Argentina	(1) 12
OCEAN	Océano Atlántico L	Yvonne Moncada	Sales Agent	Ing. Gustavo M	Buenos Aires		1010	Argentina	(1) 13
CACTU	Cactus Comidas pa	Patricio Simpson	Sales Agent	Cerrito 333	Buenos Aires		1010	Argentina	(1) 11
GROSR	GROSELLA-Resta	Manuel Pereira	Owner	5ª Ave. Los Pa	Caracas	DF	1081	Venezuela	(2) 28
RICSU	Richter Supermarkt	Michael Holz	Sales Manager	Grenzacherwe	Genève		1203	Switzerland	0897
FURIB	Furia Bacalhau e Fi	Lino Rodriguez	Sales Manager	Jardim das rosi	Lisboa		1675	Portugal	(1) 38
SIMOB	Simons bistro	Jytte Petersen	Owner	Vinbæltet 34	København		1734	Denmark	31 12
PRINI	Princesa Isabel Vin	Isabel de Castro	Sales Representative	Estrada da saú	Lisboa		1756	Portugal	(1) 38
CHOPS	Chop-suey Chinese	Yang Wang	Owner	Hauptstr. 29	Bern		3012	Switzerland	0452
LILAS	LILA-Supermercado	Carlos González	Accounting Manager	Carrera 52 con	Barquisimeto	Lara	3508	Venezuela	(9) 33
SANTG	Santé Gourmet	Jonas Bergulfsen	Owner	Erling Skakkes	Stavern		4110	Norway	07-98
LINOD	LIND-Delicatesses	Felipe Izquierdo	Owner	Ave. 5 de Mayo	I. de Margarita	Nueva Esparta	4980	Venezuela	(8) 34
PICCO	Piccolo und mehr	Georg Pippas	Sales Manager	Geislweg 14	Salzburg		5020	Austria	6562
HILAA	HILARIÓN-Abastos	Carlos Hernández	Sales Representative	Carrera 22 con	San Cristóbal	Táchira	5022	Venezuela	(5) 58
ERNSH	Ernst Handel	Roland Mendel	Sales Manager	Kirchgasse 6	Graz		8010	Austria	7675
VAFFE	Vaffeljernet	Palle Ibsen	Sales Manager	Smagsløget 45	Århus		8200	Denmark	86 2
Northwind	[Employees]	[ReportsTo]	<NULL>				<NULL>	CHA	
Northwind	[Order De	[OrderID]	10248				11077	CHA	

74 items Opened : 11:52:16 Loaded: 11:52:18 AM

User[Mike Neely] Mode[Normal] Logged[11:52:09 AM] Server[localhost]

Start i no... 12:44 PM

Metadata Refinement with Imperative Analyst (1 of 3)



After profiling the data, we use Imperative Analyst to review the results for each table and field, and then define or modify the following types of rules:

- **Table definitions (meanings)**
- **Field definitions (meanings)**
- **Data types and Lengths**
- **NOT NULL rules**
- **NOT BLANK rules**
- **NOT ZERO rules**
- **Minimum allowable values**
- **Maximum allowable values**
- **Allowable value sets**
- **Allowable pattern/format sets**
- **Primary Keys**
- **Special values and meanings, e.g., “99/99/9999”**



Metadata Refinement with Imperative Analyst (2 of 3)

The screenshot shows the Imperative Analyst software interface. The main window is titled "Imperative - [Master : Form]". It features a menu bar (File, Edit, Format, Records, Reports, Window, Help) and a toolbar with various icons. Below the menu bar are several tabs: "Setup Analysis", "Develop Data Rules", "Rule Change Statistics", "Summary Statistics By Table", "Statistics By Field", and "Data Assessment Details".

The interface is divided into two main sections: "Tables" and "Fields".

Tables Section:

Name	Definition	Include	Report
[Categories]	Product categories master table	<input checked="" type="checkbox"/>	<input type="checkbox"/>
[Customers]	Customers table	<input checked="" type="checkbox"/>	<input type="checkbox"/>
[Employees]	Employees table	<input checked="" type="checkbox"/>	<input type="checkbox"/>
[Order Det...]	Order line items	<input type="checkbox"/>	<input type="checkbox"/>
[Ord...]	Orders	<input type="checkbox"/>	<input type="checkbox"/>
[Ord...]	Orders available to order	<input type="checkbox"/>	<input type="checkbox"/>

Fields Section:

Field:	Definition:	Prim Key	Include	Priority	Unused
[Address]	Number, Street, Unit	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
[City]	City name	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
[CompanyName]	Corporate customer	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
[ContactName]	Customer Contact	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
[ContactTitle]	Contact title	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
[Country]	Country of registration	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
[CustomerID]	Customer ID--Internal Co	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
[CustomerFaxNumber]	Customer fax number	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
[CustomerPhoneNumber]	Customer phone number	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
[CustomerPostalCode]	Customer postal code	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
[CustomerType]	Customer type	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

At the bottom of the interface, there are several control buttons:

- Include All
- Exclude All
- Report All
- Report None
- Include All
- Exclude All
- Prioritize All
- Prioritize None

After Importing Profiling Results:

- Document Table Definitions (as needed)
- Specify which tables to include in analysis

For Each Included Table:

- Document Field Definitions (as needed)
- Specify / Refine Primary Key Definition
- Specify which fields to include in analysis
- Specify priority fields for analysis



Metadata Refinement with Imperative Analyst (3 of 4)

Other features of *Imperative Analyst* facilitate the use of data profiling results to define or refine the following types of metadata:

- **Table and Field Definitions**
- **Intended Data Types, Mins, Maxs**
- **Allowable Value Lists**
- **Allowable Pattern Lists**
- **Not Null, Not Blank, and Not Zero Rules**
- **“Special”, i.e., Control Values**

As the rules are defined, Analyst generates statistics and details regarding non-conformant data values. This helps with the the further refinement of the rules, and the development of data cleansing rules.



Metadata Refinement with Imperative Analyst (4 of 4)

Imperative - [Master : Form]

File Edit View Insert Format Records Tools Window Help

Type a question for help

Setup Analysis | Develop Data Rules | Rule Change Statistics | Summary Statistics By Table | Statistics By Field | Data Assessment Details

Analysis: All Included Used Fields | Table: [Orders] | Role: Customer orders involving one or more line items | Report | No. Records: 830

Field: [ShipRegion] | UniqueRatio: 0.0240964

Definition: State, province or region for shipping order

Primary Key Include
 Unused Priority

Documented	Inferred	Intended	Data Type Set
Data Type: VARCHAR	VARCHAR(13)	Character	Generic
Min Value:	AK		
Max Value:	WY		

Observed Values

Value:	Frequenc
AK	10
BC	17
CA	4
Co. Cork	19
DF	2
Essex	13
ID	31
Isle of Wight	10
Lara	14
MT	3
NM	18
Nueva Esparta	12
OR	28
Québec	13
RJ	34
SP	49
Táchira	18
WA	19
WY	9

Special Value Meaning

Special Value	Meaning

NbrNulls: 0 No NULLs Allowed
NbrBlanks: 0 No Blanks Allowed
NbrZeroes: 0 No Zeroes Allowed
Total: 0

Allowable Values

Value	M
AK	Arkans
BC	British C
CA	Californi
Co. Cork	County
DF	
Essex	
ID	
Isle of Wight	
Lara	
MT	Montan:
NM	New Me
Nueva Esparta	

Allowable Patterns

Pattern:

Observed Patterns

Pattern:	Freq
C(2)	12
COc(5)	1
Cc(3)	1
Cc(3)Wc(2)WCc(1
Cc(4)	1
Cc(4)WCc(6)	1
Cc.WCc(3)	1
CcOc(3)	1
	1

Record: 11 of 14

Record: 4 of 6



Metadata Dictionary

System	Definition
Sales	Manages orders, product pricing, customers, invoicing

System	Table	Definition
Sales	Orders	Master table of customer orders, dates, total price, discount, status

System	Table	Field	Definition	Data Type	Not			Allowable Values	Allowable Patterns	Value Limits	
					Null	Blank	Zero			Minimum	Maximum
Sales	Orders	<u>OrderID</u>	Unique identifier for each order	Integer	Yes	Yes	Yes				
Sales	Orders	<u>OrderDate</u>	Date order was placed	<u>DateTime</u>	Yes	Yes	Yes		99/99/9999	01/01/1988	12/31/2099
Sales	Orders	<u>CustomerID</u>	Foreign key to Customer	Integer	Yes	Yes	Yes				
Sales	Orders	Status	Status of Order								

Imperative Analyst generates a metadata dictionary documenting the definitions and rules developed for files, tables and fields. This is a key resource for the data conversion and migration team.



Assessing Data Quality Using Imperative Analyst

The screenshot shows the Imperative Analyst software interface. The window title is "Imperative - [Master : Form]". The menu bar includes File, Edit, View, Insert, Format, Records, Tools, Window, and Help. The toolbar contains various icons for file operations, editing, and analysis. The main window has several tabs: Setup Analysis, Develop Data Rules, Rule Change Statistics, Summary Statistics By Table, Statistics By Field, and Data Assessment Details. The "View Rule Violations" section has radio buttons for different rule types: Not Null/Blank/Zero, Allowable Values, Allowable Patterns, Allowable Min, Allowable Max, Valid Date, and Valid Number. Below this, a table shows the analysis results for the "[Customers]" table. The table has columns for Field, Definition, Num Vals, % Records, and Not In Allowed Patterns. The "PostalCode" field shows 90 values, with 98.9010989% of records in allowed patterns and 14 values not in allowed patterns. To the right, a "Retrieve Values" section shows a list of violating values and their frequencies. The "Auto Retrieve" checkbox is checked. At the bottom, a record navigation bar shows "Record: 1 of 6".

Table: [Customers] Definition: Master list of customers and associated informati Nbr Records: 91

Field	Definition	Num Vals	% Records	Not In Allowed Patterns
[Phone]	Phone number for custome	91	100	72
[PostalCode]	Postal code for surface ma	90	98.9010989	14

Retrieve Values Auto Retrieve

Value:	Frequency:
B-1180	1
B-6000	1
EC2 5NT	1
H1J 1C3	1
OX15 4NB	1
PO31 7PJ	1
S-844 67	1
S-958 22	1
SW7 1RZ	1
T2F 8M4	1
V3F 2K1	1
WA1 1DP	1
WX1 6LT	1
WX3 6FW	1

Record: 1 of 6

Once the data rules are defined, *Imperative Analyst* generates statistics showing the numbers of values violating each of the various types of rules for each field. Imperative Analyst also supports “drilling down” to the violating values. Inspection of these values may indicate the need for data cleansing, modifications to application programs, or further refinement of the data rules.



Data Report Card

Metadata Refinement Statistics

System	Num Fields	Defined	Data Typed	Not Null	Not Blank	Not Zero	w/ Value Set	w/ Pattern Set	w/ Min	w/ Max
Products	33	91%	94%	55%	55%	0%	0%	0%	0%	0%
Product Categories	148	75%	97%	1%	1%	0%	0%	0%	0%	0%
Warehouses	74	100%	99%	46%	95%	59%	0%	0%	0%	0%
Customers	77	99%	95%	18%	18%	0%	0%	8%	0%	0%
Suppliers	109	50%	50%	15%	5%	5%	0%	0%	2%	2%
Supplier Organizations	42	100%	86%	33%	86%	81%	2%	2%	0%	0%
Customer Organizations	37	92%	100%	30%	30%	0%	0%	8%	0%	0%
Orders	99	49%	100%	1%	1%	0%	0%	0%	0%	0%
Order Details	52	83%	100%	52%	21%	21%	0%	0%	0%	0%
Invoices	47	72%	100%	62%	62%	21%	0%	0%	6%	6%
Payments	166	93%	6%	6%	58%	54%	7%	14%	0%	0%
Shipments	62	100%	97%	29%	29%	0%	0%	5%	0%	0%
Shippers	2	100%	100%	50%	50%	0%	0%	0%	0%	0%
Customer Discounts								3%	0%	0%
Supplier Discounts								9%	3%	3%
Customer Demographics								9%	0%	0%
Promotions								5%	0%	0%
Promotion Responses								9%	0%	0%

Imperative Analyst generates statistics showing the percentages of tables and fields for which various types of metadata and rules have been defined. This may indicate where additional effort may be required to more completely define the data rules, as indicated here by red shading.

Data Profiling / Quality Assessment for a Large Airline

- One of South America's Largest Airlines
- Customer Data Repository Project
- *The Imperative Suite* profiled, documented, and assessed the quality of:

✓ **18 Systems**

✓ **44 Tables**

✓ **1241 Fields**

... in only **one person-month!**

Benefits of Data Quality and Integrity

Assuring the quality and integrity of its data is a crucial strategic initiative, for several reasons:

- 1) Quality data supports **better business intelligence**, which in turn means better decisions.
- 2) Quality data **reduces operations costs** by eliminating duplicate mailings, and other inefficiencies.
- 3) Quality data can **enhance sales** by enabling, for example, targeted cross-marketing, and household-level marketing.
- 4) Quality data means **fewer operational disruptions** and **less manual intervention**.