



My Journey with DataFlux

- Garry D'Lima
Business Solutions Administrator
December 13, 2013

HELPING OUR CLIENTS
FIND AND KEEP
GOOD CUSTOMERS

Content

- Introduction
- Objectives set by the management
- My Learning's
- Our Success
- Recommendations and Best Practices

Introduction

- Sharing with you my journey in resolving the **objectives** set by the management
- What **Dataflux** can do
- How we **solved** these objectives using Dataflux
- How we **implemented** some key functionalities through dfPower Studio

Need for Address Accuracy software

- Capable of handling the following features
 - Identifying data patterns (Profiling)
 - Cleansing
 - Standardizing
 - Address Accuracy
 - Match codes generation

Need for Address Accuracy software

- Different formats / styles of names and address

First Name	Last Name	Address	City	Province	Postal Code
JOHN	DOE	123 Main Street Unit \$ 101	Toronto	ON	M4E 2V9
J	DOE	101 - 123 Main Street		ON	M4E 2V9

First Name	Last Name	Address	City	Province	Postal Code
JOHN	DOE	101= 123 Main Street	Toronto	ON	M4E 2V9

Full Name	Address	City	Province	Postal Code
JOHN DOE	123 Main Street Apt 101	Toronto	ON	M4E 2V9
J DOE	101 * 123 Main St		Ontario	M4E 1V9

Objectives set by the management

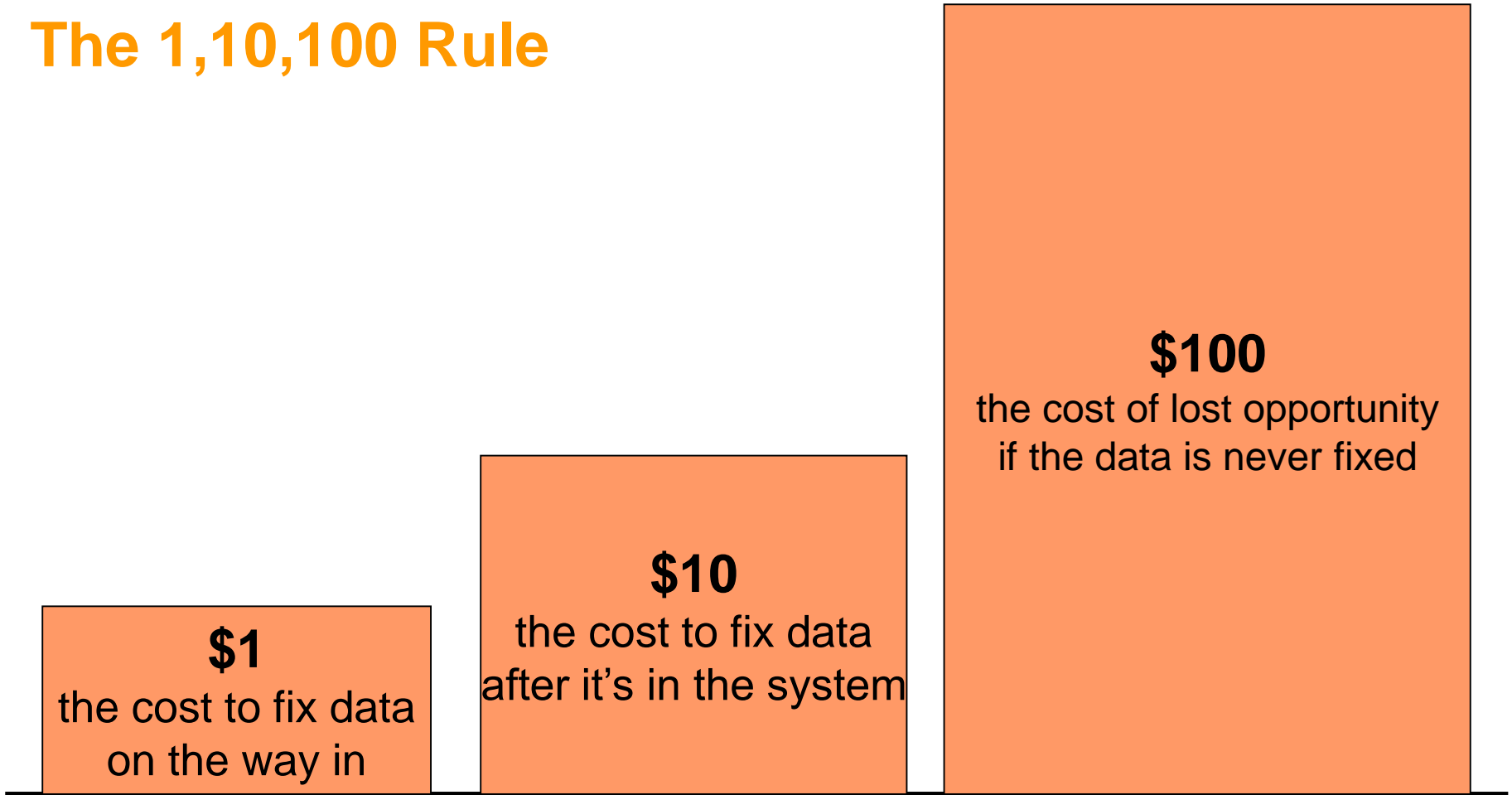
- **Transition** of Address Accuracy software to Dataflux
- Address Accuracy software to be **replaced** in the production environment in a short period of time
- Process to handle **multiple** input files from **various** sources with **different** layouts

Importance of Data Quality

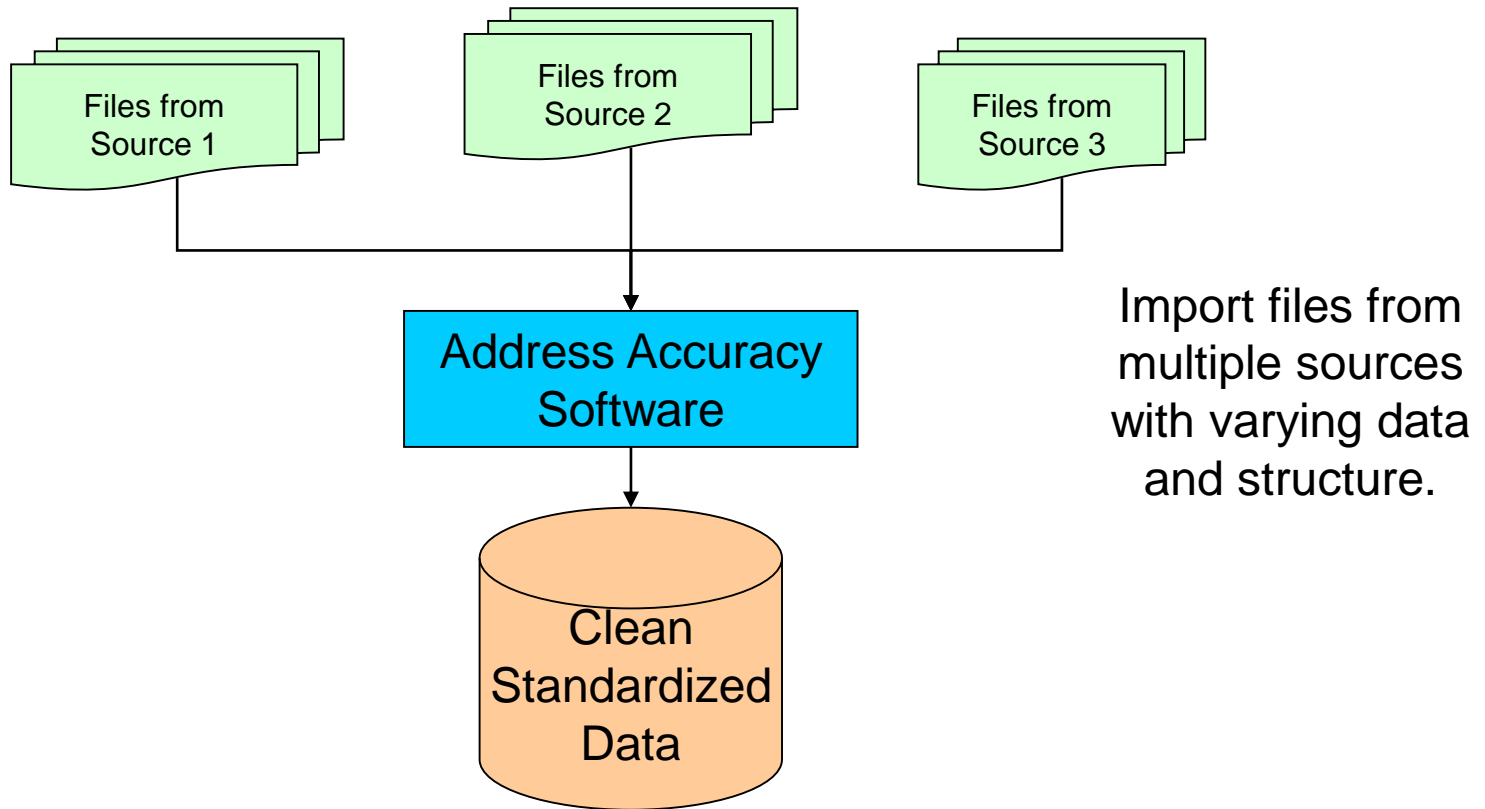
- The simple fact is that bad data **costs money**
- **Stop** bad data in its tracks before it gets out of hand
- Believe it; a bite is worse than a bark



The 1,10,100 Rule



Vision Flow diagram



My Learning's

- Why Dataflux?
- Dataflux Methodology
- What is dfPower Studio?
- dfPower Studio Architect

My Learning's

- Why DataFlux?
 - A **leader** in data quality (Gartner magic quadrant)
 - Provides data management capabilities, with a **focus** on data quality.
 - Enables organizations to **analyze, improve and control** their data through an integrated technology platform.

DataFlux Methodology

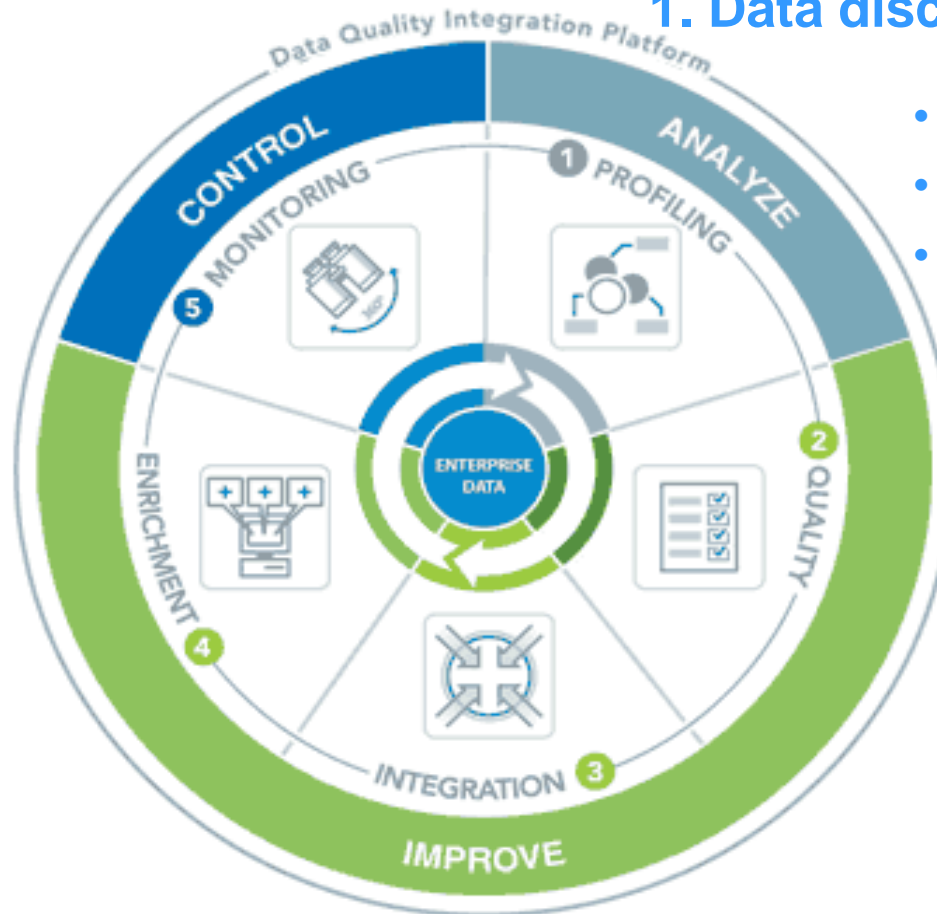
- 5 step approach for Data Improvement



DataFlux Methodology (5 step approach)

1. Data discovery or data auditing

- Structure discovery
- Data discovery
- Relationship discovery



DataFlux Methodology (5 step approach)



2. Data Quality

- Parsing
- Standardization
- Match Coding

DataFlux Methodology (5 step approach)



3. Data Integration

- Join
- Merge
- Householding

DataFlux Methodology (5 step approach)



4. Enriching Data

- Address Verification
- Phone Validation
- Geocoding

DataFlux Methodology (5 step approach)

5. Data Monitoring

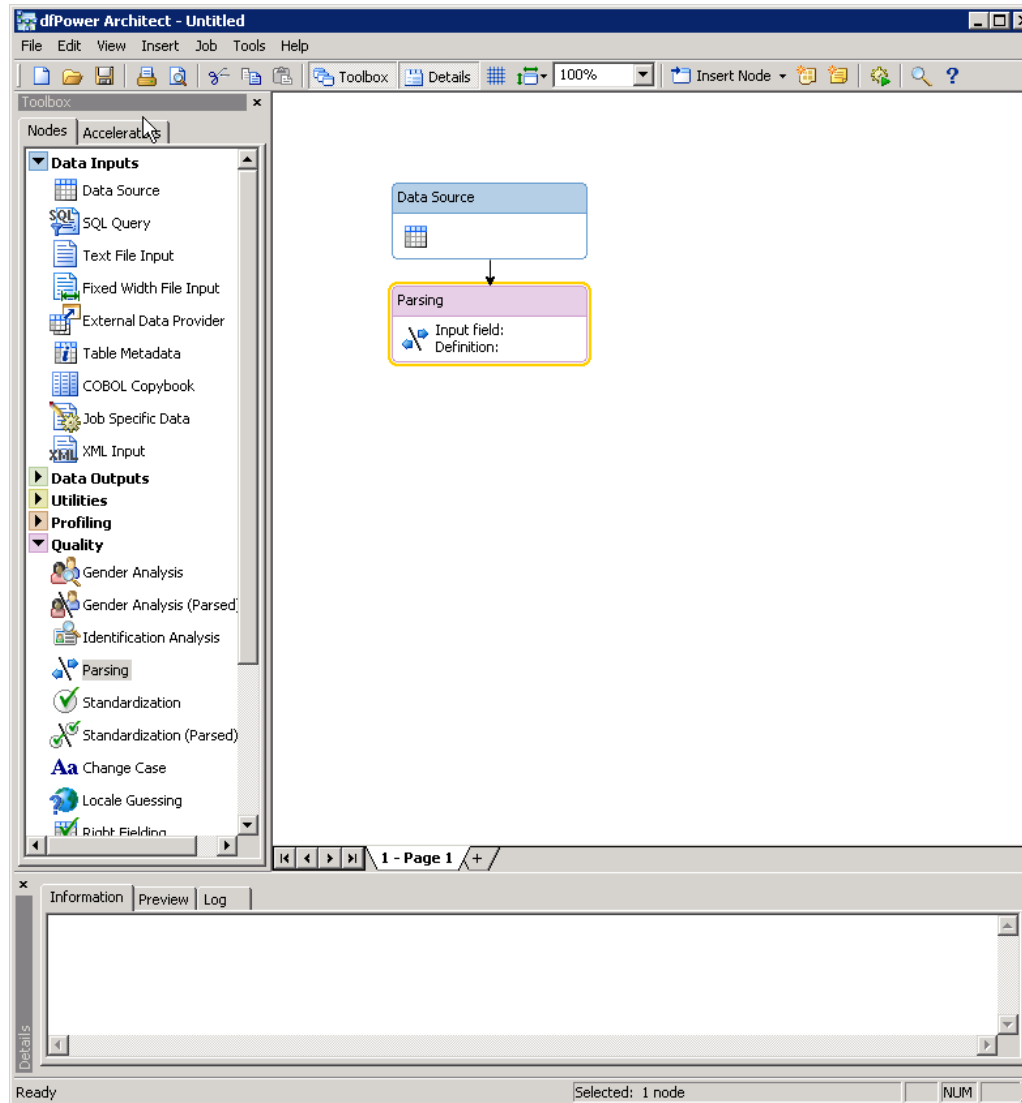
- Auditing
- Alerts



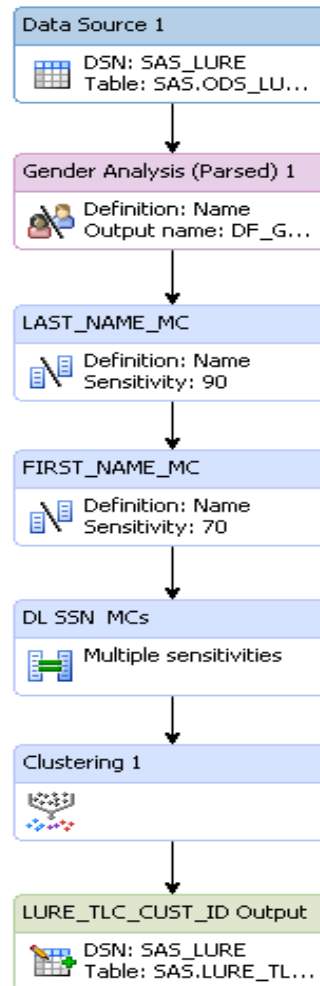
DataFlux dfPower Studio

- Connects to virtually any data source
- Design data quality rules and workflows
- Built up Quality Knowledge Base
- Use dfPower Studio to profile, cleanse, integrate, enrich, monitor, and otherwise improve data quality throughout the enterprise.
- Using dfPower Architect, an innovative job flow builder in dfPower Studio, enables users to build complex management workflows quickly and logically.

dfPower Architect



Job created using dfPower Architect



Open data set or text file

Add Gender field

Add match code on Last Name field

Add match code on First Name field

Add match codes on additional fields

Group rows based on match codes

Create output file

Our Success

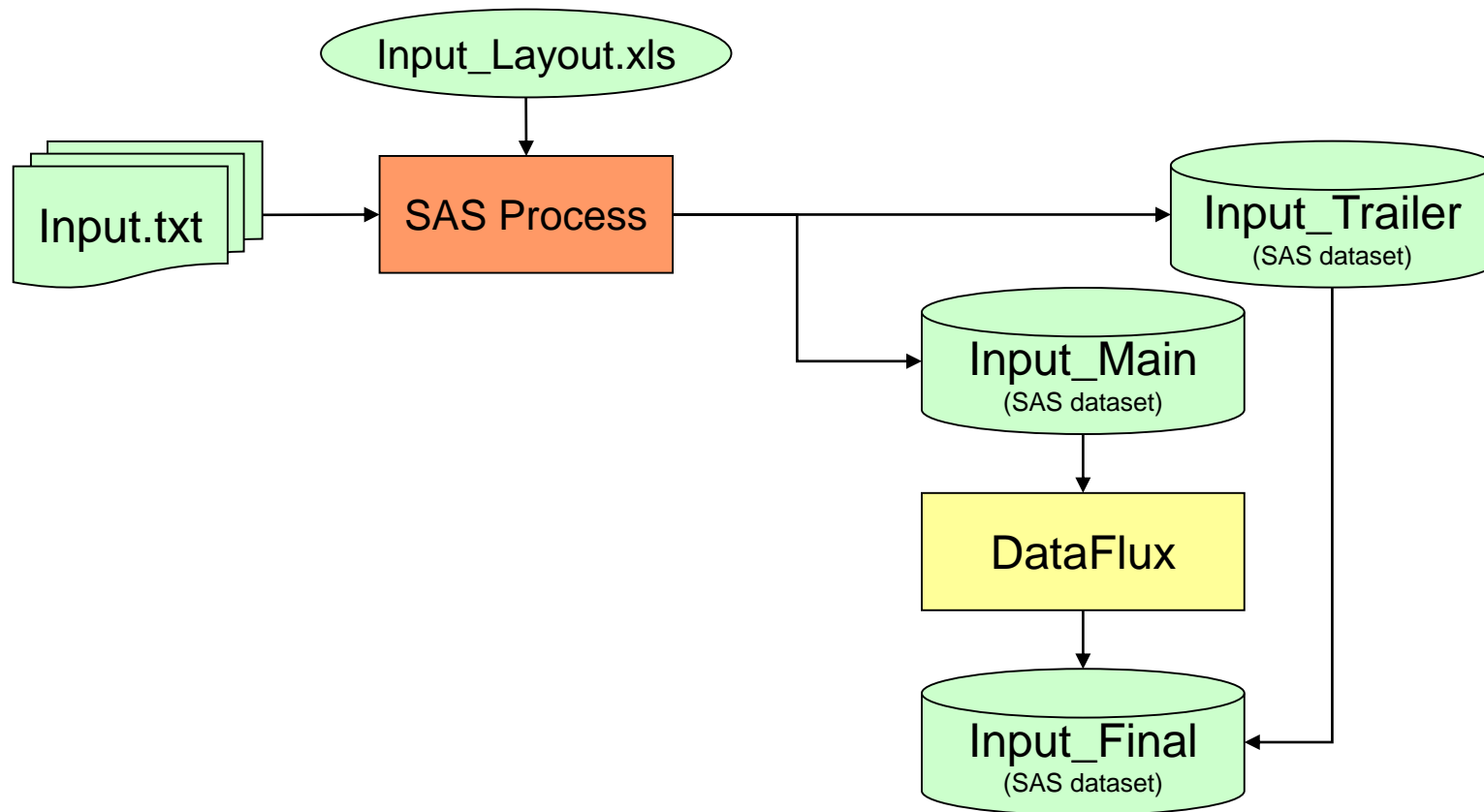
- Dataflux helped to create more **accurate** and **reliable** data through a suite of data quality, data integration, data governance and data management solutions
- With accurate and reliable data; helps making **better** and **faster** decisions
- Its flexibility to loosen and tighten the matching criteria to fit customer needs

Recommendations and Best Practices

- Our solution to handle multiple files and layouts
- Coolest feature
- Some pain points

Recommendations and Best Practices

- Our solution to handle multiple files and layouts



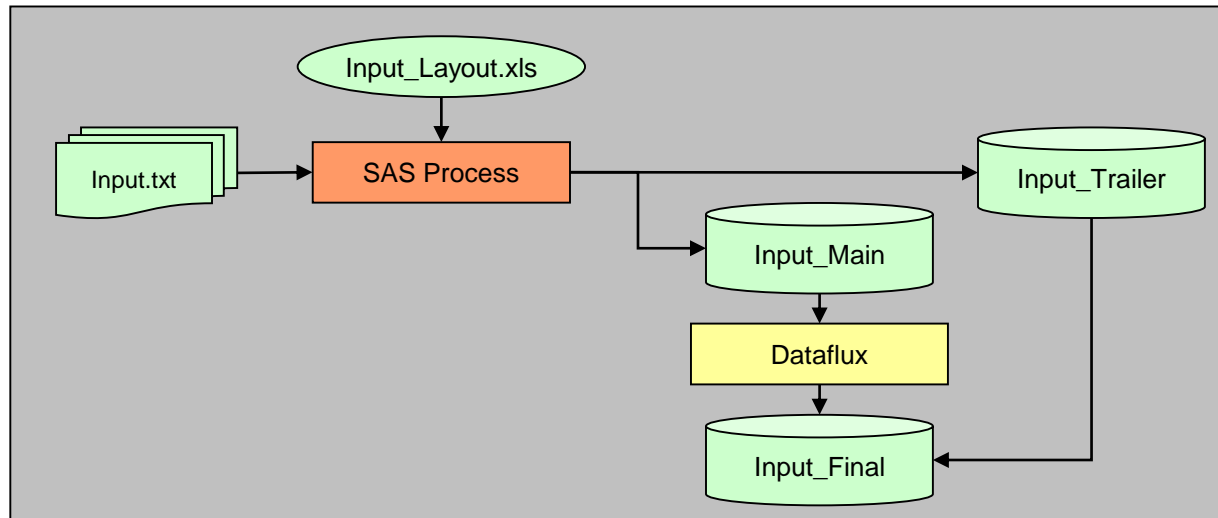
Recommendations and Best Practices

Input_Layout.xls

Filename	Format	Delimiter	Data Has Header	Data Has Trailer	Split Names
Input.txt	Fixed		NO	No	YES

Field	Type	Start	Length	Finish	Category	Order
MEMBER	CHAR	1	18	18	Trailer	Trailer1
FIRSTNAME	CHAR	19	15	33	Main	Name1
LASTNAME	CHAR	34	14	47	Main	Name2
ADDRESS1	CHAR	48	34	81	Main	Address1
ADDRESS2	CHAR	82	29	110	Main	Address2
CITY	CHAR	111	19	129	Main	City
PROVINCE	CHAR	130	2	131	Main	Province
POSTAL	CHAR	132	7	138	Main	PostalCode
LANGUAGE	CHAR	139	1	139	Trailer	Trailer2

Recommendations and Best Practices



Field	Type	Start	Length	Finish	Category	Order
MEMBER	CHAR	1	18	18	Trailer	Trailer1
FIRSTNAME	CHAR	19	15	33	Main	Name1
LASTNAME	CHAR	34	14	47	Main	Name2
ADDRESS1	CHAR	48	34	81	Main	Address1
ADDRESS2	CHAR	82	29	110	Main	Address2
CITY	CHAR	111	19	129	Main	City
PROVINCE	CHAR	130	2	131	Main	Province
POSTAL	CHAR	132	7	138	Main	Postal Code
LANGUAGE	CHAR	139	1	139	Trailer	Trailer2

Coollest feature

- Sensitivity of Match codes
 - Used during the creation of match codes
 - Changing the sensitivity level you can control what is considered a match
 - Sensitivity values ranges from 50 to 95
 - The default value is 85
 - For e.g.

Full Name
Patty Fielding
Patricia Feelding
Patricia J. Fielding

Some pain points

- Parsing of Full Names
 - Name switching

Full Name	First Name	Middle Name	Last Name
Paul Ryan	Paul		Ryan
Paul A Ryan	A	Ryan	Paul

Summary

- Where can DataFlux help you?
 - ✓ Proper definition of data and data values
 - ✓ Establishment and application of standards
 - ✓ Elimination of redundant data
 - ✓ Elimination of false null values
 - ✓ Validation of data during input
 - ✓ Resolution of data conflicts

Thank you....