Features of IBM DataWorks Forge in IBM Bluemix

After you complete this unit, you should understand:

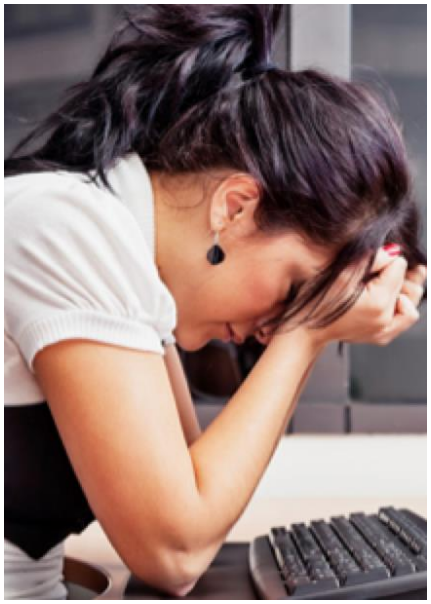
- Key characteristics of IBM DataWorks Forge Service in Bluemix
- How DataWorks Forge improves data collection and analysis

IBM DataWorks Forge: the IBM data refinery

- Provides a lightweight, cloud-based set of data refinement and access services
  - Makes fit-for-purpose data quickly and easily available to everyone across the enterprise.
- Ingests raw data, and provides services to enable you to perform:
  - Data preparation
  - Data movement and delivery
    - While also meeting associated privacy and security concerns before distributing refined data to end users.

Why Forge? How does Carol, a business analyst, go about analyzing data?

- Must ask IT for data and wait for it
- Ends up with raw data of undetermined quality
- If the data is the wrong data or the data sets are incomplete, she must wrangle with IT (more waiting)
- Culls through and joins data
- Gets the prepared data sets into a database or into analytics applications such as IBM Watson Analytics



IBM DataWorks Forge revolutionizes the experience for knowledge workers

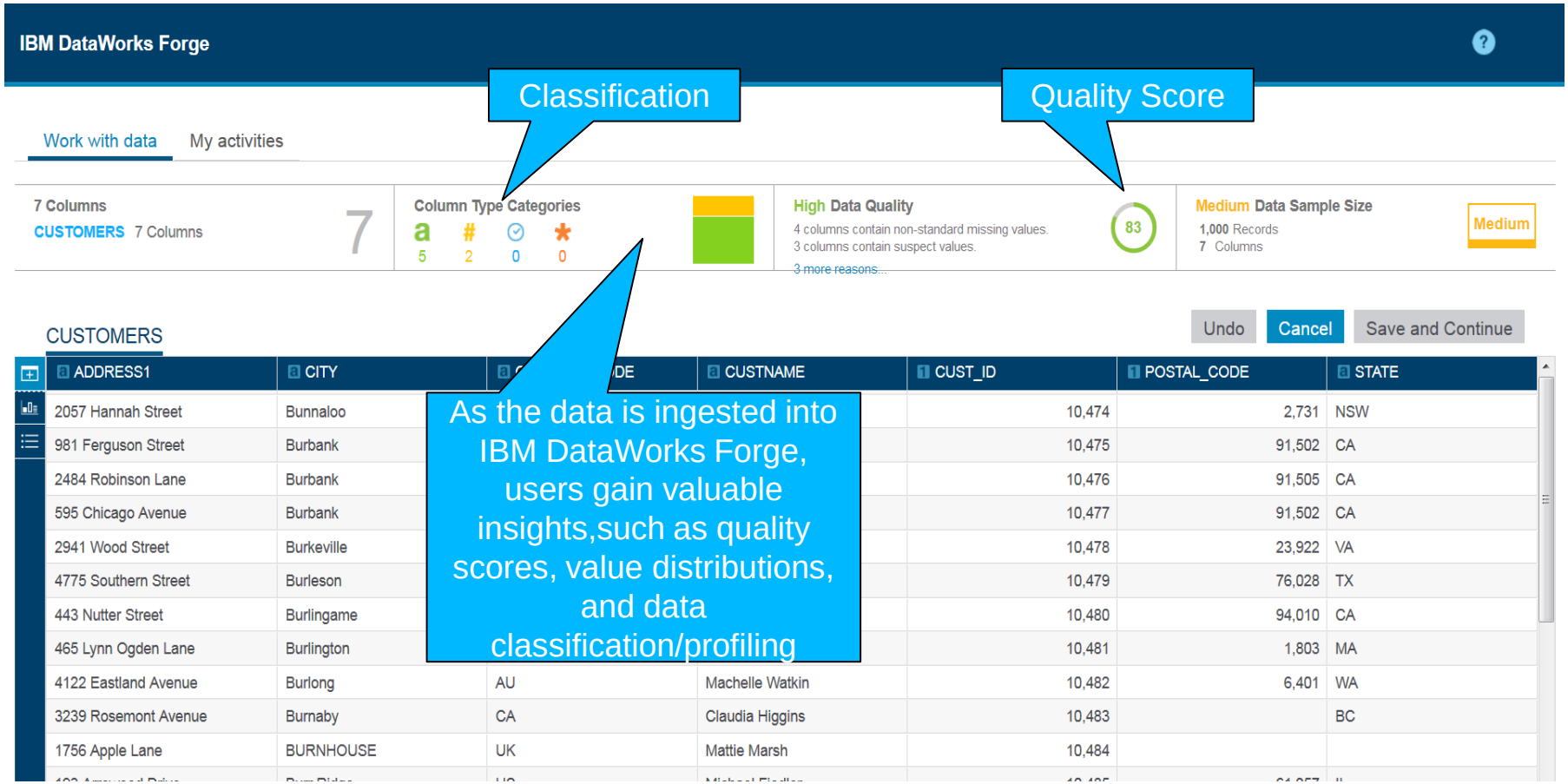
The better way:
IBM DataWorks Forge

- Carol is empowered to quickly find relevant data herself.
- Provides automatically profiled and classified data
- Delivers quality scores and value distributions so that Carol can visualize and understand the data
- Allows Carol to quickly and easily shape, enrich, and improve quality by joining, standardizing, filtering, removing duplicates, and so on
- Allows Carol to apply her changes and deliver the data to her chosen target or targets

IBM DataWorks Forge supported sources and targets

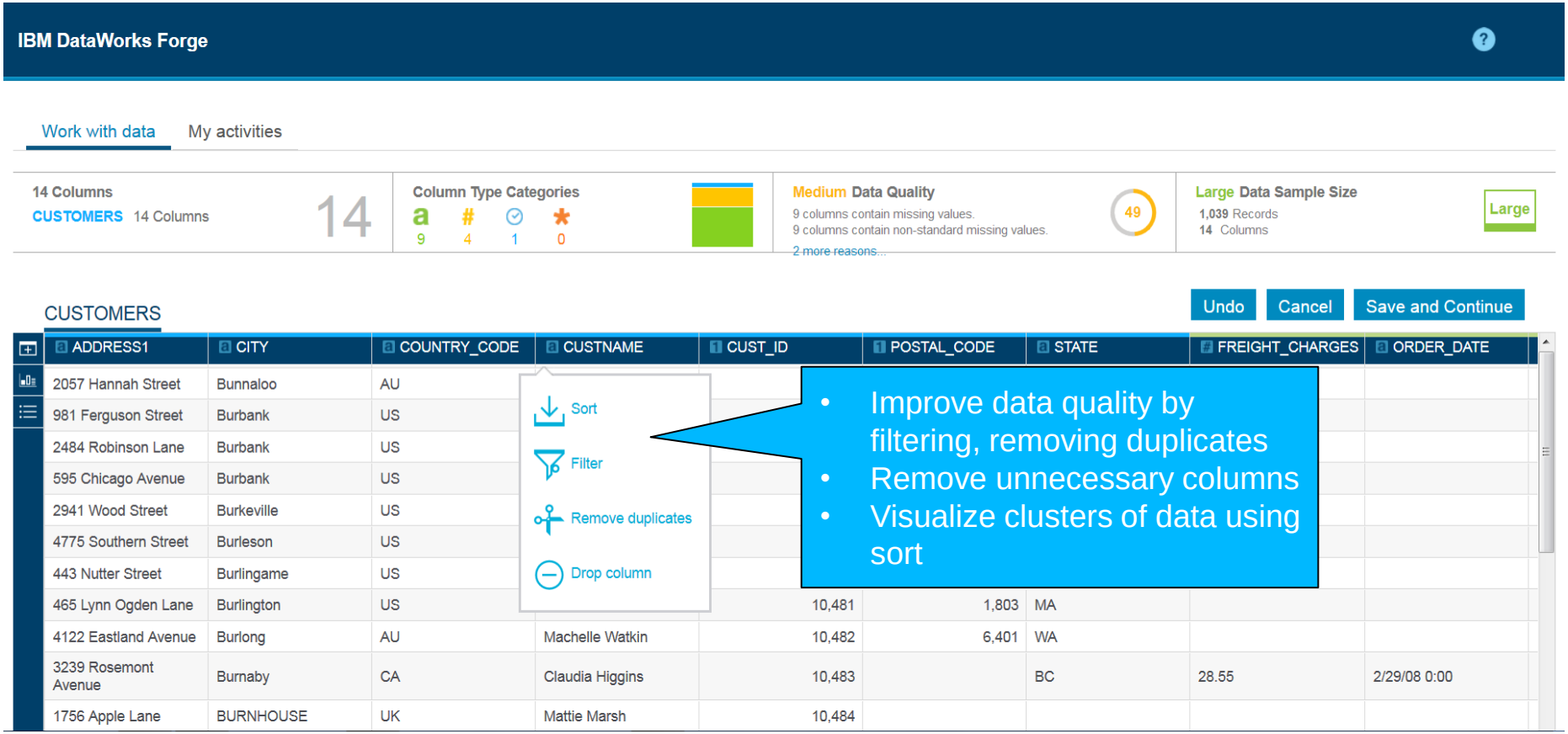
Sources	Targets
dashDB	Cloudant NoSQL DB
IBM DB2	dashDB
Oracle	IBM Watson Analytics
Salesforce.com	SQL Database
SQL Database	
Cloudera Impala	
Amazon Redshift	
IBM Informix	
IBM Netezza	
Microsoft Azure	
Microsoft SQL Server	
MySQL	
Oracle	
Pivotal Greenplum	
PostgreSQL	
Sybase	
Sybase IQ	

IBM DataWorks Forge at work



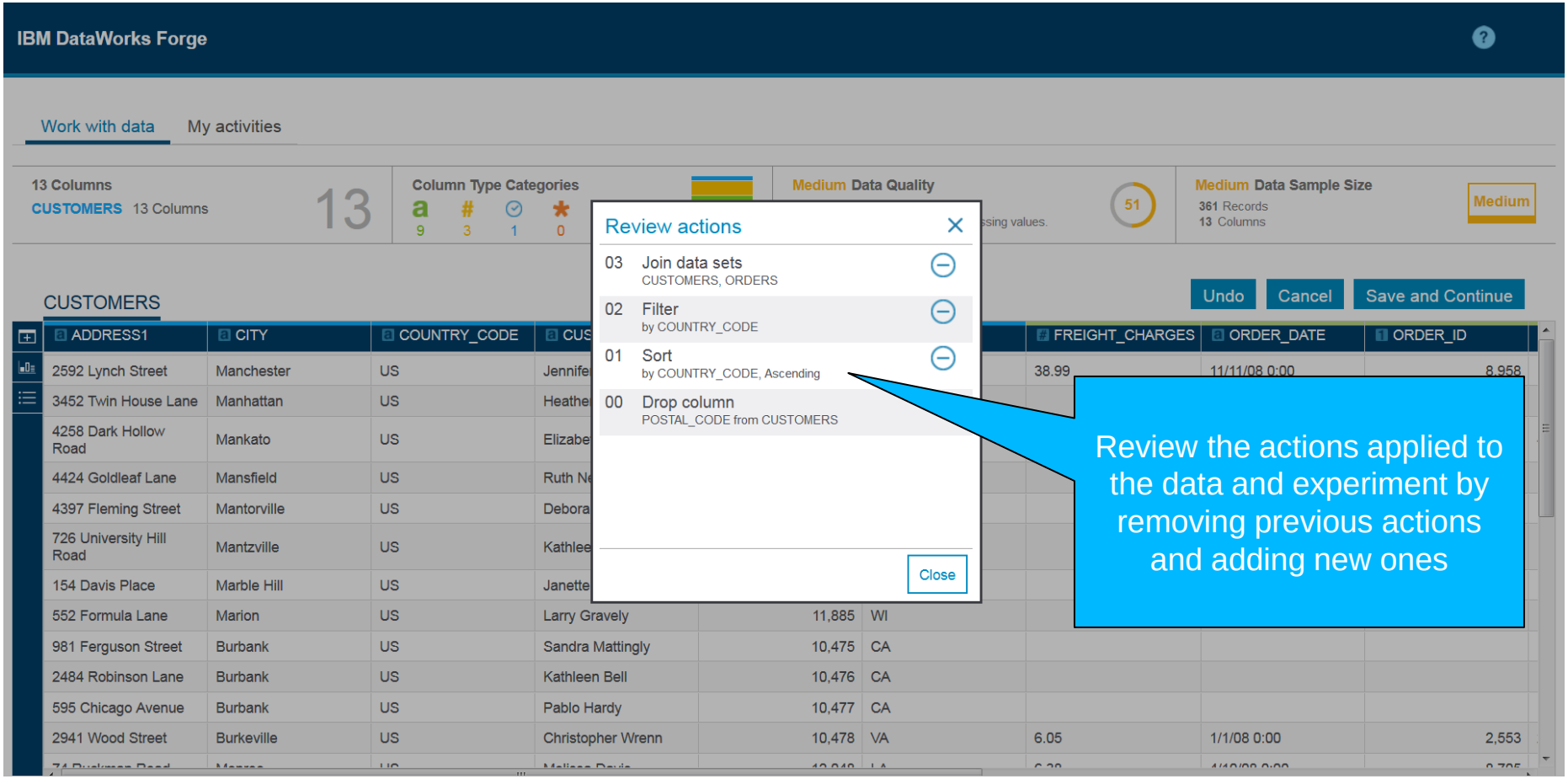
ADDRESS1	CITY	CUSTNAME	CUST_ID	POSTAL_CODE	STATE
2057 Hannah Street	Bunnaloo		10,474	2,731	NSW
981 Ferguson Street	Burbank		10,475	91,502	CA
2484 Robinson Lane	Burbank		10,476	91,505	CA
595 Chicago Avenue	Burbank		10,477	91,502	CA
2941 Wood Street	Burkeville		10,478	23,922	VA
4775 Southern Street	Burleson		10,479	76,028	TX
443 Nutter Street	Burlingame		10,480	94,010	CA
465 Lynn Ogden Lane	Burlington		10,481	1,803	MA
4122 Eastland Avenue	Burlong	AU	Machelle Watkin	10,482	6,401	WA
3239 Rosemont Avenue	Burnaby	CA	Claudia Higgins	10,483		BC
1756 Apple Lane	BURNHOUSE	UK	Mattie Marsh	10,484		

Improve data quality with IBM DataWorks Forge



ADDRESS1	CITY	COUNTRY_CODE	CUSTNAME	CUST_ID	POSTAL_CODE	STATE	FREIGHT_CHARGES	ORDER_DATE
2057 Hannah Street	Bunnaloo	AU						
981 Ferguson Street	Burbank	US						
2484 Robinson Lane	Burbank	US						
595 Chicago Avenue	Burbank	US						
2941 Wood Street	Burkeville	US						
4775 Southern Street	Burleson	US						
443 Nutter Street	Burlingame	US						
465 Lynn Ogden Lane	Burlington	US		10,481	1,803	MA		
4122 Eastland Avenue	Burlong	AU	Machelle Watkin	10,482	6,401	WA		
3239 Rosemont Avenue	Burnaby	CA	Claudia Higgins	10,483		BC	28.55	2/29/08 0:00
1756 Apple Lane	BURNHOUSE	UK	Mattie Marsh	10,484				

Enrich and experiment with data with IBM DataWorks Forge



Summary

- IBM DataWorks Forge service in Bluemix allows you to:
  - Identify relevant data
  - Transform the Data to suit your needs
  - Load it to a system for use