From Files and Web Services

School of Information Studies
Syracuse University

Extraction From File Systems

- Not all OLTP sources are RDBMS or can be extracted via query.

- Use metadata columns in stage table to keep track of which files were processed when.

Order No	Order Date	Order Amt	Processed Date	Processed File
101425	12/14/2016	$450.02	1/1/2017	o9203.dat
101426	12/20/2016	$380.55	1/1/2017	o9203.dat
101427	12/30/2016	$1,968.90	1/1/2017	o9203.dat
101428	1/5/2017	$192.40	2/1/2017	o9204.dat
101429	1/12/2017	$500.25	2/1/2017	o9204.dat

o9203.dat

o9203.dat

School of Information Studies
Syracuse University

Extract From Web Services or Web Scraping

- Web has a lot of useful data for the data warehouse, especially for data mining and machine learning.
  - Weather: store visits impacted by weather?
  - Product data of competitors
  - Social media: what are people saying about us?

- Custom programs to extract the data.

- Store on file system first. Web is notorious for being here today and gone tomorrow.

- Then use the file system approach.