

September 2019

# How Urban Uses Big Data Systems to Democratize Access to LED Data

Vivian Zheng, Data Science Analyst, Office of Technology and Data Science



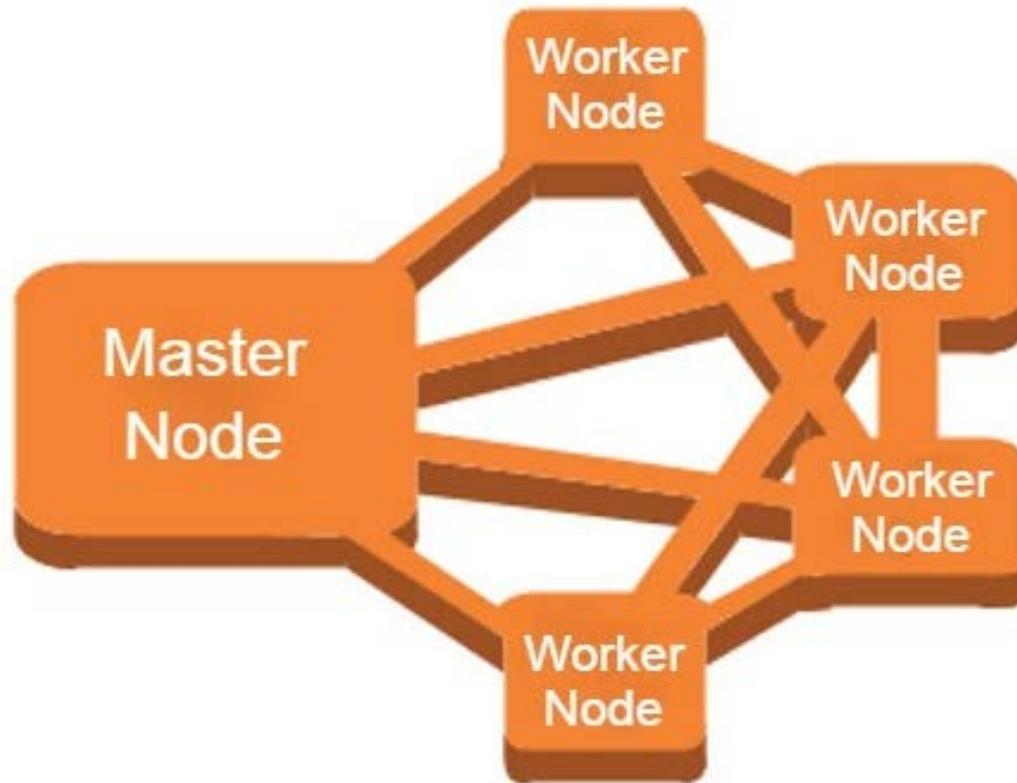
# Easy and Trusted Access to Data

# LEHD Origin-Destination Employment Statistics (LODES)

- Block-level Economy Information
  - 11 million census blocks; over 75,000 zipped CSV files
  - Simplify data downloading and processing
- Tract-level and place-level LODES Data
  - Available at Urban's Data Catalog (<https://datacatalog.urban.org/search/type/dataset>)

# How to Be More Efficient - Technology

## On-Demand Cloud Clusters



Reduce processing time: 500 hours → 10min

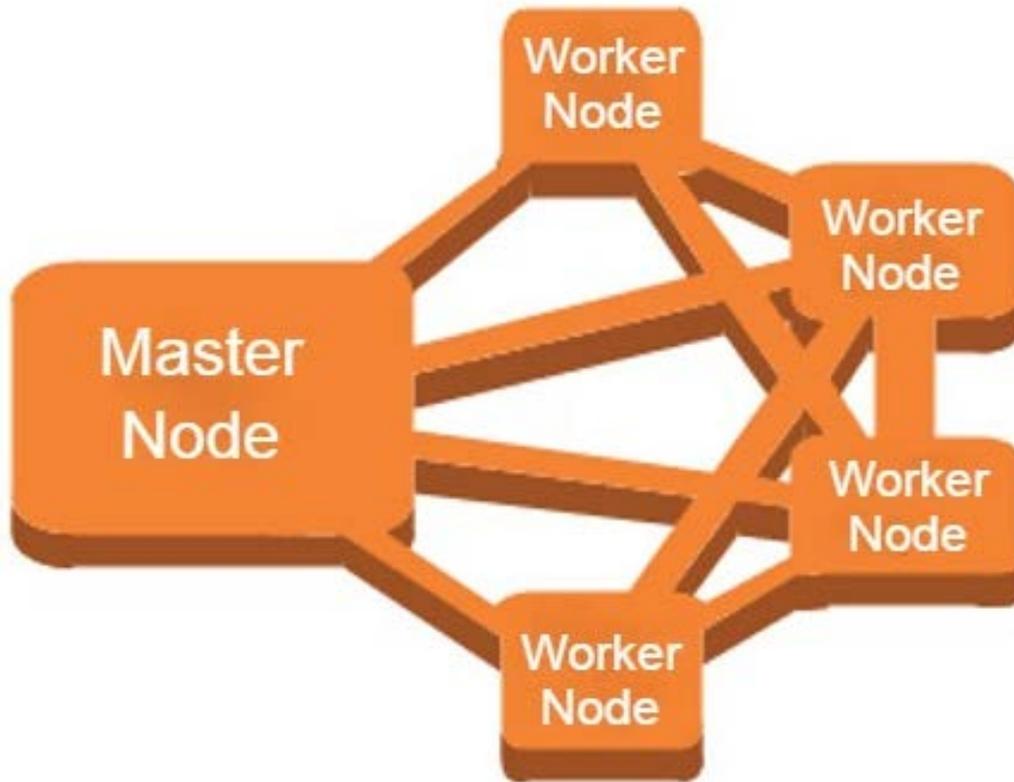
Save money: \$2/hour

# How to be efficient - Technology

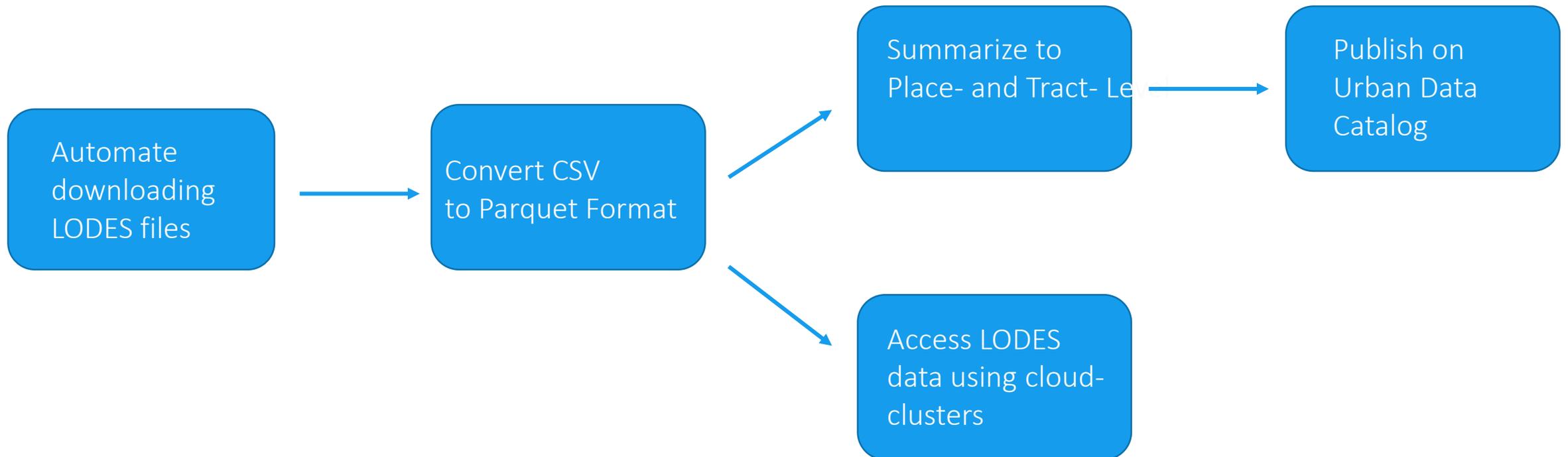
On-demand Cloud Clusters

+

Python/R



# Pipeline for Processing LODES Data



# 1. Automate downloading LODES files

## LEHD Origin-Destination Employment Statistics (LODES)

LEHD Origin-Destination Employment Statistics (LODES) used by [OnTheMap](#) are available for download below. Version 7 of LODES was enumerated by 2010 census blocks. Previous versions of LODES were enumerated with 2000 census blocks.

Data files are state-based and organized into three types: Origin-Destination (OD), Residence Area Characteristics (RAC), and Workplace Area Characteristics (WAC), all at census block geographic detail. Data is available for most states for the years 2002–2017.

To browse the LODES data files in their directory structure or to access them with a FTP program (must be able to access HTTP), go to [lehd.ces.census.gov/data/lodes/](http://lehd.ces.census.gov/data/lodes/).

### Download LODES data\*:

Version:  State/Territory:

Type:

[View Files](#) [Close](#)

[Metadata for DC](#) | [Geography crosswalk for DC](#) | [DC sha256sum file](#)

File Name	Date	Size
<a href="#">dc_rac_S000_JT00_2002.csv.gz</a>	22 Sep 2017 01:30	72 KB
<a href="#">dc_rac_S000_JT00_2003.csv.gz</a>	22 Sep 2017 01:30	68 KB
<a href="#">dc_rac_S000_JT00_2004.csv.gz</a>	22 Sep 2017 01:30	69 KB
<a href="#">dc_rac_S000_JT00_2005.csv.gz</a>	22 Sep 2017 01:30	70 KB
<a href="#">dc_rac_S000_JT00_2006.csv.gz</a>	22 Sep 2017 01:30	71 KB
<a href="#">dc_rac_S000_JT00_2007.csv.gz</a>	22 Sep 2017 01:30	72 KB
<a href="#">dc_rac_S000_JT00_2008.csv.gz</a>	22 Sep 2017 01:30	73 KB
<a href="#">dc_rac_S000_JT00_2009.csv.gz</a>	22 Sep 2017 01:30	109 KB
<a href="#">dc_rac_S000_JT00_2010.csv.gz</a>	22 Sep 2017 01:30	146 KB
<a href="#">dc_rac_S000_JT00_2011.csv.gz</a>	22 Sep 2017 01:30	139 KB
<a href="#">dc_rac_S000_JT00_2012.csv.gz</a>	22 Sep 2017 01:30	140 KB
<a href="#">dc_rac_S000_JT00_2013.csv.gz</a>	22 Sep 2017 01:30	140 KB

# Index of /data/lodes/LODES7

Name	Last modified	Size	Description
<a href="#">Parent Directory</a>		-	
<a href="#">LODESTechDoc7.0.pdf</a>	2017-09-25 09:04	60K	
<a href="#">LODESTechDoc7.1.pdf</a>	2017-09-25 09:04	61K	
<a href="#">LODESTechDoc7.2.pdf</a>	2017-09-25 09:04	60K	
<a href="#">LODESTechDoc7.3.pdf</a>	2017-09-21 21:25	61K	
<a href="#">ak/</a>	2017-09-21 21:34	-	
<a href="#">al/</a>	2017-09-21 21:44	-	
<a href="#">ar/</a>	2017-09-21 21:39	-	
<a href="#">az/</a>	2017-09-21 21:33	-	
<a href="#">ca/</a>	2017-09-21 21:37	-	
<a href="#">co/</a>	2017-09-21 21:51	-	
<a href="#">ct/</a>	2017-09-21 21:21	-	
<a href="#">dc/</a>	2017-09-21 21:30	-	
<a href="#">de/</a>	2017-09-21 21:45	-	
<a href="#">fl/</a>	2017-09-21 21:15	-	
<a href="#">ga/</a>	2017-09-21 21:49	-	
<a href="#">hi/</a>	2017-09-21 21:42	-	
<a href="#">ia/</a>	2017-09-21 21:00	-	
<a href="#">id/</a>	2017-09-21 21:08	-	
<a href="#">il/</a>	2017-09-21 21:10	-	
<a href="#">in/</a>	2017-09-21 21:03	-	
<a href="#">ks/</a>	2017-09-21 21:06	-	
<a href="#">ky/</a>	2017-09-21 21:40	-	
<a href="#">la/</a>	2017-09-21 21:32	-	
<a href="#">ma/</a>	2017-09-21 21:50	-	
<a href="#">md/</a>	2017-09-21 21:31	-	
<a href="#">me/</a>	2017-09-21 21:07	-	

## 2. Convert CSV to Parquet Format

- a. Compressed File
- b. Columnar Storage Format

### Row Oriented Storage Format

ID	Name	Salary
1,	X,	10000
2,	Y,	20000
3,	Z,	35000

### Columnar Storage Format

ID	1,	2,	3
Name	X,	Y,	Z
Salary	10000,	20000,	35000

This contain eight LODES census tract level summary files—four Residence Area Characteristics (RAC), and four Workplace Area Characteristics (WAC) files. RAC and WAC data each contains four files representing four job types—all jobs, primary jobs, all federal jobs, and federal primary jobs.

## Data and Resources

 <b>RAC All Jobs</b> Residence Area Characteristics (RAC), job type: all jobs	<a href="#">Preview</a>	<a href="#">Download</a>
 <b>RAC Primary Jobs</b> Residence Area Characteristics (RAC), job type: primary jobs	<a href="#">Preview</a>	<a href="#">Download</a>
 <b>WAC All Jobs</b> Workplace Area Characteristics (WAC), job type: all jobs	<a href="#">Preview</a>	<a href="#">Download</a>
 <b>WAC Primary jobs</b> Workplace Area Characteristics (WAC), job type: primary jobs	<a href="#">Preview</a>	<a href="#">Download</a>
 <b>RAC All Federal Jobs</b> Residence Area Characteristics (RAC), job type: all federal jobs	<a href="#">Preview</a>	<a href="#">Download</a>
 <b>RAC Federal Primary Jobs</b> Residence Area Characteristics (RAC), job type: federal primary jobs	<a href="#">Preview</a>	<a href="#">Download</a>
 <b>WAC All Federal Jobs</b> Workplace Area Characteristics (WAC), job type: all federal jobs	<a href="#">Preview</a>	<a href="#">Download</a>
 <b>WAC Federal Primary jobs</b> Workplace Area Characteristics (WAC), job type: federal primary jobs	<a href="#">Preview</a>	<a href="#">Download</a>

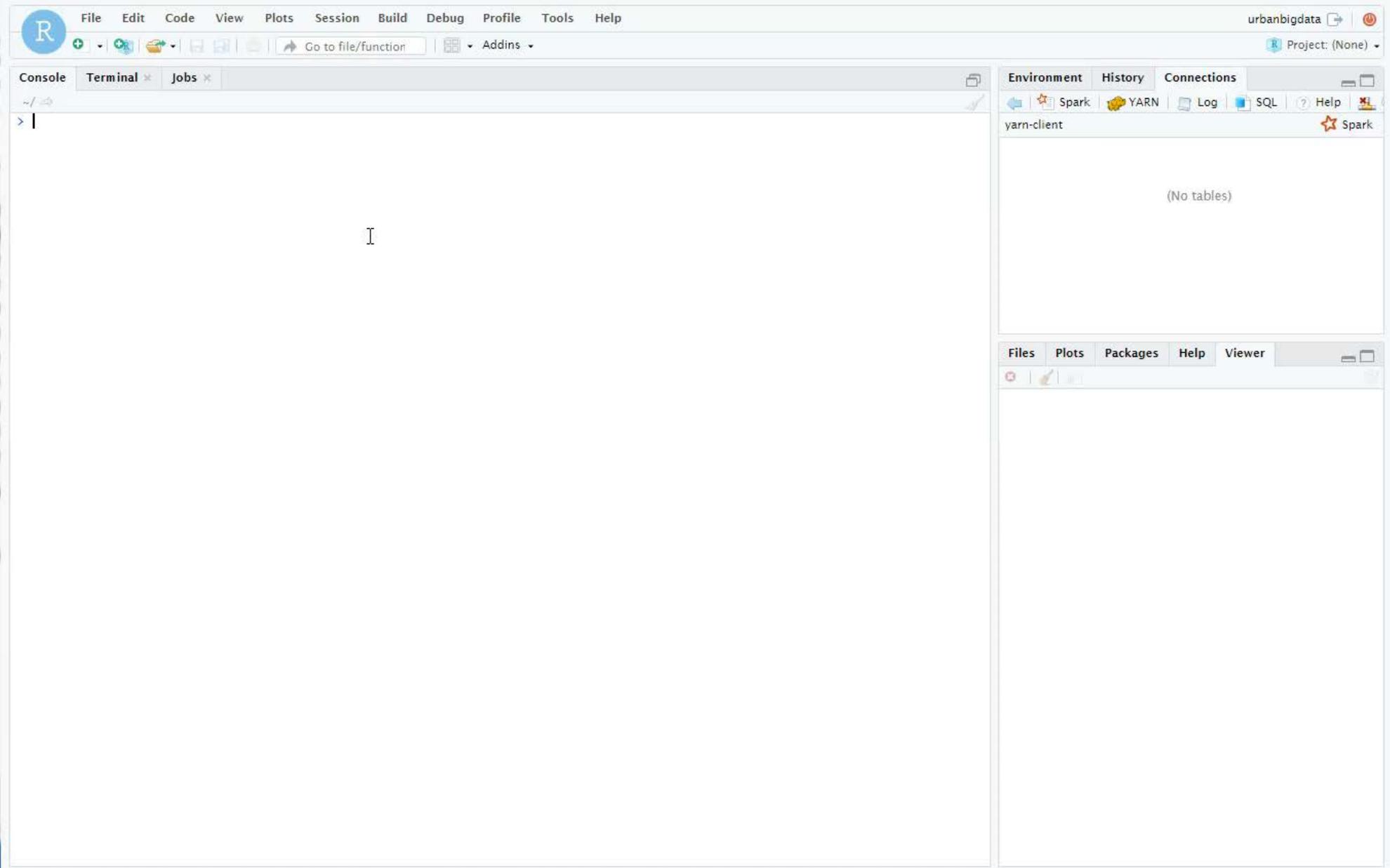
[Download All](#) 

## 3. Summarize to Place Level and Tract Level

- Merge with geography crosswalks using block ID
- Sum up the total number of workers in each block
- Parallel in processing



# Urban Big Data System - Processing LODES Data in Cloud Cluster



# Analytical Examples

- Theodos, Brett, Eric Hangen, Jay Dev, and Sierra Latham. "Coming Back from the Brink." (2017)
  - Capital Flows and Neighborhoods Patterns in Commercial, Industrial, and Multifamily Investment in Detroit
- Theodos, Brett, Brady Meixell, and Carl Hedman. "Did States Maximize Their Opportunity Zone Selections?" (2018).

# How to Access the Summarized LODES Data

- Urban's Data Catalog (<https://datacatalog.urban.org/search/type/dataset>)
  - Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES) Summary Files—Census Tract Level
  - Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES) Summary Files—Census Place Level
- Access LODES using the cloud clusters (Urban Internal)

## Future work

- Update new LODES data
- Include more job types for RAC and WAC
- Include OD
- Integrate other big datasets

Thank you!