

Overview - Scrapers

- Download market data and store it in the “scraped data” S3 bucket
- Organized hierarchically by collection and service
 - “collection” generally refers to a market (e.g., MISO, PJM)
 - “service” refers to the source of the data (e.g., a web API, an FTP file server)
- Each collection contains one or more services
- Each service provides one or more types of data
 - Fetching data for PJM alone requires accessing 6 different services

Overview - Parsers

- Extract relevant information from scraped files
- Convert data to a uniform format (CSV)
- Parsed data is stored in the “normalized” S3 bucket
- Each scraper is paired with a parser
 - Together a scraper and parser are referred to as a retriever