

List of dicts

Let us see an example of how we can read data from a file into **list of dicts** using Python as programming language.

- When we read data from a file into a **list**, typically each element in the **list** will be of type binary or string.
- We can convert the element into **dict** to simplify the processing.
- Once each element is converted to **dict**, we can access elements in the **dict** using attribute name.
- Let us see an example to read the data from a file into **list of dicts** and access dates.

```
# Reading data from file into a list
path = '/data/retail_db/orders/part-00000'
# C:\users\itiversity\Research\data\retail_db\orders\part-00000
orders_file = open(path)
```

```
%%sh
```

```
wc -l /data/retail_db/orders/part-00000
```

```
68883 /data/retail_db/orders/part-00000
```

```
type(orders_file)
```

```
_io.TextIOWrapper
```

```
orders_raw = orders_file.read()
```

```
orders = orders_raw.splitlines()
```

```
orders[:10]
```

```
['1,2013-07-25 00:00:00.0,11599,CLOSED',
 '2,2013-07-25 00:00:00.0,256,PENDING_PAYMENT',
 '3,2013-07-25 00:00:00.0,12111,COMPLETE',
 '4,2013-07-25 00:00:00.0,8827,CLOSED',
 '5,2013-07-25 00:00:00.0,11318,COMPLETE',
 '6,2013-07-25 00:00:00.0,7130,COMPLETE',
 '7,2013-07-25 00:00:00.0,4530,COMPLETE',
 '8,2013-07-25 00:00:00.0,2911,PROCESSING',
 '9,2013-07-25 00:00:00.0,5657,PENDING_PAYMENT',
 '10,2013-07-25 00:00:00.0,5648,PENDING_PAYMENT']
```

```
len(orders) # same as number of records in the file
```

```
68883
```

```
def get_order_dict(order):
    order_details = order.split(',')
    order_dict = {
        'order_id': int(order_details[0]),
        'order_date': order_details[1],
        'order_customer_id': int(order_details[2]),
        'order_status': order_details[3],
    }
    return order_dict
```

```
get_order_dict(orders[0])
```

```
{'order_id': 1,
 'order_date': '2013-07-25 00:00:00.0',
 'order_customer_id': 11599,
 'order_status': 'CLOSED'}
```

```
order_dicts = [get_order_dict(order) for order in orders]
```

```
type(order_dicts)
```

```
list
```

```
type(order_dicts[0])
```

```
dict
```

```
order_dicts[0]
```

```
{'order_id': 1,  
 'order_date': '2013-07-25 00:00:00.0',  
 'order_customer_id': 11599,  
 'order_status': 'CLOSED'}
```

```
order_dicts[:3]
```

```
[{'order_id': 1,  
  'order_date': '2013-07-25 00:00:00.0',  
  'order_customer_id': 11599,  
  'order_status': 'CLOSED'},  
 {'order_id': 2,  
  'order_date': '2013-07-25 00:00:00.0',  
  'order_customer_id': 256,  
  'order_status': 'PENDING_PAYMENT'},  
 {'order_id': 3,  
  'order_date': '2013-07-25 00:00:00.0',  
  'order_customer_id': 12111,  
  'order_status': 'COMPLETE'}]
```

```
len(order_dicts)
```

```
68883
```

```
order_dates = [order['order_date'] for order in order_dicts]
```

```
order_dates[:3]
```

```
['2013-07-25 00:00:00.0', '2013-07-25 00:00:00.0', '2013-07-25 00:00:00.0']
```

```
len(order_dates)
```

```
68883
```

```
set(order_dates)
```

```
{ '2013-07-25 00:00:00.0',  
  '2013-07-26 00:00:00.0',  
  '2013-07-27 00:00:00.0',  
  '2013-07-28 00:00:00.0',  
  '2013-07-29 00:00:00.0',  
  '2013-07-30 00:00:00.0',  
  '2013-07-31 00:00:00.0',  
  '2013-08-01 00:00:00.0',  
  '2013-08-02 00:00:00.0',  
  '2013-08-03 00:00:00.0',  
  '2013-08-04 00:00:00.0',  
  '2013-08-05 00:00:00.0',  
  '2013-08-06 00:00:00.0',  
  '2013-08-07 00:00:00.0',  
  '2013-08-08 00:00:00.0',  
  '2013-08-09 00:00:00.0',  
  '2013-08-10 00:00:00.0',  
  '2013-08-11 00:00:00.0',  
  '2013-08-12 00:00:00.0',  
  '2013-08-13 00:00:00.0',  
  '2013-08-14 00:00:00.0',  
  '2013-08-15 00:00:00.0',  
  '2013-08-16 00:00:00.0',  
  '2013-08-17 00:00:00.0',  
  '2013-08-18 00:00:00.0',  
  '2013-08-19 00:00:00.0',  
  '2013-08-20 00:00:00.0',  
  '2013-08-21 00:00:00.0',  
  '2013-08-22 00:00:00.0',  
  '2013-08-23 00:00:00.0',  
  '2013-08-24 00:00:00.0',  
  '2013-08-25 00:00:00.0',  
  '2013-08-26 00:00:00.0',  
  '2013-08-27 00:00:00.0',  
  '2013-08-28 00:00:00.0',  
  '2013-08-29 00:00:00.0',  
  '2013-08-30 00:00:00.0',  
  '2013-08-31 00:00:00.0',  
  '2013-09-01 00:00:00.0',  
  '2013-09-02 00:00:00.0',  
  '2013-09-03 00:00:00.0',  
  '2013-09-04 00:00:00.0',  
  '2013-09-05 00:00:00.0',  
  '2013-09-06 00:00:00.0',  
  '2013-09-07 00:00:00.0',  
  '2013-09-08 00:00:00.0',  
  '2013-09-09 00:00:00.0',  
  '2013-09-10 00:00:00.0',  
  '2013-09-11 00:00:00.0',  
  '2013-09-12 00:00:00.0',  
  '2013-09-13 00:00:00.0',  
  '2013-09-14 00:00:00.0',  
  '2013-09-15 00:00:00.0',  
  '2013-09-16 00:00:00.0',  
  '2013-09-17 00:00:00.0',  
  '2013-09-18 00:00:00.0',  
  '2013-09-19 00:00:00.0',  
  '2013-09-20 00:00:00.0',  
  '2013-09-21 00:00:00.0',  
  '2013-09-22 00:00:00.0',  
  '2013-09-23 00:00:00.0',  
  '2013-09-24 00:00:00.0',  
  '2013-09-25 00:00:00.0',  
  '2013-09-26 00:00:00.0',  
  '2013-09-27 00:00:00.0',  
  '2013-09-28 00:00:00.0',  
  '2013-09-29 00:00:00.0',  
  '2013-09-30 00:00:00.0',  
  '2013-10-01 00:00:00.0',  
  '2013-10-02 00:00:00.0',  
  '2013-10-03 00:00:00.0',  
  '2013-10-04 00:00:00.0',  
  '2013-10-05 00:00:00.0',  
  '2013-10-06 00:00:00.0',  
  '2013-10-07 00:00:00.0',  
  '2013-10-08 00:00:00.0',  
  '2013-10-09 00:00:00.0',  
  '2013-10-10 00:00:00.0',  
  '2013-10-11 00:00:00.0',  
  '2013-10-12 00:00:00.0',  
  '2013-10-13 00:00:00.0',  
  '2013-10-14 00:00:00.0',  
  '2013-10-15 00:00:00.0',  
  '2013-10-16 00:00:00.0',  
  '2013-10-17 00:00:00.0',  
  '2013-10-18 00:00:00.0',  
  '2013-10-19 00:00:00.0',  
  '2013-10-20 00:00:00.0',
```

'2013-10-21 00:00:00.0',
'2013-10-22 00:00:00.0',
'2013-10-23 00:00:00.0',
'2013-10-24 00:00:00.0',
'2013-10-25 00:00:00.0',
'2013-10-26 00:00:00.0',
'2013-10-27 00:00:00.0',
'2013-10-28 00:00:00.0',
'2013-10-29 00:00:00.0',
'2013-10-30 00:00:00.0',
'2013-10-31 00:00:00.0',
'2013-11-01 00:00:00.0',
'2013-11-02 00:00:00.0',
'2013-11-03 00:00:00.0',
'2013-11-04 00:00:00.0',
'2013-11-05 00:00:00.0',
'2013-11-06 00:00:00.0',
'2013-11-07 00:00:00.0',
'2013-11-08 00:00:00.0',
'2013-11-09 00:00:00.0',
'2013-11-10 00:00:00.0',
'2013-11-11 00:00:00.0',
'2013-11-12 00:00:00.0',
'2013-11-13 00:00:00.0',
'2013-11-14 00:00:00.0',
'2013-11-15 00:00:00.0',
'2013-11-16 00:00:00.0',
'2013-11-17 00:00:00.0',
'2013-11-18 00:00:00.0',
'2013-11-19 00:00:00.0',
'2013-11-20 00:00:00.0',
'2013-11-21 00:00:00.0',
'2013-11-22 00:00:00.0',
'2013-11-23 00:00:00.0',
'2013-11-24 00:00:00.0',
'2013-11-25 00:00:00.0',
'2013-11-26 00:00:00.0',
'2013-11-27 00:00:00.0',
'2013-11-28 00:00:00.0',
'2013-11-29 00:00:00.0',
'2013-11-30 00:00:00.0',
'2013-12-01 00:00:00.0',
'2013-12-02 00:00:00.0',
'2013-12-03 00:00:00.0',
'2013-12-04 00:00:00.0',
'2013-12-05 00:00:00.0',
'2013-12-06 00:00:00.0',
'2013-12-07 00:00:00.0',
'2013-12-08 00:00:00.0',
'2013-12-09 00:00:00.0',
'2013-12-10 00:00:00.0',
'2013-12-11 00:00:00.0',
'2013-12-12 00:00:00.0',
'2013-12-13 00:00:00.0',
'2013-12-14 00:00:00.0',
'2013-12-15 00:00:00.0',
'2013-12-16 00:00:00.0',
'2013-12-17 00:00:00.0',
'2013-12-18 00:00:00.0',
'2013-12-19 00:00:00.0',
'2013-12-20 00:00:00.0',
'2013-12-21 00:00:00.0',
'2013-12-22 00:00:00.0',
'2013-12-23 00:00:00.0',
'2013-12-24 00:00:00.0',
'2013-12-25 00:00:00.0',
'2013-12-26 00:00:00.0',
'2013-12-27 00:00:00.0',
'2013-12-28 00:00:00.0',
'2013-12-29 00:00:00.0',
'2013-12-30 00:00:00.0',
'2013-12-31 00:00:00.0',
'2014-01-01 00:00:00.0',
'2014-01-02 00:00:00.0',
'2014-01-03 00:00:00.0',
'2014-01-04 00:00:00.0',
'2014-01-05 00:00:00.0',
'2014-01-06 00:00:00.0',
'2014-01-07 00:00:00.0',
'2014-01-08 00:00:00.0',
'2014-01-09 00:00:00.0',
'2014-01-10 00:00:00.0',
'2014-01-11 00:00:00.0',
'2014-01-12 00:00:00.0',
'2014-01-13 00:00:00.0',
'2014-01-14 00:00:00.0',
'2014-01-15 00:00:00.0',
'2014-01-16 00:00:00.0',
'2014-01-17 00:00:00.0',

'2014-01-18 00:00:00.0',
'2014-01-19 00:00:00.0',
'2014-01-20 00:00:00.0',
'2014-01-21 00:00:00.0',
'2014-01-22 00:00:00.0',
'2014-01-23 00:00:00.0',
'2014-01-24 00:00:00.0',
'2014-01-25 00:00:00.0',
'2014-01-26 00:00:00.0',
'2014-01-27 00:00:00.0',
'2014-01-28 00:00:00.0',
'2014-01-29 00:00:00.0',
'2014-01-30 00:00:00.0',
'2014-01-31 00:00:00.0',
'2014-02-01 00:00:00.0',
'2014-02-02 00:00:00.0',
'2014-02-03 00:00:00.0',
'2014-02-04 00:00:00.0',
'2014-02-05 00:00:00.0',
'2014-02-06 00:00:00.0',
'2014-02-07 00:00:00.0',
'2014-02-08 00:00:00.0',
'2014-02-09 00:00:00.0',
'2014-02-10 00:00:00.0',
'2014-02-11 00:00:00.0',
'2014-02-12 00:00:00.0',
'2014-02-13 00:00:00.0',
'2014-02-14 00:00:00.0',
'2014-02-15 00:00:00.0',
'2014-02-16 00:00:00.0',
'2014-02-17 00:00:00.0',
'2014-02-18 00:00:00.0',
'2014-02-19 00:00:00.0',
'2014-02-20 00:00:00.0',
'2014-02-21 00:00:00.0',
'2014-02-22 00:00:00.0',
'2014-02-23 00:00:00.0',
'2014-02-24 00:00:00.0',
'2014-02-25 00:00:00.0',
'2014-02-26 00:00:00.0',
'2014-02-27 00:00:00.0',
'2014-02-28 00:00:00.0',
'2014-03-01 00:00:00.0',
'2014-03-02 00:00:00.0',
'2014-03-03 00:00:00.0',
'2014-03-04 00:00:00.0',
'2014-03-05 00:00:00.0',
'2014-03-06 00:00:00.0',
'2014-03-07 00:00:00.0',
'2014-03-08 00:00:00.0',
'2014-03-10 00:00:00.0',
'2014-03-11 00:00:00.0',
'2014-03-12 00:00:00.0',
'2014-03-13 00:00:00.0',
'2014-03-14 00:00:00.0',
'2014-03-15 00:00:00.0',
'2014-03-16 00:00:00.0',
'2014-03-17 00:00:00.0',
'2014-03-18 00:00:00.0',
'2014-03-19 00:00:00.0',
'2014-03-20 00:00:00.0',
'2014-03-21 00:00:00.0',
'2014-03-22 00:00:00.0',
'2014-03-23 00:00:00.0',
'2014-03-24 00:00:00.0',
'2014-03-25 00:00:00.0',
'2014-03-26 00:00:00.0',
'2014-03-27 00:00:00.0',
'2014-03-28 00:00:00.0',
'2014-03-29 00:00:00.0',
'2014-03-30 00:00:00.0',
'2014-03-31 00:00:00.0',
'2014-04-01 00:00:00.0',
'2014-04-02 00:00:00.0',
'2014-04-03 00:00:00.0',
'2014-04-04 00:00:00.0',
'2014-04-05 00:00:00.0',
'2014-04-06 00:00:00.0',
'2014-04-07 00:00:00.0',
'2014-04-08 00:00:00.0',
'2014-04-09 00:00:00.0',
'2014-04-10 00:00:00.0',
'2014-04-11 00:00:00.0',
'2014-04-12 00:00:00.0',
'2014-04-13 00:00:00.0',
'2014-04-14 00:00:00.0',
'2014-04-15 00:00:00.0',
'2014-04-16 00:00:00.0',
'2014-04-17 00:00:00.0',

'2014-04-18 00:00:00.0',
'2014-04-19 00:00:00.0',
'2014-04-20 00:00:00.0',
'2014-04-21 00:00:00.0',
'2014-04-22 00:00:00.0',
'2014-04-23 00:00:00.0',
'2014-04-24 00:00:00.0',
'2014-04-25 00:00:00.0',
'2014-04-26 00:00:00.0',
'2014-04-27 00:00:00.0',
'2014-04-28 00:00:00.0',
'2014-04-29 00:00:00.0',
'2014-04-30 00:00:00.0',
'2014-05-01 00:00:00.0',
'2014-05-02 00:00:00.0',
'2014-05-03 00:00:00.0',
'2014-05-04 00:00:00.0',
'2014-05-05 00:00:00.0',
'2014-05-06 00:00:00.0',
'2014-05-07 00:00:00.0',
'2014-05-08 00:00:00.0',
'2014-05-09 00:00:00.0',
'2014-05-10 00:00:00.0',
'2014-05-11 00:00:00.0',
'2014-05-12 00:00:00.0',
'2014-05-13 00:00:00.0',
'2014-05-14 00:00:00.0',
'2014-05-15 00:00:00.0',
'2014-05-16 00:00:00.0',
'2014-05-17 00:00:00.0',
'2014-05-18 00:00:00.0',
'2014-05-19 00:00:00.0',
'2014-05-20 00:00:00.0',
'2014-05-21 00:00:00.0',
'2014-05-22 00:00:00.0',
'2014-05-23 00:00:00.0',
'2014-05-24 00:00:00.0',
'2014-05-25 00:00:00.0',
'2014-05-26 00:00:00.0',
'2014-05-27 00:00:00.0',
'2014-05-28 00:00:00.0',
'2014-05-29 00:00:00.0',
'2014-05-30 00:00:00.0',
'2014-05-31 00:00:00.0',
'2014-06-01 00:00:00.0',
'2014-06-02 00:00:00.0',
'2014-06-03 00:00:00.0',
'2014-06-04 00:00:00.0',
'2014-06-05 00:00:00.0',
'2014-06-06 00:00:00.0',
'2014-06-07 00:00:00.0',
'2014-06-08 00:00:00.0',
'2014-06-09 00:00:00.0',
'2014-06-10 00:00:00.0',
'2014-06-11 00:00:00.0',
'2014-06-12 00:00:00.0',
'2014-06-13 00:00:00.0',
'2014-06-14 00:00:00.0',
'2014-06-15 00:00:00.0',
'2014-06-16 00:00:00.0',
'2014-06-17 00:00:00.0',
'2014-06-18 00:00:00.0',
'2014-06-19 00:00:00.0',
'2014-06-20 00:00:00.0',
'2014-06-21 00:00:00.0',
'2014-06-22 00:00:00.0',
'2014-06-23 00:00:00.0',
'2014-06-24 00:00:00.0',
'2014-06-25 00:00:00.0',
'2014-06-26 00:00:00.0',
'2014-06-27 00:00:00.0',
'2014-06-28 00:00:00.0',
'2014-06-29 00:00:00.0',
'2014-06-30 00:00:00.0',
'2014-07-01 00:00:00.0',
'2014-07-02 00:00:00.0',
'2014-07-03 00:00:00.0',
'2014-07-04 00:00:00.0',
'2014-07-05 00:00:00.0',
'2014-07-06 00:00:00.0',
'2014-07-07 00:00:00.0',
'2014-07-08 00:00:00.0',
'2014-07-09 00:00:00.0',
'2014-07-10 00:00:00.0',
'2014-07-11 00:00:00.0',
'2014-07-12 00:00:00.0',
'2014-07-13 00:00:00.0',
'2014-07-14 00:00:00.0',
'2014-07-15 00:00:00.0',

```
'2014-07-16 00:00:00.0',
'2014-07-17 00:00:00.0',
'2014-07-18 00:00:00.0',
'2014-07-19 00:00:00.0',
'2014-07-20 00:00:00.0',
'2014-07-21 00:00:00.0',
'2014-07-22 00:00:00.0',
'2014-07-23 00:00:00.0',
'2014-07-24 00:00:00.0']
```

```
order_customer_ids = [order['order_customer_id'] for order in order_dicts]
```

```
order_customer_ids[:3]
```

```
[11599, 256, 12111]
```

```
type(order_customer_ids[0])
```

```
int
```

```
def get_order_dict(order):
    order_details = order.split(',')
    order_dict = {
        'order_id': int(order_details[0]),
        'order_date': order_details[1],
        'order_customer_id': int(order_details[2]),
        'order_status': order_details[3],
    }
    return order_dict
```

```
# Reading data from file into a list
path = '/data/retail_db/orders/part-00000'
# C:\users\itiversity\Research\data\retail_db\orders\part-00000
orders_file = open(path)
orders_raw = orders_file.read()
orders = orders_raw.splitlines()
order_dicts = [get_order_dict(order) for order in orders]
order_dates = [order['order_date'] for order in order_dicts]
```

```
order_dates[:3]
```

```
['2013-07-25 00:00:00.0', '2013-07-25 00:00:00.0', '2013-07-25 00:00:00.0']
```

```
len(order_dates)
```

```
68883
```