Data Processing

Many data sets can be processed sequentially:
- The set of all Twitter posts
- Votes cast in an election
- Sensor readings of an airplane
- The positive integers: 1, 2, 3, ...

However, the sequence interface we used before does not always apply
- A sequence has a finite, known length
- A sequence allows element selection for any element

Some important ideas in big data processing:
- Implicit representations of streams of sequential data
- Declarative programming languages to manipulate and transform data
- Distributed computing

Iterators

A container can provide an iterator that provides access to its elements in some order

```python
iter(iterable): Return an iterator over the elements of an iterable value
next(iterator): Return the next element in an iterator
```

Iterators are always ordered, even if the container that produced them is not

```python
# Python 3.7
>>> k = {'one': 1, 'two': 2, 'three': 3}
>>> v = k.values()
>>> next(k) 'one'
>>> next(k) 'two'
>>> next(k) 'three'
```

Keys and values are iterated over in an arbitrary order which is non-random, varies across Python implementations, and depends on the dictionary's history of insertions and deletions. If keys, values and item views are iterated over with no intervening modifications to the dictionary, the order of items will directly correspond.

For Statements

```python
for <name> in <expression>
  <suite>
```

1. Evaluate the header `<expression>`, which must evaluate to an iterable object
2. For each element in that sequence, in order:
   A. `Binds <name>` to that element in the first frame of the current environment
   B. Execute the `<suite>`

When executing a `for` statement, `iter` returns an iterator and `next` provides each item:

```python
>>> counts = [1, 2, 3]
>>> for item in counts:
...    print(item)
1
2
3
```
Processing Iterators

A `StopIteration` exception is raised whenever `next` is called on an empty iterator.

```python
def contains(a, b):
    ai = iter(a)
    for x in b:
        while next(ai) != x:
            pass  # do nothing
        except StopIteration:
            return False
    return True
```

Built-In Iterator Functions

Built-in Functions for Iteration

Many built-in Python sequence operations return iterators that compute results lazily:

- `map(func, iterable)`: Iterate over `func(x)` for `x` in `iterable`
- `filter(func, iterable)`: Iterate over `x` in `iterable` if `func(x)`
- `zip(first_iter, second_iter)`: Iterate over co-indexed `(x, y)` pairs
- `reversed(sequence)`: Iterate over `x` in a sequence in reverse order

To view the contents of an iterator, place the resulting elements into a container:

- `list(iterable)`: Create a list containing all `x` in `iterable`
- `tuple(iterable)`: Create a tuple containing all `x` in `iterable`
- `sorted(iterable)`: Create a sorted list containing `x` in `iterable`

(Demo)

Generators

**Generators and Generator Functions**

```python
>>> def plus_minus(x):
...     yield x
...     yield -x

>>> t = plus_minus(3)
```

A generator is a function that yields values instead of returning them. A normal function returns once; a generator function can yield multiple times.

When a generator function is called, it returns a generator that iterates over its yields.

(Demo)

Iterables and User-Defined Classes

The special method `__iter__` is called by the built-in `iter()` & should return an iterator:

```python
class Countdown:
    def __init__(self, start):
        self.start = start
    def __iter__(self):
        v = self.start
        while v > 0:
            yield v
            v -= 1
```

Generators can Yield from Iterators

A `yield` statement yields all values from an iterator or iterable (Python 3.3):

```python
def a_then_b(a, b):
    yield from a
    yield from b
```

```python
def countdown(k):
    if k > 0:
        yield k
    yield from countdown(k - 1)
```

(Demo)