

11/12/18 Notes

Transforms

- **ParDo** - low-level Map function
 - Applies a function to one element, and returns any number of elements
 - User code inherits from DoFn
 - Returns a PCollection
 - Can take optional side arguments (additional PCollections or values, depending on use case)

- **GroupByKey** - groups PCollections of (key, value) pairs to a (key, [value1, value2, ... , valuen])

○

```
Input:
people = [
    ("jason", 1),
    ("robby", 6),
    ("jason", 7)
]
```

○

```
Output:
("jason", [1, 7])
("robby", [6])
```

- **CoGroupByKey** - groups two or more PCollections by key, with dictionary values

○

```
Input one:
number = [
    ("jason", 1),
    ("robby", 6),
    ("frank", 7)
]
Input two:
email = [
    ("jason", "jason@jason.jason"),
    ("robby", "robby@robby.robby"),
    ("frank", "frank@gmail.com")
]
```

○

```
Output:
("jason", {
    "number": 1,
    "email": "jason@jason.jason"
})
("robby", {
    "number": 6,
    "email": "robby@robby.robby"
```

```
})  
("frank", {  
  "number": 7,  
  "email": "frank@gmail.com"  
})
```