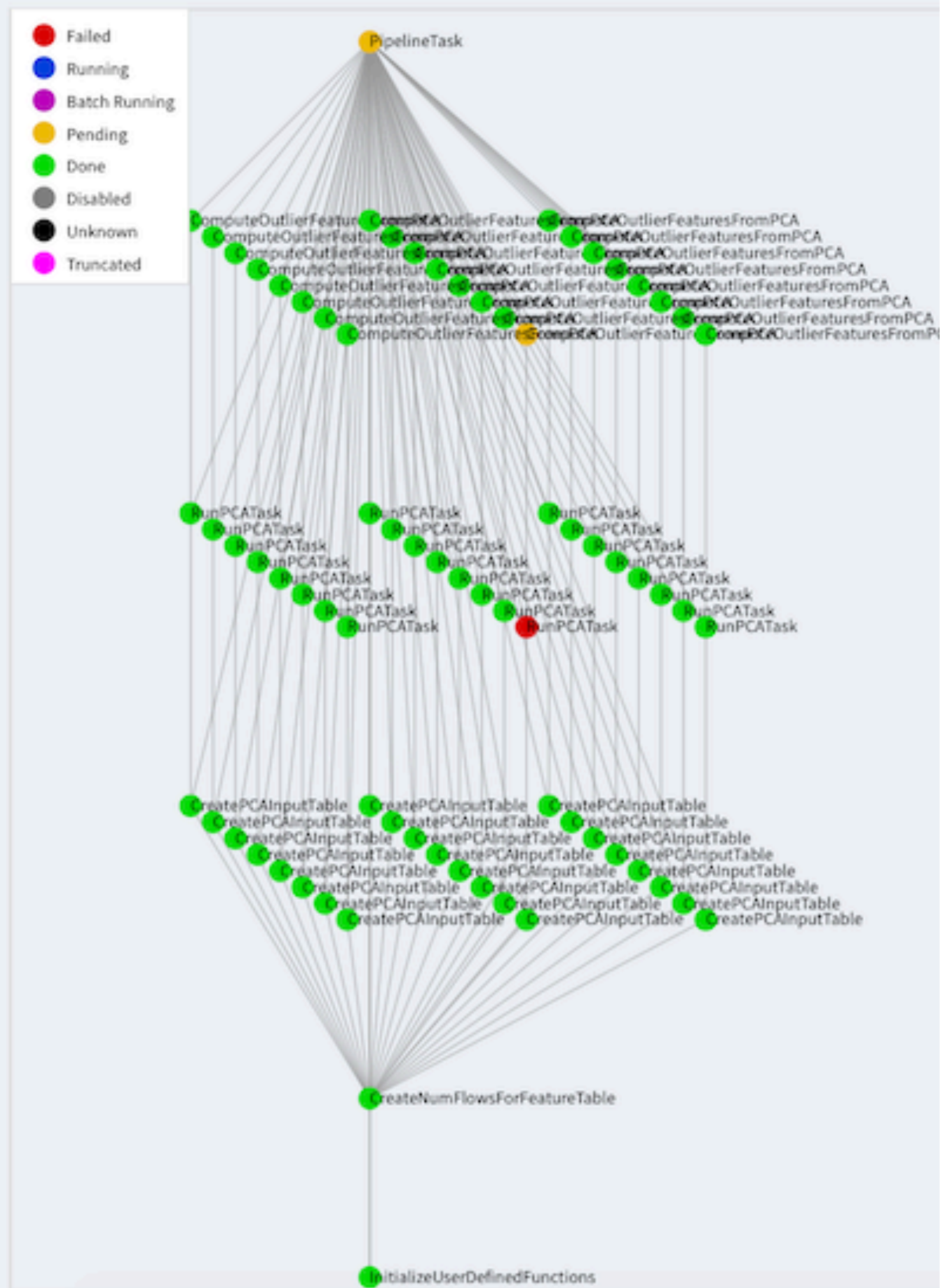


PipelineTask(date=2017-02-07)

Dependency Graph



WenereadarenewETL

Existing solutions were too much:

- Complex set-ups
- Too flexible
 - code based
 - unstructured - relies on conventions
 - no resource guarantees
- Too powerful
 - Too many features lead to steep learning curve
- Optimized for a different working point.

'few sources - deep processing'

VS.

'many sources - shallow processing'

Enter Data Packages

“Data Package is a simple container format used to describe and package a collection of data.

The format provides a simple contract for data interoperability that supports frictionless delivery, installation and management of data.”

frictionlessdata.io

```
{
  "name": "israel-budget-data",
  "resources": [
    {
      "name": "budget-2018",
      "path": "data/budget-2018.csv",
      "encoding": "windows-1255",
      "schema": {
        "fields": [
          {
            "name": "year",
            "type": "integer"
          },
          # ...
        ]
      }
    }
  ]
}
```

We need a new ETL

Existing solutions were too much:

- Complex set-ups
- Too flexible
 - code based
 - unstructured - relies on conventions
 - no resource guarantees
- Too powerful
 - Too many features lead to steep learning curve
- Optimized for a different working point.

'few sources - deep processing'

VS.

'many sources - shallow processing'

PipelineTask(date=2017-02-07)

Dependency Graph

