Schema handling

Row-based

```
yield Row(id=..., val=...)
```

- Column names inferred from field names
- Schema inference may trigger a small job

Tuple-based (recommended)

Define schema

```
schema = StructType([
    StructField("id",   StringType(),   False),
    StructField("val",  IntegerType(),  True),
    ...
])
```

Yield tuples (order matters)

```
def f(it):
    for r in it:
        yield (_id, _val, ...)

df = spark.createDataFrame(rdd, schema=schema)
```

Or yield from dictionary

```
rows["id"] = ...
rows["val"] = ...
yield row
```

- No schema inference
- Faster
- Predictable types