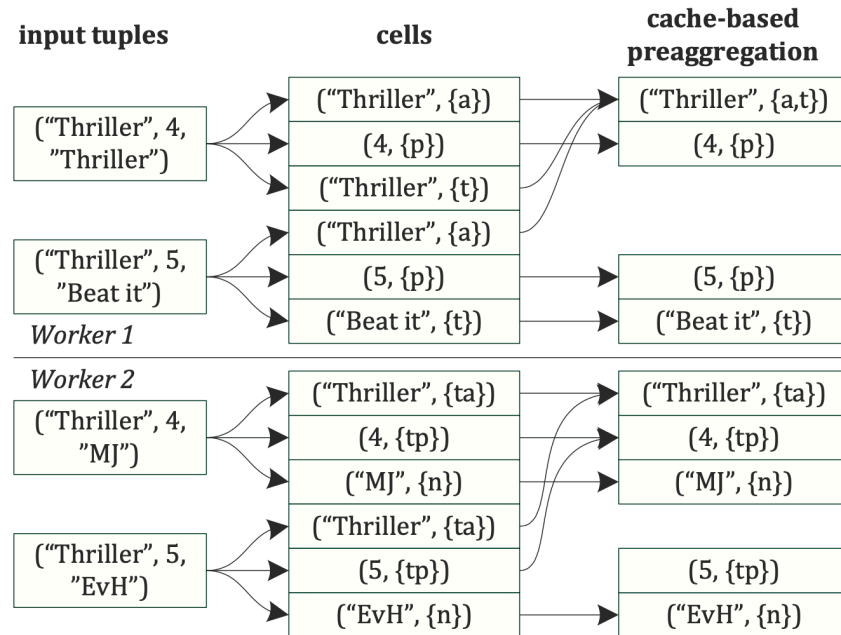


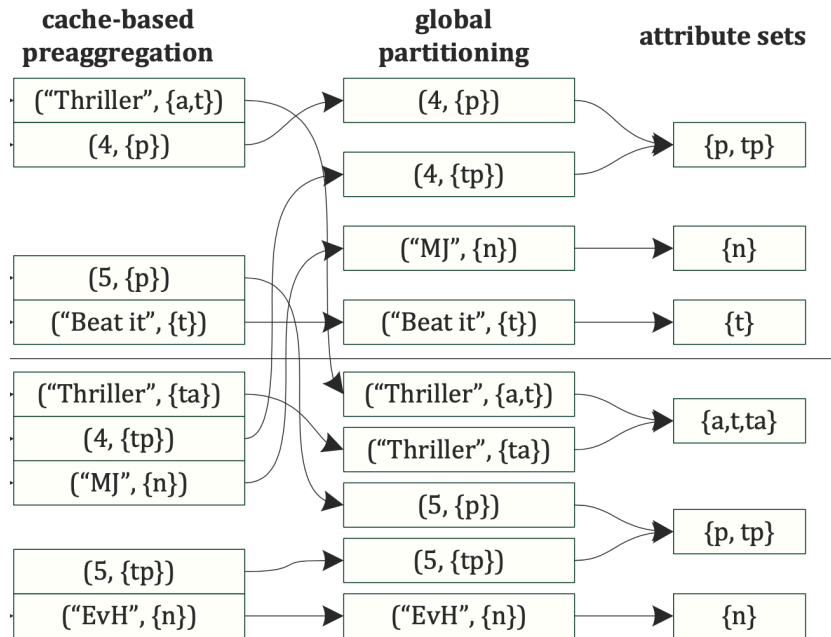
Group “ElixirEvangelists”

Assignment 4: ddm-idd – Full Outer Join in Spark



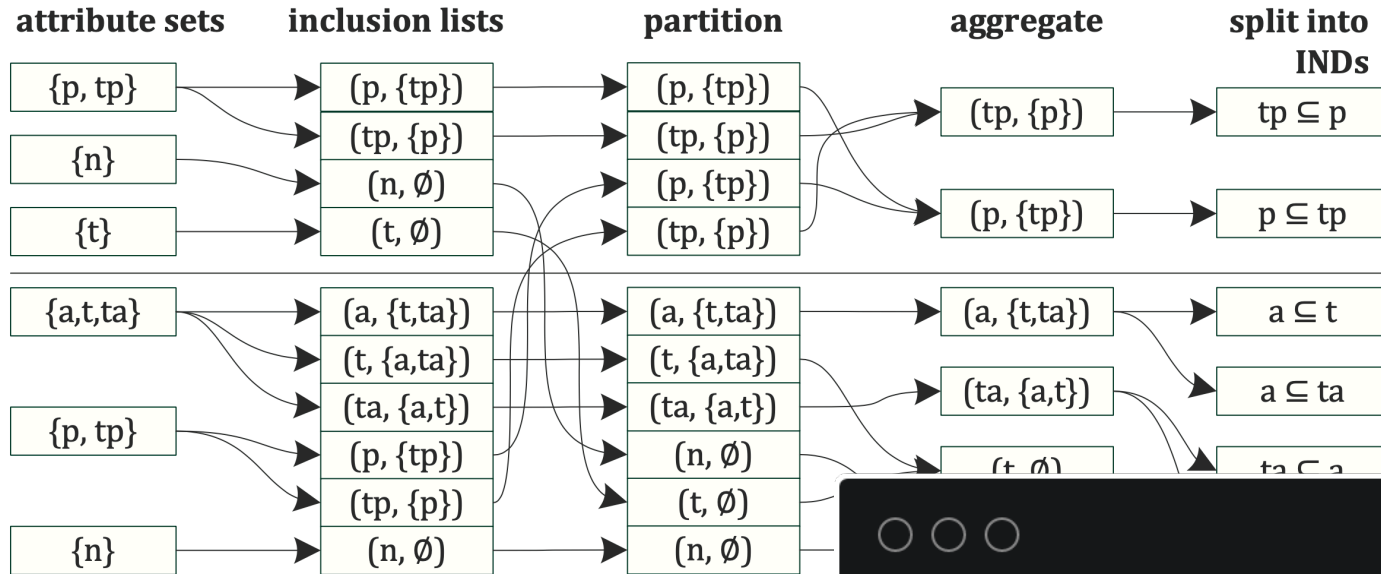
```
1 frames.map(frame => {
2   frame
3   .flatMap(row => row.toSeq.map(p => String.valueOf(p)).zip(frame.columns))
4   .groupByKey(p => p._1)
5   .mapValues { case (_key, values) => Set(values) }
6   .reduceGroups((storage, value) => storage ++ value)
7   .map { case (key, value) => (key, value.seq.toList.distinct) }
8 })
9 .reduce((acc, frame) => acc.union(frame))
10 .repartition($"_1")
```

Assignment 4: ddm-idd – Full Outer Join in Spark



```
1 global_repartitioned
2 .groupByKey(_. _1)
3 .mapValues({ case (key, value) => value })
4 .reduceGroups((acc, value) => acc ++ value)
5 .map((value) => value._2)
```

Assignment 4: ddm-idd – Derive ID from Outer Join



```

1 attribute_sets.flatMap((set) => {
2   set.map(e => (e, set.filter(el => !el.equals(e))))
3 })
4 .repartition($"_1")
5 .groupBy(_._1)
6 .reduceGroups((acc, value) => (acc._1, acc._2.intersect(value._2)))
7 .map(p => p._2)
8 .filter(data => data._2.length > 0)
9 .sort("_1")

```

Assignment 4: ddm-idd – DAG created with the Spark History Server

