

# HyPerInsight: Data Exploration Deep Inside HyPer

Nina Hubig, Linnea Passing, Maximilian E. Schüle, Dimitri Vorona, Alfons Kemper, Thomas Neumann  
 {hubig, passing, schuele, vorona, kemper, neumann}@in.tum.de

**data mining operators**

- Apriori
- DBscan
- k-Means
- k-Modes
- NaiveBayes
- PageRank

**OpenFlights dataset**

- Airlines
- Airports
- Routes

**query result**

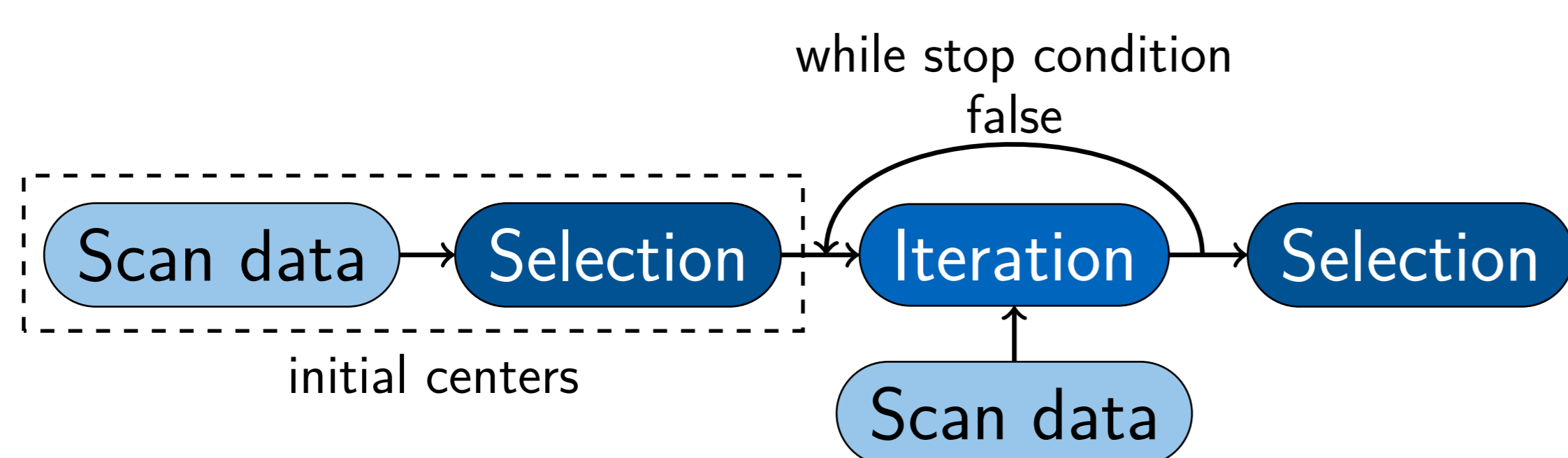
- Airline Prediction
- Airplane Prediction
- Airport Clustering/Rank

```

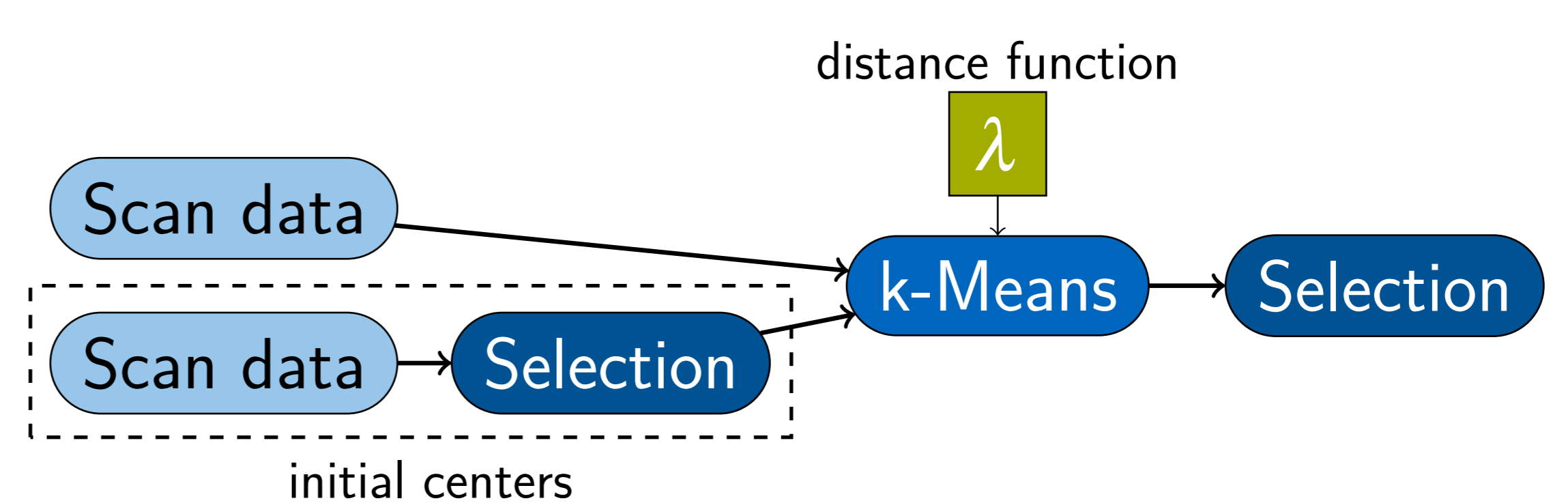
1 select lng, lat, cluster, airport_name
2 from kmeans(
3   (select (longitude / 180 * pi()) as lng, (latitude / 180 * pi()) as lat,
4     airport_name from airports),
5   lambda(a,b) (2 * atan2(sqrt(sin((b.lat-a.lat)/2) ^ 2 + cos(a.lat) * cos(b.lat) *
6     (sin((b.lng - a.lng) / 2) ^ 2)), sqrt(1-sin((b.lat-a.lat)/2) ^ 2 + cos(a.lat) *
7     cos(b.lat) * (sin((b.lng - a.lng) / 2) ^ 2))),
8   10
9 )
10 order by lng, lat
        
```

LNG	LAT	CLUSTER	AIRPORT_NAME
-3.13945	-0.29131	9	Matei Airport
-3.13065	1.20198	1	Mys Shmidta Airport
-3.13011	-0.30968	9	Cicia Airport
-3.12372	-0.30140	9	Vanua Balavu Airport
-3.12095	-0.31764	9	Lakeba Island Airport

## SQL



## Operators + Lambdas



- + Leverages sophisticated SQL optimizer
- + Reduces memory footprint compared to recursive CTEs
- User has to implement complex algorithms
- Invariants and assumptions are opaque to the optimizer
- Non-standard SQL extension

- + Seamless integration with SQL
  - + Fully optimizable
  - + Allows re-using of large parts of the query execution
  - Requires deep knowledge of database implementation
  - Every variant and parametrization has to be implemented
- This is solved by introducing lambda functions

