

Ge Song, Frédéric Magoulès (Supervisor), Fabrice Huet
(Co-Supervisor)
Lab MICS
CentraleSupélec
Université Paris-Saclay

Parallel and Continuous Join Processing for Data Stream

Thèse pour l'obtention du grade de Docteur

Table of Contents

Introduction

Part I: Data Driven Stream Join (kNN)

Part II: Query Driven Stream Join (RDF)

Conclusion and Future Work

Background



Part I: Data Driven Stream Join (kNN)



Outline

- Related Work
- Parallel Workflow
- Theoretical Analysis
- Continuous kNN
- Experiment Result
- Conclusion

Outline

- Related Work
- Parallel Workflow
- Theoretical Analysis
- Continuous kNN
- Experiment Result
- Conclusion

Definition: kNN

Given a set of query points R and a set of reference points S , a k **nearest neighbor join** is an operation which, for each point in R , discovers the k nearest neighbors in S .

-

Outline

- Related Work
- Query Decomposition and Distribution
- Data Partition and Assignment
- Parallel and Distributed Query Plan
- Continuous Join
- Analysis
- Implementation
- Experiment Result
- Conclusion

Outline

- Related Work
- Query Decomposition and Distribution
- Data Partition and Assignment
- Parallel and Distributed Query Planner
- Continuous Join
- Analysis
- Implementation
- Experiment Result
- Conclusion

Thank You!