

# Parallel DBSCAN Clustering Algorithm using Apache Spark

Anousheh Shahmirza  
School of Computer Science  
Carleton University  
Ottawa, Canada K1S 5B6  
*Anoushehshahmirza@cmail.carleton.ca*

November 13, 2019

## 1 DBSCAN Algorithm

- Needs
- Explanation
- How the algorithm works
- Problem definition

## 2 Mapreduce algorithms for big data analysis

- How it works
- Hadoop
- Apache Spark

## 3 A novel scalable DBSCAN algorithm with Spark

- Algorithm
- Advantages

## 4 Implementation

- Explanation
- Dataset description
- Results

## 5 Conclusion

- Compare the result with the original DBSCAN
- Future work

## References