SPARK OVER HADOOP(The Distributed programming framework)

> Ankita K Wani

ntroduction

Features

History

Man Reduc

_ .

Working

Applications

Disadvantage

Conclusion

Thanks



SPARK OVER HADOOP(The Distributed programming framework)

Ankita K. Wani

SSBT'S COET, Bambhori, Jalgaon

October 8, 2015



outline

SPARK OVER HADOOP(The Distributed programming framework)

> Ankita K Wani

Introduction

C--4....

. . .

map reduce

Spa

VVOIKIIIG

Applications

Disadvantage

_. .

1 Introduction

2 Features

3 History

4 Map-Reduce

5 Spark

6 Working

7 Applications

8 Disadvantages

9 Conclusion

Introduction

SPARK OVER HADOOP(The Distributed programming framework)

Wani

Introduction

Feature:

Map-Reduc

Spark

Applications

Disadvantage

Conclusion

- Spark is an open-source distributed programming framework .
- Spark is a cluster framework that performs in-memory computing, with the goal of outperforming disk-based engines like Hadoop.
- Supports advanced Python, Scala ,SparkR ,Java8, SQL.
- Spark is a data parallel open source processing framework.
- An alternative to the traditional batch map/reduce model.
- It is used for real-time stream data processing and fast interactive queries .

Features

SPARK OVER HADOOP(The Distributed programming framework)

vvaiii

Introduction

Features

.....

lap-Reduc

c .

A . . . I' . . . I'

Disadvantage

D ISGUTUTTUGE.

- In-memory (RAM) based
- Real-time stream data processing
- Makes big data interactive.
- Parallel open source processing.
- 2-5 times less code.
- Scalability to over 8000 nodes in production.
- Support for structured and relational query processing (SQL).
- Easy Management.

History

SPARK OVER HADOOP(The Distributed programming framework)

VVanı

Introduction

Features

History

Map-Reduce

Sparl

A 11 ...

Аррисации

. . .

- Originally developed by Matei Zaharia in 2009 at UC Berkeley's AMP Lab written in Scala
- Released as open source in 2010
- In 2013, It was donated to Apache Software Foundation.
- In February 2014, Spark became a Top-Level Apache Project.

Data sharing

SPARK OVER HADOOP(The Distributed programming framework)

> Ankita K Wani

Introduction

Features

Histor

Map-Reduce

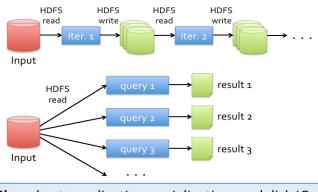
VVOIRING

Application

Conclusion

Thank

Data Sharing in MapReduce



Slow due to replication, serialization, and disk IO

Data sharing

SPARK OVER HADOOP(The Distributed programming framework)

> Ankita K Wani

ntroduction

Features

Histor

Map-Redu

Spark

Workin

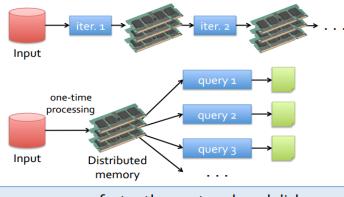
Application

_ .___

Conclusion

Thank

Data Sharing in Spark



10-100× faster than network and disk

Execution

SPARK OVER HADOOP(The Distributed programming framework)

> Ankita K Wani

ntroduction

Fasturas

Catare

Working

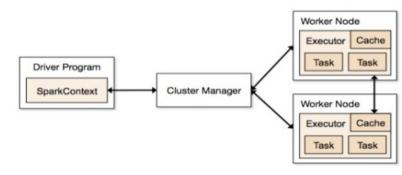
Applications

Disadvantag

Conclusion

Thank

Execution Flow



Related Terms

SPARK OVER HADOOP(The Distributed programming framework)

Wani

......

Feature

.. .

Map-Reduc

Spar

Working

Application

Disadvantag

Conclusion

Chaple

- Cluster Manager
- Executer
- Driver
- Stage

Applications

SPARK OVER HADOOP(The Distributed programming framework)

Wani Wani

Introductio

Feature

History

Man Dadiia

Map-Reduc

Spar

Applications

Disadvantag

Conclusion

- In-memory Analytics and anamoly detection.
- Traffic Estimation.
- Twitter scam classification.

Disadvantages

SPARK OVER HADOOP(The Distributed programming framework)

VVanı

Introductio

Features

Histor

Man-Reduc

Working

Application

 ${\sf Disadvantages}$

Conclusion

- Little bit less secure than Hadoop. -RDD creates some problem in security.
- In case fault tolerance it will have to start processing from begining.

Conclusion

SPARK OVER HADOOP(The Distributed programming framework)

Wanı

Introduction

Features

. . . .

Man Dadiia

map reduce

Spark

vvoi king

Application

Disadvantage

Conclusion

Thank

By making duistributed datasets a first-class primitives, spark provides a simple, efficient programming model for stateful data analytics.

RDD supports in-memory catching, fault recovery and debugging.

SPARK OVER HADOOP(The Distributed programming framework)

Ankita K

ntroduction

Feature

· catare

M. D. L.

Spair

Applications

Disadvantages

Conclusion

