PULSUS
WWW.PULSUS.COM

7th Global Conference on
DATA SCIENCE AND MACHINE LEARNING

December 13, 2022 | Webinar

Storage requirements and options for running spark on Kubernetes

Rachit Arora
Oracle Corp, India

In a world of serverless computing users tend to be frugal when it comes to expenditure on compute, storage and other resources. Paying for the same when they aren't in use becomes a significant factor. Offering Spark as service on cloud presents very unique challenges. Running Spark on Kubernetes presents a lot of challenges especially around storage and persistence. Spark workloads have very unique requirements of Storage for intermediate data, long time persistence, Share file system and requirements become very tight when it same need to be offered as a service for enterprise to mange GDPR and other compliance like ISO 27001 and HIPAA certifications. This talk covers challenges involved in providing Serverless Spark Clusters share the specific issues one can encounter when running large Kubernetes clusters in production especially covering the scenarios related to persistence. This talk will help people using Kubernetes or docker runtime in production and help them understand various storage options available and which is more suitable for running Spark workloads on Kubernetes and what more can be done

Reference

1. Yanfeng Zhang, Qinxin Gao, Lixin Gao and Cuirong Wang, "iMapReduce: A Distributed Computing Framework for Iterative Computation" (2019), Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW) 2011 IEEE International Symposium, pp. 1112-1121.

2. Yanfeng Zhang, Qinxin Gao, Lixin Gao and Cuirong Wang, "PrIter: A Distributed Framework for Prioritizing Iterative Computations Parallel and Distributed Systems" (2020), IEEE Transactions onTransactions on Prallel and Distributed Systems, vol. 24, no. 9, pp. 1884-1893.

Biography

Rachit Arora is a Consulting Member of Technical Staff, Oracle Cloud Infrastructure, IDC. He is key designer of the Oracle's offerings on Cloud for Hadoop ecosystem. He has extensive experience in architecture, design and agile development. Rachit is an expert in application development in Cloud architecture and development using hadoop and it's ecosystem. Rachit has been active speaker for BigData technologies in various conference like Information Management Technical Conference-2015, ContainerCon NA-2016, Container Camp Sydeny 2017, Microxchg Berlin 2018, DataworksSumit 2018.

e: rachit.a.arora@oracle.com