



Hybrid cloud solution

NetApp Solutions

NetApp
October 20, 2023

This PDF was generated from <https://docs.netapp.com/us-en/netapp-solutions/data-analytics/apache-spark-hybrid-cloud-solution.html> on October 20, 2023. Always check docs.netapp.com for the latest.

Table of Contents

Hybrid cloud solution 1

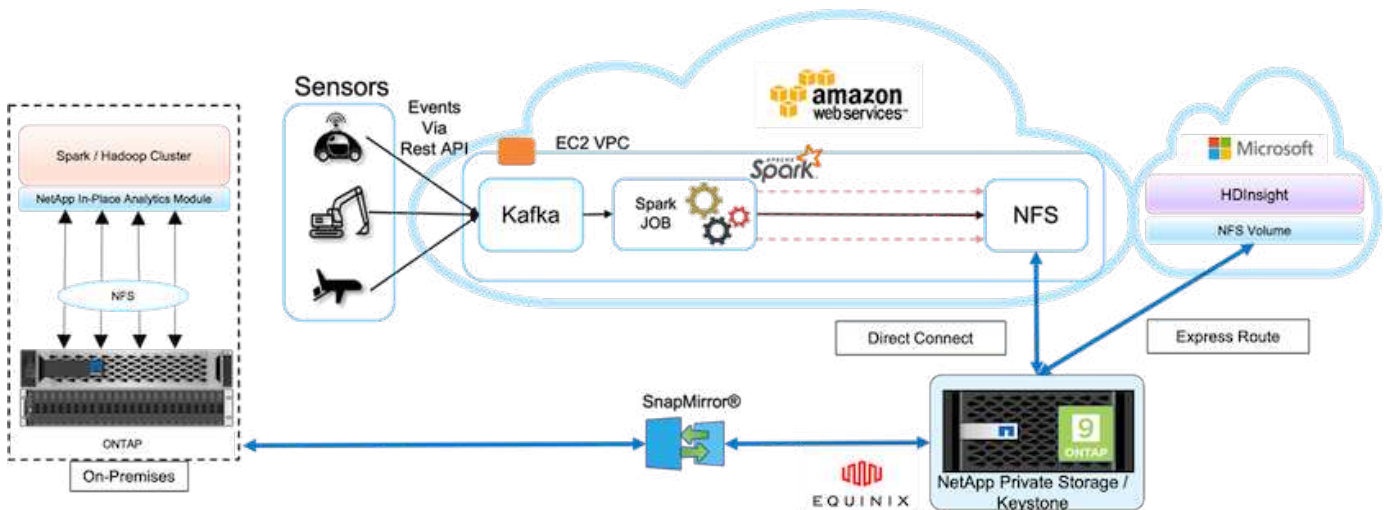
Hybrid cloud solution

[Previous: Testing results.](#)

A modern enterprise data center is a hybrid cloud that connects multiple distributed infrastructure environments through a continuous data management plane with a consistent operating model, on premises and/or in multiple public clouds. To get the most out of a hybrid cloud, you must be able to seamlessly move data between your on-premises and multi-cloud environments without the need for any data conversions or application refactoring.

Customers have indicated that they start their hybrid cloud journey either by moving secondary storage to the cloud for use cases such as data protection or by moving less business-critical workloads such as application development and DevOps to the cloud. They then move on to more critical workloads. Web and content hosting, DevOps and application development, databases, analytics, and containerized apps are among the most popular hybrid-cloud workloads. The complexity, cost, and risks of enterprise AI projects have historically hindered AI adoption from experimental stage to production.

With a NetApp hybrid-cloud solution, customers benefit from integrated security, data governance, and compliance tools with a single control panel for data and workflow management across distributed environments, while optimizing the total cost of ownership based on their consumption. The following figure is an example solution of a cloud service partner tasked with providing multi-cloud connectivity for customers' big-data-analytics data.



In this scenario, IoT data received in AWS from different sources is stored in a central location in NetApp Private Storage (NPS). The NPS storage is connected to Spark or Hadoop clusters located in AWS and Azure enabling big-data-analytics applications running in multiple clouds accessing the same data. The main requirements and challenges for this use case include the following:

- Customers want to run analytics jobs on the same data using multiple clouds.
- Data must be received from different sources such as on-premises and cloud environments through different sensors and hubs.
- The solution must be efficient and cost effective.
- The main challenge is to build a cost-effective and efficient solution that delivers hybrid analytics services between different on-premises and cloud environments.

Our data protection and multicloud connectivity solution resolves the pain point of having cloud analytics applications across multiple hyperscalers. As shown in the figure above, data from sensors is streamed and ingested into the AWS Spark cluster through Kafka. The data is stored in an NFS share residing in NPS, which is located outside of the cloud provider within an Equinix data center.

Because NetApp NPS is connected to Amazon AWS and Microsoft Azure through Direct Connect and Express Route connections respectively, customers can leverage the In-Place Analytics Module to access the data from both Amazon and AWS analytics clusters. Consequently, because both on-premises and NPS storage runs ONTAP software, [SnapMirror](#) can mirror the NPS data into the on-premises cluster, providing hybrid cloud analytics across on-premises and multiple clouds.

For the best performance, NetApp typically recommends using multiple network interfaces and direct connection or express routes to access the data from cloud instances. We have other data mover solutions including [XCP](#) and [BlueXP Copy and Sync](#) to help customers build application-aware, secure, and cost-effective hybrid-cloud Spark clusters.

[Next: Python scripts for each major use case.](#)

Copyright information

Copyright © 2023 NetApp, Inc. All Rights Reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

LIMITED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (b)(3) of the Rights in Technical Data -Noncommercial Items at DFARS 252.227-7013 (FEB 2014) and FAR 52.227-19 (DEC 2007).

Data contained herein pertains to a commercial product and/or commercial service (as defined in FAR 2.101) and is proprietary to NetApp, Inc. All NetApp technical data and computer software provided under this Agreement is commercial in nature and developed solely at private expense. The U.S. Government has a non-exclusive, non-transferrable, nonsublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b) (FEB 2014).

Trademark information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.