



# **Click-through rate prediction use case summary**

NetApp Solutions

NetApp  
August 18, 2021

This PDF was generated from [https://docs.netapp.com/us-en/netapp-solutions/ai/aks-anf\\_click-through\\_rate\\_prediction\\_use\\_case\\_summary.html](https://docs.netapp.com/us-en/netapp-solutions/ai/aks-anf_click-through_rate_prediction_use_case_summary.html) on August 18, 2021. Always check docs.netapp.com for the latest.

# Table of Contents

Click-through rate prediction use case summary ..... 1

# Click-through rate prediction use case summary

[Previous: Cloud resource requirements.](#)

This use case is based on the publicly available [Terabyte Click Logs](#) dataset from [Criteo AI Lab](#). With the recent advances in ML platforms, a lot of attention is now on learning at scale. In this context, Terabyte Click Logs is now the dataset of reference in assessing the scalability of ML platforms and algorithms. By predicting the click-through rate, an advertiser can select the most plausible visitors who are most likely to respond to the ads, analyzing their browsing history, and showing the most relevant ads based on the interests of the user.

This solution provided in this technical report highlights the following benefits:

- Azure NetApp Files advantages in distributed or large-scale training
- RAPIDS CUDA-enabled data processing (cuDF, cuPy, and so on) and ML algorithms (cuML)
- Dask parallel computing framework for distributed training

The collaborative relationship between Dask and Azure NetApp Files demonstrates the drastic improvement in random forest model training time by two orders of magnitude. This improvement is comparable to the conventional Pandas approach when dealing with real-world click logs data with 45GB of text (on average) each day.

[Next: Install and set up the aks cluster.](#)

## Copyright Information

Copyright © 2021 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.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

## 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.