ARC·TS ADVANCED RESEARCH COMPUTING TECHNOLOGY SERVICES UNIVERSITY OF MICHIGAN

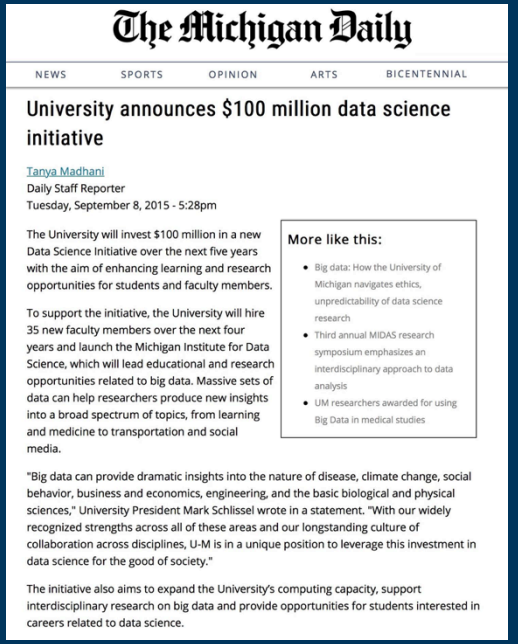
Advanced Research Computing — Technology Services provides access to and support for advanced computing resources. ARC-TS facilitates new and more powerful approaches to research challenges in fields ranging from physics to linguistics, and from engineering to medicine.

- Big Data Systems Administrator @ ARC-TS
- HPC admin
- UNIX admin
- Fan!

Matt McLean
mattmc@umich.edu

The DSI

In 2015, U-M announced the Data Science Initiative, investing $100 million over five years to enhance opportunities for student and faculty researchers across the university to tap into the enormous potential of big data. The initiative included the goal of expanding U-M's research computing capacity.



The Michigan Daily

NEWS SPORTS OPINION ARTS BICENTENNIAL

University announces $100 million data science initiative

Tanya Madhani
Daily Staff Reporter
Tuesday, September 8, 2015 - 5:28pm

The University will invest $100 million in a new Data Science Initiative over the next five years with the aim of enhancing learning and research opportunities for students and faculty members.

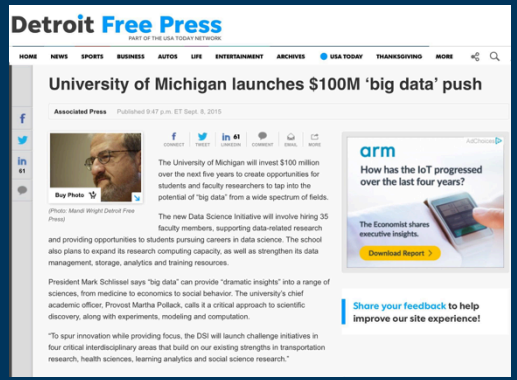
To support the initiative, the University will hire 35 new faculty members over the next four years and launch the Michigan Institute for Data Science, which will lead educational and research opportunities related to big data. Massive sets of data can help researchers produce new insights into a broad spectrum of topics, from learning and medicine to transportation and social media.

More like this:

- Big data: How the University of Michigan navigates ethics, unpredictability of data science research
- Third annual MIDAS research symposium emphasizes an interdisciplinary approach to data analysis
- UM researchers awarded for using Big Data in medical studies

"Big data can provide dramatic insights into the nature of disease, climate change, social behavior, business and economics, engineering, and the basic biological and physical sciences," University President Mark Schlissel wrote in a statement. "With our widely recognized strengths across all of these areas and our longstanding culture of collaboration across disciplines, U-M is in a unique position to leverage this investment in data science for the good of society."

The initiative also aims to expand the University's computing capacity, support interdisciplinary research on big data and provide opportunities for students interested in careers related to data science.



Detroit Free Press
PART OF THE USA TODAY NETWORK

HOME NEWS SPORTS BUSINESS AUTOS LIFE ENTERTAINMENT ARCHIVES USA TODAY THANKSGIVING MORE

University of Michigan launches $100M 'big data' push

Associated Press Published 9:47 p.m. ET Sept. 8, 2015

Buy Photo

(Photo: Mandi Wright Detroit Free Press)

The University of Michigan will invest $100 million over the next five years to create opportunities for students and faculty researchers to tap into the potential of "big data" from a wide spectrum of fields.

The new Data Science Initiative will involve hiring 35 faculty members, supporting data-related research and providing opportunities to students pursuing careers in data science. The school also plans to expand its research computing capacity, as well as strengthen its data management, storage, analytics and training resources.

President Mark Schlissel says "big data" can provide "dramatic insights" into a range of sciences, from medicine to economics to social behavior. The university's chief academic officer, Provost Martha Pollack, calls it a critical approach to scientific discovery, along with experiments, modeling and computation.

"To spur innovation while providing focus, the DSI will launch challenge initiatives in four critical interdisciplinary areas that build on our existing strengths in transportation research, health sciences, learning analytics and social science research."

Share your feedback to help improve our site experience!

ARC-TS, in furtherance of that goal, set out to design and implement a big data Hadoop cluster for use by researchers across campus.

UNIVERSITY OF MICHIGAN

Cavium Partnership

In late 2016, Cavium, Inc. (NASDAQ: CAVM) agreed to gift U-M 100+ ThunderX servers valued at $3.5 million to build a large scale Hadoop cluster.

The gift is structured in three phases, and by year three the cluster will have over 10,000 cores, 55TB of memory, and 9PB of storage.

Because of this gift, ARC-TS will be able to provide a large scale Hadoop ecosystem **for free** to any researcher at U-M.

MARKETSINSIDER PORTFOLIO

MARKETS STOCKS INDICES COMMODITIES CURRENCIES MUTUAL FUNDS ETFS BONDS NEWS

University of Michigan Partners with Cavium on Big Data Computing Platform for U-M Researchers

PRESS RELEASE PR Newswire
Nov. 7, 2017, 12:13 PM

SHARE

SAN JOSE, Calif., Nov. 7, 2017 /PRNewswire/ -- Cavium™, Inc. (NASDAQ: CAVM), a leading provider of semiconductor products that enable secure and intelligent processing for enterprise, data center, wired and wireless networking and the University of Michigan today announced a new partnership that will position U-M as a leader in data-intensive scientific research by creating a powerful Big Data computing cluster using dual socket servers powered by Cavium's ThunderX® ARMv8-A workload optimized processors.

The University RECORD
News for Faculty and Staff

November 7, 2017

U-M, Cavium partner on big data research computing platform

By Dan Meisler
Advanced Research Computing

Topic: Information Technology

Like 64 Tweet Share print

A new partnership between the University of Michigan and Cavium Inc., a San Jose-based provider of semiconductor products, will create a powerful new big data computing cluster available to all U-M researchers.

The $3.5 million ThunderX computing cluster will enable U-M researchers to, for example, process massive amounts of data generated by remote sensors in distributed manufacturing environments, or by test fleets of automated and connected vehicles.

The cluster will run the Hortonworks Data Platform providing Spark, Hadoop MapReduce, and other tools for large-scale data processing.

+ MORE INFORMATION
- Cavium ThunderX Cluster at

UNIVERSITY OF MICHIGAN

CAVIUM

Hortonworks HDP

Java OpenJDK on ARM is ready today.

While the Hadoop ecosystem is written in Java, some components still have x86_64 dependencies or optimizations.

In mid 2017, ARC-TS partnered with HortonWorks, a leading provider of curated and tested Hadoop Ecosystem components to develop a version of its signature Hadoop software bundle HDP for use on ThunderX.

UNIVERSITY OF MICHIGAN

At U-M, the core of Hadoop is successfully running on ARM

What we have running today

- HDFS - HA
- YARN - HA
- Hive 2 - HA
- Spark 2
- Mapreduce
- Pig

```
[root@cavium-dn0037 ~]# hdfs dfs -df -h
Filesystem              Size    Used  Available  Use%
hdfs://cavium-hadoop   2.9 P  16.2 T      2.7 P    1%
```

```
[root@cavium-dn0037 ~]# /usr/hdp/2.6.2.0-205/spark2/bin/spark-submit --version
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 2.1.1.2.6.2.0-205
      /_/

Using Scala version 2.11.8, OpenJDK 64-Bit Server VM, 1.8.0_121
```

```
[root@cavium-dn0037 ~]# hadoop version
Hadoop 2.7.3.2.6.2.0-205
Subversion https://github.com/hortonworks/hadoop-release.git -r fd80ec475
Compiled by root on 2017-09-08T01:52Z
Compiled with protoc 2.5.0
From source with checksum 90b73c4c185645c1f47b61f942230
```

```
[root@cavium-dn0037 ~]# cat /etc/redhat-release
Red Hat Enterprise Linux Server release 7.3 (Maipo)
[root@cavium-dn0037 ~]# uname -io
aarch64 GNU/Linux
```

UNIVERSITY OF MICHIGAN