Week 2 Article: Big Data Trends in 2019

[https://www.dataversity.net/big-data-trends-2019/#](https://www.dataversity.net/big-data-trends-2019/)

I found the concept of data analysis occurring as it is being streamed very interesting. One would have to know what they were looking for to use this methodology. I guess it is the same thing as real-time data monitoring which is being used in applications in the aviation industry to pro-actively determine if any issues occur during flight to minimize downtime.

Week 3 Article: High Level Overview of Apache Spark

<https://hackernoon.com/high-level-overview-of-apache-spark-c225a0a162e9>

Good overview of what Spark is and what it’s strength are. The coding sections were a little too much for someone with my none technical background but the explanations of how it operates and structured was a good introduction for me.

Week 4 Article: The Future of Data Storage

<https://www.dataversity.net/future-data-storage/>

Good article on the tidal wave of data that will need to be stored over the next data and how companies are trying to set themselves up to deal with it. The most interesting idea in this article is the linkage with machine learning and that fact that we, in many cases, do not know what data we will need to feed machine learning in the future. Bottom line, save it all.

Week 5 Article: Big Data 2018: Cloud storage becomes the de facto data lake

<https://www.zdnet.com/article/big-data-2018-cloud-storage-becomes-the-de-facto-data-lake/>

Although the article focuses on cloud based storage the most interesting element was the paragraph treating streaming analytics and processing the data while “in motion”. Love the concept.

Week 6 article: Internet of things challenges in storage and data

<https://www.computerweekly.com/news/252450705/Internet-of-things-challenges-in-storage-and-data>

Two really good concepts to remember from this article.

Edge processing: Not pushing are data into the main datacenter but processing it at the edge and only pushing critical/key processed data.

Information Lifecycle: lifecycles extended because we don’t know what data will be useful for machine learning and AI applications in the future.

Week 7 article: TOP 5 CHALLENGES IN BIG DATA & ANALYTICS

<https://www.globaltechcouncil.org/big-data/top-5-challenges-in-big-data-analytics/>

First time I have been faced with the reality that most big data is unstructured and that NoSQL capability is the key to being able to extract the most out of big data.

Week 8 article: IoT, Big Data, and Cloud: 5 Notable Cloud Storage Trends Coming in 2019

<https://ubidots.com/blog/iot-big-data-and-cloud-5-notable-cloud-storage-trends-coming-in-2019/>

First time I have seen the term Dark Data: information that businesses collect during regular activities but fail to use for other purposes. This data could contain value, however, it is stored in an unstructured format and may not be accessible via a query (e.g. information in scanned documents).

Week 9 article: Data Fabrics for Big Data

<https://tdwi.org/articles/2018/06/20/ta-all-data-fabrics-for-big-data.aspx>

The term big data fabric is loosely defined at present, representing a need rather than a specific solution. A big data fabric is a system that provides seamless, real-time integration and access across the multiple data silos of a big data system. I have also hear this concept referred to as the data tread.

Week 10 article: How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh

<https://martinfowler.com/articles/data-monolith-to-mesh.html>

A good article on a new 3rd generation of data architecture called data mesh. I like that it starts with the failings of the first two generations, data warehousing and data lakes.