Data Migration To The Cloud

Patrick Waters

11/22/2021

Quick Introduction..

- Graduated from Ulster University, Ireland, in 2017 with an undergraduate degree in Computer Science.
- Spent 4 years working in the tech industry as a Consultant, specializing in Data Migration.
- In my first-year as PhD student at UNR.

Lesson Objective

The overall objective for this lesson is to demonstrate how topics taught in CS446/646 apply to real-world projects, problems and software solutions.

Specifically, the lesson will cover the following topics:

- File Systems and File Severs
- Limitations of on-premises storage solutions.
- How Cloud computing can provide the solution.
- Digital Transformation and Data Migration to the Cloud.

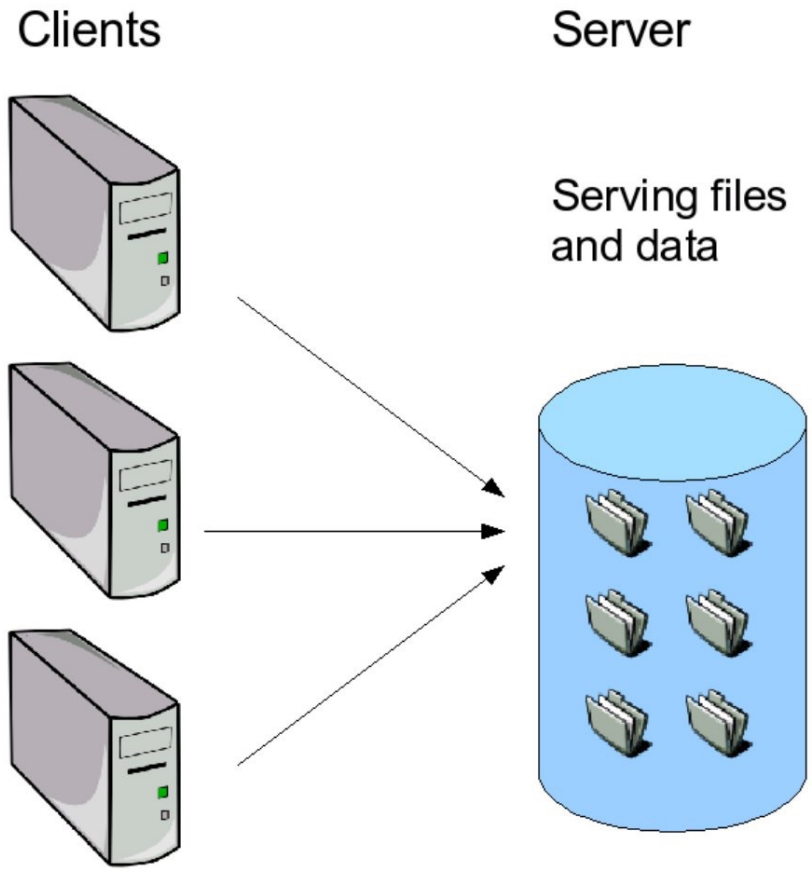
My role in industry was to deliver large-scale cloud migration projects. I'm going to go through one of those project and show how many of the technical challenges encountered can be linked back to the underlying components of operating systems.

Review: File System and File Severs

A ***file system*** is a method and data structure that the operating system uses to control how data is stored and retrieved. Each operating system tends to use its own file system.

A ***file server,*** a computer or device that is responsible for the central storage and management of data files so that other computers on the same network can access them. They share information. It is an example of an on-premises storage solution.

On-premises Storage: The server is hosted and running on the premises of the organization using the software.



The Data Storage Problem…

Traditionally, companies have used on-premises solutions to store and manage their data. These solutions have their advantages but with the exponential growth of business data, as well as a changing work culture, they are becoming a problem for many organizations.

Risk	Impact
Storage devices are inflexible and are running out of space	Employees are unable to perform their jobs
Poor Performance from heavily-used storage devices	Employees become frustrated and revert to storing data on unsecure local devices.
Backups too slow or not completing	Core business data is lost due to failed backups
Anti-Virus scans are too slow or not completing	Undetected threats could result in data breaches and data compliance liability

The root cause of a lot of these problems can be explained by the topics we have covered in this class. For example; memory management, file system performance and hardware constraints are all contributing factors. Improve those and you'll be one step closer to solving the problem.

A Real-Life Example

In 2019 one of the largest marketing agency in the world merged with another, forming a new company with 200 offices and 20,000 employees across the world. This also meant the merging of their data, and with an '*estimated*' 2PB (Petabyte) it was a not a simple task. Many of the offices operated independently in that they had they're own set storage solutions and internal processes, and as a result nobody truly knew how much data the company was holding, where it was being stored, who was accessing it, and even how much it was costing. The top priority for the CEO was to solve this problem and take control of all the unstructured data, but how?

Cloud Computing

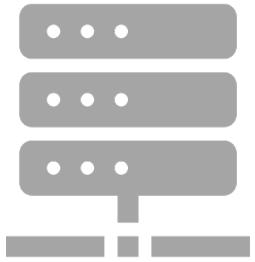
1. Perform a deep metadata-based analysis of all the data.
2. Identify data for deletion and for migration.
3. Transfer or migrated the required data into a cloud-based SaaS solution E.g. Microsoft 365.

Review: Virtualization and Cloud Computing

Virtualization refers to a piece of software that behaves and functions like a physical hardware object. A **cloud environment** is one in which all this virtualization is housed somewhere else, and a vendor provides the service to clients on a fee basis.

IaaS (*Infrastructure-as-a-Service*):

Provides access to resources such as physical machines, virtual storage, etc.



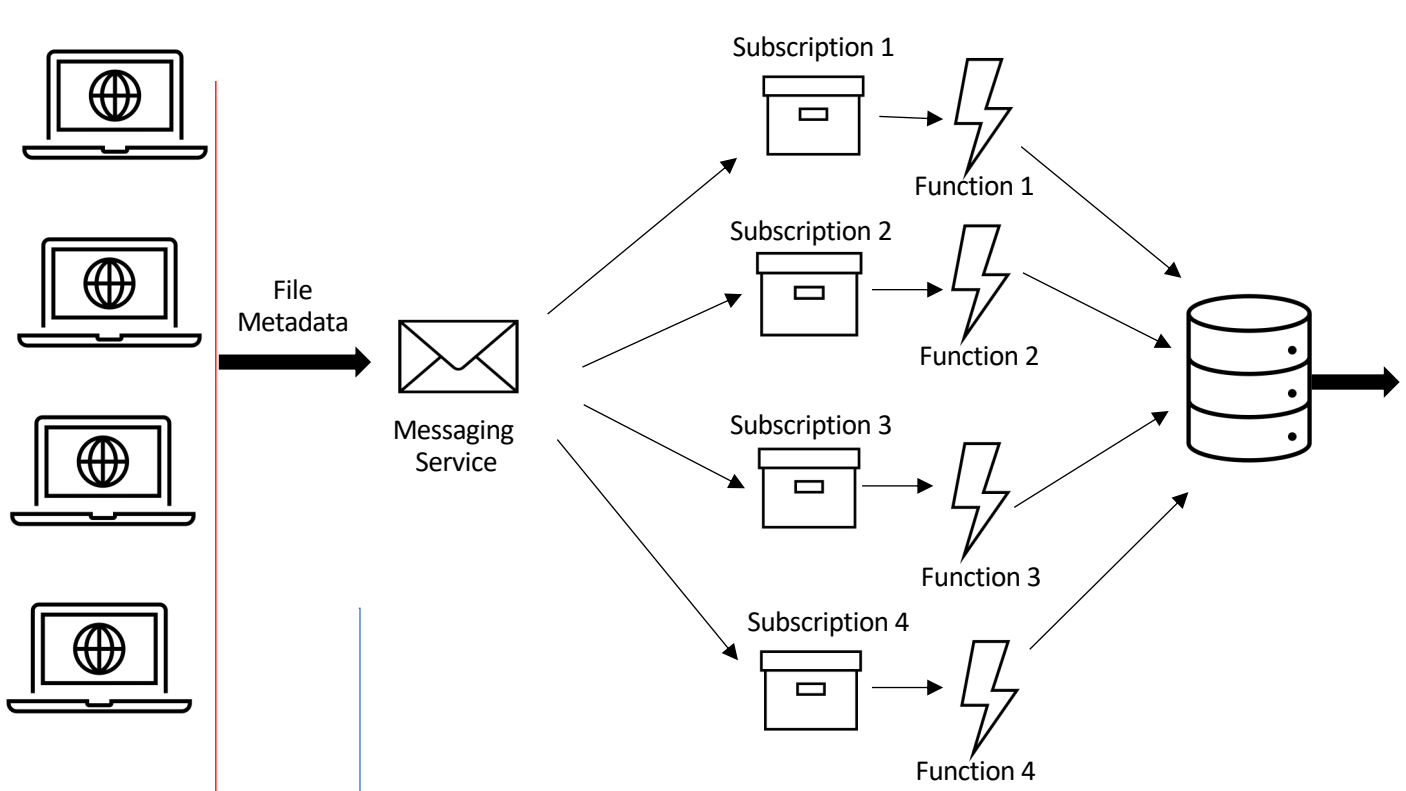
PaaS (*Platform-as-a-Service*):

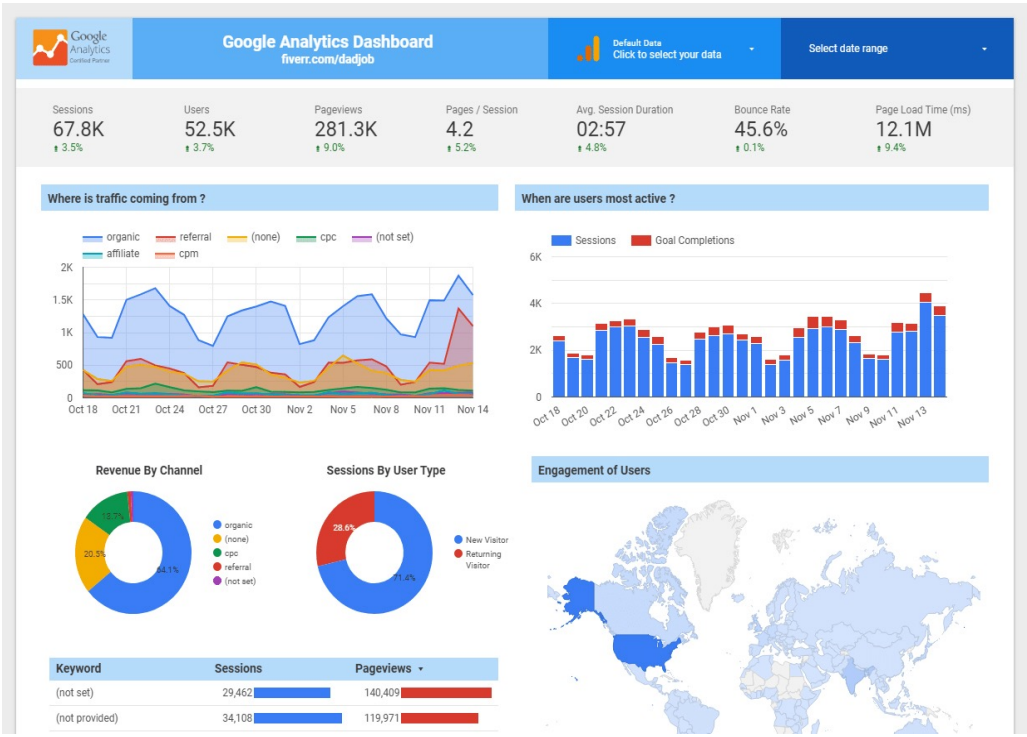
Supports the complete lifecycle of building and developing web applications entirely from the internet.

SaaS (*Software-as-a Service*):

A way of delivering applications over the Internet—as a service.

Example: Using the Cloud for Metadata Analysis





On-Premises Systems

Cloud Environment

Using SaaS To Resolve File Sever Risks

File Sever Risk	SaaS Solution (E.g., Microsoft 365 – SharePoint Online)
Storage devices are running out of space	Flexible storage limits that can be increased or decreased on demand.
Poor Performance from heavily-used storage devices	Providing you have a good internet connection performance should not be an issue. The responsibility is with the service provider who can scale resources accordingly.
Backups are too slow or not completing.	Service providers offer advanced data-loss prevention options as well as fast backup retrieval.
Anti-Virus scans are too slow or not completing	Anti-Virus scans are no longer necessary.

Thank you for listening.

Any questions?