**Cloud Concepts and Storage**

**What problems can arise when using a Virtual Machine?**

* A VM (Virtual Machine) 🡪 Depends on infrastructure (this can be the laptop or anything else).
* So, if problems arise with the infrastructure 🡪 Then you cannot use the VM.
* How to overcome this problem with a VM and the infrastructure of it?

Use a **cloud storage/cloud service provider**.

**What is a cloud storage/cloud service provider?**

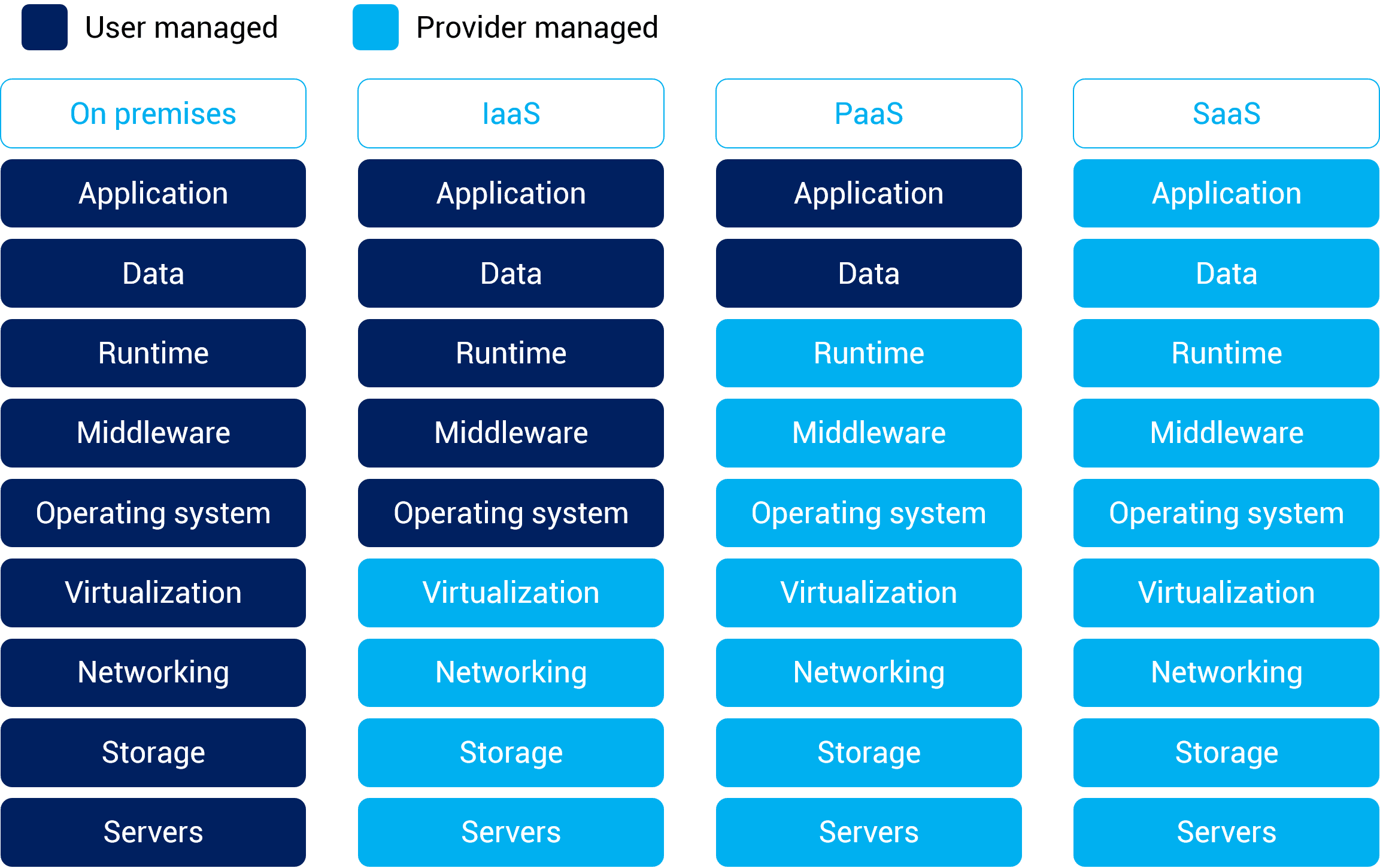
* A **cloud storage/cloud service provider** 🡪 Has everything that you need, i.e. all the infrastructure that you need.
* This infrastructure includes storage, network, VM, etc.
* So 🡪 Can run the VM and other things as well solely using the cloud storage.

Examples – The 3 main cloud providers

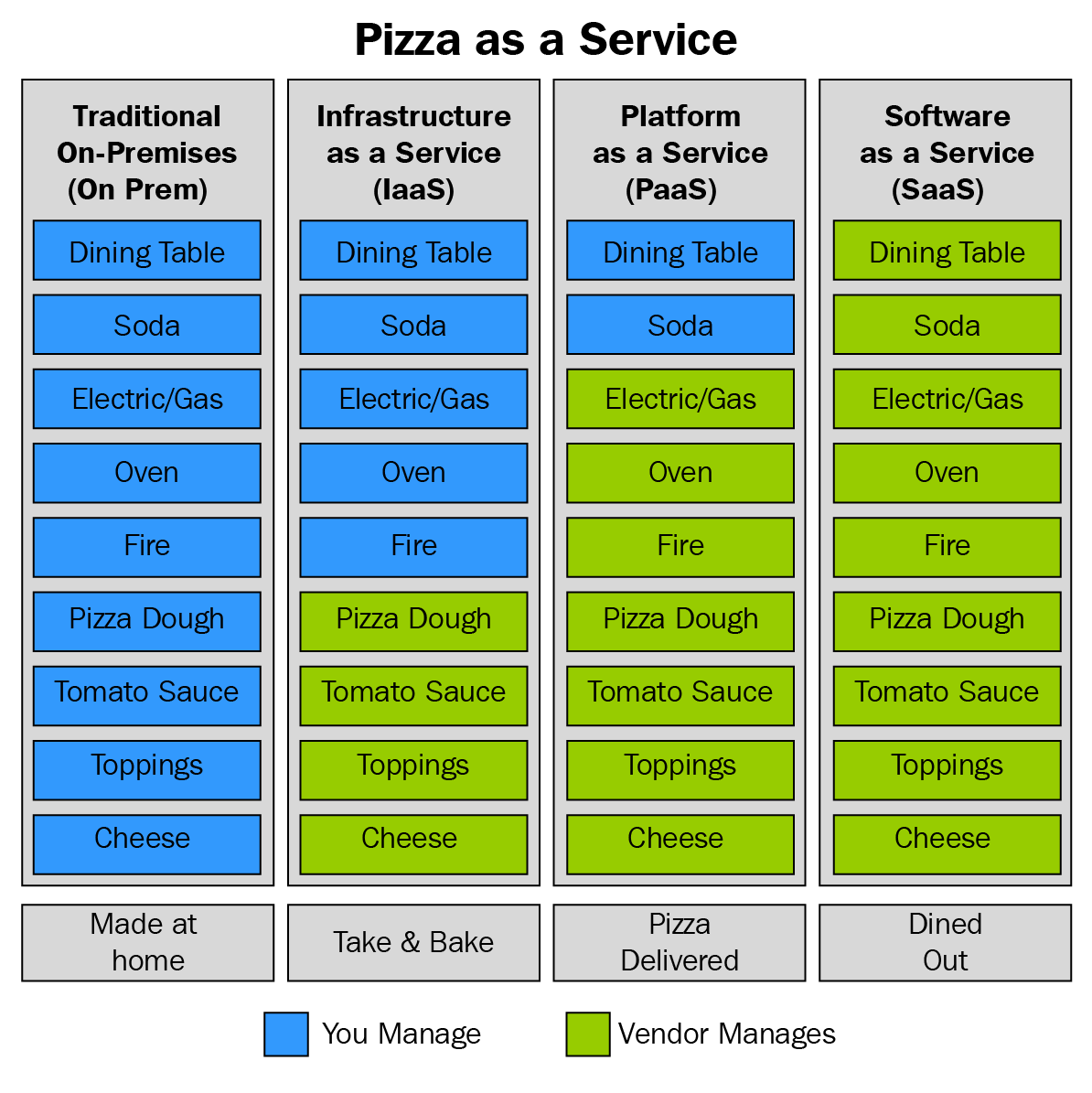
1. AWS (Amazon Web Services) (main one used)
2. GCP (Google Cloud Platform)
3. Agile Cloud

|  |  |
| --- | --- |
| **The 3 Kinds of Cloud Services** | |
| 1. IAAS   (Infrastructure As A Service) | Examples   1. DigitalOcean 2. Linode 3. Rackspace 4. Amazon Web Services (AWS) 5. Cisco Metapod 6. Microsoft Azure 7. Google Compute Engine (GCE) |
| 1. PAAS   (Platform As A Service) | Examples   1. AWS EMR (Amazon Web Services Elastic MapReduce)  * Has all the components to work with big data 🡪 Has HDFS, Hadoop, Spark.  1. AWS Elastic Beanstalk 2. Windows Azure 3. Heroku 4. Force.com 5. Google App Engine 6. Apache Stratos 7. OpenShift. |
| 1. SAAS   (Software As A Service) | Examples   1. Google Apps  * Google Docs 🡪 Just type 🡪 Don’t maintain anything.  1. Dropbox 2. Salesforce 3. Cisco WebEx 4. Concur 5. GoToMeeting |
| **Nowadays moving: From IAAS 🡪 PAAS 🡪 SAAS** | |

**Summary of Key Differences between On-Premises, IAAS, PAAS and SAAS**



**Pizza as a Cloud Service Example**



Need to be aware of **ALL** components in GCP, AWS, Agile:

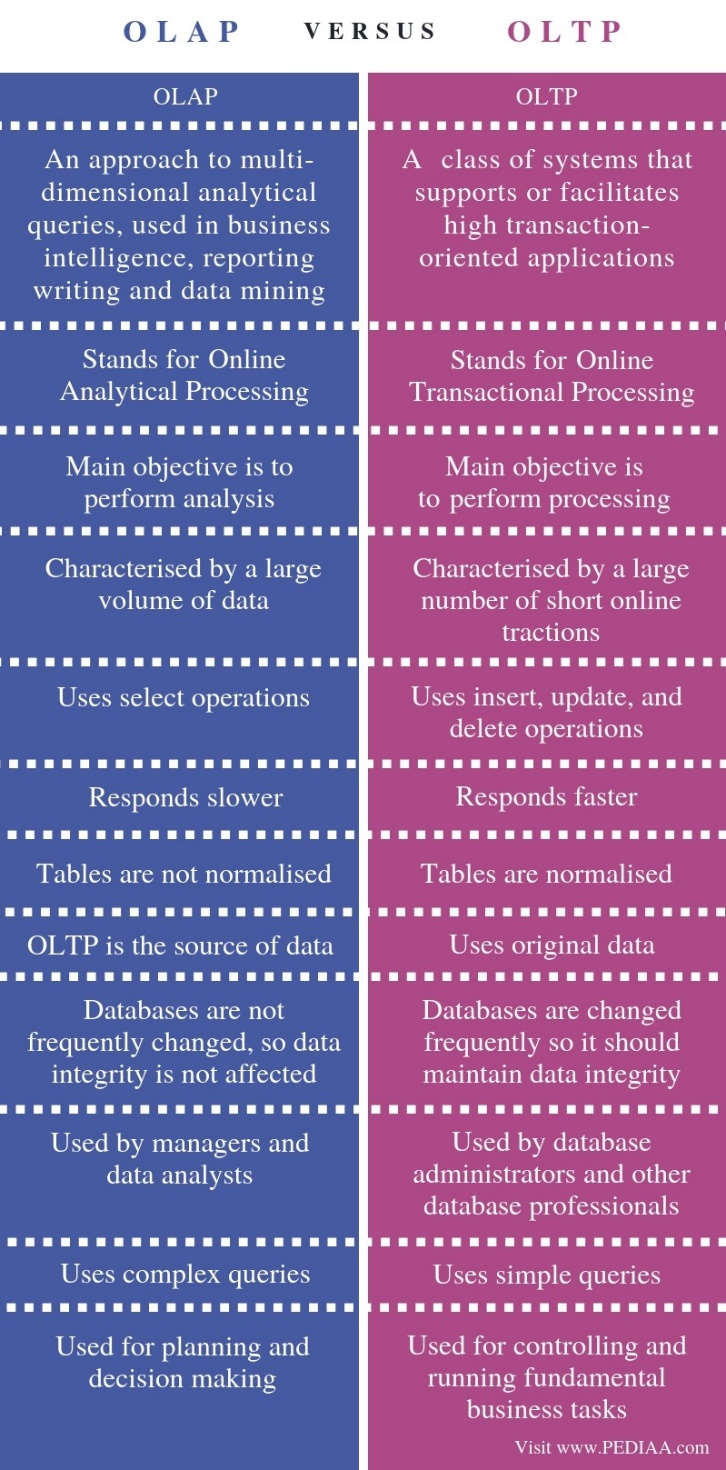
* Orchestration Tool: Dataflow 🡪 GCP
* Orchestration Tool: Stepfunction 🡪 AWS
* Orchestration Tool: 🡪 Agiles

**Big Data**

Data sets that are so large or complex that traditional data processing software or simple code is inadequate to deal with them.

* Big Data 🡪 Requires specialised tools and software to ingest, clean, manipulate, and extract trends and other relevant info.
* Examples
  + Netflix 🡪 Uses big data to recommend other content.
  + Facebook 🡪 Needs to cache images for billions of users around the world.
  + Google 🡪 Needs to process millions of search queries per second.
  + PayPal 🡪 Uses big data to predict fraud across millions of transactions.
* Big data is part of the world now, and rapidly becoming an expected part of many businesses.
* Big data 🡪 Becoming so much more than merely large data sets. In today's world, it represents an entire ecosystem of data sets, tools, and applications.
* Discuss the tools that are available, and why you might want to use them.
* Think of Big Data as:
  + The data sets that are being processed.
  + The infrastructure that’s needed to support such analysis.
* Big Data technologies have been developed and are subject to continuous improvement

**The Two Types of Processing of Big Data (Look and Read more into)**



Links to look at:

* <https://www.youtube.com/watch?v=BQpCGyUzTHo> - OLAP vs OLTP | Online Transaction Processing vs Online Analytical Processing | Intellipaat
* <https://www.youtube.com/watch?v=GkZre_zkJJ0> - OLTP vs OLAP | Online Transaction Processing vs Online Analytical Processing | Edureka
* <https://www.educba.com/oltp-vs-olap/>
* <https://techdifferences.com/difference-between-oltp-and-olap.html#:~:text=OLAP%20is%20an%20online%20system%20that%20reports%20to,OLAP%20is%20an%20online%20database%20query%20answering%20system.>
* <https://www.guru99.com/oltp-vs-olap.html>

Examples of OTLP and OLAP

**2 Types of Processing**

|  |  |
| --- | --- |
| **Type of Processing** | **Examples** |
| 1. Batch Processing |  |
| 1. Stream Processing |  |

Links to look at:

* <https://www.bmc.com/blogs/batch-processing-stream-processing-real-time/>
* <https://rivery.io/batch-vs-stream-processing-pros-and-cons-2/>
* <https://blog.philipp-brunenberg.de/entries/batch-stream/>
* <https://hazelcast.com/glossary/stream-processing/>

**Scalability**

Links to look at:

* <https://www.informit.com/articles/article.aspx?p=2258416>
* <https://www.ngdata.com/the-importance-of-scalability-in-big-data-processing/>
* <https://www.analyticsinsight.net/big-data-visualization-scalable-computing/#:~:text=%20Big%20Data%20with%20Visualization%20and%20Scalable%20Computing,Trace%20out%20areas%20that%20need%20enhancement.%20More%20>
* Horizontal vs Vertical - <https://pediaa.com/what-is-the-difference-between-horizontal-and-vertical-scaling/>