E4
COMPUTER
ENGINEERING

KAPTAIN

ITS|UMBRIA
SMART ACADEMY

ITS UMBRIA SMART ACADEMY

CASE STUDY

WHEN PERFORMANCE MATTERS

CUSTOMER PROFILE

ITS Umbria is an **Academy for Applied Sciences and Technologies**, an excellent opportunity for young graduates who wish to acquire high technological skills, thus creating the conditions for an easy and qualified entry into the labor market.

An objective that can be achieved by attending the **two-year courses promoted by the Ministry of Education and the Umbria Region**, with the participation of the University and of the main companies in the area.

The didactic and curricular flexibility, the strong dynamism, the technological and digital footprint, the active participation of companies and the certified levels of employment achieved, make **ITS one of the most effective and performing education channels in the country**, certainly the more modern and contemporary.

In fact, it integrates high schools, universities, companies and experts from the labor market in a very effective contamination of knowledge and skills. At the end of the training courses, people gain the State Diploma of Tertiary Education, equal to the 5th level of the European Qualifications Framework – EQF, issued by the Ministry of Education.

OFFICES: PERUGIA, TERNI, FOLIGNO
CURRENT STUDENTS 380

ACADEMIC ROUTES IN BIM, CIRCULAR ECONOMY, AGRI-FOOD, MECHATRONIC, TOURISM & MARKETING

80% OF THE STUDENTS FIND EMPLOYMENT WITHIN 12 MONTHS AFTER THE DIPLOMA

1° PLACE IN THE NATIONAL RANKING DRAWN UP BY THE MINISTRY OF EDUCATION IN 2020

REQUEST

"Need for a high-performance, turnkey infrastructure that could host large-scale data analysis platforms".

ITS Umbria Smart Academy, with the aim of growing and develop its teaching laboratory, **needed to implement a high-performance infrastructure** that could host various platforms and **large-scale data analysis frameworks**.

The customer also needed a "turnkey" and easy to use solution, which, through container orchestration, would allow maximum flexibility in the use of the cluster's computing and storage resources.

SOLUTION

To best meet these needs E4 proposed **Kaptain**, a **Kubernetes cluster** with highly reliable core services, designed to host the most expensive workloads and therefore **suitable for hosting platforms and frameworks for Big Data analysis**, thanks to the integration of high-end servers and a high-performance network, based on Mellanox Ethernet RDMA 25+ Gigabit/s technology, used for communication between containers and for accessing storage resources. Regardless of the specific configuration, Kaptain always includes **a high-performance Distributed Block Storage service**, on which the Kubernetes Default Storage Class is configured, and a web-accessible UI, both for the administrator and for the end user, extremely powerful and also simple to use. Kaptain: a high-performance Kubernetes cluster, ready to use!
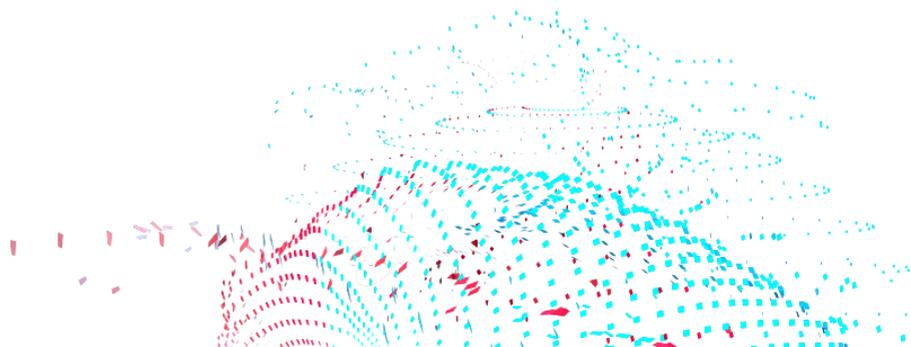
INFRASTRUCTURE DESCRIPTION

The budget available to ITS Umbria required the design of a custom version of Kaptain, according to the diagram in the figure: the **3 Master-Storage Nodes** host the **Kubernetes ControlPlane** and provide the **Cloud Native Distributed Block Storage** servic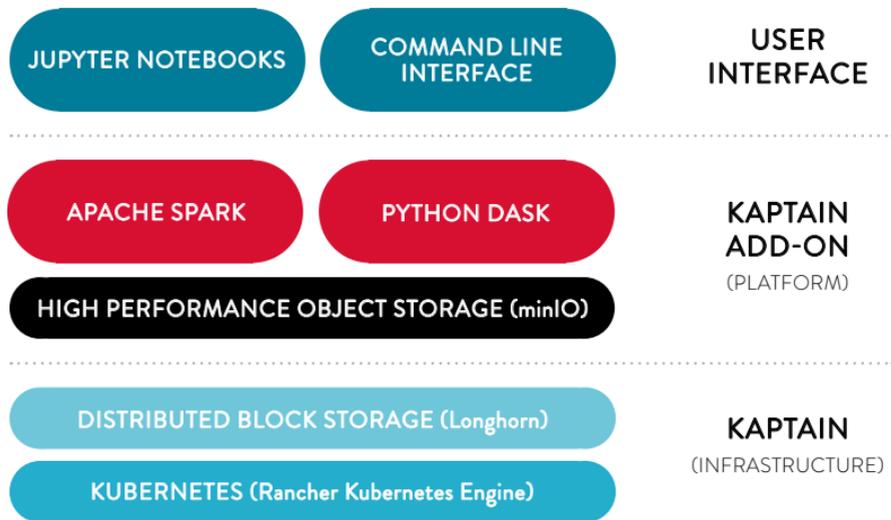es, aggregating the 3 additional 1.9 TB SSDs present on each node, in a single pool of approximately 18 TB gross, managed by Longhorn. The related Kubernetes Storage Class provides Persistent Volumes to containers running across the entire infrastructure; Longhorn's services use the Mellanox RDMA network for both access and volume replication operations, necessary to ensure data security. The **3 Worker Nodes guarantee 120 cores and 576 GB of RAM** to be dedicated entirely to the Data Analysis platforms and frameworks.
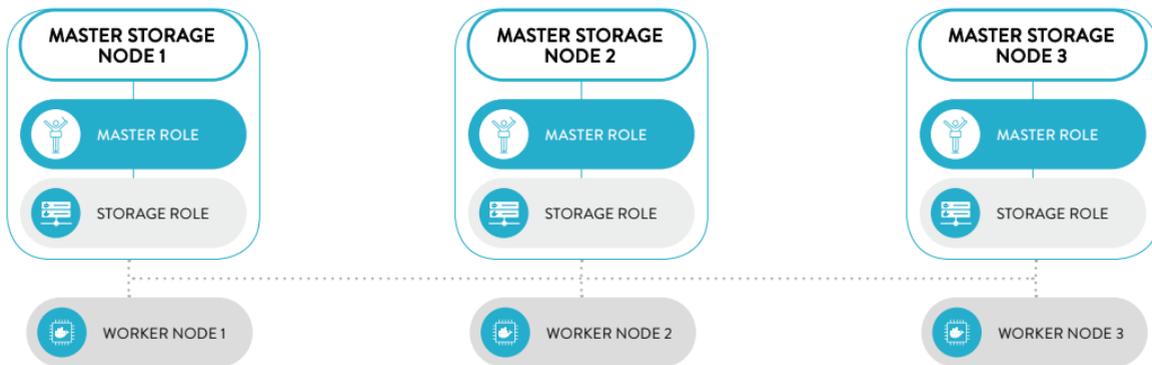
With this basic configuration, a Object Storage service distributed through MinIO, a **Cloud Native Distributed Object Storage solution oriented towards Data Analysis**, was set up.

The level of specific platforms and frameworks for Data Analytics has been implemented through the deployment of version 3.0.1 of **Spark for Kubernetes** and a multicore **Python DASK** environment that can be scaled to multiple containers, depending on the size of the dataset to be analyzed, and accessible from the web through the JupyterLab interface.

JUPYTER NOTEBOOKS	COMMAND LINE INTERFACE	USER INTERFACE

APACHE SPARK	PYTHON DASK	KAPTAIN ADD-ON (PLATFORM)
HIGH PERFORMANCE OBJECT STORAGE (minIO)		

DISTRIBUTED BLOCK STORAGE (Longhorn)	KAPTAIN (INFRASTRUCTURE)
KUBERNETES (Rancher Kubernetes Engine)	

BENEFITS

Kaptain is a solution that allows you to deal with the analysis of large amounts of data with the **Cloud Native Data Science** paradigm, offering the possibility of **combining, on the same system, platforms and frameworks** that, in other scenarios, would require dedicated infrastructures (physical and/or virtual): for example, the deployment of a distributed Data Architecture based on Apache Spark can coexist with a solution for large-scale data analysis based on DASK and with Kubeflow, the toolkit for the deployment of scalable workflows of **Machine Learning**, which incorporates **Deep Learning frameworks** such as Tensorflow, pyTorch or MxNet.

Furthermore, a part of the block storage resources was used to configure a **high-performance Object Storage service** based on minIO, to structure a fast, **secure and accessible data repository from all the Data Analysis** tools that can be installed on Kaptain.

Finally, the presence of the Apps & Marketplace catalog, included in Rancher Server, then allows the end user **to instantiate, with a few mouse clicks, a wide range of applications and services**, naturally using only the sub-set of computing and storage resources, that the infrastructure administrator made available to him: workflow manager to manage the analysis pipelines, tools to implement CI/CD in the Data Science field, as well as the main open-source DBMS systems, both SQL and noSQL, configurable also in high reliability.

KAPTAIN: THE PERFECT MEETING POINT

"ITS Umbria Smart Academy offers courses for young graduates who wish to acquire high technological skills, thus creating the conditions for an easy and qualified entry into the labour market. A few months ago, we were commissioned by ITS Umbria Smart Academy to study which was the best technology to invest in for the development of training curricula with specialization in digital technologies. **The main objective was to equip the educational laboratory with a high-performance infrastructure that could host large-scale data analysis platforms***. The technology had to be flexible enough to be exploited in different educational contexts, from data science to software production, from Machine Learning to Cloud Computing. Furthermore, we wanted to acquire an infrastructure that would also allow the implementation of projects with industrial partners.*

It was not easy to adequately cover all needs*. In addition to the objectives listed above, another has been added: the need to have a "turnkey" hardware and software platform, easy to use and maintain.* **E4's Kaptain solution represented the perfect meeting point***: a* **hardware platform that can be calibrated to our budget and our computing needs***, a* **customizable software** *stack,* **flexible resource management***, a large and modern application catalog. Not only that, the professionalism and availability of the E4 staff meant that all the surrounding needs emerged sufficiently in advance and could be addressed in the best way. The delivery times were thus fully respected, and in a short time we were able to take advantage of the power of Kaptain."*

Prof. Emilio Di Giacomo, Prof. Fabrizio Montecchiani