

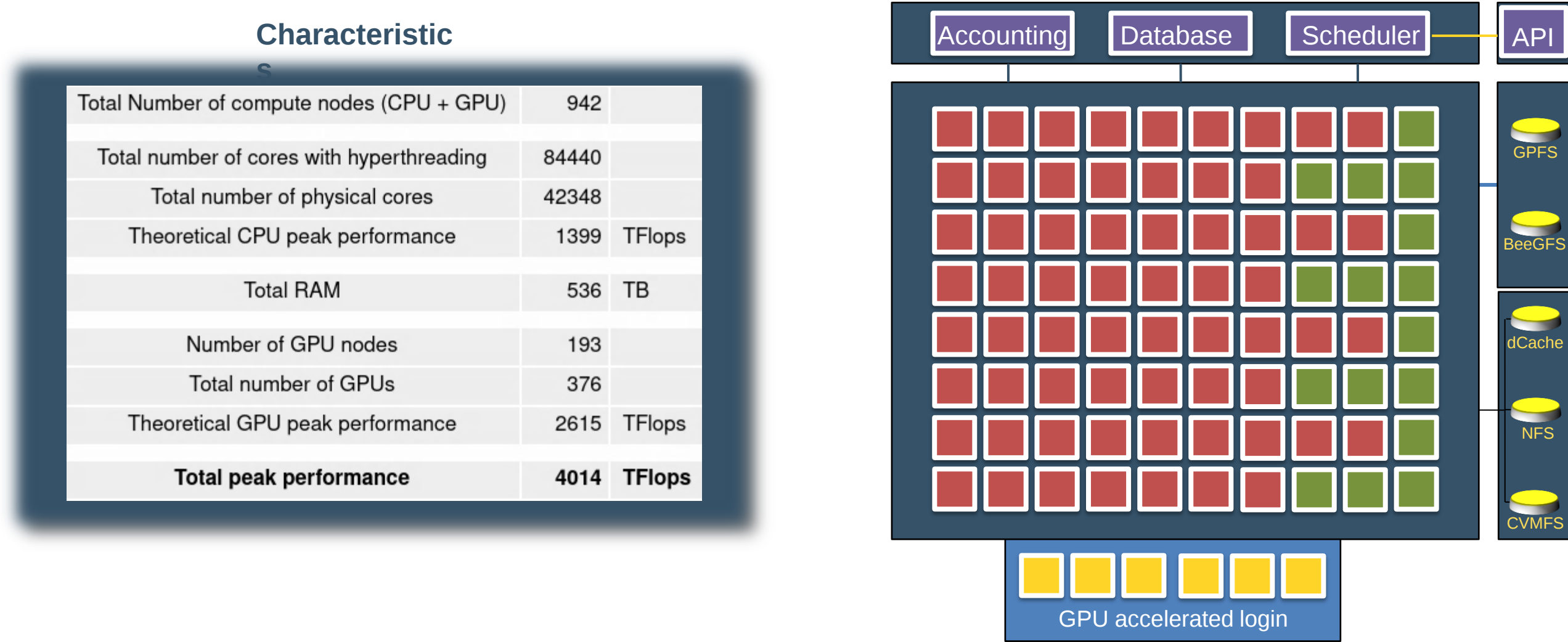
Leveraging Jupyter on Maxwell HPC: joyful, visual and green computing



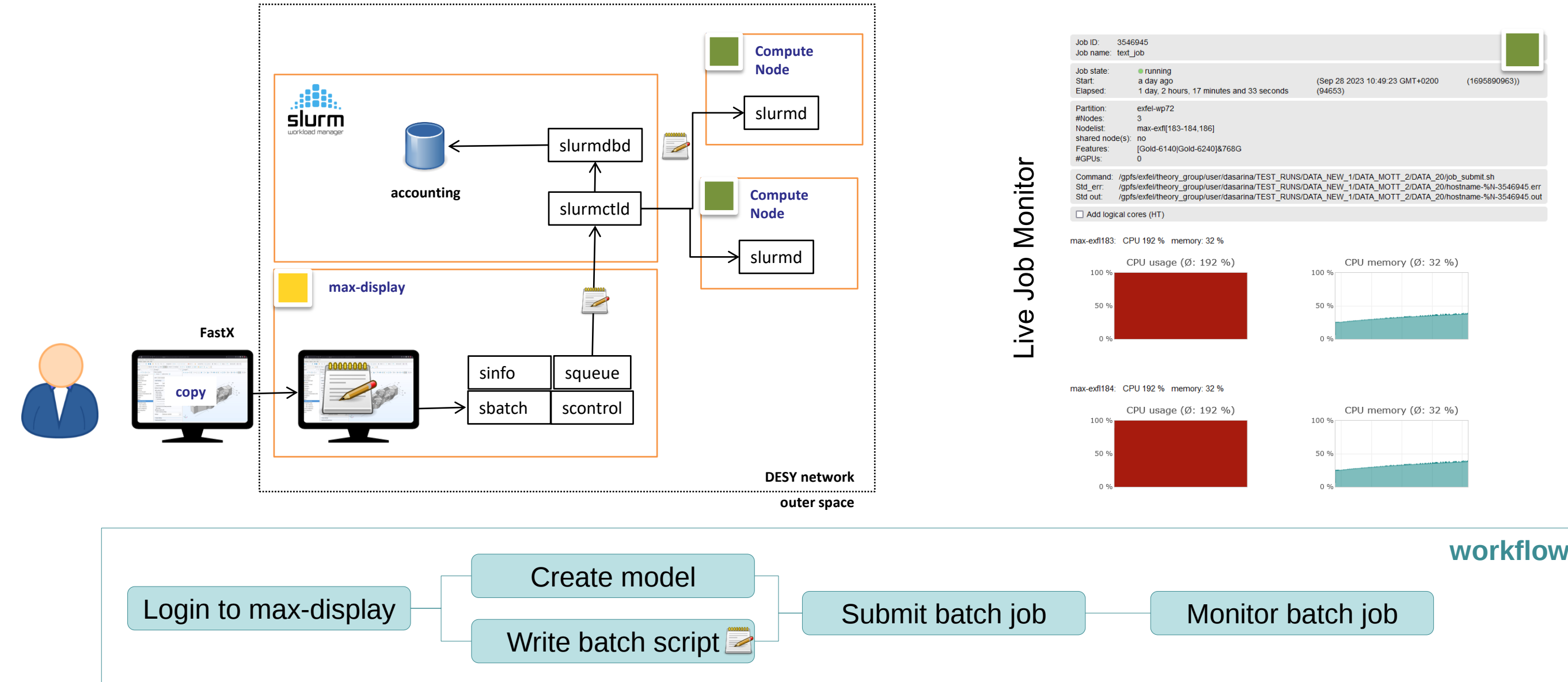
Yves Kemp, Arlena Mills-Marzoli, Neele Rahmlow, Sven Sternberger, Axel Wichmann, Frank Schlünzen

Abstract: Jupyter notebooks are great tools to mitigate the complexities of (heterogeneous) HPC systems, like the Maxwell cluster at DESY which serves the computational needs of all user facilities on campus, as well as a wide variety of applications ranging from plasma accelerators to quantum chemistry. We aim to expand the Jupyter ecosystem using frameworks like streamlit to provide application environments tailored to the needs of less experienced users, including real-time visualization capabilities. On this basis we are implementing for example Jupyter-driven remote desktops, user-friendly dashboards to compose or monitor batch-jobs, and visual frontends for data catalogues like SciCat. The implementations are accompanied by visual tools for resource utilization and CO2 footprints suitable both for users as well as admins.

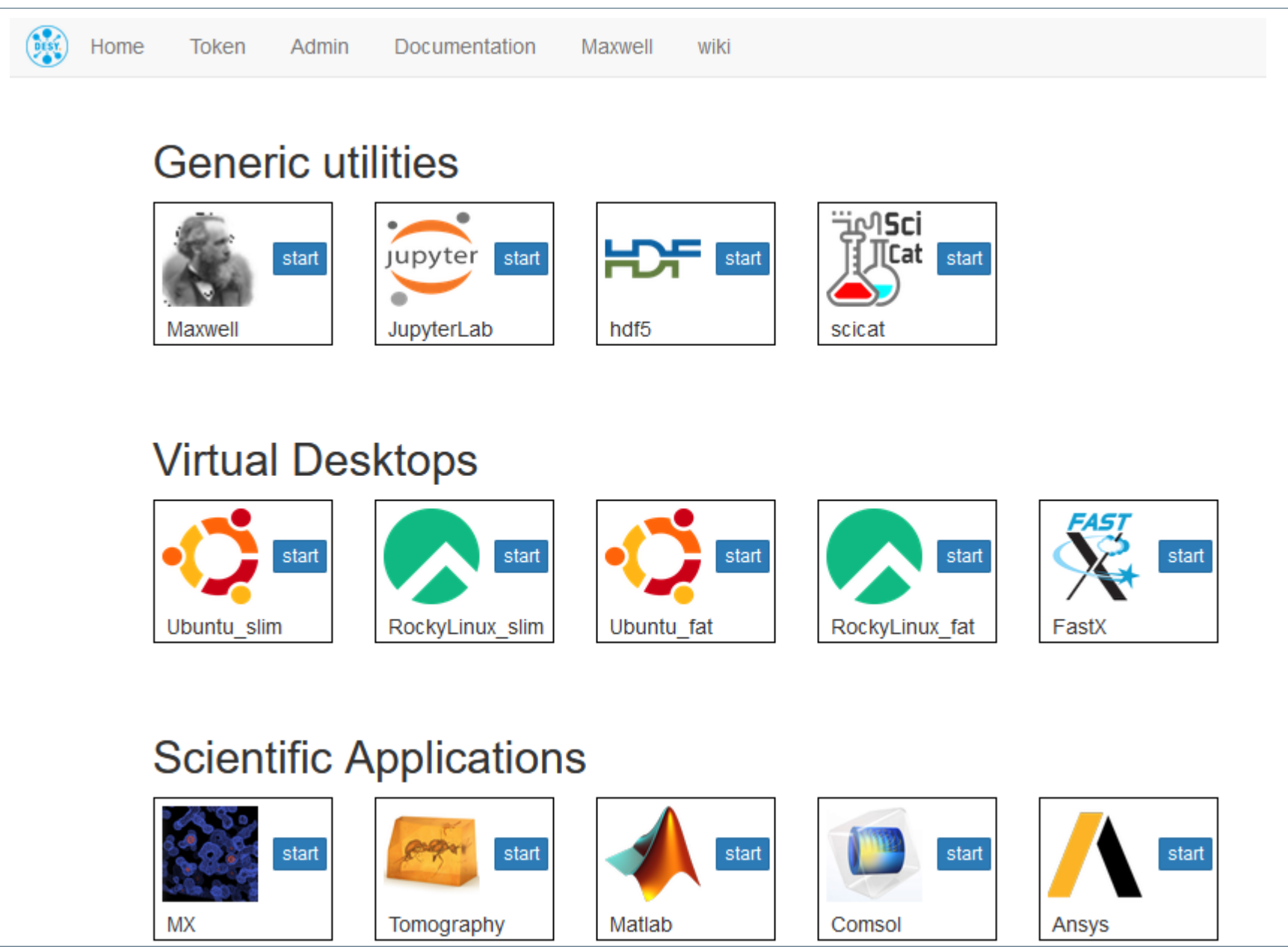
Maxwell HPC Cluster



The regular Workflow

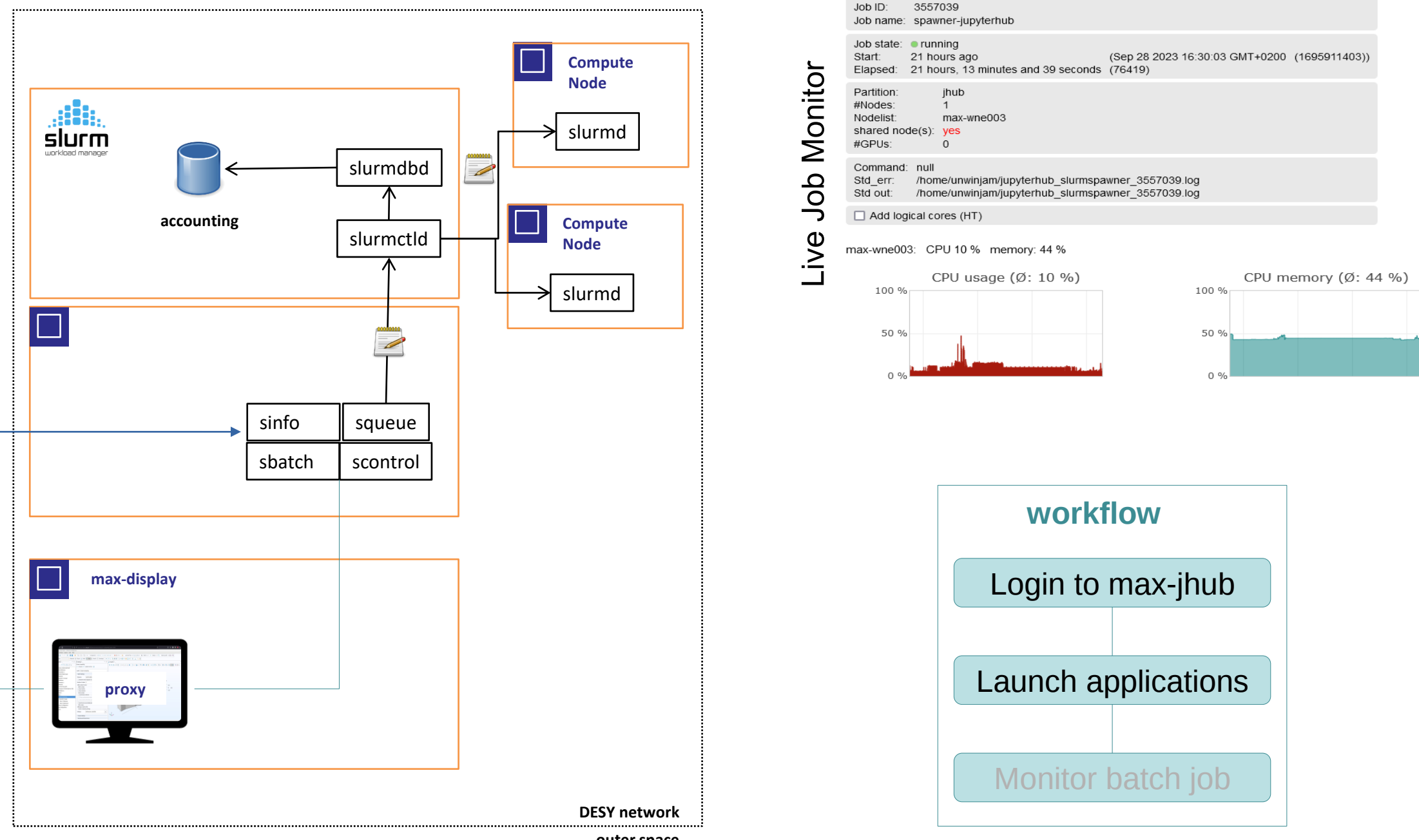
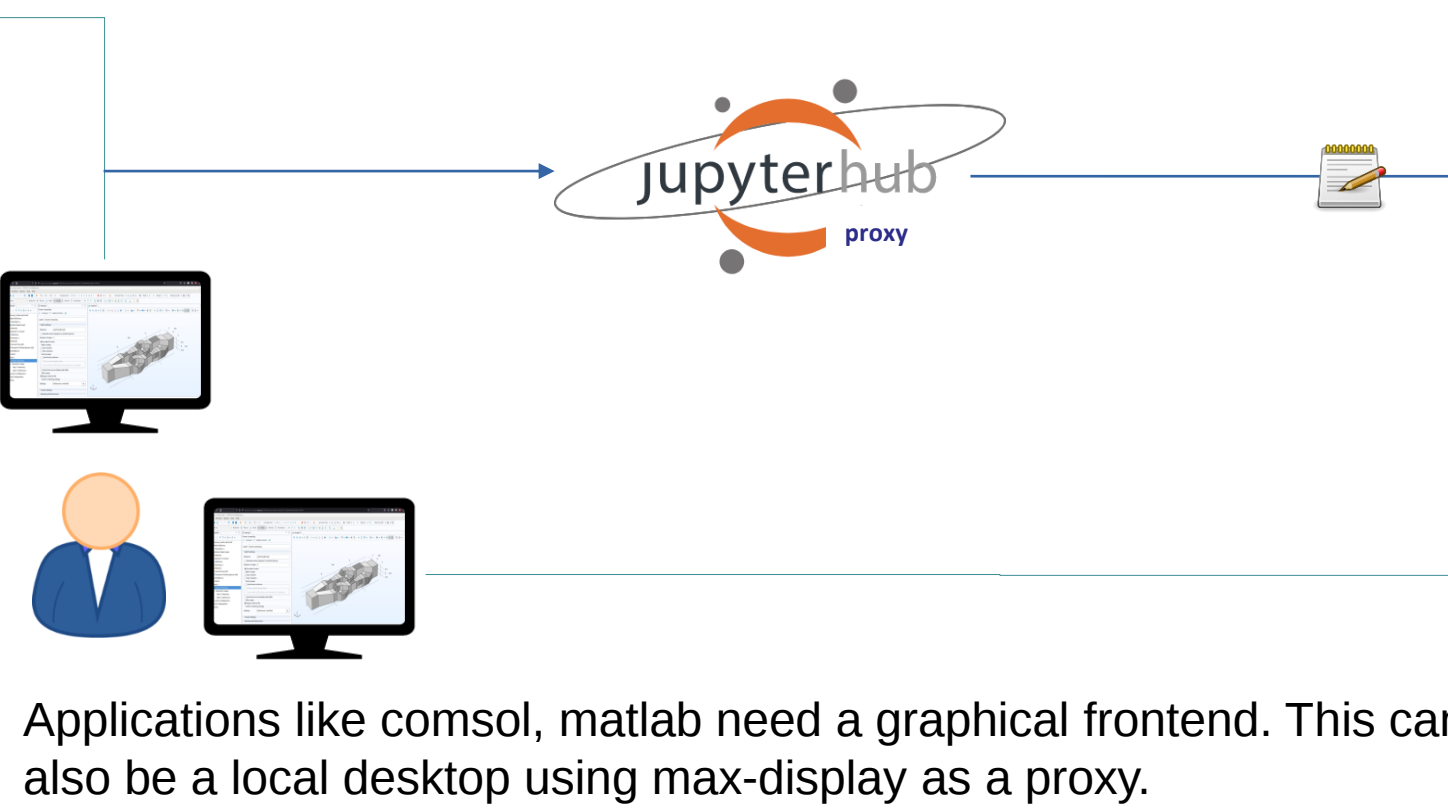


Most users will use the graphical login nodes (max-display) for graphical work and submission of batch-jobs. The login nodes are clustered and well equipped with GPGPUs and memory allowing for example CAD modeling. The nodes can be reached via ssh, a FastX client or a web-browser from anywhere in the internet, but are completely isolated from the DESY internal network for security reasons, and even root privileges are not sufficient to modify any files on the cluster filesystem (root squashing). The impact of compromised account will be minimal.

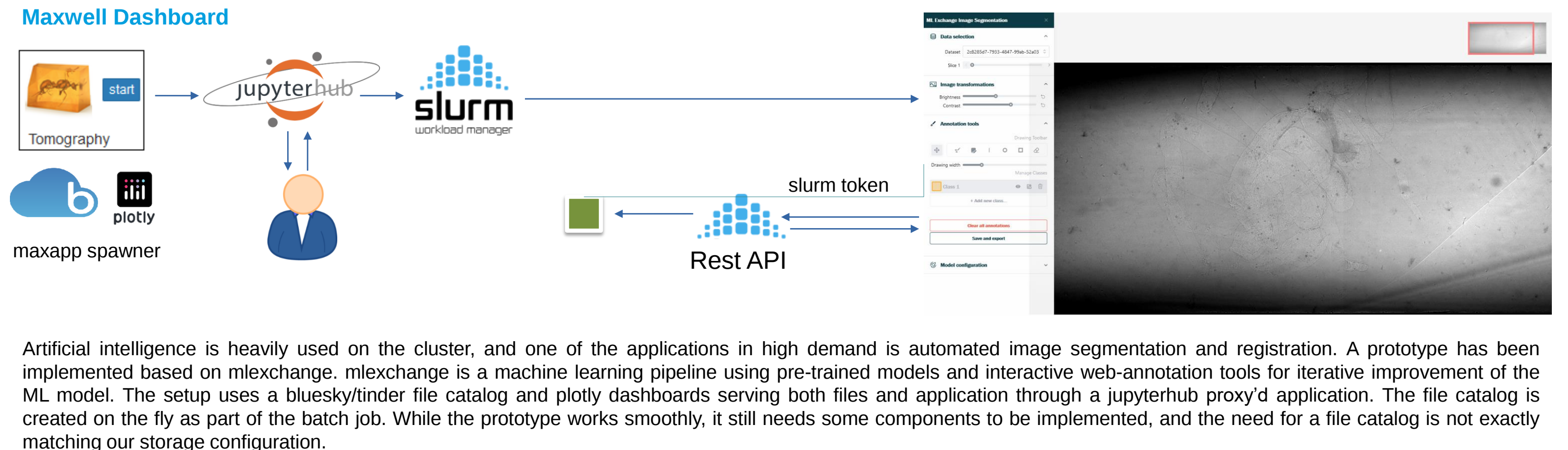
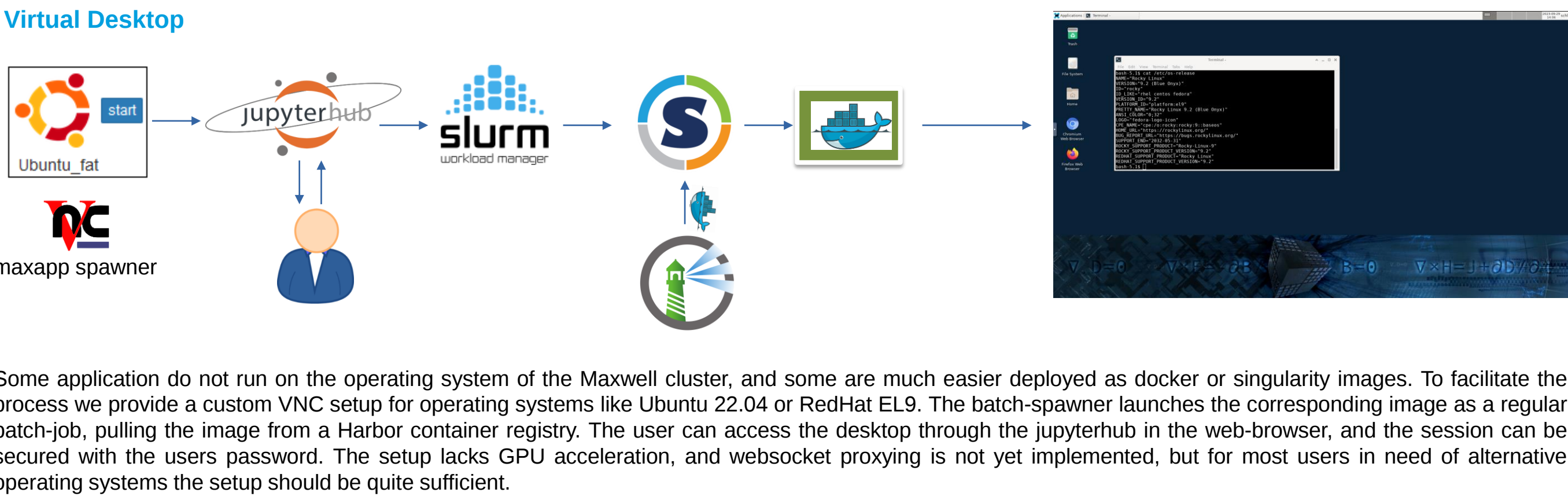
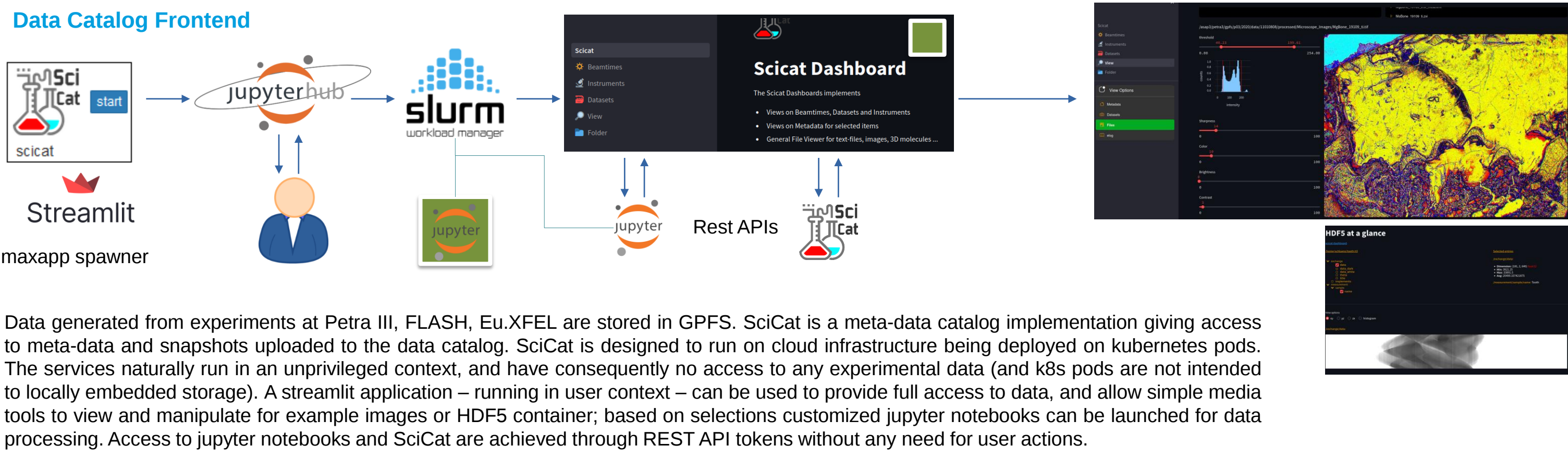
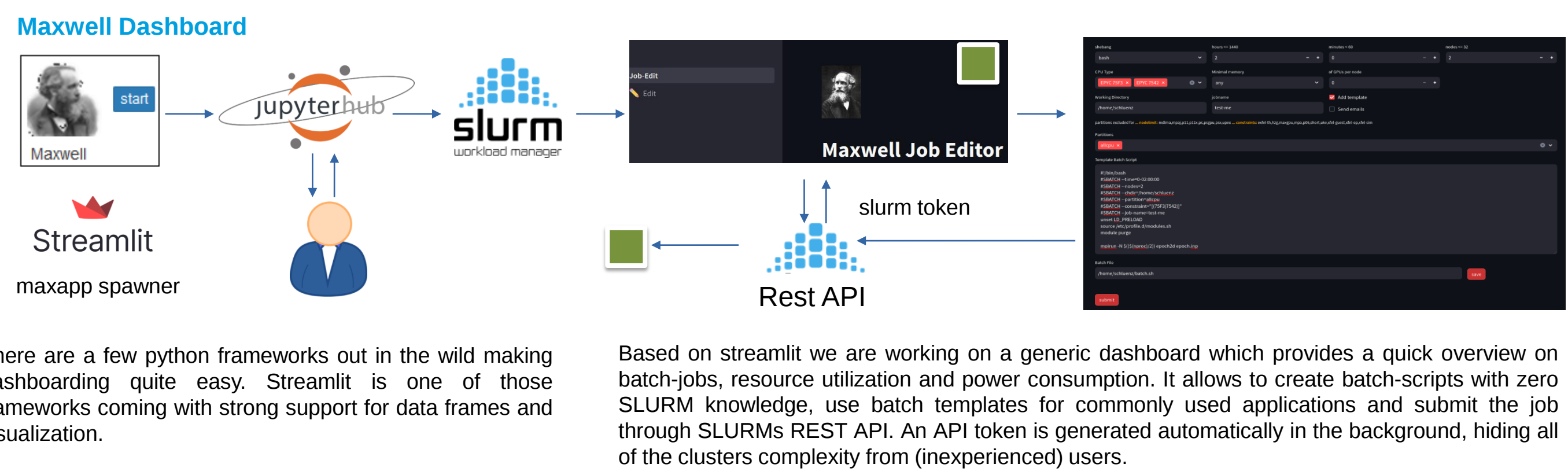


JupyterHub as a proxy

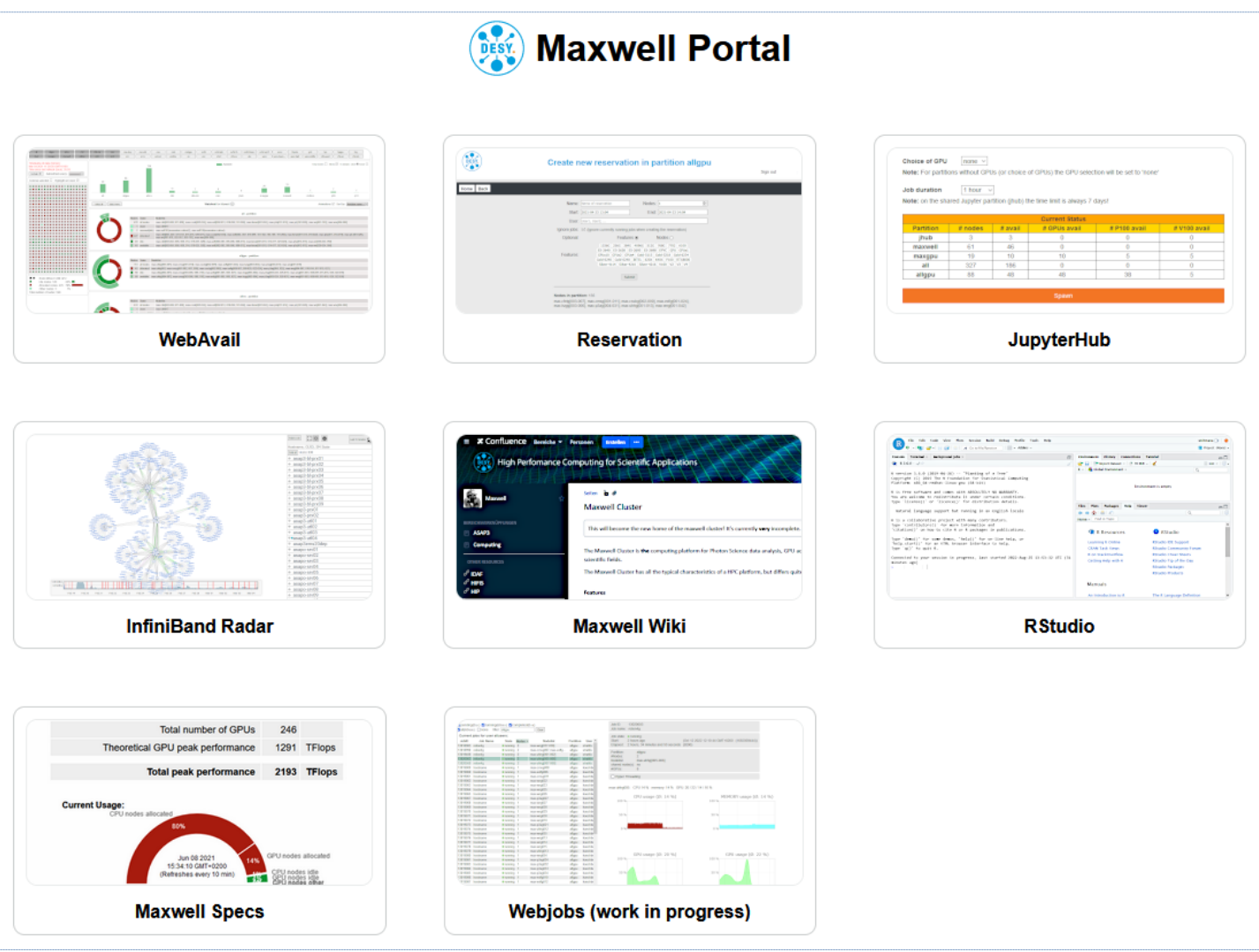
JupyterHub can proxy almost every kind of application. Our batch-spawner implementation (maxapp-spawner) creates and submits the batch scripts, defining the proxied ports and application specific compute requirements. The proxied application runs as a batch job entirely in user space, which makes it very simple to allow access for example to storage, avoiding the nightmare of implementing GPFS extended ACLs in a generic web-service. User need to remember only a single access point – the jupyterhub – which controls and keeps track of applications.



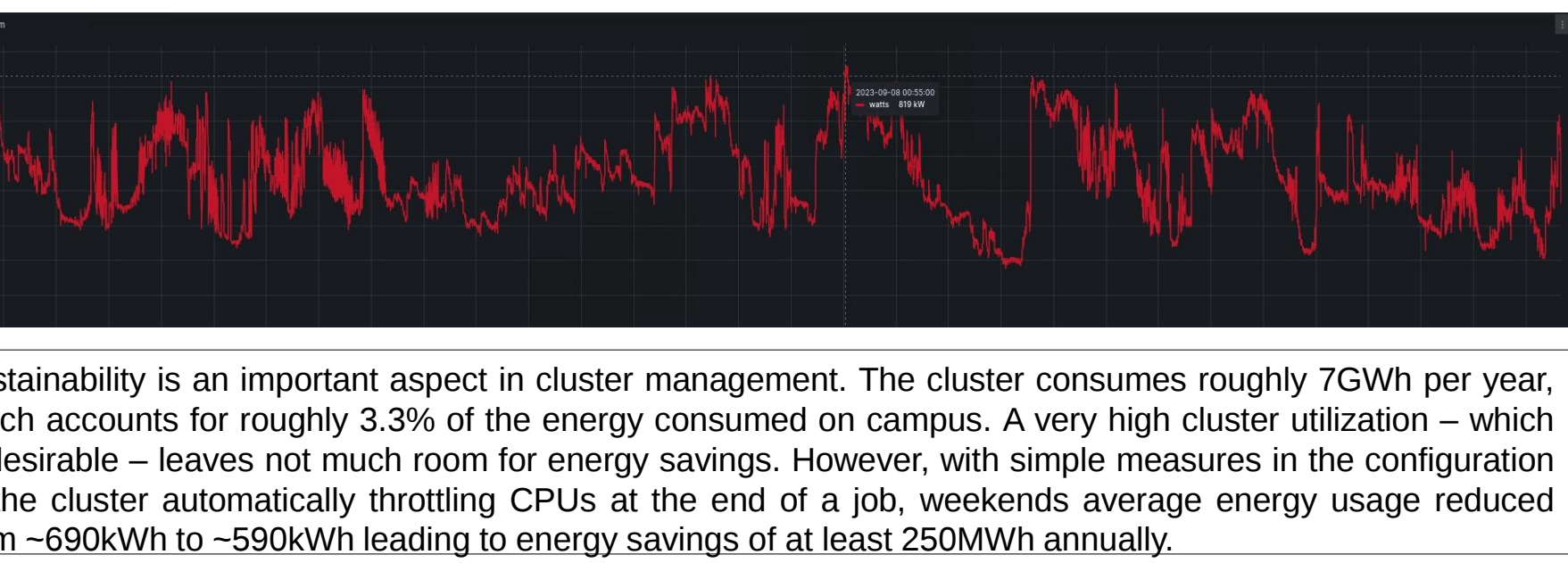
Proxied applications



Services



The maxwell portal is the entry-point to the most prominent services on the cluster. This includes access to the jupyterhub, R-Studio, cluster reservation tools, and the often demanded one-click visualization of characteristics and utilization of the cluster.



Cluster users are not always aware of its intricate nature, which quite frequently leads to poor utilization of resources, and a very uneven distribution of the workload on the compute nodes. The webjobs service – also an in-house development – visualizes the distribution at a glance, and can alert users of poor or uneven resource utilization, which helps to maximize throughput and hence to minimize energy consumption

