

WP5 – Extreme Health Use Cases

Use-Case	New Data Connectors	KPIs	Description
Variant-Interactions (BSC)	Data functional interpretation Data merger Data partitioner Data predictor Data preparator Data selector HPC connector Lithops Test of association	KPI-1 - Throughput Improvements KPI-3 - Resource Auto-scaling	MPI version shows a speed-up of 5x compared to the Apache Spark version. The adoption of a GPU component in our HPC data connector improves performance by 2.1 times.
			Integration with Lithops
Surgery (NCT)	Federated Learning Pravega clients Pravega GStreamer	KPI-2 - Data Speed Improvements KPI-3 - Resource Auto-scaling KPI-4 - Confidential computing KPI-5 - Simplicity and Productivity	With Pravega's integration, end-to-end IO latency is reduced by 45%.
			Integration with SCONE.
			Integration with Pravega.
			Use of TEEs
			Combined deployment and data management time for video analytics is reduced by 50%.



WP5 – Extreme Health Use Cases

Use-Case	New Data Connectors	KPIs	Description
Transcriptomics (SANO)	Federated Learning Dataplug connector Lithops connector	KPI-1 - Throughput Improvements	The new generated STAR index resulted in 12-times faster. The FL workflow yields a significant data transfer reduction and improved data ingestion rates compared to centralized approaches.
		KPI-3 - Resource Auto-scaling	The study of the most optimal AWS EC2 spot instances together with the index distribution solution reduces compute costs by around 50%.
		KPI-4 - Confidential computing	Integration with SCONE.
		KPI-5 - Simplicity and Productivity	Increased regulatory compliance and data security due to FL nature increases the productivity of the platform.
Genomics (UKHS)	Data loader Data merger Data partitioner Data shuffling Near-Data shuffling	KPI-1 - Throughput Improvements	Integration with Dataplug reduces data partitioning, data transfer (by 200%) and data duplication.
		KPI-3 - Resource Auto-scaling	Integration with Lithops resulted in x37.46 times faster than the HPC version. Glider integration reduces execution time by 36%.



WP5 – Extreme Health Use Cases

Use-Case	New Data Connectors	KPIs	Description
Metabolomics (EMBL)	Data merger Data partitioner Data selector Data submitter FDR estimator Feature computer Lithops connector Results downloader	KPI-1 - Throughput Improvements KPI-3 - Resource Auto-scaling	ML inference can be performed directly in METASPÂCE.
			The implemented ML-version of metabolite identification allows for resource autoscaling for datasets of the size ranging from under 1 GB to 20 GB.
		KPI-4 - Confidential computing KPI-5 - Simplicity and Productivity	Integration with SCONE.
			The ML-based metabolite identification is already available to users on the production version of METASPACE and is already used by the METASPACE users.

