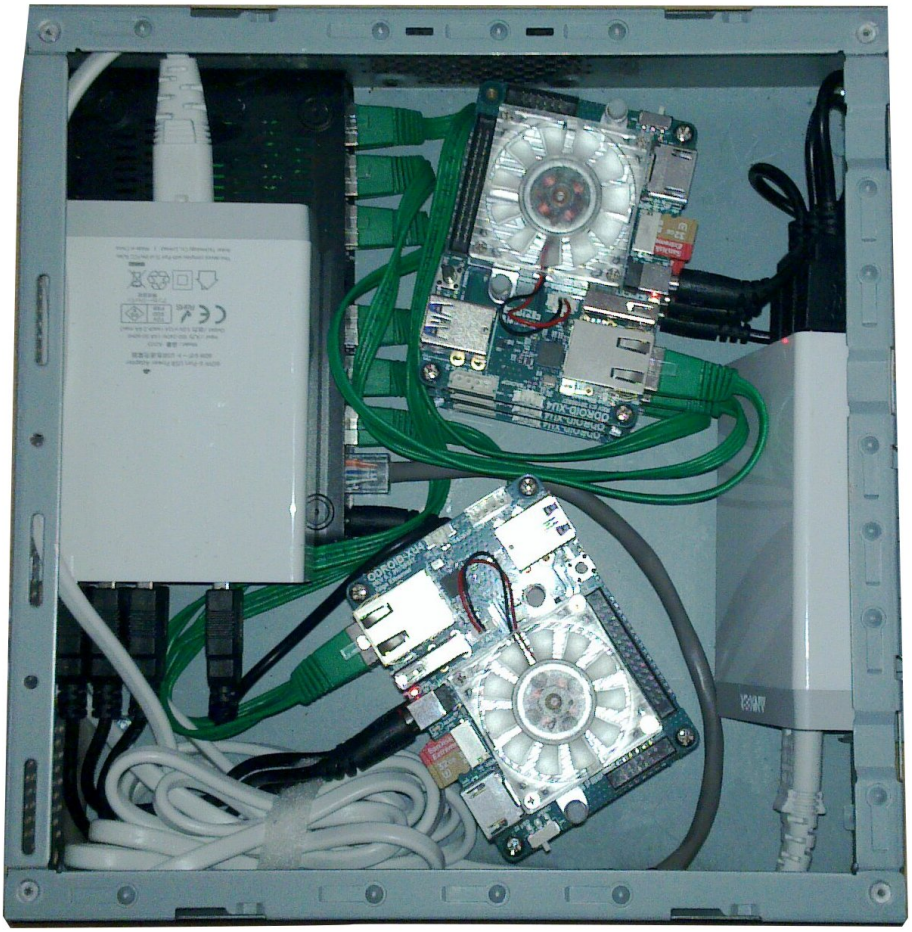
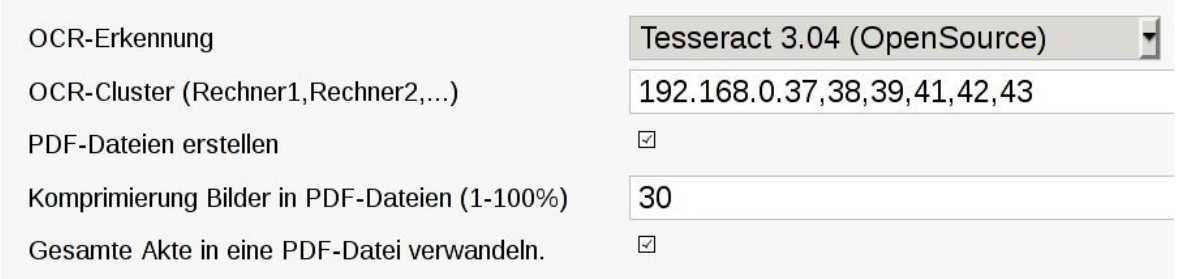
Up to 10 million pages of text recognition per day with the ArchivistaBox OCR Cluster

Egg, 20th November 2015: *the ArchivistaBox OCR Cluster (computer network) allows image files to be automatically converted into searchable PDFs or text files by means of text recognition (OCR). Thanks to its 24 to 1920 processor (CPU core) scalable cluster technology, the ArchivistaBox OCR Cluster is capable of converting between 120,000 and 10 million image files per day into searchable text files (OCR).*



The OCR Cluster uses power-saving ARM processors (CPUs). So, a **48-CPU cluster would easily fit into a three litre mITX housing and require just 75 watts of power when under load.** Such a unit is capable of processing 180 pages per minute. That is equivalent to an **output of 250,000 pages per day.** The OCR Cluster administration is web-based. The required IP addresses of the various nodes are entered prior to delivery and configurations such as language, text layout, scan profile and network drives are carried out via a web interface, pursuant to customer requirements.



An API (Application Programming Interface) with HTTP call-up is available as an optional extra for text recognition control. Text recognition can also be started and monitored at the console. Documents can be called up for processing via FTP (file upload), SMB (network drive), HTTP or HTTPS (web) and also, if connected, via a document scanner.
The Tesseract 3.0x-based text recognition system recognises **over 50 different languages, including old fonts, such as Black Letter script and Gothic.** Additional languages and/or special fonts can be integrated as and when required. Delivery of the recognised text is carried out by the integrated ArchivistaDMS (Document Management System). If necessary, searchable PDF files can be exported directly to an external drive.



The OCR Cluster is either delivered in the form of mini computers (each weighing around 100 grams), or (optionally) installed in a range of “classic” casings, including a rack-mounted configuration. OCR Clusters are priced according to the number of CPU cores they contain. A single node contains eight CPUs (processors) and corresponds to an ArchivistaBox with the required performance scope. As an example, an **OCR Cluster with 24 CPU cores and a performance rating of 120,000 pages per day is currently priced at 981.18 euros (3 x ArchivistaBox Dolder).** The nodes (ArchivistaBoxes) required for the OCR Cluster can be ordered at **shop.archivista.ch.**