Helmholtz Centre
POTSDAM

GFZ GNSS Operational Data Center

Status Quo and Future Directions

M. Bradke, T. Nischan

Contact: bradke@gfz-potsdam.de

HELMHOLTZ CENTRE POTSDAM
GFZ GERMAN RESEARCH CENTRE FOR GEOSCIENCES

Introduction

With the setup of global and regional GNSS networks and the IGS Analysis Center in the early 1990s, GFZ has been implemented an **Operational Data Center (ODC)** to handle the huge amount of data. It is the backbone for all GNSS analysis activities at GFZ.

With the steadily increasing number of networks, stations and data, the current GNSS Data Center software of the GFZ has reached the end of its life cycle. Developed for the sequential processing of RINEX v.2 data for a limited number of stations, it can hardly keep up with the daily amount of data that passes the processing line. The introduction of RINEX v.3 (Multi-GNSS with additional observation types) into operation in 2012 added to this overload.

Therefore GFZ needed to take on the tremendous task to redesign and rewrite most of the old software. The plan was to develop a system for the pre-processing, monitoring, management and quality control of GNSS data in a more sophisticated way.

The software handles all available data filetypes in the field of GNSS. This includes receiver vendor (binary) data from all major manufacturers used within the IGS (e.g. Leica, Trimble, Javad, Septentrio, Topcon, etc.) as well as RINEX v.2/v.3 observation, navigation and meteorological files. All relevant metadata from RINEX observation files are determined by using a GFZ internal quality check tool. All information is stored in a PostgreSQL database and can be accessed for a variety of applications.

Architecture, Processing Workflow and Monitoring

Statistics and requirements

Currently a total of ~ 1.100 stations and ~ 50.000 observation files from a broad range of networks (daily, hourly, high-rate, RINEX v.2, RINEX v.3) are processed at our data center each day. Due to this high amount of data and the specifications set by GFZ's "Near Real-time (NRT)" applications, the ODC software is designed to meet the requirements of an environment with high requirements regarding reliability, availability and latency, as defined in the box below:

Availability:	99.85% (Downtime max. 12 hours per year)
Data latency:	< 60 seconds after incoming
Data backup:	every 6 hours,
	2 identical "online" archives + 1 raw data tape archive
Database backup:	streaming replication

Based on the experiences with the old ODC software system(s), the main focus was on the configuration, flexibility and scalability of the software.

Architecture

The backbone of the whole system is the server-central PostgreSQL database. It contains all relevant information for the data processing chain:

- Logging of all system and thread information,
- Process information for each individual process,
- Queue information for the conditional transmission of files (e.g. send daily files when completed),
- Information about collectable/downloadable files from external data centers.

Furthermore we developed a database which contains the metadata and quality metrics of the processed RINEX files. This database is running on a separate server with replication.
The whole software is written in an object-oriented style and consists out of main packages and plugins, which can be connected via JSON configuration blocks in a customized manner. This way we are able to develop special plugins for individual tasks. The backend is written in Perl and consists out of three major monitors. Figure 1 shows the architecture of the ODC in a simplified way.

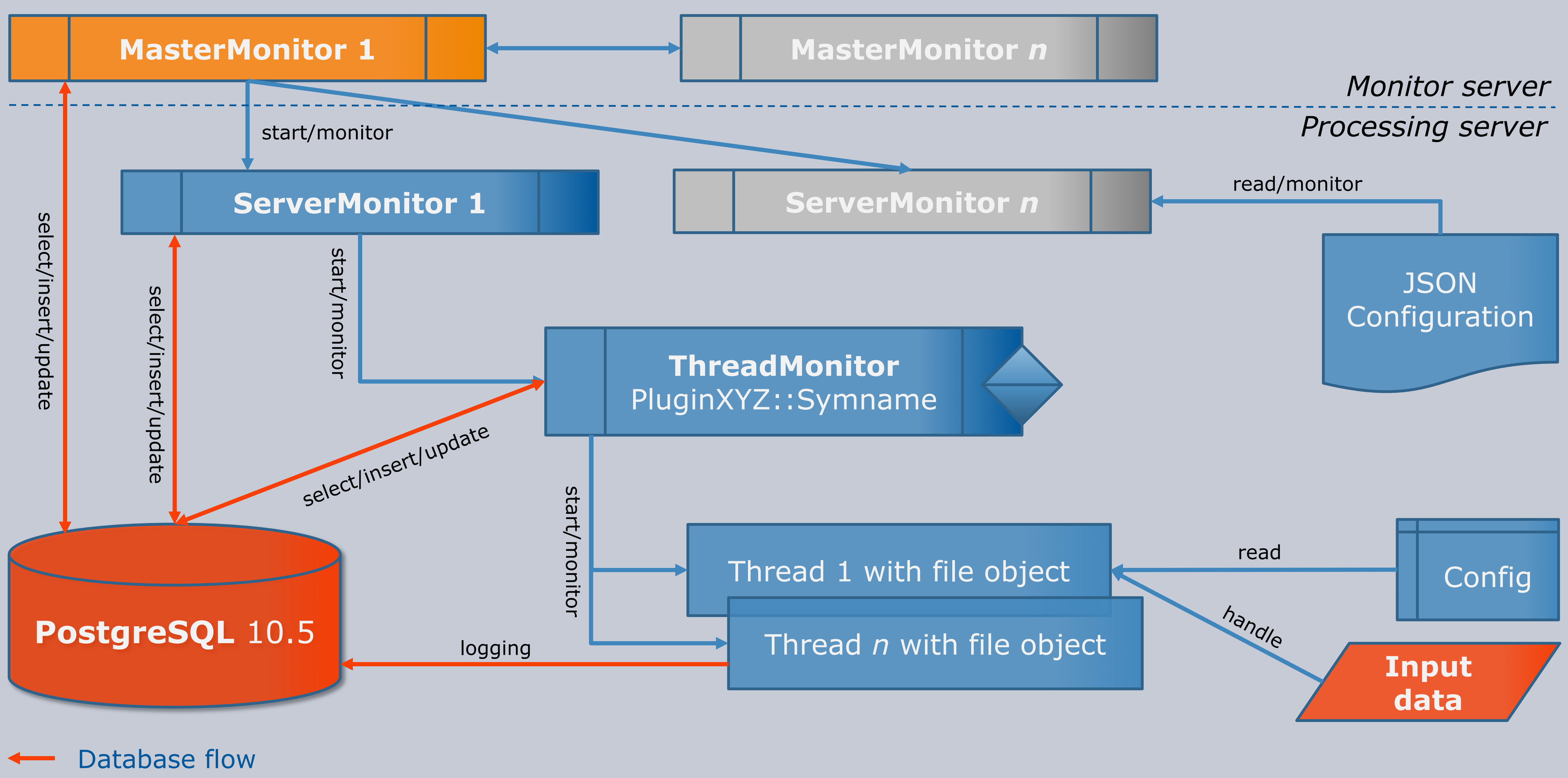

Fig. 1 Simplified view on the software architecture of the new ODC

The **MasterMonitor** is designed to monitor the system environment and processes from a remote host. It also arranges automatic failovers to backup servers. The **ServerMonitor** is mainly responsible to start and monitor the plugins and processes. It keeps track of the environment, checks the server vitality and reacts to failures in a defined way. The task of the **ThreadMonitor** is to start and monitor thread queues, create a list of sorted files (e.g. by RINEX version, descending date, etc.) and to pass each file to a threaded plugin.

Plugins and processing workflow

The whole ODC processing is based upon two major developments in the past years: **SEMISYS** (Sensor Meta Information System) and **GFZRNX**. SEMISYS is needed as input to create RINEX headers for stations who send raw data files as well as for the metadata check to external RINEX observation files. Every station is defined by a unique database ID, which is used in our internal filename convention to ensure that the data file belongs to the correct station. GFZRNX is extensively used for header editing, file checking and splice operations as well as to extract basic quality check metrics. In combination with GFZ's internal QC tool, it is possible to extract a full set of quality check metadata, that is stored in the database and gives the user the opportunity to perform a broad range of statistical analysis for stations and networks. Figure 2 describes the data processing schema from a station or data center to the GFZ archive and distribution to external data centers.

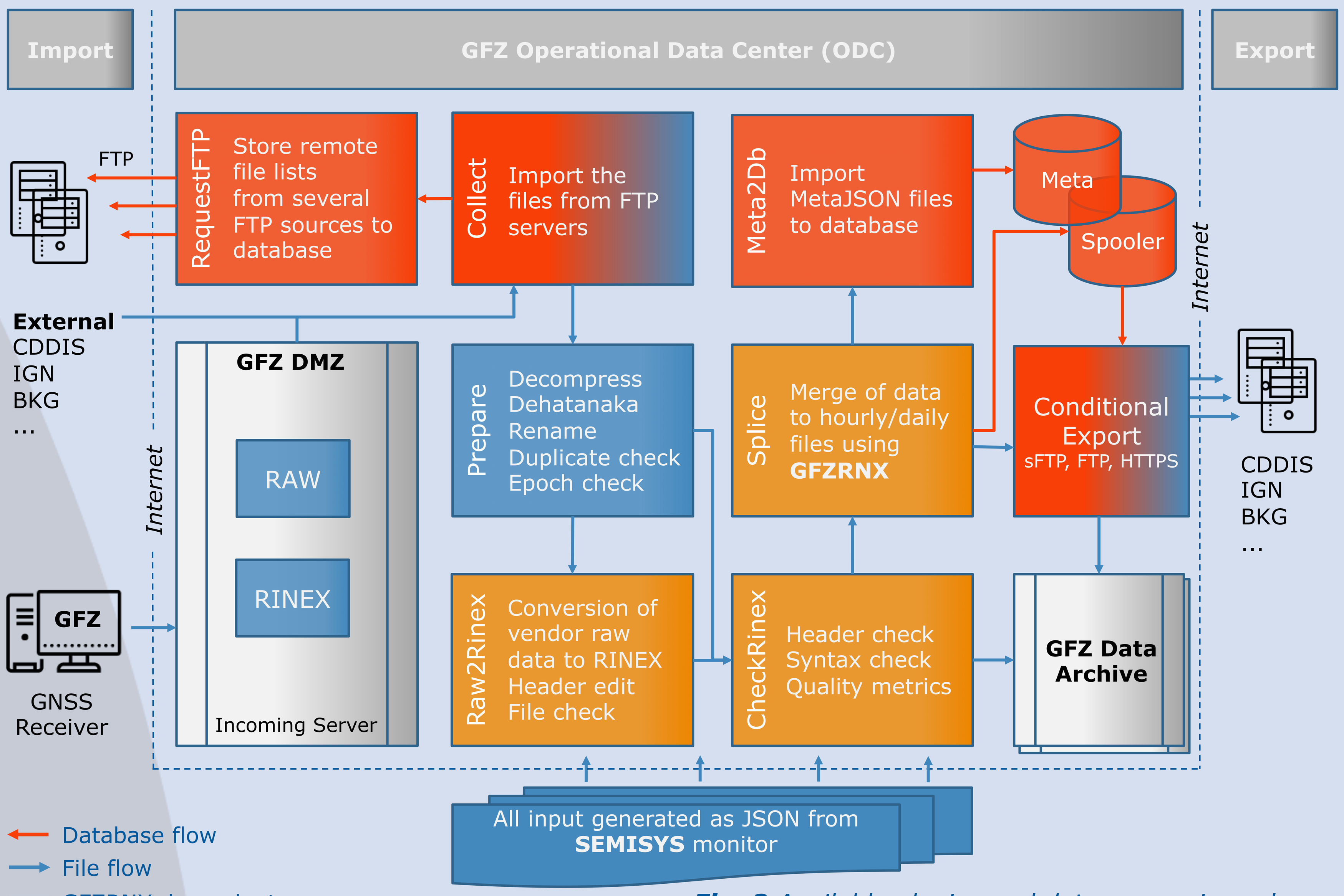

Fig. 2 Available plugins and data processing schema

System monitoring

As a temporary but effective solution, we use self-developed ASCII terminal plots to keep track of processes, file transfers and basic data availability and latency metrics. Since we store every process log entry as well as data availability and quality check information in the database, we can easily combine certain types of errors to a ticket. Based on the ticket system, we generate automatic emails in case of problems or outages. Figure 3 shows some examples of the ODC ASCII plots.

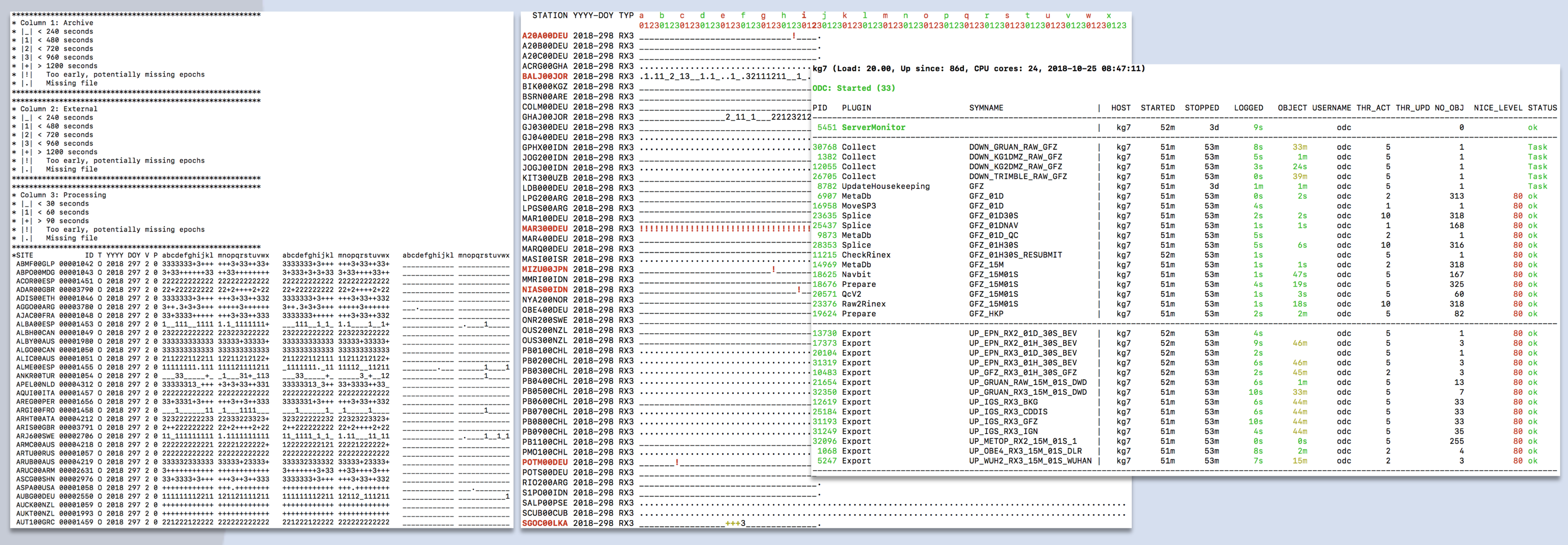

Fig. 3 Process and data monitoring in a terminal
Left: EPOS.P8 *EPOCH_INF* file, containing availability and latency information
Middle: High-rate data latency/completeness
Right: Process overview

Summary and Future Developments

GFZ has been undertaken a major effort to put the Operational Data Center (ODC) on a new foundation. The development of the backend of this system has been completed within one year and is fully operational since August 2018. GFZ's analysis software EPOS.P8 has been changed accordingly to work within this new environment. With the new system, GFZ reaches a higher level of GNSS data pre-processing, monitoring and quality control, which builds the base for high quality analysis products.

We are currently checking the compliance with the specifications and focussing on the development of the frontend to ensure a comfortable human-machine interaction. The Graphical User Interface (GUI) is designed as a web-based dashboard and covers the monitoring of processes, single stations and networks as well as a ticket system. Future directions also cover the development of GNSS web services, the integration of real-time streams into the ODC and a database to store GNSS observations.

See also

GFZRNX

SEMISYS