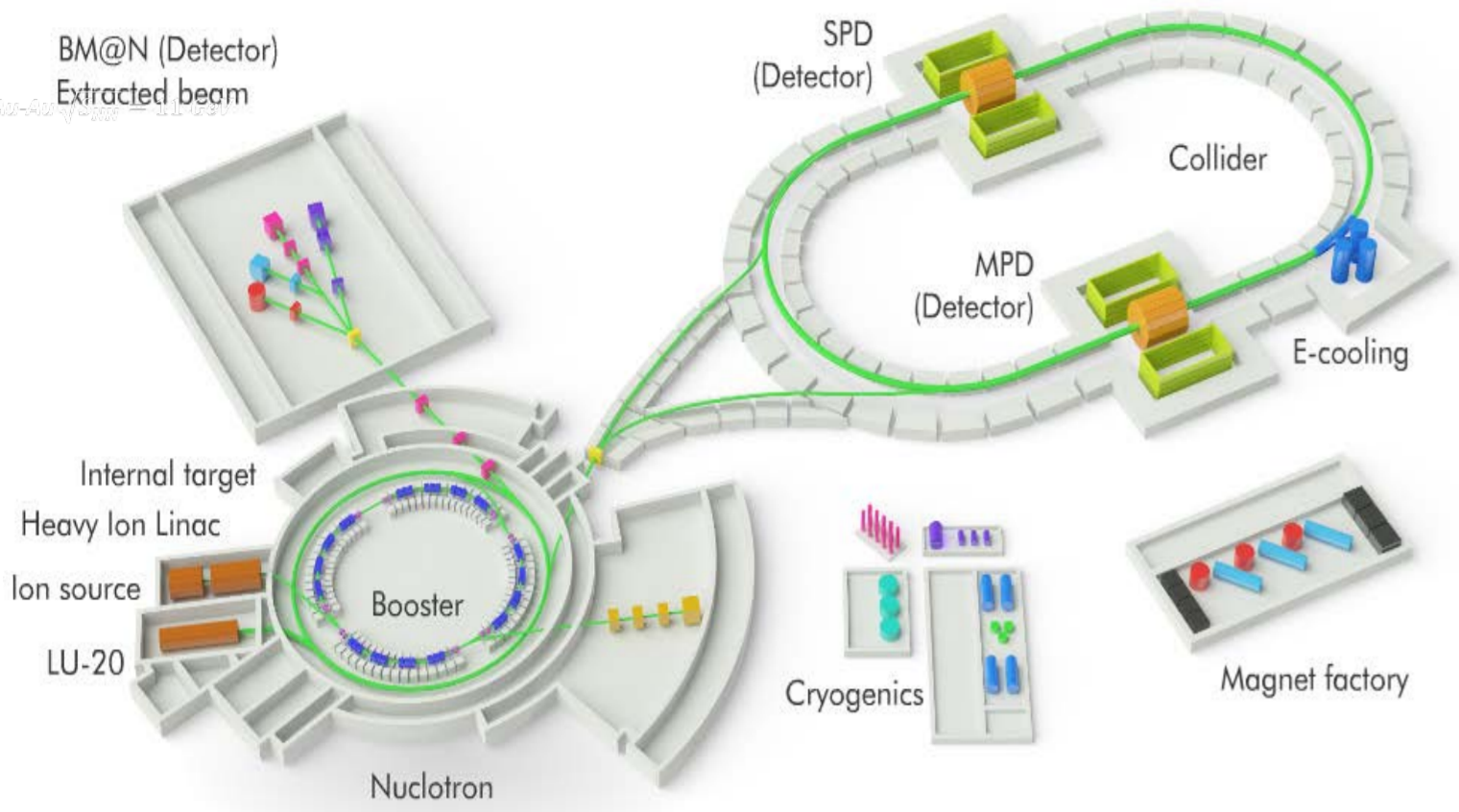# Event Metadata System for the Experiments of the NICA Complex

E. Alexandrov[1], I. Alexandrov[1], A. Chebotov[1], A. Degtyarev[2], I. Filozova[1], K. Gertsenberger[1], P. Klimai[2,3], A. Yakovlev[1]

(1) Joint Institute for Nuclear Research, Dubna   (2) Moscow Institute of Physics and Technology   (3) Institute for Nuclear Research, Moscow

Nuclotron-based Ion Collider fAcility

- Beams: from $p, d^\uparrow$ to Au^{79+}
- Luminosity: 10^{27} (Au^{79+}), 10^{32} (p) $cm^{-2}s^{-1}$
- Collision energy: $\sqrt{S_{NN_{Au}}} = 4 - 11$ Gev $E_{lab} = 1 - 6$ $AGev$
- Fixed target experiment: BM@N (2018)
- 2 interaction points: MPD (2023) & SPD (>2025)

Official site: nica.jinr.ru

Event Metadata System (EMS) for NICA

Prerequisites of the Event Metadata System:
- high interaction rate (up to 50 kHz for BM@N),
- NICA large data stream (expected about 10-20 PB per year),
- particular physics analysis requires only small part of all the events distributed over a large set of obtained files,
- event indexing and event tag systems help to accelerate achieving physics results in large experiments on particle collisions.

Goals of the EMS for the NICA experiments:
- Index all reconstructed events stored in ROOT DST files.
- Store necessary event metadata, e.g., number of primary and all tracks, number of +/- charged particles, primary and secondary particles found, number of hits by detectors, total input and output charge in the event, software version, together with reference to the event.
- Search and select only a set of events required for a particular physics analysis to speed up processing and reduce the load of the computing infrastructure.
- Convenient access to metadata for users and systems (Web UI, REST API, C++)
- Flexibly tune per experiment (BM@N, BM@N SRC, MPD, SPD)

Requirements

EMS requirements:
- Scalability. BM@N has recorded about 0.5B experimental events, and billions of events per year are expected for the NICA experiments.
- Performance. There are not too many requests per seconds, but heavy ones.
- Availability and fail safety.
- Role-based access control (LDAP or database)
  - Event Consumer
  - EMD Writer
  - EMD Administrator
- Integration with other experiment systems
  - Run metadata are stored in the Condition DB.
  - CERN ROOT and FairRoot-based frameworks (BmnRoot, MpdRoot, SPDroot).

EMS Architecture

Main functions are summary description of collision events and their identifiers to select events for a desired analysis goal; recording and storing necessary event metadata, their management and convenient access; organization of online and offline interfaces for selecting events of interest.

It is also responsible for providing statistics, as wells as, creating, maintaining and checking the quality of the catalogue of the physics events.


EMS Architecture
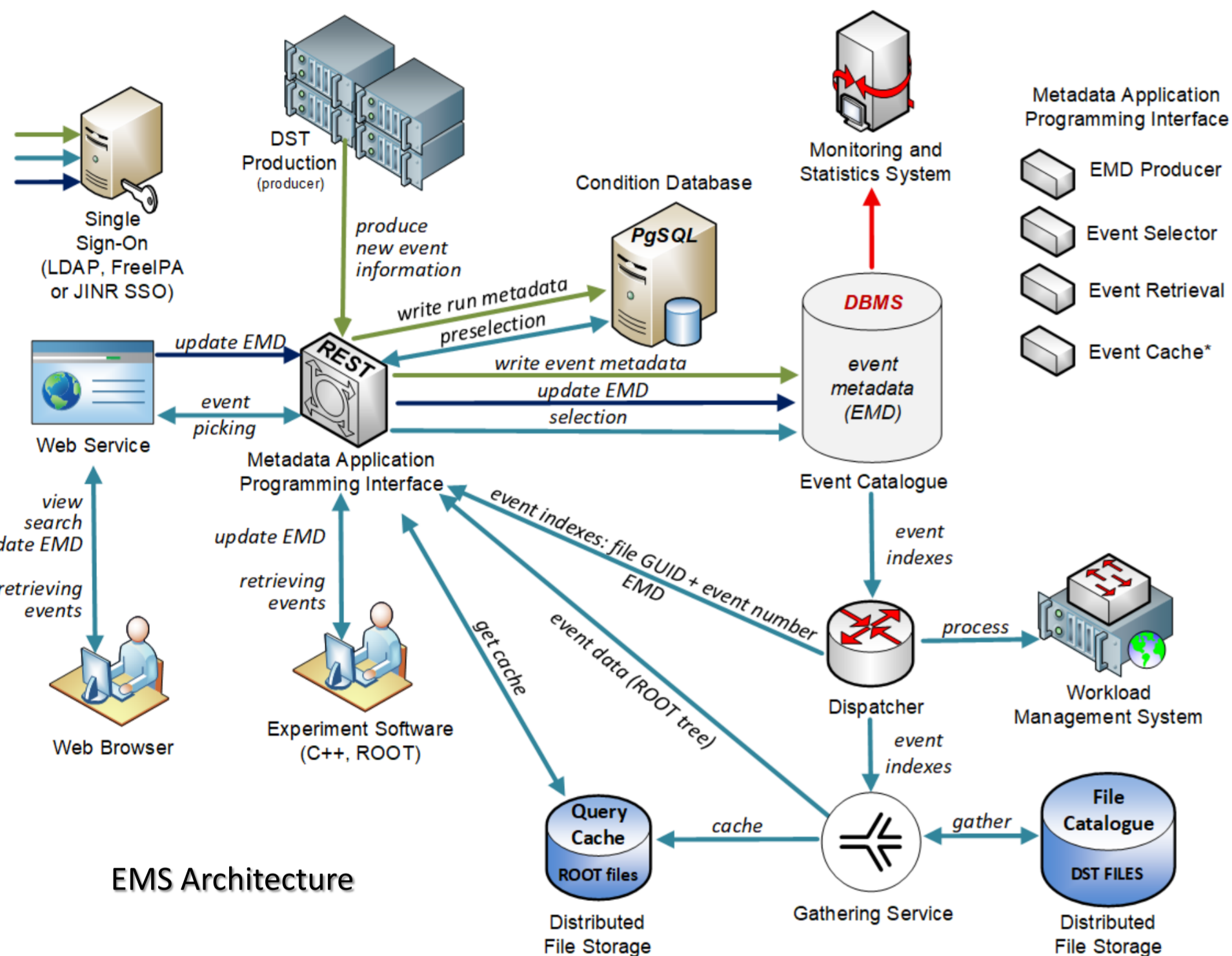
Event Catalogue as a central component

The EMS is based on the Event Database called Event Catalogue, which contains summary properties of collision events, allowing user to quickly search for a required set of events by various criteria.

The Catalogue identifies stored events with a file pointer defining a reference to the data file and event number in the file, which uniquely match every event.

Four prototypes (SQL and NoSQL) of the Event Catalogue have been deployed with PostgreSQL, Hadoop HBase and Apache Cassandra DBMS to choose the optimal solution (up to 3B events).
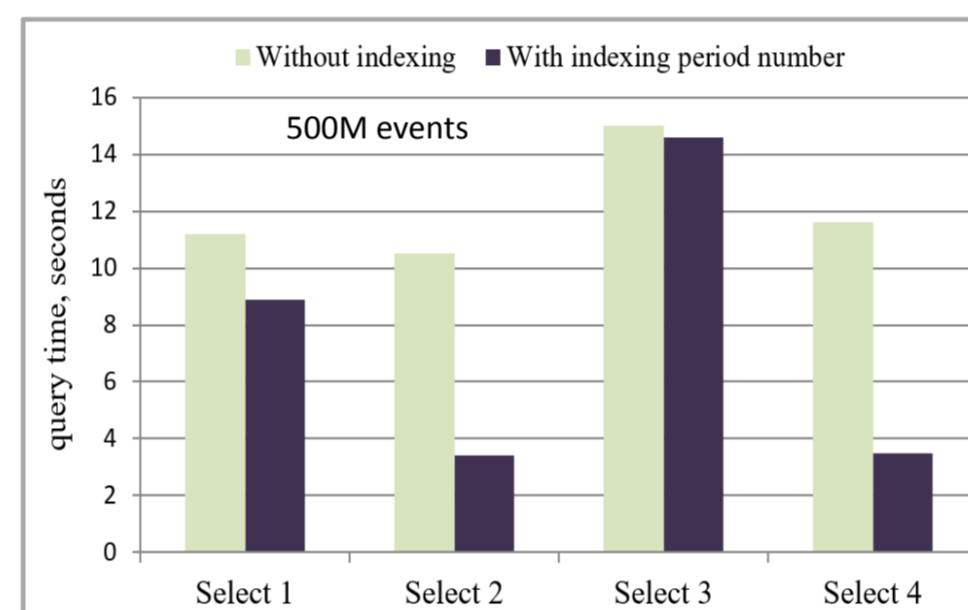
Performance tests for the PostgreSQL prototype of the Event Catalogue

The PostgreSQL database management system has been chosen to implement the Event Catalogue [1].

In particular, Apache Cassandra showed performance comparable to PostgreSQL but has complex configuration and non-universality for all queries.

The metadata structure of the Event Catalogue for the BM@N experiment has been approved by the collaboration, where the event metadata are written only if primary vertex is found in the event.

configurable metadata
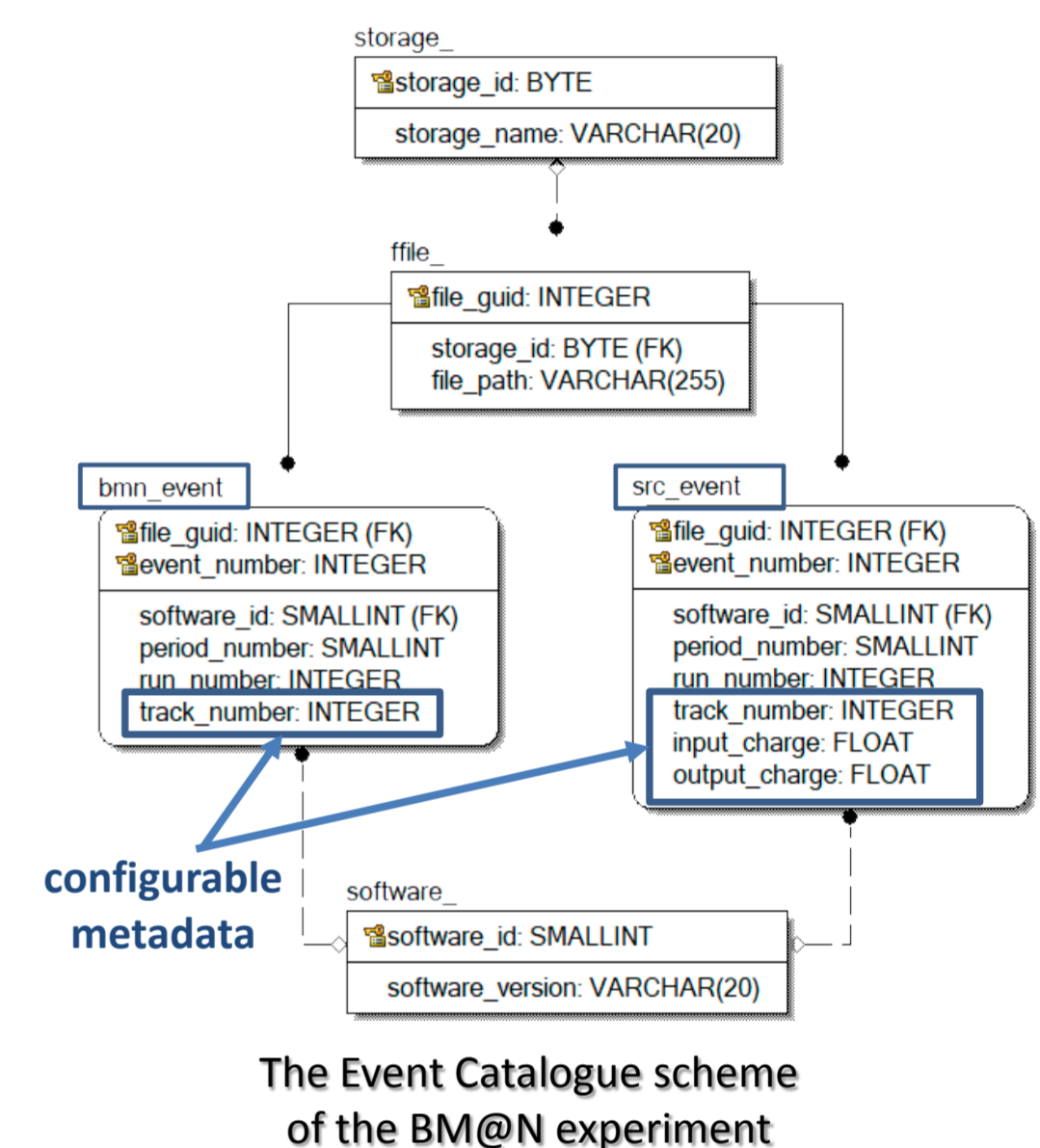
The Event Catalogue scheme of the BM@N experiment

A specific configurable set of event metadata attributes is defined at the deployment step of the Event Metadata System for the NICA experiments using a configuration file.

Implemented Interfaces and Services for EMS

Web Service (User Web Interface)
- Enables users to browse and search for event metadata stored in the Event Catalogue and retrieve events, which satisfy given parameters
- Provides events according to the selected event metadata and run metadata of the Condition Database, such as period and run numbers, beam particles and its energy, trigger info, magnetic field, etc.

Web Service deployed for BM@N

event metadata in JSON format

REST API service (Metadata Application Programming Interface)
- Ensures writing new metadata to the Event Catalogue during data processing and requesting events by other experiment systems for chosen criteria, e.g. for physics analysis in the ROOT-based frameworks
- Provides HTTP-based API using JSON formatting: *HTTP POST* command to create event metadata in the catalogue, *HTTP GET* request to obtain event records by criteria, *HTTP DELETE* to delete event metadata
- FreeIPA/LDAP protocol is supported for authentication (*admin*, *writer* and *consumer* groups)
- Uses the same selection criteria as in the web service including range support, for instance:
GET /emd?period_number=7&run_number=5000+&software_version=20.08.0&track_number=10-15

GET /eventFile[?parameters]   *GET /eventFileRef[?parameters]*

C++/ROOT Interface (additional interface for the ROOT-based frameworks)
- A special C++ interface class (*EventCatalogue*) has been developed to view, add, modify event metadata and to request events according to given parameters.

Deployment

- The Event Metadata System has been developed as a universal configurable solution that can be deployed for any NICA experiment.
- The configuration for the experiment is defined at the deployment step via YAML file, which specifies database connection attributes (both for the Event Catalogue and Condition Database), parameters of the Web and REST API services, and custom event metadata fields.
- The common configuration and deployment system for the EMS is under implementation, which uses Docker containers and shell scripts.
- The EMS has been already deployed for the first NICA experiment, BM@N.

A part of the BM@N configuration file for the EMS deployment system

```
event_db:   # condition_db - configured
  host: ***          # similarly, if used
  port: ***
  db_name: ***
  user: ***
  password: ***
user_auth:
  ldap_server: bmn-ipa.jinr.ru
  ldap_port: 389
  user_dn_format: "uid=%s,cn=users,…,dc=ru"
  ldap_username: *****
  ldap_password: *****
  writer_group_dn: "cn=bmneventwriter,…,dc=ru"
  admin_group_dn: "cn=bmneventadmin,…,dc=ru"
title: "Event Index Main Page"
pages:
  - name: "BM@N Events"
    api_url: "/event_api/v1/bmn"
    web_url: "/event_web/bmn"
    db_table_name: "bmn_event" …
  …
```

Conclusions

- The Event Metadata System based on the Event Catalogue has been developed to index events of the NICA experiments and store their metadata.
- The event information system managing physics event metadata makes it possible to quickly search by required parameters used in various physics analyses for a set of physics events to use in further event data processing.
- The EMS provides Web and REST API services, and a dedicated C++/ROOT interface to be used by collaboration members and other experiment systems.
- The access based on FreeIPA/LDAP or built-in database authentication has been implemented.
- The Event Catalogue on PostgreSQL provides acceptable response times for the expected amount of event data.
- The first version of the Event Metadata System has been deployed for the BM@N experiment.

References

1. Alexandrov E., Alexandrov I., Degtyarev A., et.al. Design of the Event Metadata System for the Experiments at NICA // Physics of Particles and Nuclei Letters. 2021. Vol. 18. pp. 603-616.

Contact Information

If you have any questions, please email:
pklimai@gmail.com