

Blue Brain Nexus Demo

2022-01-11

Agenda

Overview of Blue Brain Nexus platform

Data transformation to achieve consistent data representation

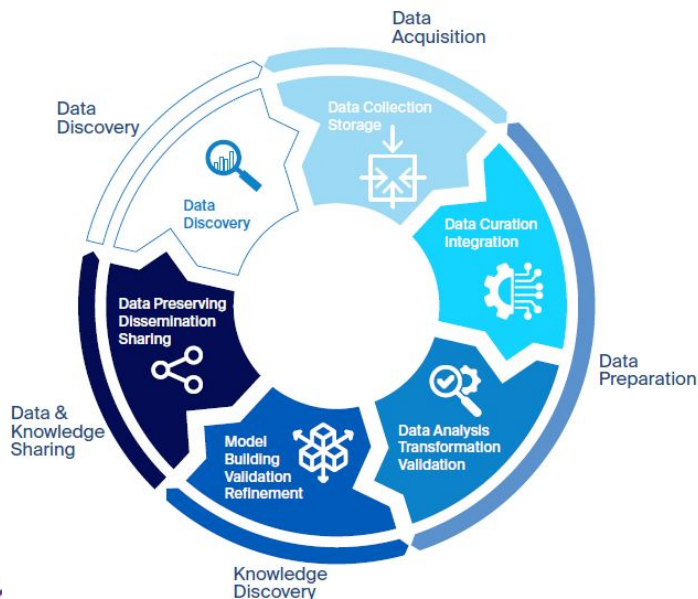
Data sharing

Querying data

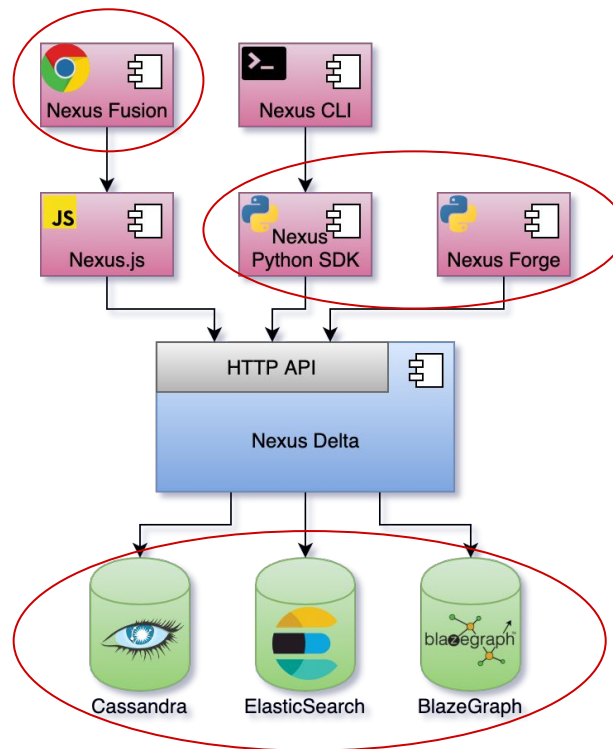
Overview of Platform

User capabilities provided by BBN products

Blue Brain Nexus (BBN) is an ecosystem containing the components on the right. BBN helps us support analysis for individual patients as well as population-level (aggregate) analysis.



High-level BBN Architecture



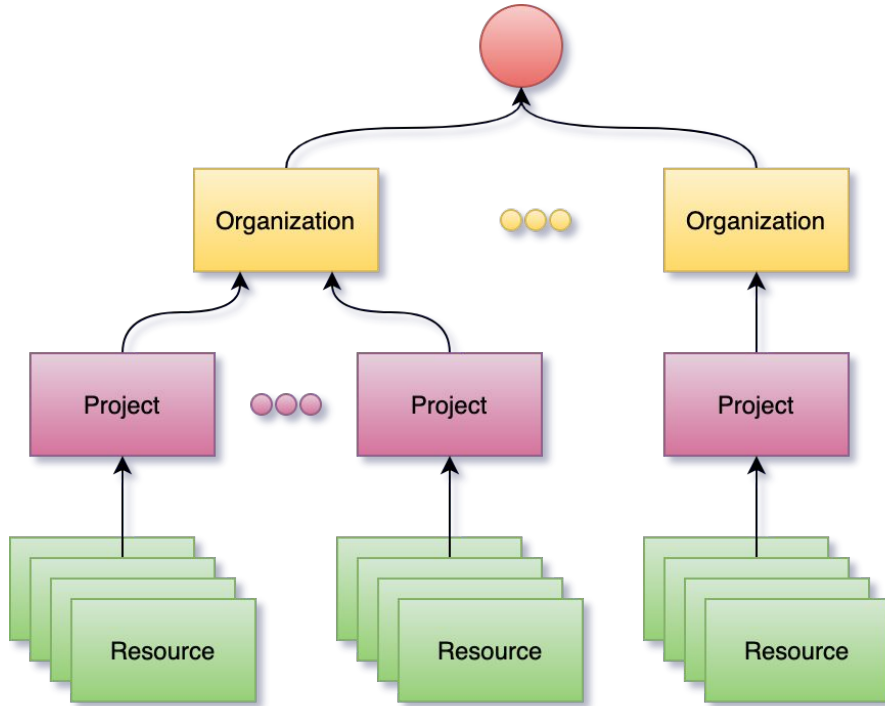
Web interface for users to interact with the Knowledge Graph

Python SDK's and libraries to make it easier to interact with Blue Brain Nexus

Nexus Delta is a scalable and secure service to store the data. Apache Cassandra acts as the primary event store. Data is then automatically indexed into default and custom Elasticsearch and BlazeGraph views.

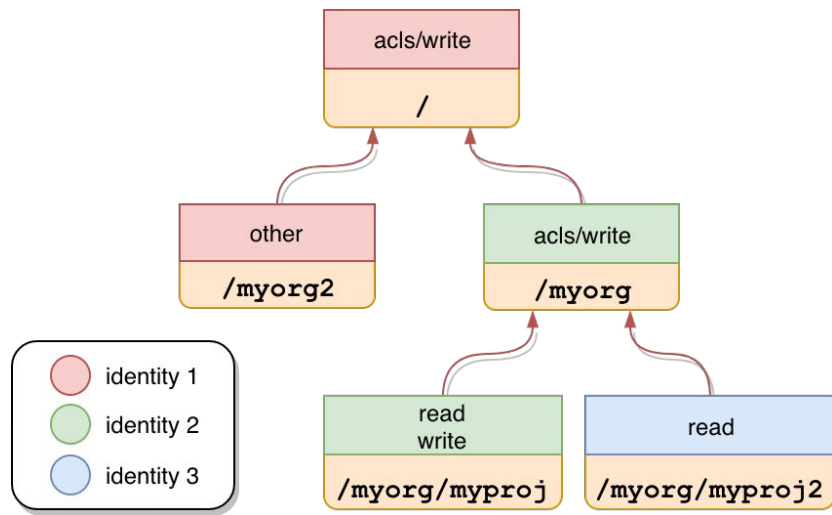
ETL Process - Consistent Data Representation

Project Structure



User data is represented as sub-resources to projects which in turn are sub-resources of organizations. Organization and project resources provide logical grouping and isolation allowing for variation in configuration and access control policies.

Access Control Lists



An ACL defines the applications' data access restriction using the following three parameters:

- permission: the value used to limit a client (user, group) access to resources.
- identity: a client identity reference, e.g. a certain user, a group, an anonymous user or someone who is authenticated to a certain realm.
- path: the location where to apply the restrictions. Examples of paths are: /, /myorg or /myorg/myproject

PUT

{{deployment}}/acls/{{org}}/{{project}}?rev=1

Params

Auth

Headers (10)

Body

Pre-req.

Tests

Settings

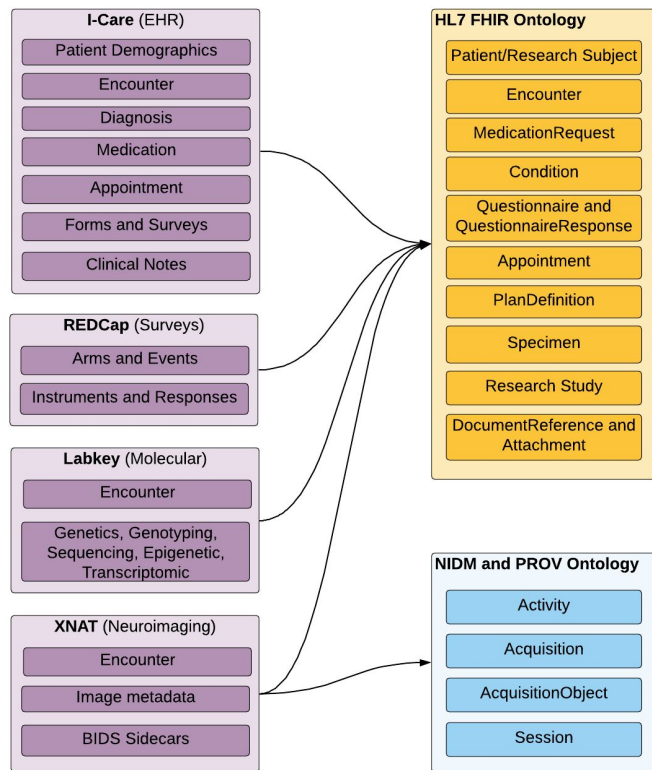
raw

JSON

Beautify

```
1 {
2   .... "acl": [
3     .... {
4       .... "identity": {
5         .... "subject": "service-account-nexus-sa",
6         .... "realm": "serviceaccounts"
7       },
8       .... "permissions": [
9         .... "schemas/write",
10        .... "views/write",
11        .... "files/write",
12        .... "permissions/write",
13        .... "acls/write",
14        .... "realms/write",
15        .... "projects/read",
16        .... "acls/read",
17        .... "organizations/create",
18        .... "organizations/write",
19        .... "resources/write",
20        .... "realms/read",
21        .... "projects/create",
22        .... "permissions/read",
23        .... "resources/read",
24        .... "organizations/read",
25        .... "resolvers/write",
26        .... "events/read",
27        .... "views/query",
28        .... "projects/write"
29      ]
30     },
31   ]
32 }
```

Data Modeling - ontology selection and mapping



Criteria:

- Prominent ontologies with good coverage for data sources
- Well maintained by an established team
- Roadmap aligned with our plans
- Easy to understand documentation
- Aligned with internal understanding of data

Ontology Links:

- [FHIR](#)
- [NIDM](#)
- [BIDS](#)
- [PROV-O](#)
-

Modeling notes:

- Knowledge graph contains metadata about files and points to file location (images, attachments, notes, etc.)

Data Model Example - FHIR Patient Representation

8.1.2 Resource Content

| Structure | UML | XML | JSON | Turtle | R3 Diff | All |
|------------------------|----------|-------|-----------------|---|---------|-----|
| Structure | | | | | | |
| Name | Flags | Card. | Type | Description & Constraints | | |
| Ⓜ Patient | N | | DomainResource | Information about an individual or an Elements defined in Ancestors: id, me | | |
| 📁 identifier | Σ | 0..* | Identifier | An identifier for this patient | | |
| 📁 active | ?! Σ | 0..1 | boolean | Whether this patient's record is in act | | |
| 📁 name | Σ | 0..* | HumanName | A name associated with the patient | | |
| 📁 telecom | Σ | 0..* | ContactPoint | A contact detail for the individual | | |
| 📁 gender | Σ | 0..1 | code | male female other unknown AdministrativeGender (Required) | | |
| 📁 birthDate | Σ | 0..1 | date | The date of birth for the individual | | |
| 🔍 deceased[x] | ?! Σ | 0..1 | | Indicates if the individual is deceased | | |
| 📁 deceasedBoolean | | | boolean | | | |
| 📁 deceasedDateTime | | | dateTime | | | |
| 📁 address | Σ | 0..* | Address | An address for the individual | | |
| 📁 maritalStatus | | 0..1 | CodeableConcept | Marital (civil) status of a patient MaritalStatus (Extensible) | | |
| 🔍 multipleBirth[x] | | 0..1 | | Whether patient is part of a multiple | | |
| 📁 multipleBirthBoolean | | | boolean | | | |
| 📁 multipleBirthInteger | | | integer | | | |
| 📁 photo | | 0..* | Attachment | Image of the patient | | |
| 📁 contact | I | 0..* | BackboneElement | A contact party (e.g. guardian, partne + Rule: SHALL at least contain a cont | | |
| 📁 relationship | | 0..* | CodeableConcept | The kind of relationship Patient Contact Relationship (Extensib | | |
| 📁 name | | 0..1 | HumanName | A name associated with the contact p | | |

Link: <https://www.hl7.org/fhir/patient.html>

Well maintained specification for describing the common entities in healthcare.

Entities can also be extended and new entities can be described and shared through: <https://simplifier.net/>

Benefits of a graph model is that the logical model and physical model are the same. Therefore, the technical team, healthcare professionals, and business teams can communicate using a consistent vocabulary and understanding of subject domains.

ETL - Patient data transformation

Currently, there is no library that automatically converts Cerner EHR data to FHIR JSON-LD representation. Therefore, transformation process is manual.

| EHR Field Name | EHR Value | FHIR Patient |
|----------------|-----------|-------------------------|
| PatientID | 123456789 | Patient.identifier |
| Gender | Female | Patient.gender |
| PostalCode | A1A1A1 | Patient.contact.address |



```
{
  "@context": "https://nexus-clinical.camh.ca/v1/resources/camh_clinical_prod/shared_resources/_/project_context",
  "@type": [
    "fhir:Patient",
    "prov:Person"
  ],
  "fhir:Patient.address": {
    "@type": "fhir:Address",
    "fhir:Address.postalCode": [
      {
        "@type": "fhir:string",
        "fhir:value": "A1A1A1"
      }
    ]
  },
  "fhir:Patient.gender": [
    {
      "@type": "fhir:code",
      "fhir:value": "Female"
    }
  ],
  "fhir:Patient.identifier": [
    {
      "@type": "fhir:Identifier",
      "fhir:Identifier.type": {
        "@type": "fhir:CodeableConcept",
        "fhir:CodeableConcept.text": {
          "fhir:value": "PatientID"
        }
      },
      "fhir:Identifier.value": {
        "fhir:value": 123456789
      }
    }
  ]
}
```

ETL Demo - Loading data to Blue Brain Nexus

Adherence to FAIR data principles.

Resource assigned a UUID

Vocabulary maintained and link can be followed to documentation

Provenance information automatically generated

```

1 {
2   "id": "https://reservoir.global/v1/resources/Test/adeel/_/14efb689-7d54-4ec8-a571-d8ee19d2a8aa",
3   "type": "http://hl7.org/fhir/Patient",
4   "http://bluebrain.github.io/nexus/vocabulary/constrainedBy": {
5     "id": "https://bluebrain.github.io/nexus/schemas/unconstrainedR1.json"
6   },
7   "http://hl7.org/fhir/Patient.address": {
8     "@type": "http://hl7.org/fhir/Address",
9     "http://hl7.org/fhir/Address.postalCode": {
10       "@type": "http://hl7.org/fhir/string",
11       "http://hl7.org/fhir/value": "A1A1A1"
12     }
13   },
14   "http://hl7.org/fhir/Patient.birthDate": {
15     "@type": "http://hl7.org/fhir/date",
16     "http://hl7.org/fhir/value": "1988"
17   },
18   "http://hl7.org/fhir/Patient.gender": {
19     "@type": "http://hl7.org/fhir/code",
20     "http://hl7.org/fhir/value": "Female"
21   },
22   "http://hl7.org/fhir/Patient.identifier": {
23     "@type": "http://hl7.org/fhir/Identifier",
24     "http://hl7.org/fhir/Identifier.type": {
25       "@type": "http://hl7.org/fhir/CodeableConcept",
26       "http://hl7.org/fhir/CodeableConcept.text": {
27         "http://hl7.org/fhir/value": "PatientID"
28       }
29     },
30     "http://hl7.org/fhir/Identifier.value": {
31       "http://hl7.org/fhir/value": "123456789"
32     }
33   },
34   "https://bluebrain.github.io/nexus/vocabulary/constrainedBy": {
35     "id": "https://bluebrain.github.io/nexus/schemas/unconstrainedR1.json"
36   },
37   "https://bluebrain.github.io/nexus/vocabulary/createdAt": "2021-07-14T17:23:49.986219Z",
38   "https://bluebrain.github.io/nexus/vocabulary/createdBy": "https://reservoir.global/v1/realms/orcid/users/0000-0001-6173-9384",
39   "https://bluebrain.github.io/nexus/vocabulary/deprecated": false,
40   "https://bluebrain.github.io/nexus/vocabulary/incoming": "https://reservoir.global/v1/resources/Test/adeel/_/14efb689-7d54-4ec8-a571-d8ee19d2a8aa/incoming",
41   "https://bluebrain.github.io/nexus/vocabulary/outgoing": "https://reservoir.global/v1/resources/Test/adeel/_/14efb689-7d54-4ec8-a571-d8ee19d2a8aa/outgoing",
42   "https://bluebrain.github.io/nexus/vocabulary/project": {
43     "id": "https://reservoir.global/v1/projects/Test/adeel"
44   },
45   "https://bluebrain.github.io/nexus/vocabulary/rev": 1,
46   "https://bluebrain.github.io/nexus/vocabulary/self": "https://reservoir.global/v1/resources/Test/adeel/_/14efb689-7d54-4ec8-a571-d8ee19d2a8aa",
47   "https://bluebrain.github.io/nexus/vocabulary/updatedAt": "2021-07-14T17:23:49.986219Z",
48   "https://bluebrain.github.io/nexus/vocabulary/updatedBy": "https://reservoir.global/v1/realms/orcid/users/0000-0001-6173-9384"
49 }
```

en.wikipedia.org/wiki/FAIR_data

Findable [edit]

The first step in (re)using data is to find them. **Metadata** and data should be easy to find for both humans and computers. **Machine-readable** metadata are essential for automatic **discovery** of datasets and services, so this is an essential component of the FAIRification process.

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource^[2]

Accessible [edit]

Once the user finds the required data, they need to know how they can be accessed, possibly including **authentication** and **authorisation**.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available^[2]

Interoperable [edit]

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for **analysis**, **storage**, and **processing**.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use **vocabularies** that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data^[2]

Reusable [edit]

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

The principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure. For instance, principle F4 defines that both metadata and data are registered or indexed in a searchable resource (the infrastructure component).^[2]

Validation for FHIR data

| Constraint type | Example |
|-----------------|--|
| Value type | Patient address is of type fhir:Address |
| Cardinality | Patient name has exactly one first name |
| Value range | Age is between 0-120 |
| String-based | Patient identifier is exactly nine characters long |
| Property-pair | Patient birthDate <= deceasedDate |
| Logical | Patient first name or given name have a value. |
| Shape-based | Patient address meets all the constraints defined for type fhir:Address. |

Typically during data integration, validation rules are embedded in the code.

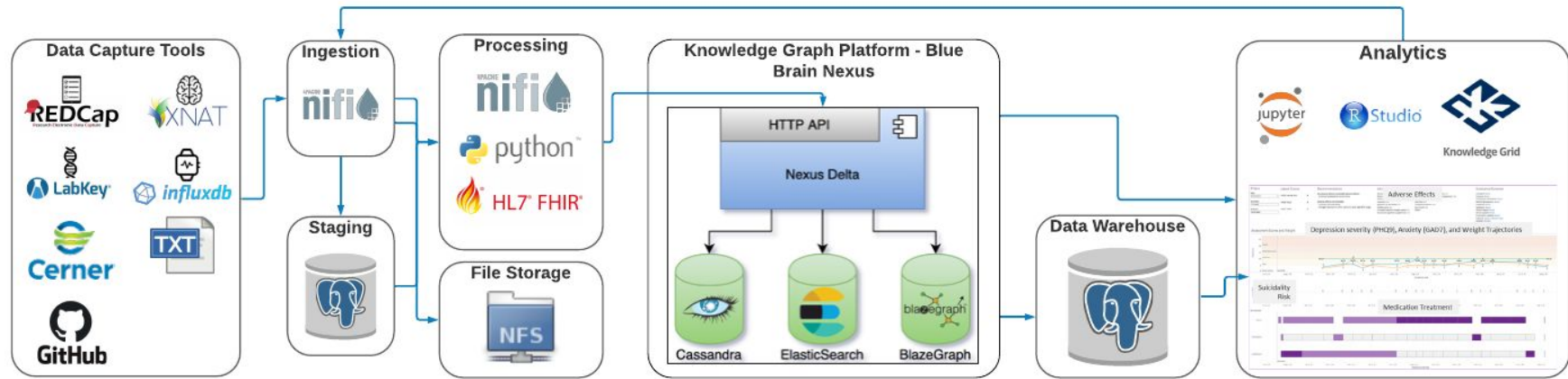
In Nexus, we can define resources that can be used to validate that resources meet a set of data quality conditions.

The benefit of defining the resources in Nexus is that the rules can be modified by users without modifying the backend code.

The table on the left describes the rules that can be defined in Nexus.

```
ex:PatientShape
  a sh:NodeShape ;
  sh:targetClass fhir:Patient ;
  sh:property [
    sh:path fhir:firstName ;
    sh:maxCount 1 ;
    sh:datatype xsd:string ;
  ] .
```

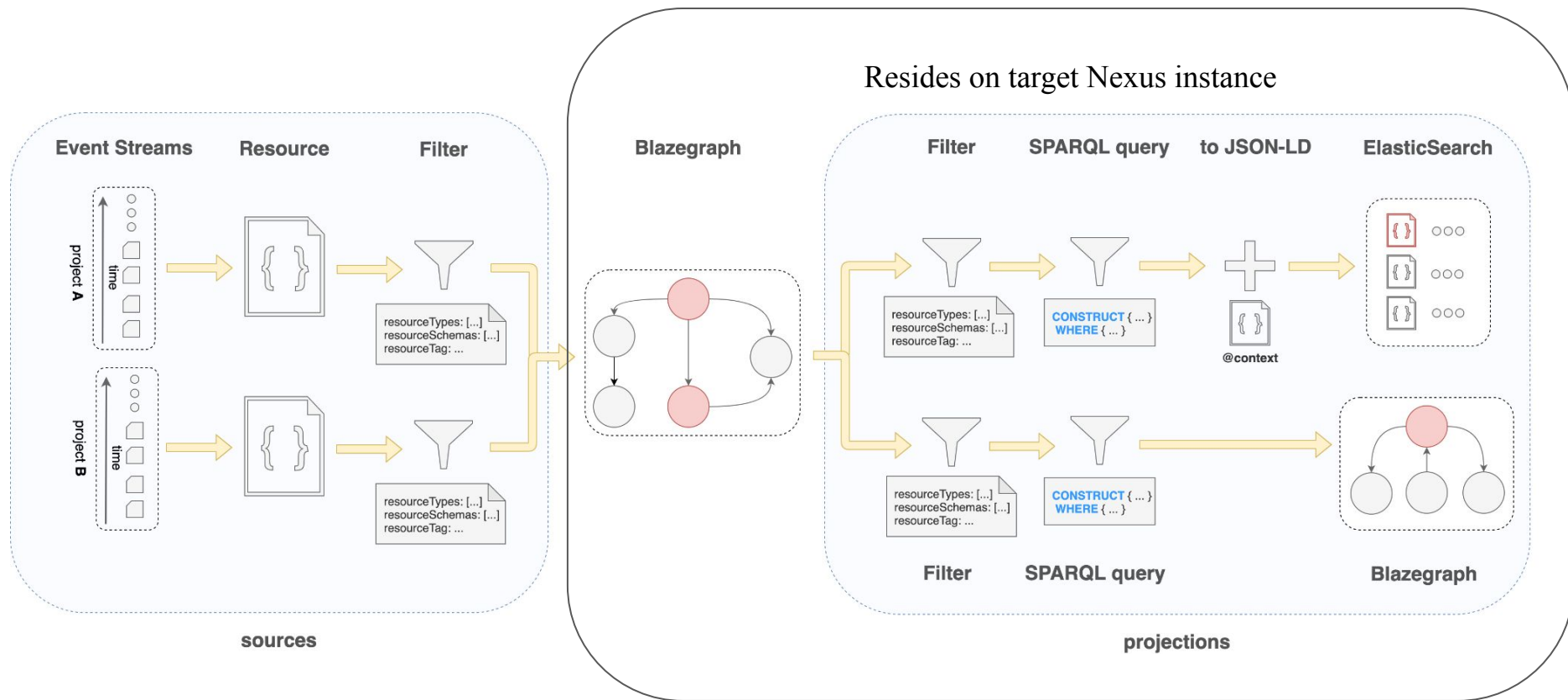
ETL - Pipeline Automation



1. Pull data from source systems using Apache NiFi and load as-is to staging
2. Process raw data from staging using Apache NiFi and Python and load to Blue Brain Nexus (Knowledge graph)
3. Create additional BlazeGraph and Elasticsearch indexes as required
4. Create projections from Elasticsearch to Postgres as required for data warehouse support
5. Make data available to be accessed through Analytics tools and services
6. Write back analytics results into the Knowledge Graph for continuous learning in the Graph.

Data Sharing

BBN Composite Views - Data Flow



BBN Composite View - Sources

| Source Name | Description |
|--------------------------|--|
| ProjectEventStream | Data from the same project is used as a source for the view. |
| CrossProjectEventStream | Data from another project in the same organization is used as a source for the view. |
| RemoteProjectEventStream | Data from another project in a remote Nexus instance is used as a source for the view. |

Authorization for Composite Views

For RemoteProjectEventStream, the view requires a token to be passed as a parameter.

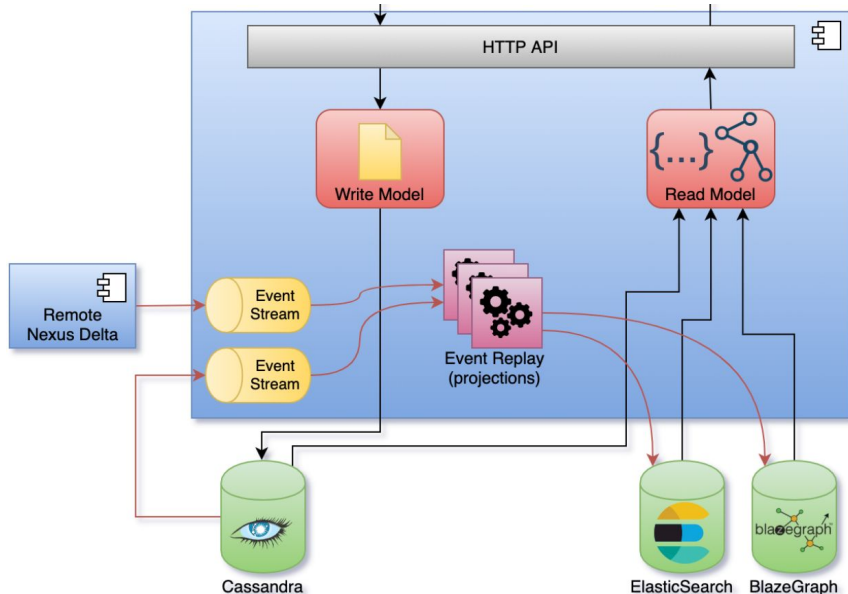
```
{
  "sources": [
    {
      "@id": "{sourceId}",
      "@type":
"RemoteProjectEventStream",
      "project": "{project}",
      "endpoint": "{endpoint}",
      "token": "{token}",
      "resourceSchemas": [
"{resourceSchema}", ...],
      "resourceTypes": [
"{resourceType}", ...],
      "resourceTag": "{tag}"
    }
  ],
  ...
}
```

where...

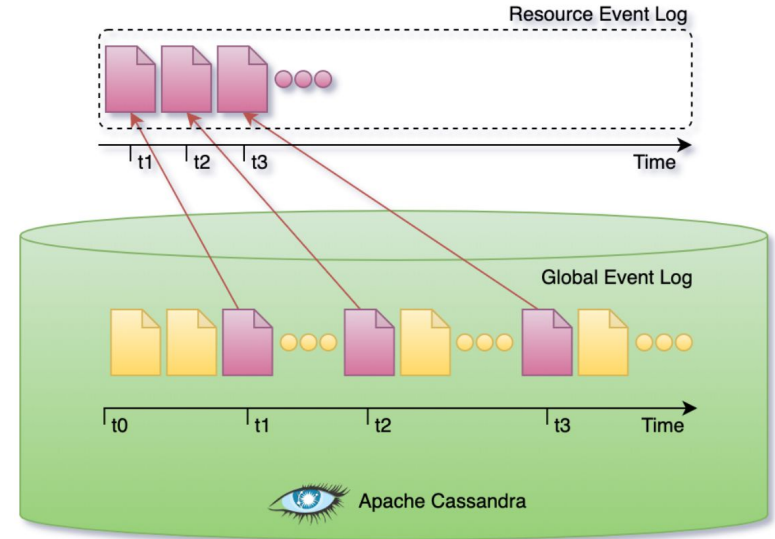
- {sourceId}: Iri - The identifier of the source. This field is optional. When missing, a randomly generated Iri will be assigned.
- {project}: String - the remote project (in the format 'myorg/myproject').
- {endpoint}: Iri - the Nexus deployment endpoint.
- **{token}: String - the Nexus deployment token. This field is optional. When missing, the Nexus endpoint will be accessed without authentication.**
- {resourceSchema}: Iri - Selects only resources that are validated against the provided schema Iri. This field is optional.
- {resourceType}: Iri - Select only resources of the provided type Iri. This field is optional.
- {tag}: String - Selects only resources with the provided tag. This field is optional.

BBN Event Stream

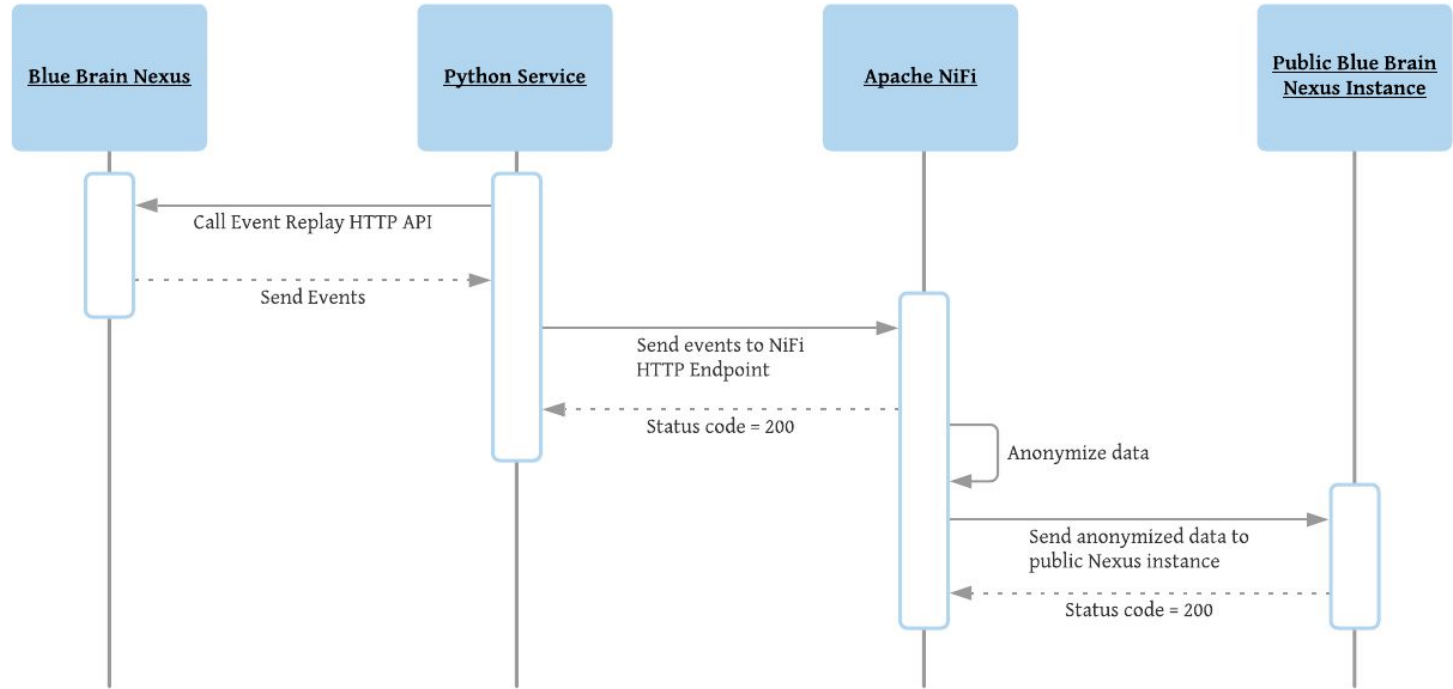
Nexus records all operations performed on resources in Apache Cassandra as events. This image shows how the events are used to construct the views in Elasticsearch and BlazeGraph. The events can also be accessed by HTTP API to push the events to another data store.



This image describes how all the resource events are stored in a single Global Event Log in Apache Cassandra. Therefore, when you access the HTTP API, you will receive events for all resources on the HTTP path.



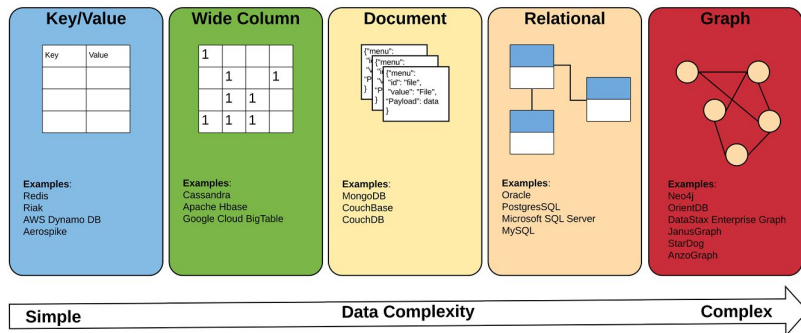
BBN Event Stream - Demo






Not shown in demo but described for use-case completeness.

Supporting Analysis - Querying Data

Querying the data - supporting different use cases



<https://livebook.manning.com/book/graph-databases-in-action/chapter-1/v-3/>

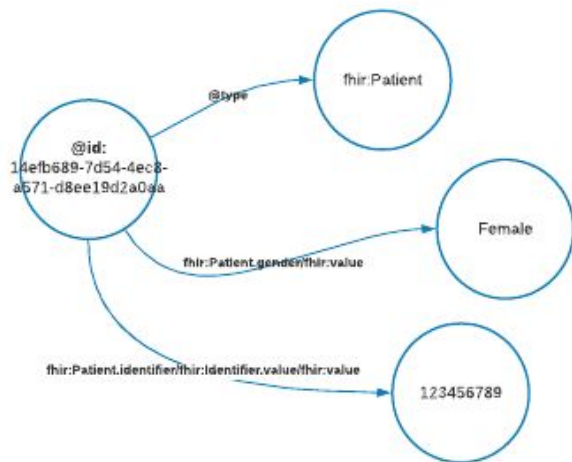
| Data Store | Usage |
|---|--|
|  | Population-level search Use for search-engine like queries - text based searches that returns hits based on relevance. |
|  | Patient-level data exploration Use when you want to retrieve different types of data for a particular node and you want to analyze the relationships between the node and other data points. |
|  | Population-level analysis Primary use case is to support our dashboards and reports. Returns large amounts of data in a consistent format. |

Querying the data - SPARQL Query

SPARQL Query

In a SPARQL query, imagine performing a graph traversal. You're starting at a particular node and tell your query which edges to traverse.

The syntax for a graph query follows a 'subject → predicate → object' pattern or 'starting node → edge → ending node'.



Test | adeel | nxv:defaultSparqlIndex ✓

```
1 prefix fhir: <http://hl7.org/fhir/>
2
3 select ?p_identifier where {
4   ?p a fhir:Patient .
5   ?p fhir:Patient.identifier/fhir:Identifier.value/fhir:value ?p_identifier .
6   ?p fhir:Patient.gender/fhir:value 'Female' .
7 }
```

Execute SPARQL query

p_identifier

123456789

Querying the data - Elasticsearch query

```
1 {  
2   "from" : 0, "size" : 100,  
3   "query": {  
4     "match_phrase": {  
5       "note_content": {  
6         "query": "ADD concerta", "slop": 10  
7       }  
8     }  
9   }  
10 }
```

```
6     "total": 0  
7   },  
8   "hits": {  
9     "hits": [  
10      {  
11        "_score": 0.739864,  
12        "_id": "https://www.camh.ca/kcniel/clinical_note_3",  
13        "_index": "delta_7d9c0a67-71fe-4ee9-abaf-53195a555484_d65c0a99-9134-4978-bc7a-9801f3ab150b_3",  
14        "_source": {  
15          "@id": "clinical_note_3",  
16          "@type": "clinical_note",  
17          "note_content": "Contact Information for the Referral Source [~~~] [~~~] 455 [~~~] [~~~] [~~~], ON [~~~]  
Phone: [~~~] [~~~] Dr. [~~~], Thank you for referring Mr. [~~~] [~~~] to us. Today I saw [~~~] at  
the Mood and Anxiety outpatient clinic at [~~~] and reviewed [~~~] case with Dr. [~~~], staff  
psychiatrist. As you know [~~~] is a [~~~] year [~~~] [~~~] who lives in [~~~] by [~~~]. [~~~] works  
as a mechanical engineer. [~~~] is currently not in any relationships and does not have any [~~~] or  
pets. Reason for Referral/Presenting Problem/Chief Complaint [~~~] was referred to us by yourself  
for management of symptoms of depression. History of Present Illness Today [~~~] stated that [~~~]  
symptoms of depression started about 4 years ago. At that time [~~~] was experiencing low mood,  
anhedonia, low energy, low motivation and poor concentration. [~~~] went to Toronto [~~~] General  
hospital at that time and was followed-up by a psychiatrist and a counsellor for couple of months  
and got started on sertraline and bupropion which were helpful. [~~~] stopped [~~~] medications  
after 2 years because [~~~] was feeling 'numb' and although [~~~] did not feel sad anymore, [~~~]  
was not able to enjoy anything. [~~~] started feeling depressed again in [~~~] and Dr. [~~~] started  
[~~~] on citalopram which was gradually increased to 30mg po daily. [~~~] has recently been started  
on concerta for possible diagnosis of ADD. [~~~] has not found citalopram helpful. [~~~] stated that  
[~~~] mood is sad most of the time. [~~~] used to enjoy music and reading which [~~~] is not  
interested in anymore. [~~~] has low motivation and [~~~] concentration is poor. [~~~] has problems  
falling asleep and has fatigue and low energy during the day. [~~~] sometimes thinks that would have  
preferred to be dead but doesn't have any suicidal plan. [~~~] stated when [~~~] got started on  
sertraline after about a month [~~~] experienced an episode where [~~~] only needed 1-2 hours of  
sleep. During that time [~~~] mood was elevated and [~~~] energy level was high. [~~~] thoughts were  
racing and [~~~] freinds were telling [~~~] that [~~~] was talking faster than usual. [~~~] was
```