



EUROPEAN CENTRAL BANK

EUROSYSTEM

SDMX meets Big Data technologies

ECB implementation



SDMX Expert Group
27/01/2021

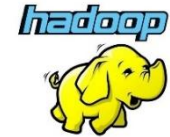
Almir Delic, Stefano Pambianco

ECB statistical data platform: moving to Big-Data techs

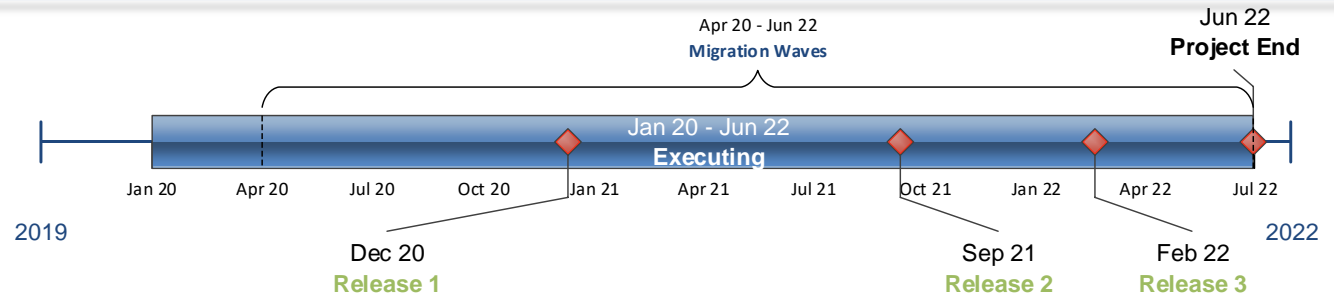
- ECB launched a project in 2019 to replace the current (26 years old) macroeconomic statistics production platform



- After POCs and investigations during 2019, Hadoop technological stack has been selected for implementing the new platform

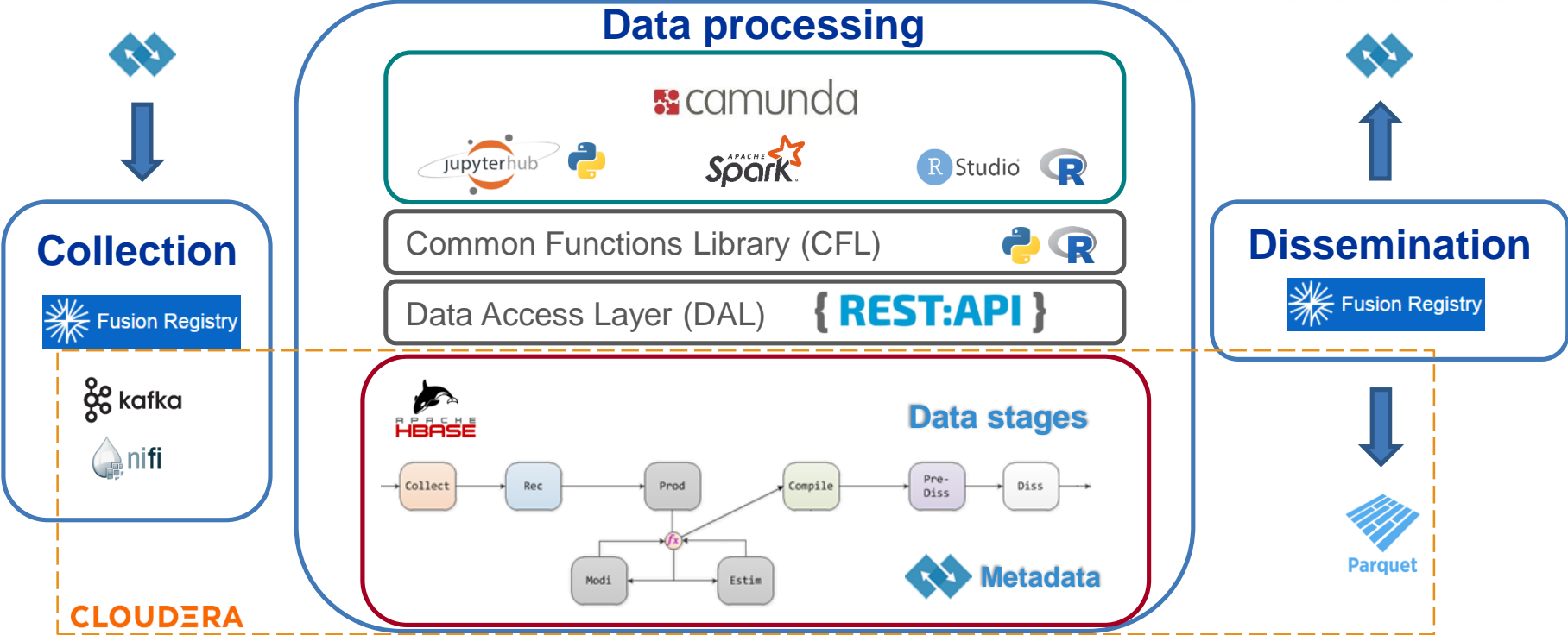


- Project timeline →

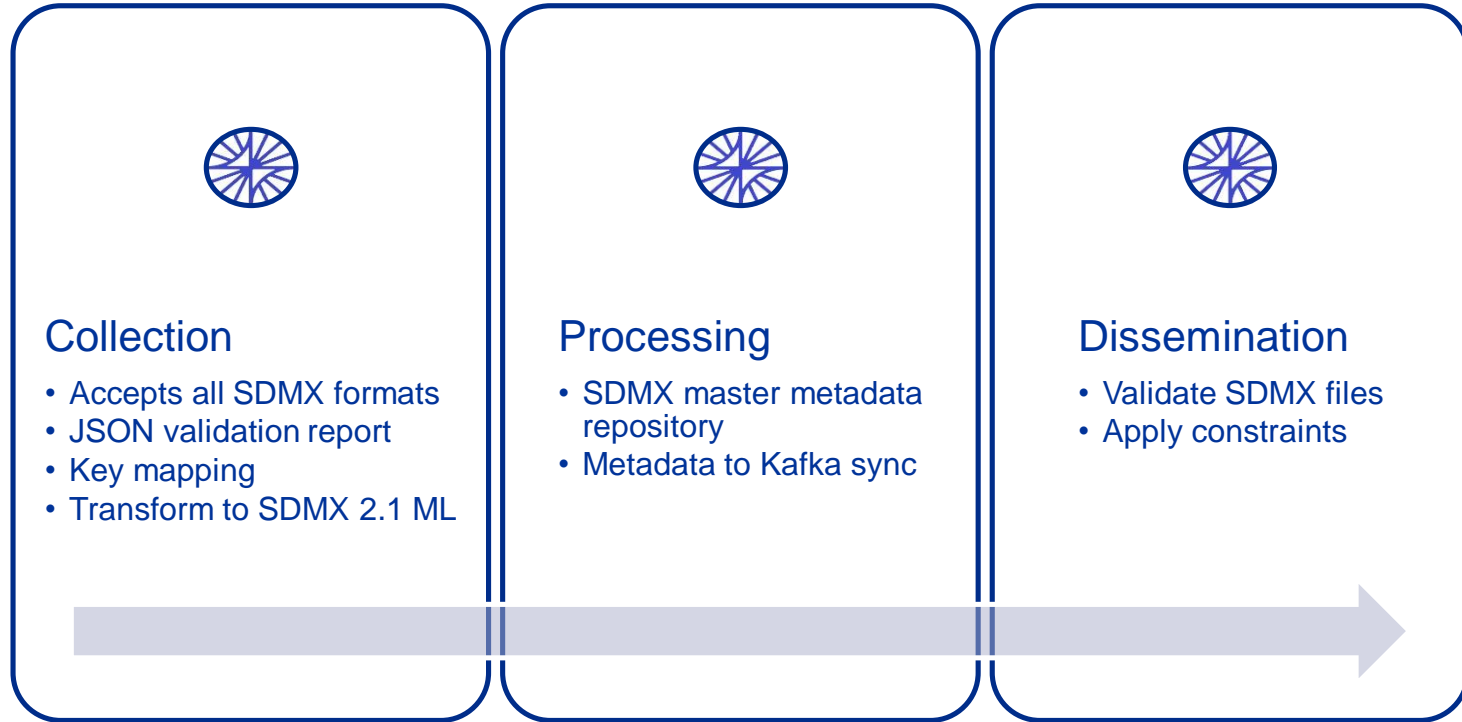
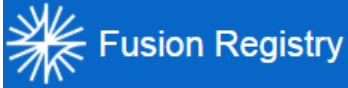


SDMX data and metadata on big data stack

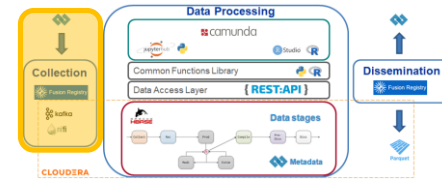
ECB statistical platform



The role of the Fusion Registry



Collection



- ✓ Fusion Registry output: SDMX 2.1 ML + validation report JSON
- ✓ Each observation data + metadata is output in separate JSON message
- ✓ JSON messages are pushed into Kafka queue and flow downstream



SDMX in HBase - Physical data model



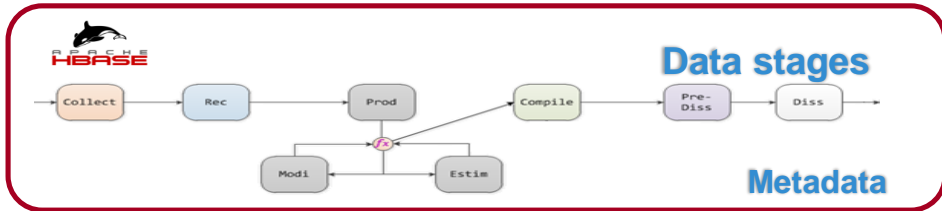
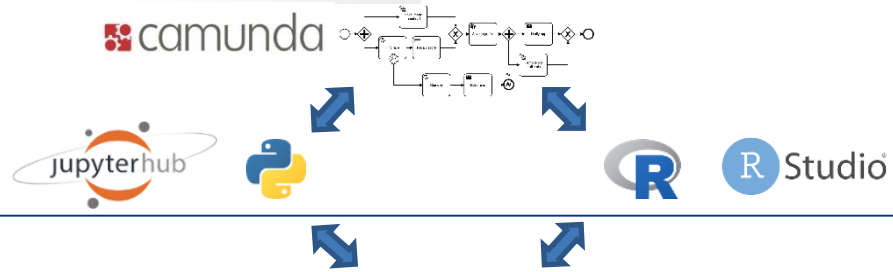
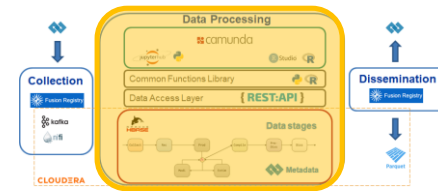
```
{"attributes": [  
  {  
    "id": "OBS_STATUS",  
    "value": "A"  
  },  
  {  
    "id": "OBS_CONF",  
    "value": "F"  
  },  
  {  
    "id": "COLLECTION",  
    "value": "E"  
  },  
  {  
    "id": "TITLE_COMPL",  
    "value": "anonymized - Net Circulation - number of  
banknotes/coins in circulation (for banknotes it has to be  
calculated, and equals created notes less destroyed notes  
less stocks of the NCB) - Coins - 1 - Current currency -  
Stock - denominated in National currency"  
  }  
],
```

```
"submissions": [  
  {  
    "id": "a5da4911-e26d-4743-b2a9-554732608f13",  
    "iref": "IREF00001",  
    "rejected": false,  
    "version": "0"  
  }  
],  
"validationResults": {  
  "dataset": [],  
  "observation": [],  
  "series": [  
    {"type": "MandatoryAttributes",  
      "message": "Missing mandatory attribute 'DECIMALS'"  
    }  
  ],  
  "value": "256730.057",  
  "layer": "A0",  
  "period": "2002-03",  
  "stage": "rec",  
  "statuses": [],  
}
```

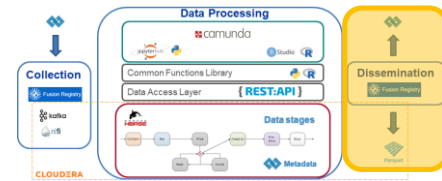
Key = {salt}|{dataflowID}|{timeseriesKey}|{periodTimestamp}|{layer}

Key = 00|BKN|M.XX.NC10.C.10P0.C0.S.N|101494080000|A0

Data processing



Dissemination



- ✓ Data is read from the DAL
- ✓ Camunda is used as scheduler / workflow manager
- ✓ Supports all SDMX formats for dissemination to NCBs / IOs
- ✓ Supports parquet dissemination for internal dissemination (with Impala tables on top)
- ✓ SDMX files validated against Fusion Registry before dissemination



