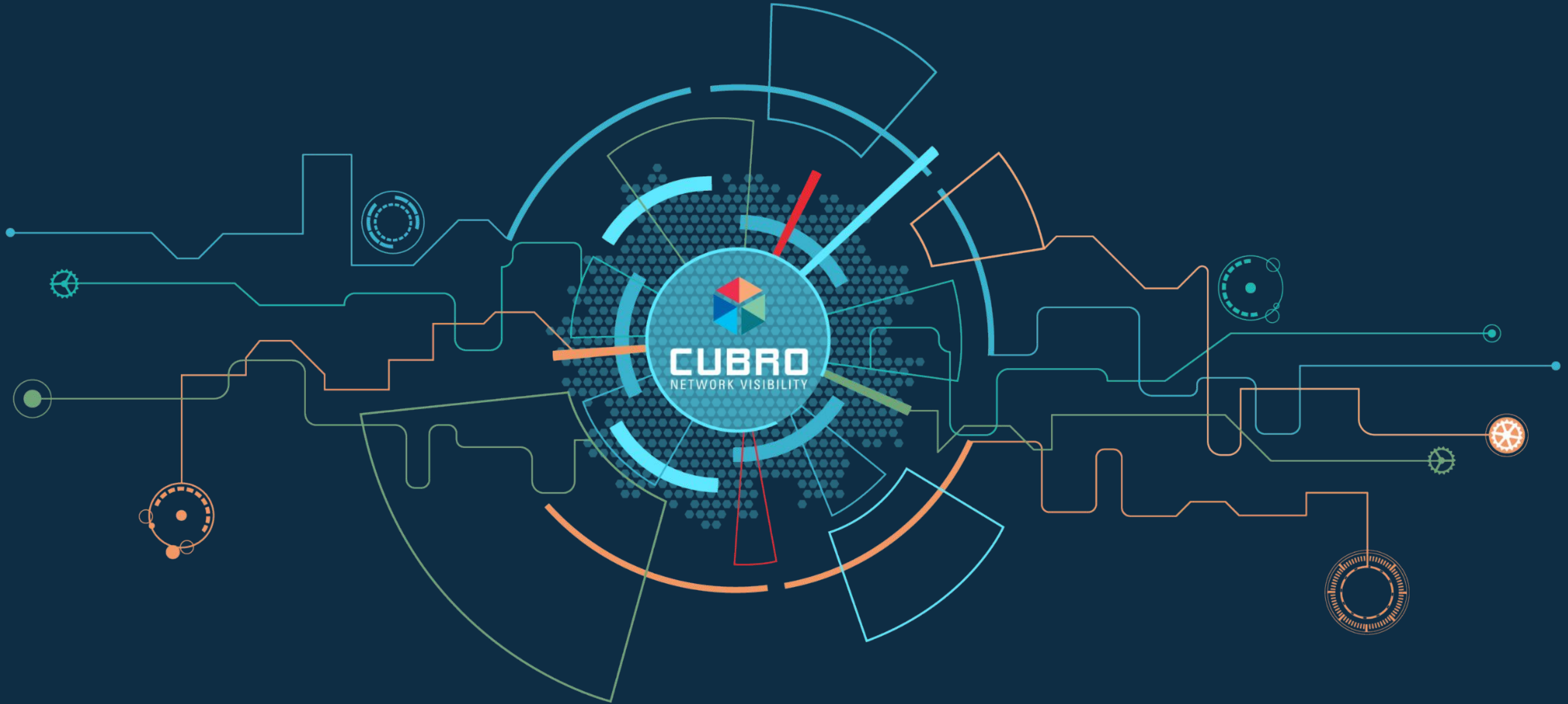


2023

Terabit Metadata extraction application solution for MNO

CUBRO
NETWORK VISIBILITY

Terabit Metadata extraction application

Challenges faced by MNOs in Managing and Analyzing Large Volumes of Data

- **Scale**: Exponential growth of mobile networks leads to massive volumes of metadata, reaching terabit levels.
- **Storage**: Storing vast amounts of metadata requires significant infrastructure investments. MNOs face challenges in finding cost-effective, scalable solutions.
- **Processing Speed:** Real-time analysis of metadata is crucial for network optimization, security, and customer insights. Traditional processing methods often fall short in handling the speed and scale of modern network metadata.
- **Compliance and Privacy:** MNOs must comply with strict data privacy regulations. Ensuring compliance adds complexity and requires robust security measures to protect user information.
- **Lack of Insights:** Despite abundant metadata, extracting actionable insights is challenging. MNOs struggle to derive meaningful patterns, trends, and correlations from the vast data pool, limiting informed decision-making.
- **Resource Constraints:** MNOs face limitations in skilled personnel, budget, and technological capabilities. These constraints hinder the effective management and analysis of large metadata volumes.

Cubro's innovative solution can minimize data volume for enhanced efficiency.

Cubro's solutions can reduce the quantity of data generated and stored for network monitoring, security and analytics use cases while retaining all the required data – this approach can reduce the capacity and footprint required of network tools and associated data lakes. It provides significant benefits in performance, CAPEX and OPEX, and power and space consumption. This document describes the associated challenges and solutions.

Overview and Advantages of Cubro's Solution

CUBRO NETWORK VISIBILITY

Step 1: Efficient Metadata Production

- The first step is to produce metadata information efficiently.
- Cubro solution outperforms common monitoring, providing superior efficiency.
- Metadata volume on an IPFIX (flow-based solution) is 3% of the raw input.
- Cubro solution achieves the same metadata quality with only 0.1% of the raw input.

Step 2: Title: Aggregation, Filtering, and Enrichment

- The second step involves aggregation, filtering, and enrichment at the processing chain's beginning.
- This step is crucial in reducing data volume and processing efforts throughout the chain.
- Proper use case planning is vital, as it leads to faster processing, cost-efficiency, and higher ROI.

Data Lake Usage Vs Database

Data Lake:

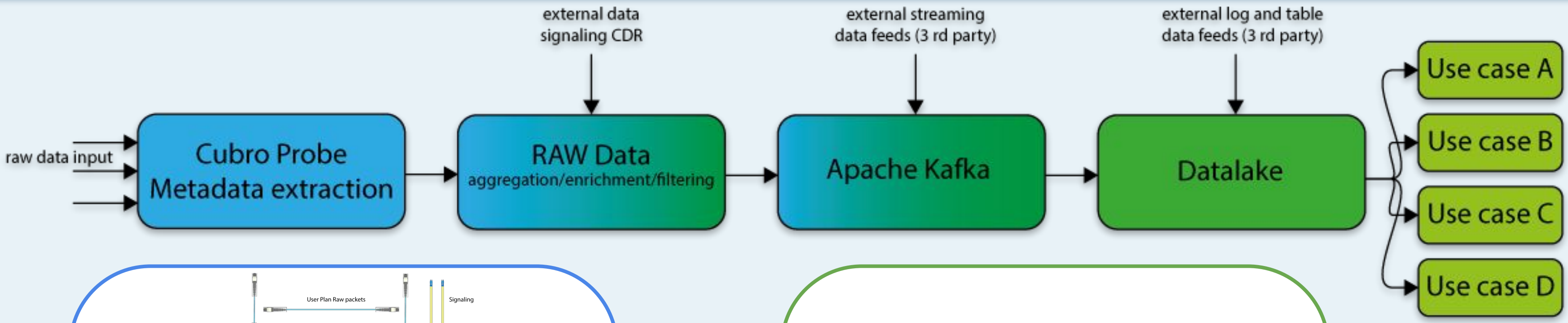
- A data lake is a centralized repository that stores large volumes of structured, semi-structured, and unstructured data in its raw, unprocessed form.
- It is designed to store vast amounts of data from various sources, such as transactional systems, social media feeds, log files, and sensor data.

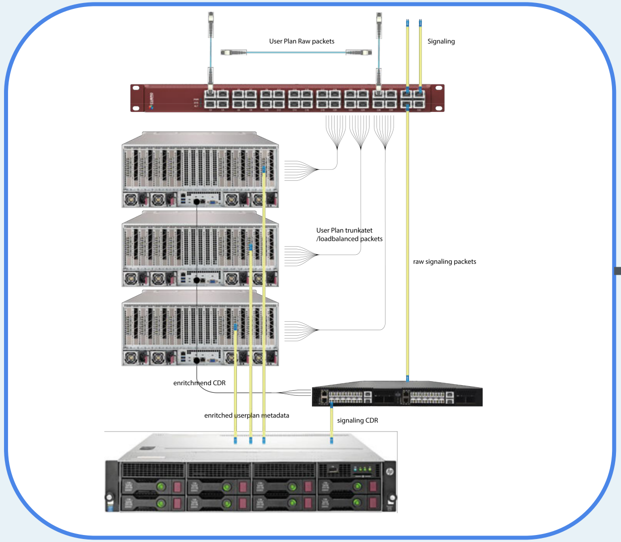
Database:

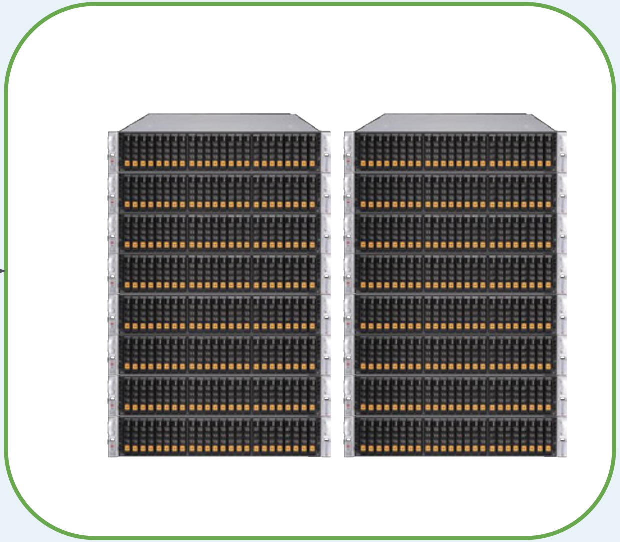
- A database is a structured collection of data organized and managed to support efficient data storage, retrieval, and manipulation.
- Databases provide a structured way to store and organize data, ensuring data integrity and consistency.

The data lake gives more options to design use case because the original idea of a data lake is to collect data from multiple sources and then produce data tables which can be consumed by different applications.

Omnia QM Solution







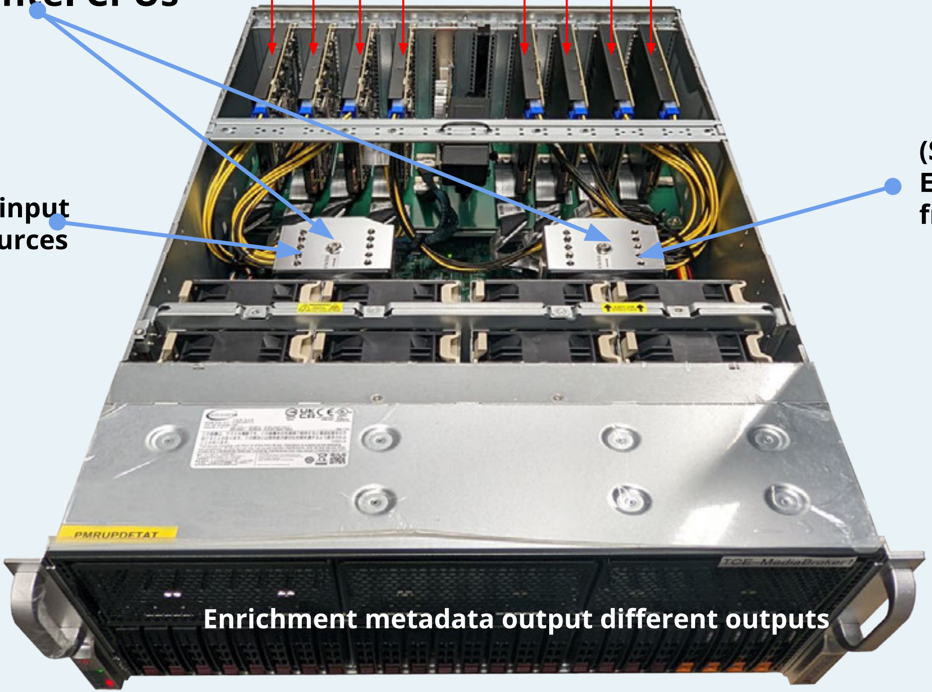
Cubro Omnic (NPU) cards in a Server

CUBRO NETWORK VISIBILITY

Enrichment and data processing on the Intel CPUs

(User plane) Raw packet input up to 40 – 85 Gbit per Smart NIC depending on the traffic type

(Signaling traffic)
Enrichment data input
from different sources

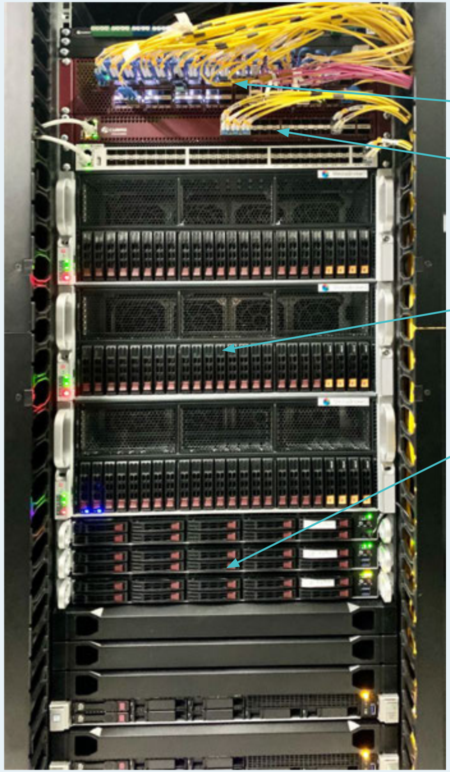


(Signaling traffic)
Enrichment data input
from different sources

Efficient Metadata generation

1–2 Tbps User Plane monitoring in an MNO (real picture)

Efficient Metadata generation

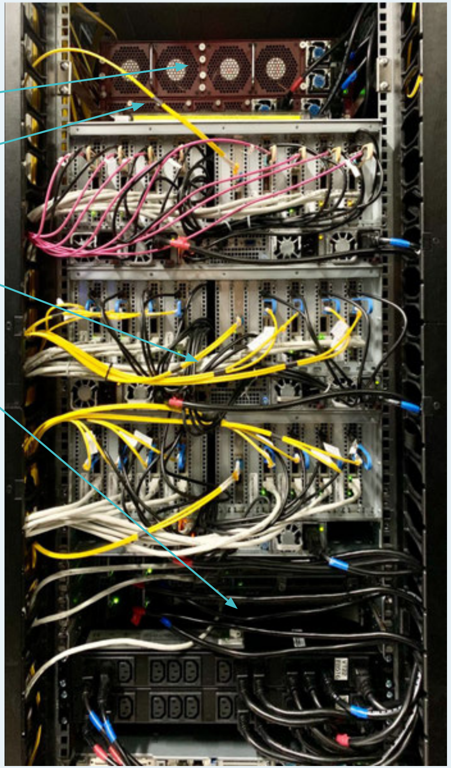


NPB EXA64100

Signaling probe

3 Server with
24 NPU smart NIC

3 Kafka server for
Metadata handling

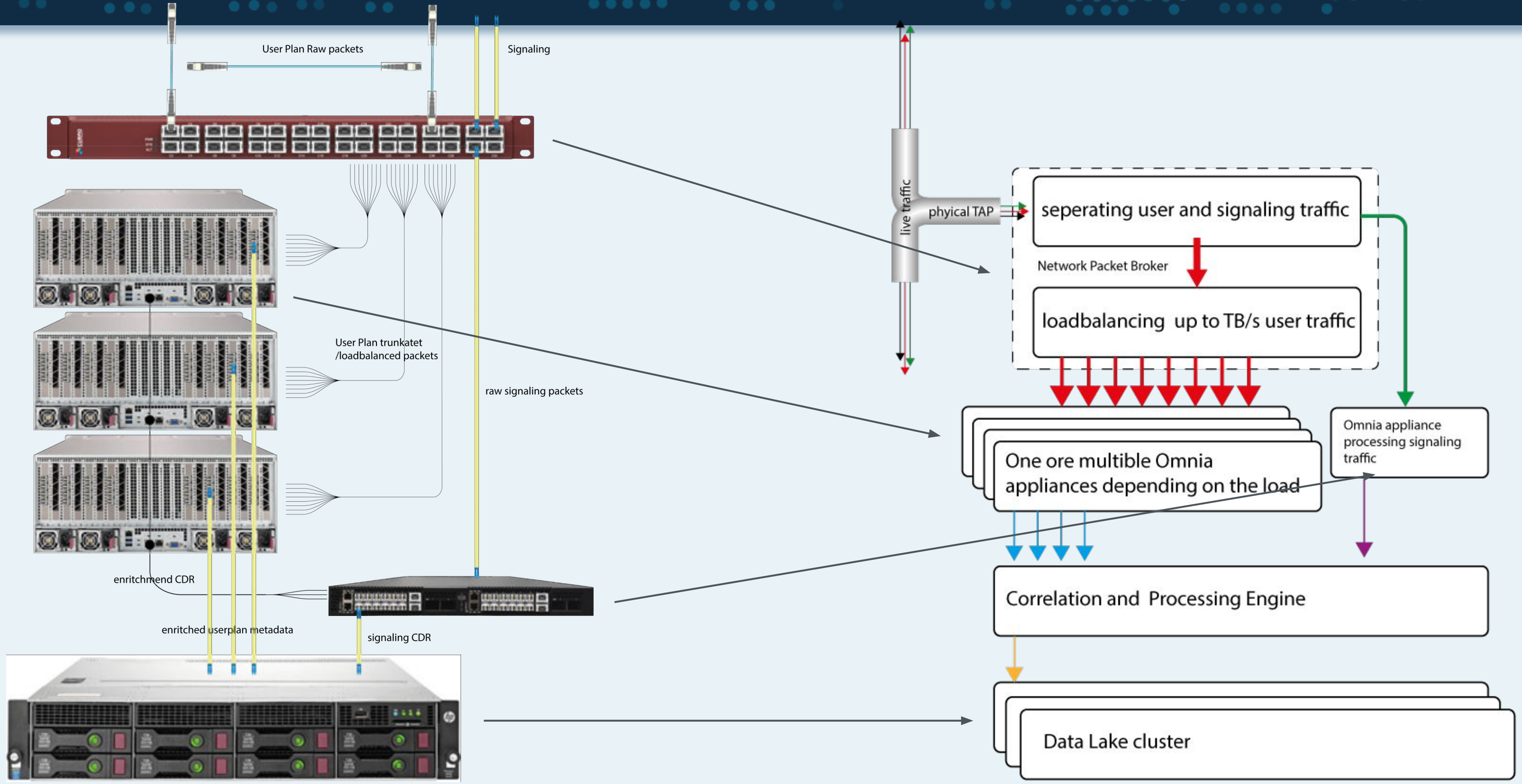


only 19 U
836 mm

System Schema for 1 - 1.2 TB of user traffic

CUBRO
NETWORK VISIBILITY

aggregation, filtering and enrichment at the beginning of the processing chain

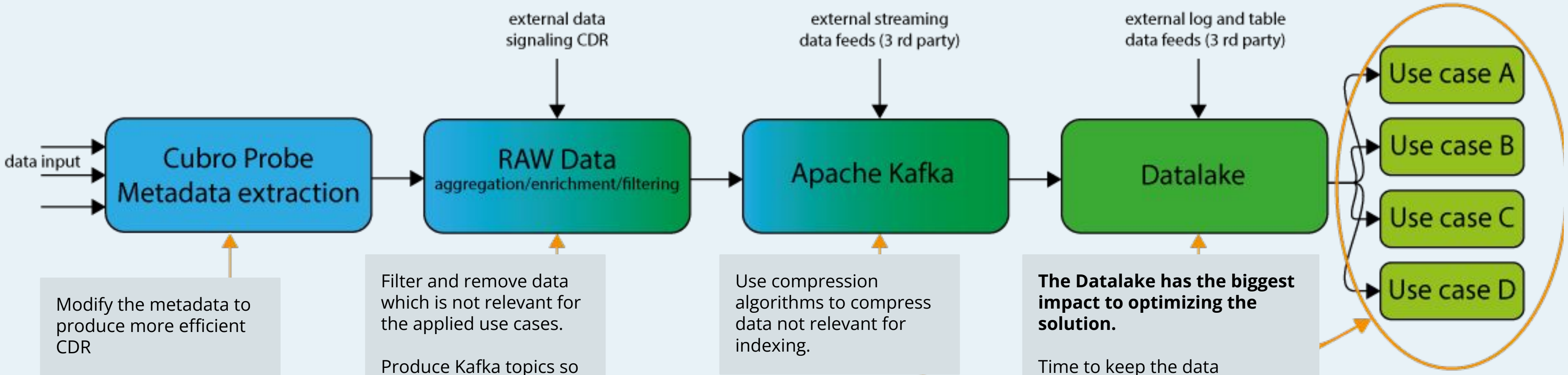


The amount of data challenge

Volume per Hour in GByte	300	600	1200	1800	2400	3000	3600	4200	4800	5400	6000
Volume per Day in TByte	7,031	14,063	28,125	42,188	56,250	70,313	84,375	98,438	112,500	126,563	140,625
Volume per Week in PByte	0,048	0,096	0,192	0,288	0,385	0,481	0,577	0,673	0,769	0,865	0,961
Volume per Month in PByte	0,213	0,426	0,851	1,277	1,703	2,129	2,554	2,980	3,406	3,831	4,257
Volume per 3 Month in PByte	0,639	1,277	2,554	3,831	5,109	6,386	7,663	8,940	10,217	11,494	12,772
Number of subscribers in Mio.	0,5	1	2	3	4	5	6	7	8	9	10
CDR's per sec	8.333	16.667	33.333	50.000	66.667	83.333	100.000	116.667	133.333	150.000	166.667

This table shows the raw traffic volume versus the numbers of subscribers. This shows that only a data lake approach can handle this volume of data. This approach also gives the flexibility to develop uses cases.

The challenge is to define the use case first



Modify the metadata to produce more efficient CDR

Filter and remove data which is not relevant for the applied use cases.

Produce Kafka topics so they can be easily consumed in the data lake.

Use compression algorithms to compress data not relevant for indexing.

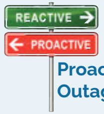
The Datalake has the biggest impact to optimizing the solution.

Time to keep the data
Aggregate data over time

The definition of the use cases is a major factor in making such a solution efficient in terms of performance and cost.
It must be a closed loop. All elements in the chain can support optimizing the data volume.

Use case examples

CUBRO NETWORK VISIBILITY



Proactive Service Outage Alerts

For this use case, only a slim table is needed, with only the total amount of traffic per application.

Xbox	2,2 Gbit/sec
Whatsapp	0,5 Gbit/sec
Office 365	0 Gbit/sec
Netflix	7,1 Gbit/sec
YouTube	6,2 Gbit/sec

Misuse of Service

For this use case, a table is needed, with only the total amount of traffic per subscriber.

2325670310	2,2 TByte/day
2325545673	1,4 TByte/day
2325645540	8,2 TByte/day
2325234784	0,2 TByte/day
2325635143	3,2 TByte/day

The data lake approach is that these tables are preprocessed so that if an application wants to consume the data, only reading these files is needed and no search ore processing is needed. If all tables for use case are produced, then the original data can be deleted. This helps to keep the storage volume small.

Reselling of geolocation data

For this use case, there are different feeds needed to build the table.
1:) Subscriber base station attachment over time

2:) a LOG/MAP file from the service provider which builds the correlation between the cell ID and the geographical data.

CDR-based information

2325670310	Cell ID 55
2325670310	Cell ID 55
2325645540	Cell ID 55
2325645540	Cell ID 56
2325635143	Cell ID 58

Static log information

Cell ID 55	Lat,Long 48.2030915,16.2076345
Cell ID 56	Lat,Long 48.1802423,16.26807372
Cell ID 57	Lat,Long 48.1614862149,16.28135
Cell ID 58	Lat,Long xxxxx
Cell ID 59	Lat,Long xxxxx

2325670310	Lat,Long 48.2030915,16.2076345
2325645540	Lat,Long 48.2030915,16.2076345
2325645540	Lat,Long 48.1802423,16.26807372

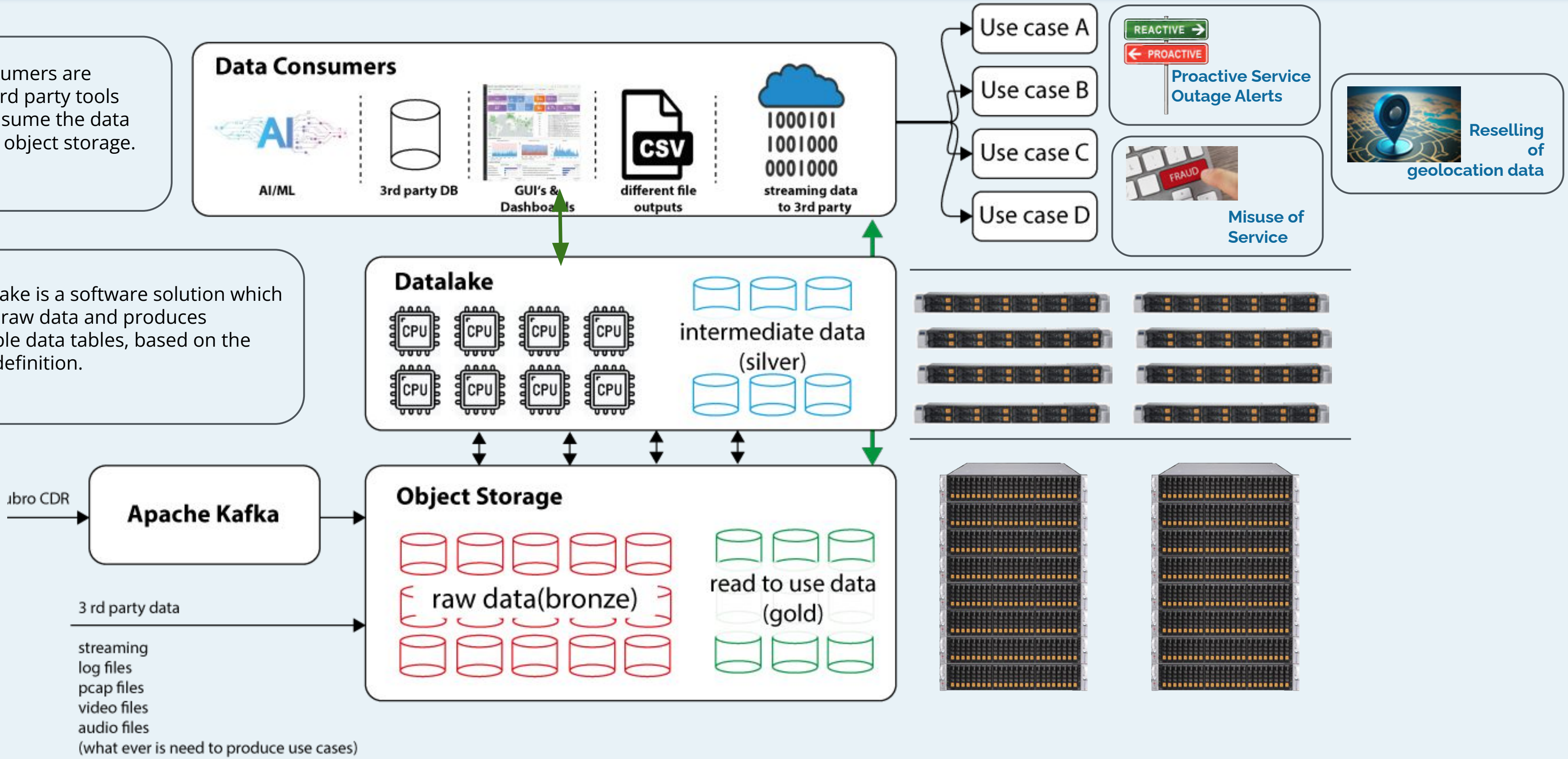
→ consumable table

The full data solution

CUBRO NETWORK VISIBILITY

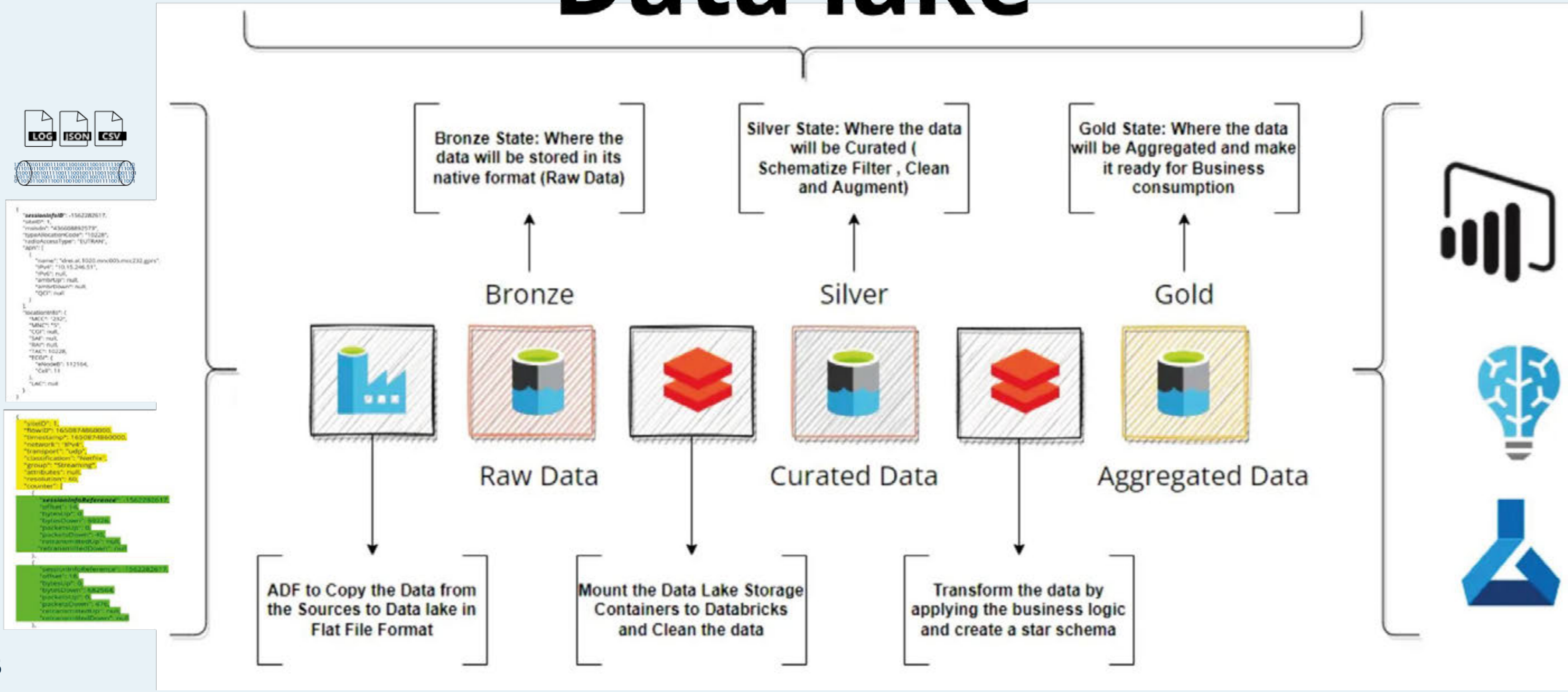
Data consumers are typically 3rd party tools which consume the data out of the object storage.

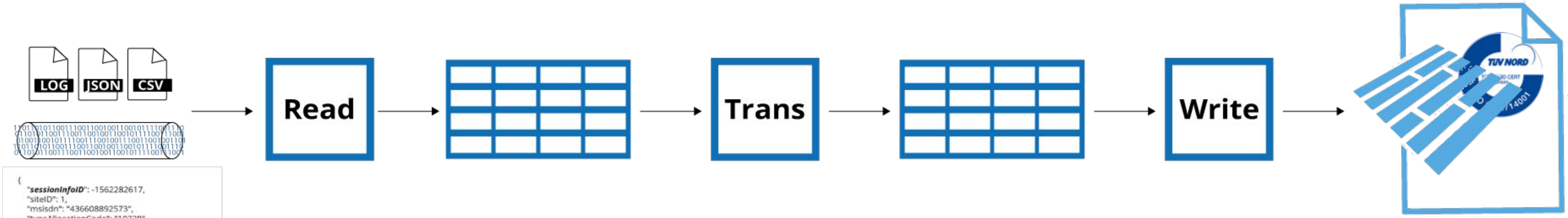
The data lake is a software solution which reads the raw data and produces consumable data tables, based on the use case definition.



CUBRO NETWORK VISIBILITY

Data lake






```
{
  "sessionInfoID": -1562282617,
  "siteID": 1,
  "msisdn": "436608892573",
  "typeAllocationCode": "10228",
  "radioAccessType": "EUTRAN",
  "apn": [
    {
      "name": "drei.at.1020.mnc005.mcc232.gprs",
      "IPv4": "10.15.246.51",
      "IPv6": null,
      "ambrUp": null,
      "ambrDown": null,
      "QCI": null
    }
  ],
  "locationInfo": {
    "MCC": "232",
    "MNC": "5",
    "CGI": null,
    "SAI": null,
    "RAI": null,
    "TAC": 10228,
    "ECGI": {
      "eNodeB": 112164,
      "Cell": 11
    },
    "LAC": null
  }
}
```

```
{
  "siteID": 1,
  "flowID": 1650874860000,
  "timestamp": 1650874860000,
  "network": "IPv4",
  "transport": "udp",
  "classification": "Netflix",
  "group": "Streaming",
  "attributes": null,
  "resolution": 60,
  "counter": [
    {
      "sessionInfoReference": -1562282617,
      "offset": 14,
      "bytesUp": 0,
      "bytesDown": 59226,
      "packetsUp": 0,
      "packetsDown": 45,
      "retransmittedUp": null,
      "retransmittedDown": null
    },
    {
      "sessionInfoReference": -1562282617,
      "offset": 18,
      "bytesUp": 0,
      "bytesDown": 682564,
      "packetsUp": 0,
      "packetsDown": 476,
      "retransmittedUp": null,
      "retransmittedDown": null
    },
```


Signaling / Control Plan

site ID	sessioninfoID	msisdn	rat	apn	cell id	MCC	MNC	LAC
1	1562282617	43 660 234 233	LTE	xxx	2	•	•	•
1	1562282618	43 660 234 224	5G	xxx	8	•	•	•
1	1562282619	43 660 565 755	3G	xxx	16	•	•	•
1	1562282620	43 660 862 934	LTE	xxx	23	•	•	•
1	1562282621	43 660 496 946	LTE	xxx	45	•	•	•
1	1562282622	43 660 888 331	5G	xxxxx	78	•	•	•
1	1562282623	43 660 405 413	5G	xxx	876	•	•	•
1	1562282624	43 660 631 179	3G	xxxxxx	3	•	•	•
1	1562282625	43 660 539 629	4G	xxx	2	•	•	•
1	1562282626	43 660 974 038	LTE	xxxxxxx	3	•	•	•
1	1562282627	43 660 642 882	LTE	xx	45	•	•	•

User Plan

site ID	sessioninfoID	User IP	Group	App	bytes up	bytes down	packets up	packets dow
1	1562282617	10.25.37.110	Streaming	Netflix	767	454	8	27
1	1562282617	10.25.37.110	Messaging	Whatsapp	188	771	14	23
1	1562282617	10.25.37.110	Social	Facebook	437	553	26	3
1	1562282617	10.25.37.110	Office	Salesforce	688	917	34	20
1	1562282618	10.35.32.200	Streaming	Youtube	291	11	10	7
1	1562282618	10.35.32.200	Social	Instagram	302	128	12	25
1	1562282618	10.35.32.200	Office	Office365	666	723	35	4
1	1562282618	10.35.32.200	Cloud	S3 bucket	410	472	38	27
1	1562282618	10.35.32.200	HTTPS	unknown	722	295	7	11
1	1562282619	10.56.19.157			238	842	19	34
	|	|	|	|	622	347	30	21
1	1562282620	10.77.35.78	|	|	293	497	37	22
	|	|	|	|	604	281	33	6
1	1562282621	10.34.77.148						
1	1562282622	10.99.23.13						
1	1562282623	10.22.19.158						
1	1562282624	10.66.34.26						
1	1562282625	10.87.43.76						
1	1562282626	10.10.123.18						
1	1562282627	10.87.12.99						

Combinde Tabel (silver)

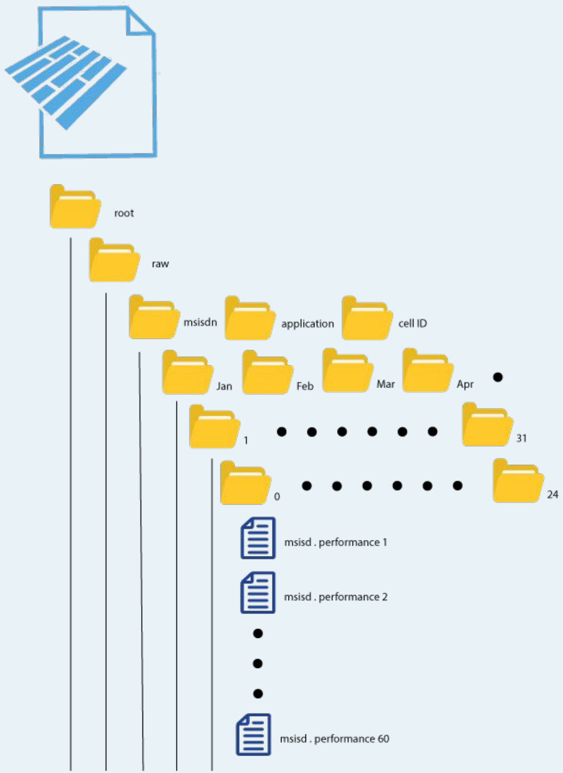
site ID	sessioninfoID	User IP	msisdn	rat	Group	App	bytes up	bytes down	packets up	packets dowi
1	1562282617	10.25.37.110	43 660 234 233	LTE	Streaming	Netflix	199	414	17	4
1	1562282617	10.25.37.110	43 660 234 233	LTE	Messaging	Whatsapp	431	724	39	34
1	1562282617	10.25.37.110	43 660 234 233	LTE	Social	Facebook	761	699	15	25
1	1562282617	10.25.37.110	43 660 234 233	LTE	Office	Salesforce	155	741	34	15
1	1562282618	10.35.32.200	43 660 234 224	5G	Streaming	Youtube	856	651	26	40
1	1562282618	10.35.32.200	43 660 234 224	5G	Social	Instagram	446	367	2	19
1	1562282618	10.35.32.200	43 660 234 224	5G	Office	Office365	226	829	13	11
1	1562282618	10.35.32.200	43 660 234 224	5G	Cloud	S3 bucket	759	611	1	5
1	1562282618	10.35.32.200	43 660 234 224	5G	HTTPS	unknown	519	400	12	29
1	1562282619	10.56.19.157	43 660 565 755	3G			183	263	39	11
	|	|			|	|	390	354	3	10
1	1562282620	10.77.35.78			|	|	296	47	18	33
	|	|			|	|	636	249	1	28
1	1562282621	10.34.77.148								
1	1562282622	10.99.23.13								
1	1562282623	10.22.19.158								
1	1562282624	10.66.34.26								
1	1562282625	10.87.43.76								
1	1562282626	10.10.123.18								
1	1562282627	10.87.12.99								

reade to use Tabel (gold)

cell id	bytes down	bytes up
2	6554	3267
8	42707	5973
16	38931	6438
23	3344	9941
45	54513	7722
78	22931	5033
876	4629	5184
3	32412	4757
2	46399	4616
3	93629	9860
45	45340	4412

msisdn	bytes down	bytes up
43 660 234 233	369554	4965
43 660 234 224	655738	3156
43 660 565 755	282191	9529
43 660 862 934	114190	4303
43 660 496 946	620065	6384
43 660 888 331	657119	4316
43 660 405 413	512471	8317
43 660 631 179	960819	5276
43 660 539 629	475265	4716
43 660 974 038	459683	4649
43 660 642 882	949917	8568

Advantages of a data lake compared to a DB



All reports and results are preprocessed

This means the access is very fast.

But it costs more storage and CPU resources than DB approach.

List of use cases

Service related

- Availability of a service in the entire network
- Availability of a service separate by RAT
- Availability of a service per base station

- Service distribution in the entire network
- Service distribution per RAT
- Service distribution per network region
- Service usage over time per day/week/month
- Service usage live view

Performance related

- Performance per subscriber (bandwidth and volume)
- Performance per base station (bandwidth and volume)
- Performance per data centre
- Performance per network segment
- Performance per RAT

Geolocation related

- Geolocation of a subscriber
- Movement of a subscriber
- How many subscribers are located per base station
- How many subscribers per sector or region

- There are much more possible use cases but cannot be disclosed to her. Call for more information.

Subscriber related

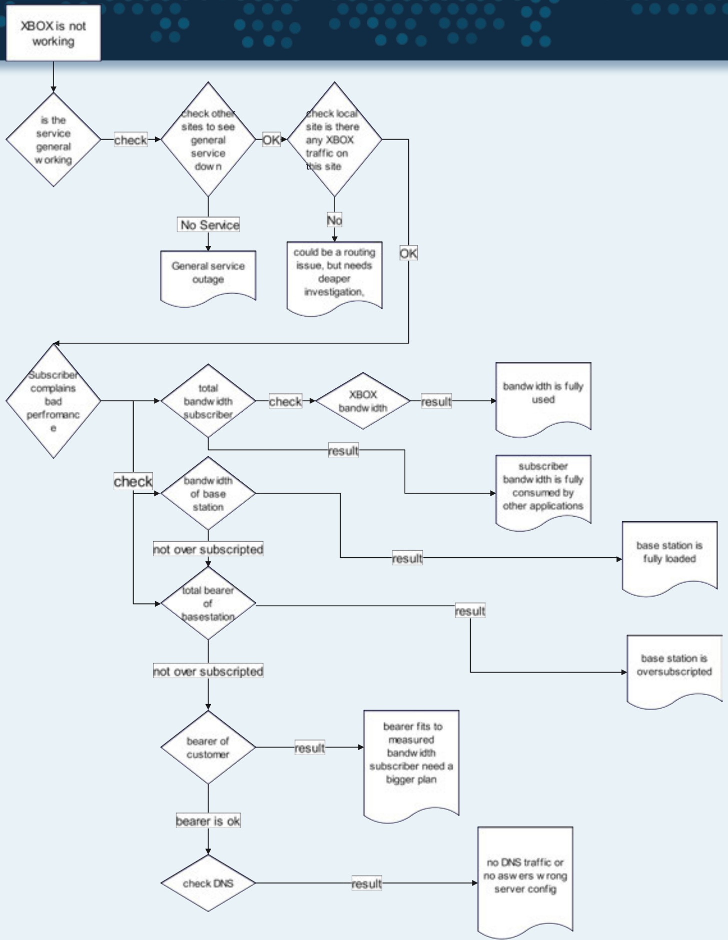
- Number of subscribers online
- Movement of a subscriber
- How many subscribers are located per base station
- How many subscribers per sector or region
- Total traffic usage per subscriber per time interval
- Bandwidth statistics per subscriber per time interval (performance)
- Service usage time per subscriber.
    - How long does a specific subscriber use a specific service (per day)
    - In which time frames a specific subscriber use a specific service

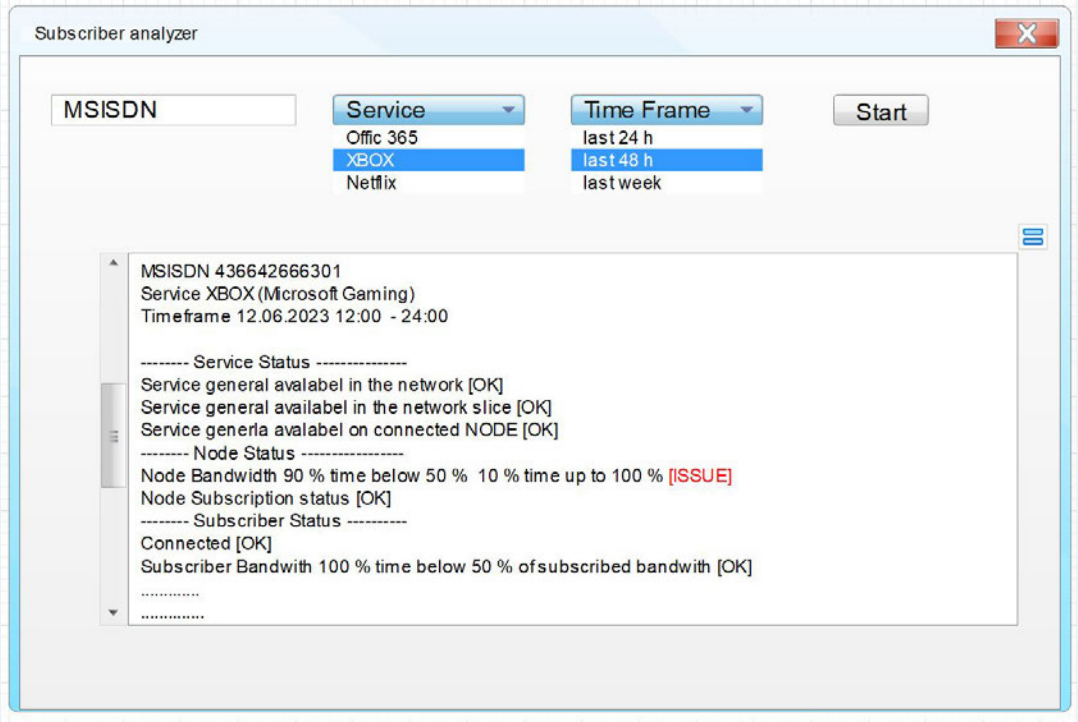
Security related

- Subscribers with significant SIM card change
- Subscribers with constant high load
- Subscribers with high load of suspicious applications (only VPN, or TOR)
-

The combination of the use cases helps also to troubleshoot customer calls to the call center. See next page.

A service related call flow (XBOX as example)





Delay as performance indicator in a Mobile network

CUBRO NETWORK VISIBILITY

Packet Delay has a massive impact on the bandwidth that can be achieved in a network

This means a high delay could be an indicator why a service is not working properly, but it is for sure not the only indicator.
The issue is now that delay or often described as RTT (round trip time) is difficult to measure because many factors have an impact.

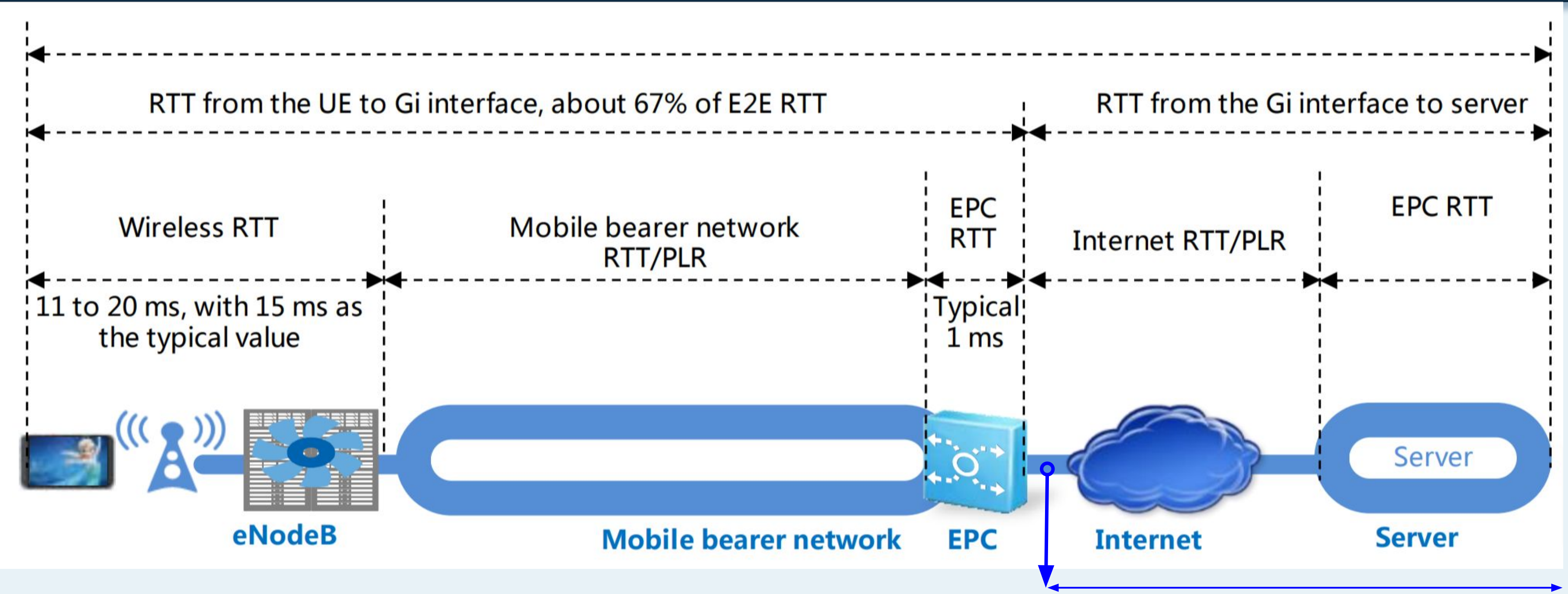
Source of Latency in a Cellular/Mobile Network

In the LTE system, the latency can be divided into two major parts: (1) user plane (U-plane) latency and (2) control plane (C-plane) latency. The U-plane latency is measured by.
One directional transmit time of a packet to become available in the IP layer between evolved UMTS terrestrial radio access network (E-UTRAN) edge/UE and UE/E-UTRAN node [28].
On the other hand, C-plane latency can be defined as the transition time of a UE to switch from an idle state to an active state. At the idle state, a UE is not connected with radio.
Resource control (RRC). After the RRC connection is being set up, the UE switches from idle state into connected state and then enters into active state after moving into dedicated.
mode. Since the application performance is dependent mainly on the U-plane latency, U-plane is the main focus of interest for low latency communication.
In the U-plane, the delay of a packet transmission in a cellular network can be contributed by the RAN, backhaul, core network, and data center/Internet. As referred in Fig. 5,
the total one-way transmission time [29] of the current LTE system can be written as

"T = TRadio + TBackhaul + TCore + TTransport"

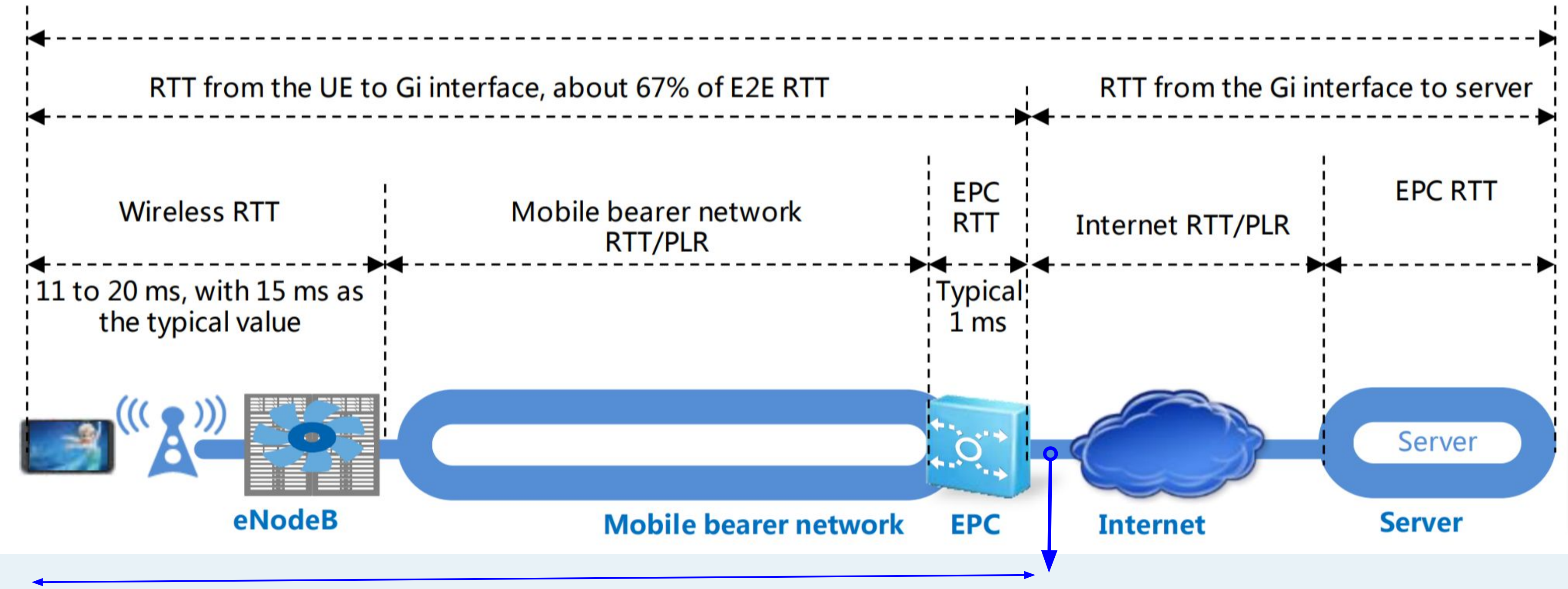
- TRadio is the packet transmission time between eNB and UEs and is mainly due to physical layer communication;
- TBackhaul is the time for building connections between eNB and the core network (i.e. EPC). Generally, the core network and eNB are connected by copper wires or microwave or optical fibers;
- TCore is the processing time taken by the core network;
- TTransport is the delay to data communication between the core network and the Internet/cloud.

E2E delay distribution in Mobile Networks



This is typically the measuring point in classical monitoring approach. The issue is only that at this point it is only possible to measure the external delay.
In fact, 67% of the delay is missed.

The Cubro solution Mobile Networks



Cubro can measure the internal delay per subscriber per second, this information can then be used to enrich this data with the subscriber CDR

We have operations in all time zones.
Reach us at: support@cubro.com