autovia

Analytics for Autonomous Driving: Large-scale sensor data processing

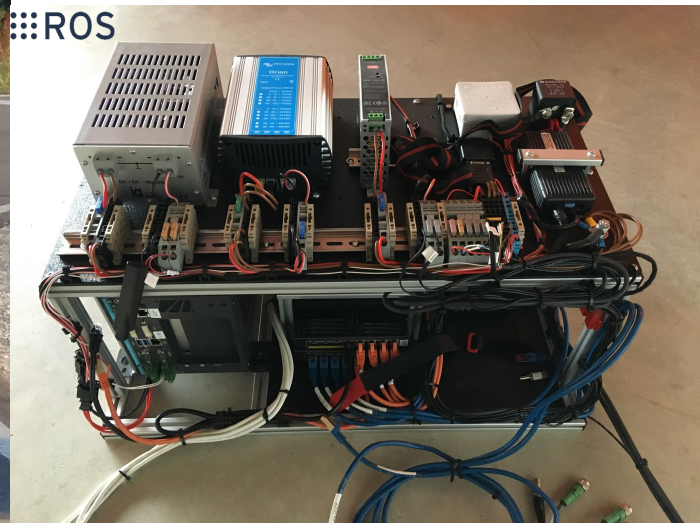
ROS-Industrial Conference 2019

Jan Wiegelmann

© Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Autonomous vehicle generates up to

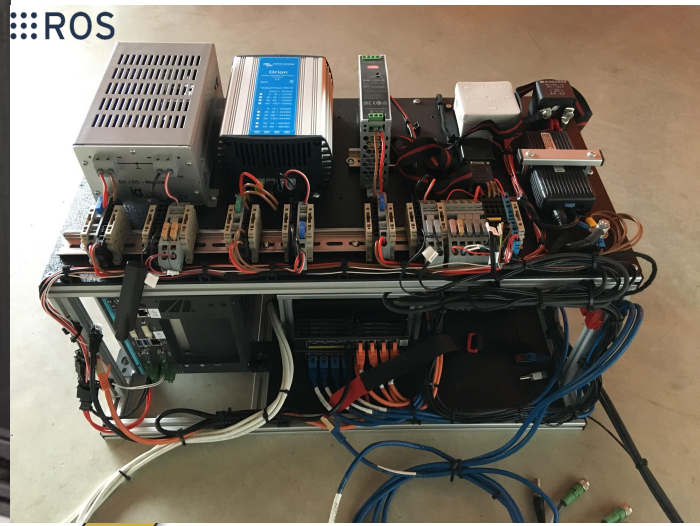
20 TB per day



MHP
A PORSCHE COMPANY

© Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Autonomous vehicle generates up to **20 TB per day**



MHP
A PORSCHE COMPANY

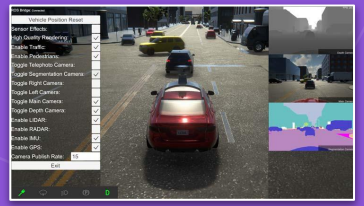
© Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

autovia

Autonomous driving development is a data challenge!

Real world driving and simulation of autonomous vehicles generates **petabytes of data per day**

Simulation



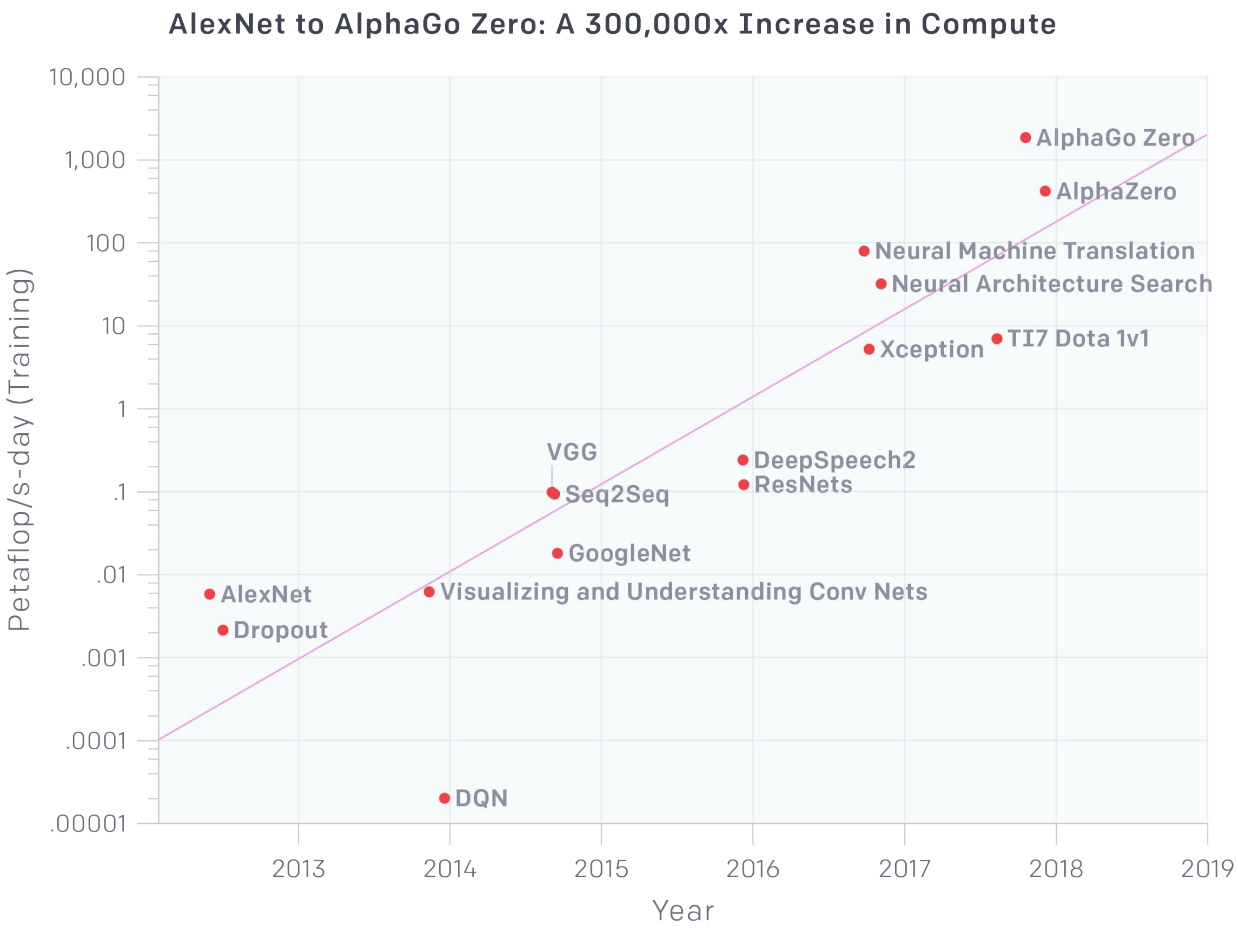
Real world driving

Petabytes of sensor data and simulation results

- **Hard to collect quality data**
- **Long running algorithms**
- **Small throughput**
- **No version control**
- **Infrastructure complexity**

Autonomous vehicle generates up to 20 TB per day

© Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai



https://openai.com/blog/ai-and-compute/

© Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

cruise

Testing at Scale

- **200,000+** hours/day
- 30,000+ VM instances
- 300,000+ CPU cores
- 5,000+ GPUs
- 3 regions, 10 zones
- 300TB/day of results

https://www.youtube.com/watch?v=NGXZ-mYqyUo

 © Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

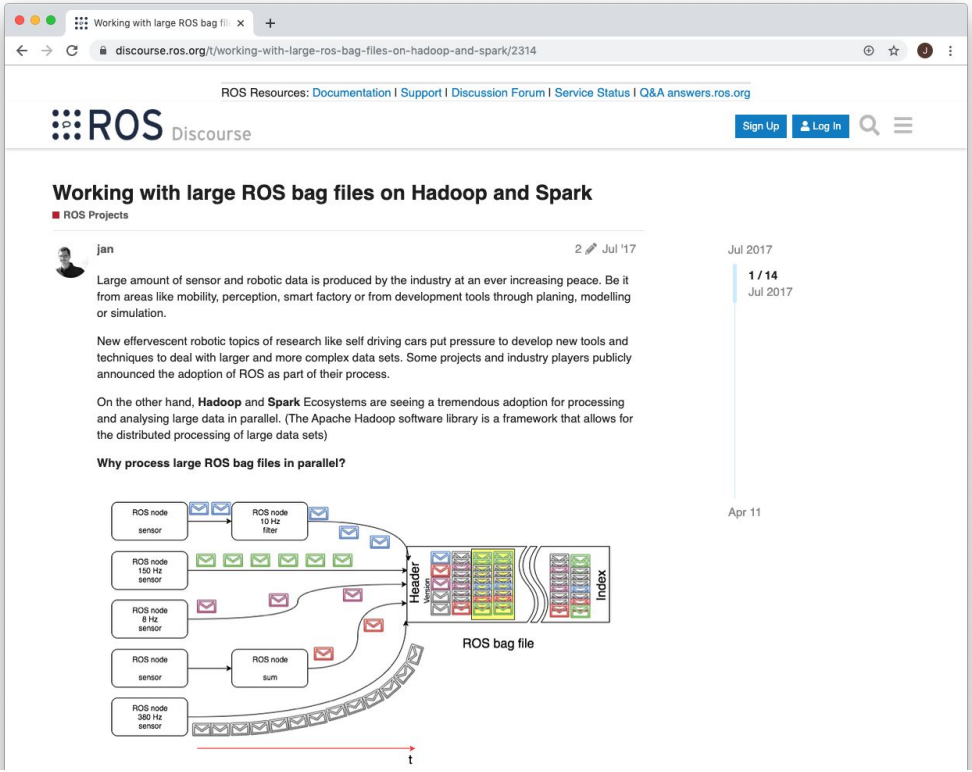
ros_hadoop

RosbagInputFormat is an open source splittable Hadoop InputFormat for the rosbag file format

Requirements

- ROS installation on each node
- Manual index extraction
- HDFS and Spark installation

Release July 2017 with Apache 2.0



created Jul '17 | last reply Apr 11 | 13 replies | 4.7k views | 8 users | 3 likes | 7 links

https://discourse.ros.org/t/working-with-large-ros-bag-files-on-hadoop-and-spark/2314

© Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

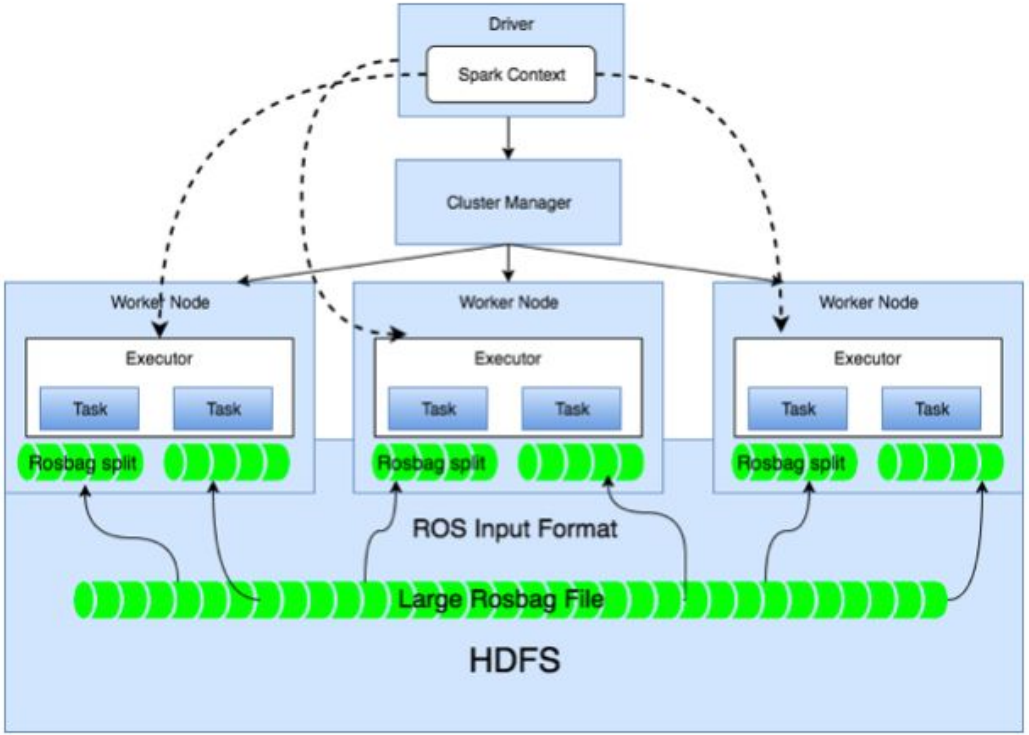
ros_hadoop

RosbagInputFormat is an open source splittable Hadoop InputFormat for the rosbag file format under Apache 2.0

Requirements

ROS installation on each node
Manual index extraction
HDFS and Spark installation

Release July 2017 with Apache 2.0

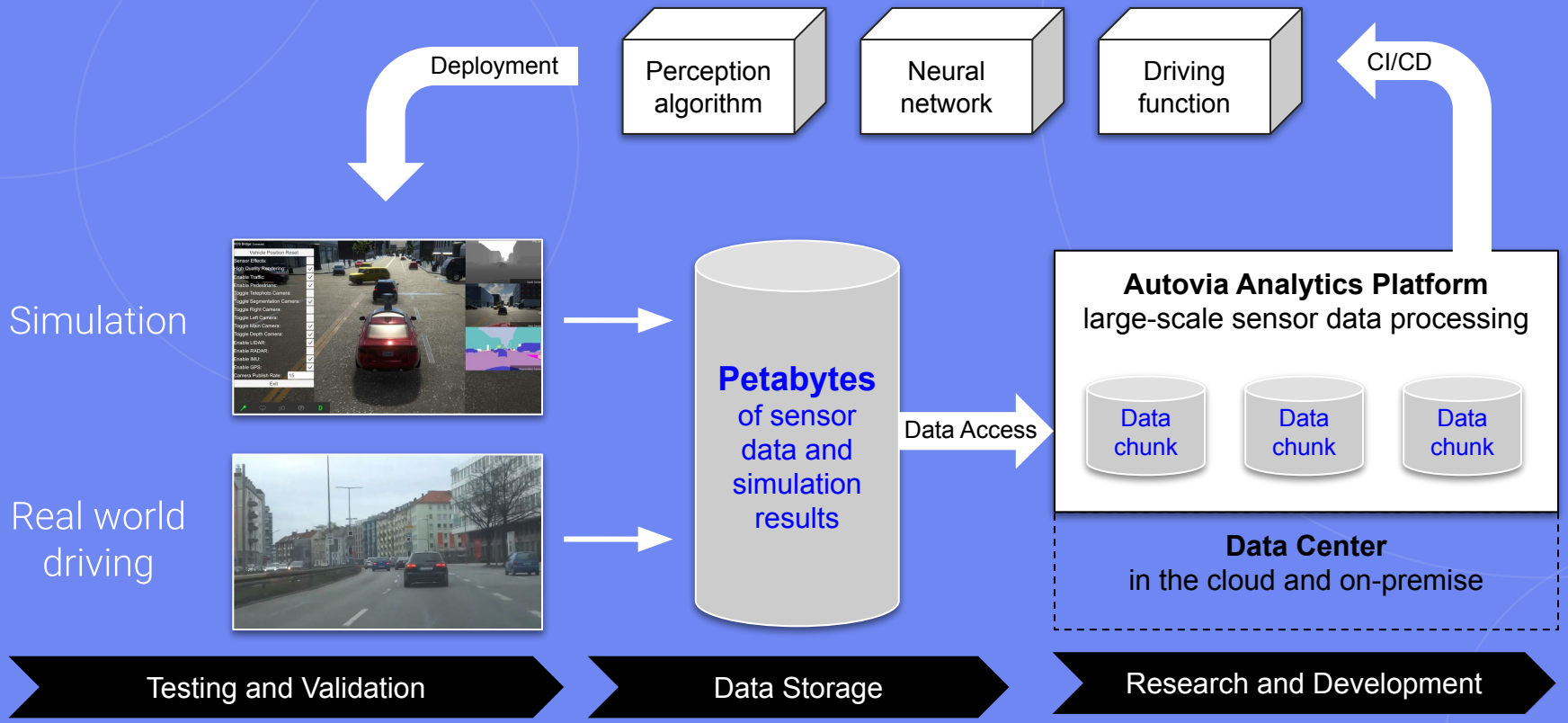


https://discourse.ros.org/t/working-with-large-ros-bag-files-on-hadoop-and-spark/2314

© Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Autovia Solution

Complete development & testing lifecycle



© Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Autovia IO

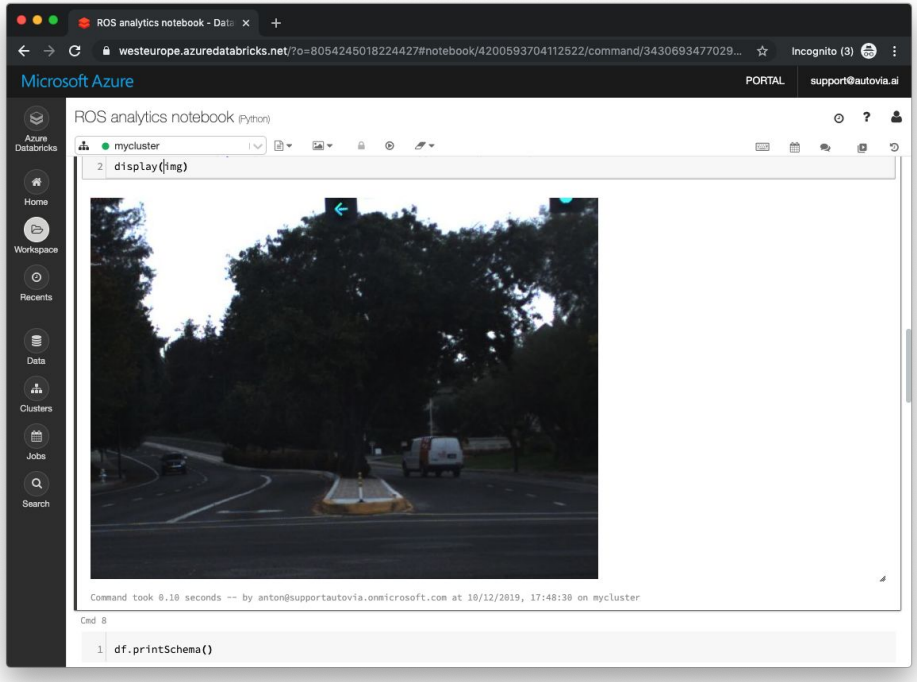
is a data access layer that enables analytics apps to read ROS bags from local and cloud storage

Not needed

- ROS installation
- Data converting & copying

Supported platforms

- AWS EMR
- Azure HDInsight
- GCP Dataproc
- Databricks
- Hadoop
  - Cloudera
  - Hortonworks
  - MapR



Autovia IO

Blob Storage

aws Azure Google Cloud

 © Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Autovia IO: Load Rosbag

Cmd 2

```
1  file_path = "/dbfs/mnt/ros/dataspeed/mkz_20151207_extra.new.bag"
2
3  msgs_path =
   "/dbfs/mnt/ros/msgs/common_msgs:/dbfs/mnt/ros/msgs/tf2_msgs:/dbfs/mnt/ros/msgs/can_msgs:/dbfs/mnt/ros/msgs/velodyne_msgs:/d
   bfs/mnt/ros/msgs/std_msgs:/dbfs/mnt/ros/msgs/dynamic_reconfigure:/dbfs/mnt/ros/dataspeed"
4
5  dfmeta = spark.read.format("ai.autovia.ros")\
6      .option("ai.autovia.fstype", "fs")\
7      .option("ai.autovia.schematype","metadata")\
8  .load(file_path)
9
10 df = spark.read.format("ai.autovia.ros")\
11                .option("ai.autovia.fstype", "fs")\
12                .option("ai.autovia.ros.msgdir", msgs_path)\
13                .load(file_path)
```

```
dfmeta: pyspark.sql.dataframe.DataFrame
  version: string
  connections: map
      key: integer
      value: struct
          conn: integer
          topic: string
          type: string
          md5Sum: string
          message_definition: string
  index: array
      element: long
```

Command took 0.28 seconds -- by adrian@supportautovia.onmicrosoft.com at 10/12/2019, 20:11:48 on mycluster

 © Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Autovia IO: List connections

Cmd 3

```
meta = dfmeta.collect()[0].asDict()
```

▸ (1) Spark Jobs

Command took 0.06 seconds -- by adrian@supportautovia.onmicrosoft.com at 10/12/2019, 20:12:37 on mycluster

Cmd 4

```
[(v.conn, v.topic) for v in meta["connections"].values()]
```

```
Out[13]: [(0, '/can_bus_dbw/can_rx'),
 (1, '/oxford_gps/gps/fix'),
 (2, '/oxford_gps/gps/vel'),
 (3, '/oxford_gps/imu/data'),
 (4, '/tf'),
 (5, '/velodyne/points')]
```

Command took 0.03 seconds -- by adrian@supportautovia.onmicrosoft.com at 10/12/2019, 20:21:17 on mycluster

 © Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Autovia IO: Filter “/oxford_gps/gps/vel” and plot data

Cmd 5

```
arr = df.filter(df.conn == 2).select(\
            "data.conn2.twist.twist.linear.x",
            "data.conn2.twist.twist.linear.y",
            "data.conn2.twist.twist.linear.z",
            )
```

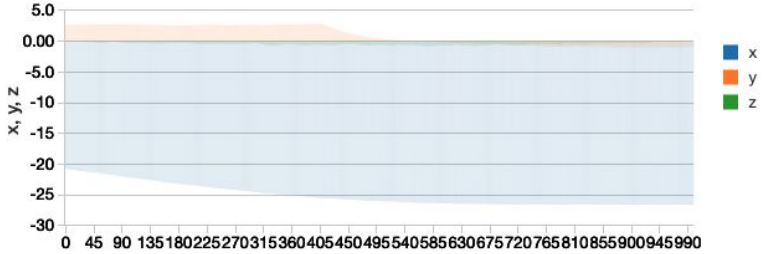
arr: pyspark.sql.dataframe.DataFrame = [x: double, y: double ... 1 more fields]

Command took 0.05 seconds -- by adrian@supportautovia.onmicrosoft.com at 10/12/2019, 21:00:08 on mycluster

Cmd 6

```
display(arr)
```

(3) Spark Jobs



Showing sample based on the first 1000 rows.

Plot Options...

Command took 35.03 seconds -- by adrian@supportautovia.onmicrosoft.com at 10/12/2019, 21:02:19 on mycluster

 © Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Autovia FS

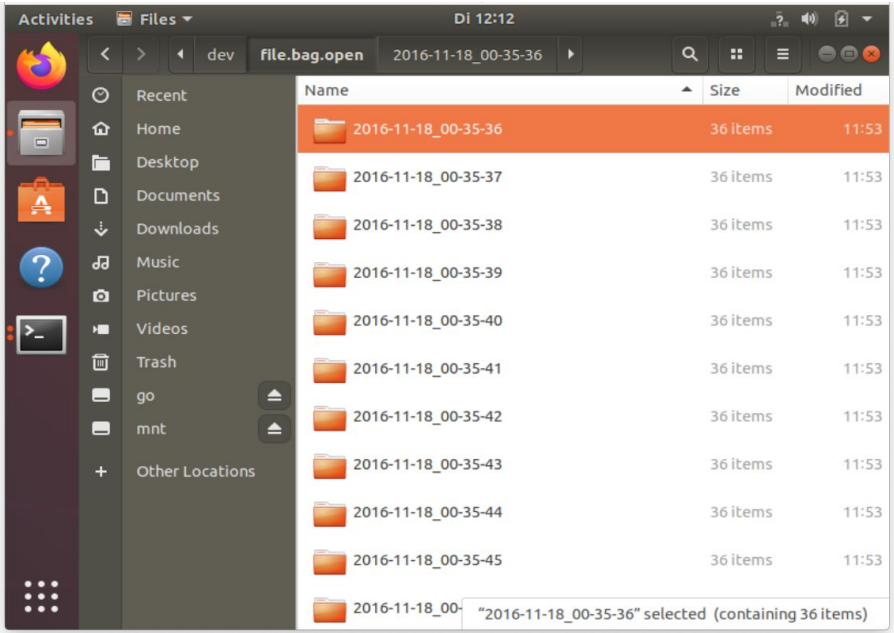
is a virtual filesystem that enables all applications to read ROS bags from local and cloud storage

Not needed

ROS installation
Data converting & copying

Start filesystem

```
$ av fs /home/<username>/mnt
```



 © Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Autovia FS

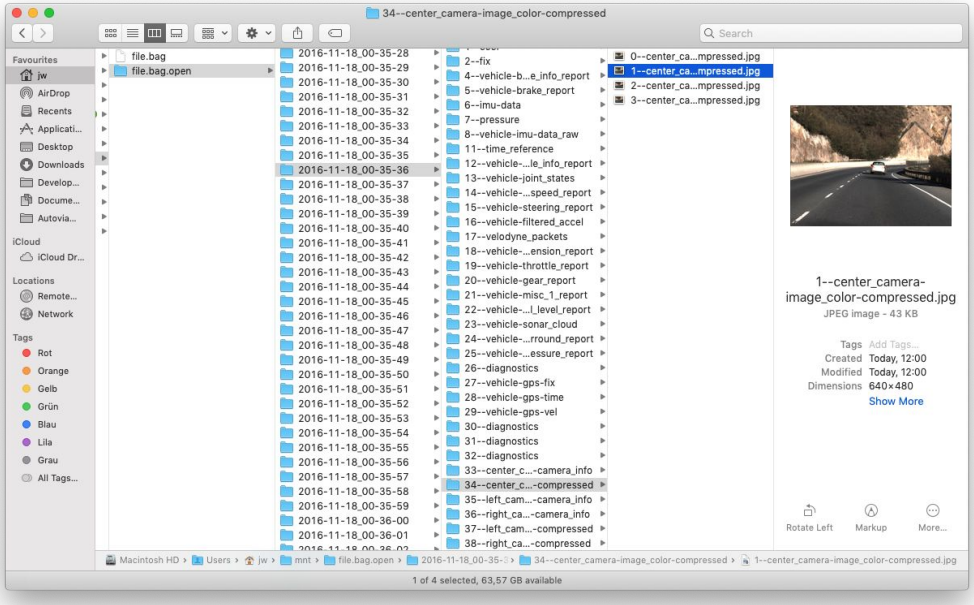
is a virtual filesystem that enables all applications to read ROS bags from local and cloud storage

Not needed

ROS installation
Data converting & copying

Start filesystem

```
$ av fs
/Users/<username>/mnt
```



 © Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Autovia FS

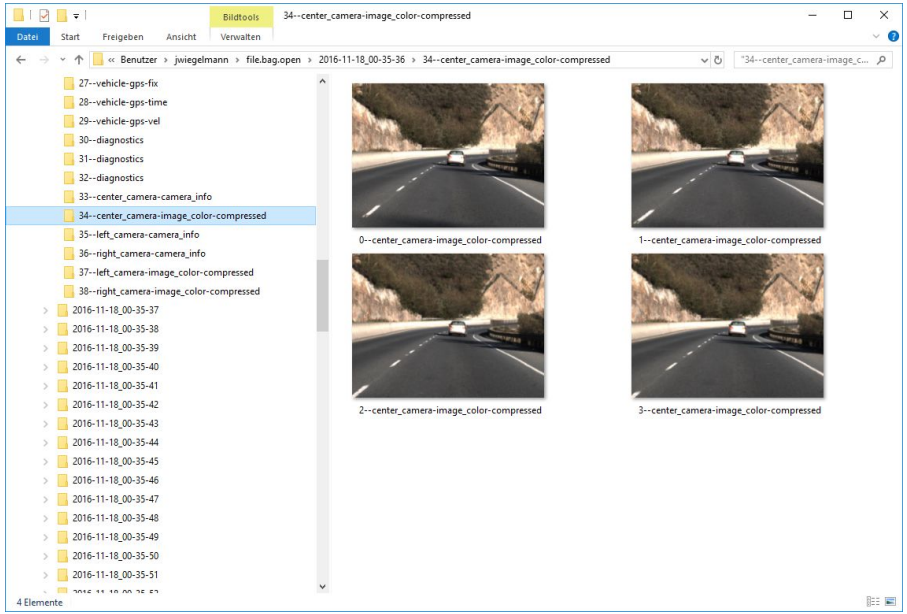
is a virtual filesystem that enables all applications to read ROS bags from local and cloud storage

Not needed

ROS installation
Data converting & copying

Start filesystem

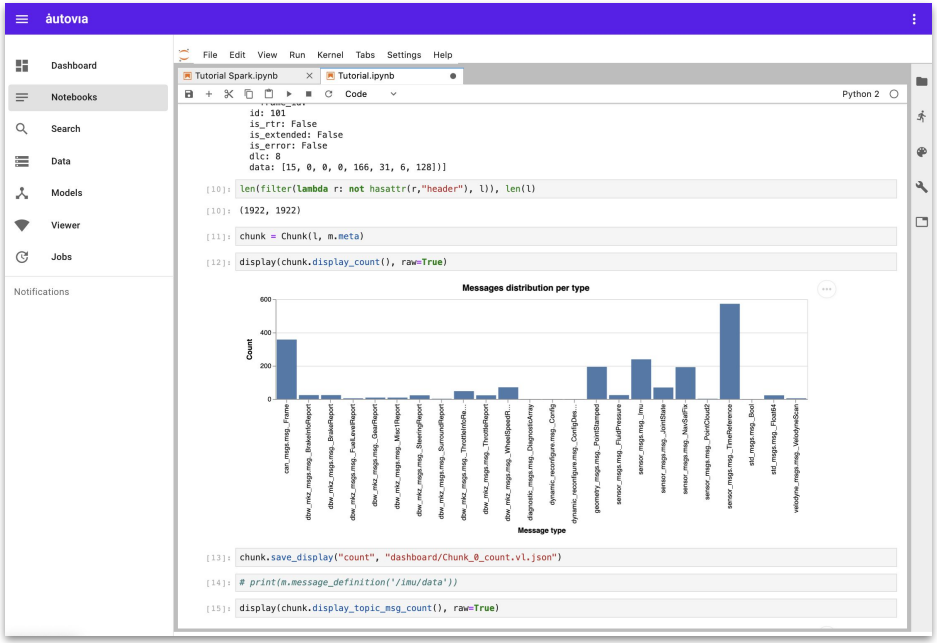
```
av.exe fs C:\Users\<username>\mnt
```



 © Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Autovia IDE

For R&D engineers in automotive and aviation industry we built an analytics toolbox running as a managed service. Autovia works on existing big data infrastructure, on bare metal or on top of existing Hadoop or Spark clusters. It can ingest R&D data directly from HDFS, S3, Azure Data Lake or any other data source.



Autovia Workspace

Your Data | Your Algorithm | Your Model

Autovia Runtime

ROS | Apache Spark | TensorFlow

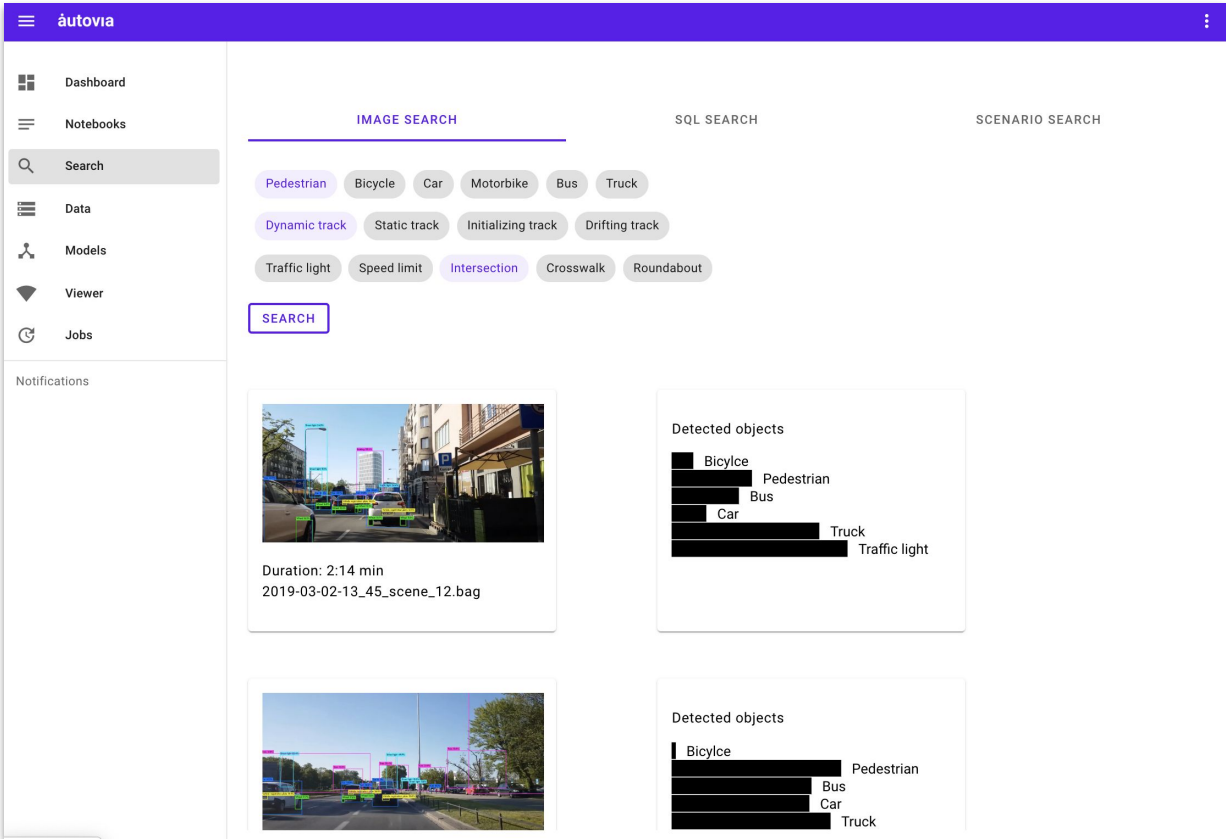
Autovia Cloud Service

aws | Azure | Google Cloud

Fast search at sensor level

Find the right scenarios for development, testing and validation

- 👍 Query results within seconds
- 👍 Search in raw sensor data
- 👍 OpenSCENARIO compatible
- 👍 Various data source support
- 👍 Find train & test data for AI

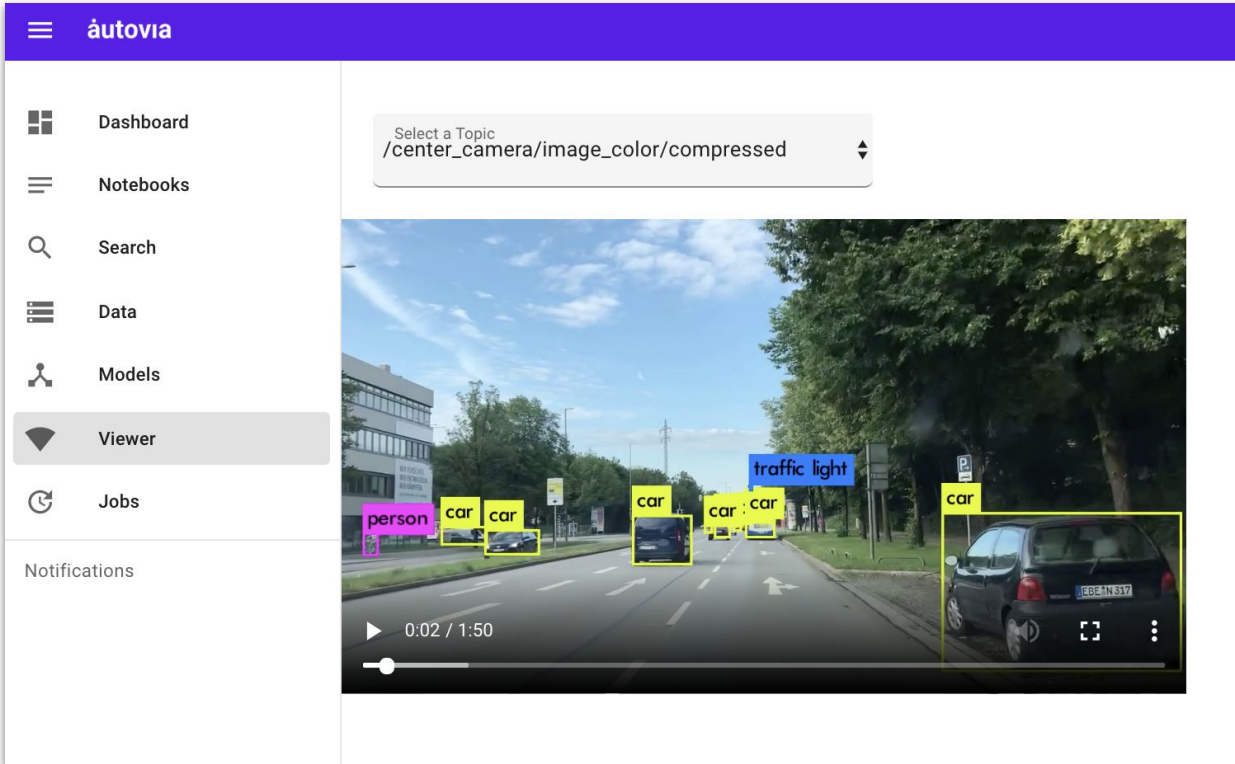


 © Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

Explore and validate data at sensor level

Test and see recorded driving scenes

- 👍 Built-in 3D point cloud viewer
- 👍 Vision output after perception
- 👍 See sensor fusion results
- 👍 Metadata management
- 👍 Validate algorithms



 © Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai

autovia

Thank you!

Jan Wiegelmann

jan@autovia.ai

http://autovia.ai

© Autovia GmbH 2019 | Confidential & Proprietary | www.autovia.ai