
APACHE AVRO INTRODUCTION CCA-175 CERTIFICATION EXAM

By www.HadoopExam.com

Note: These instructions should be used with the HadoopExam Apache Hadoop: Professional Trainings. Where it is executed and you can do hands on with trainer.

Avro File

⇒ Use of Avro files are

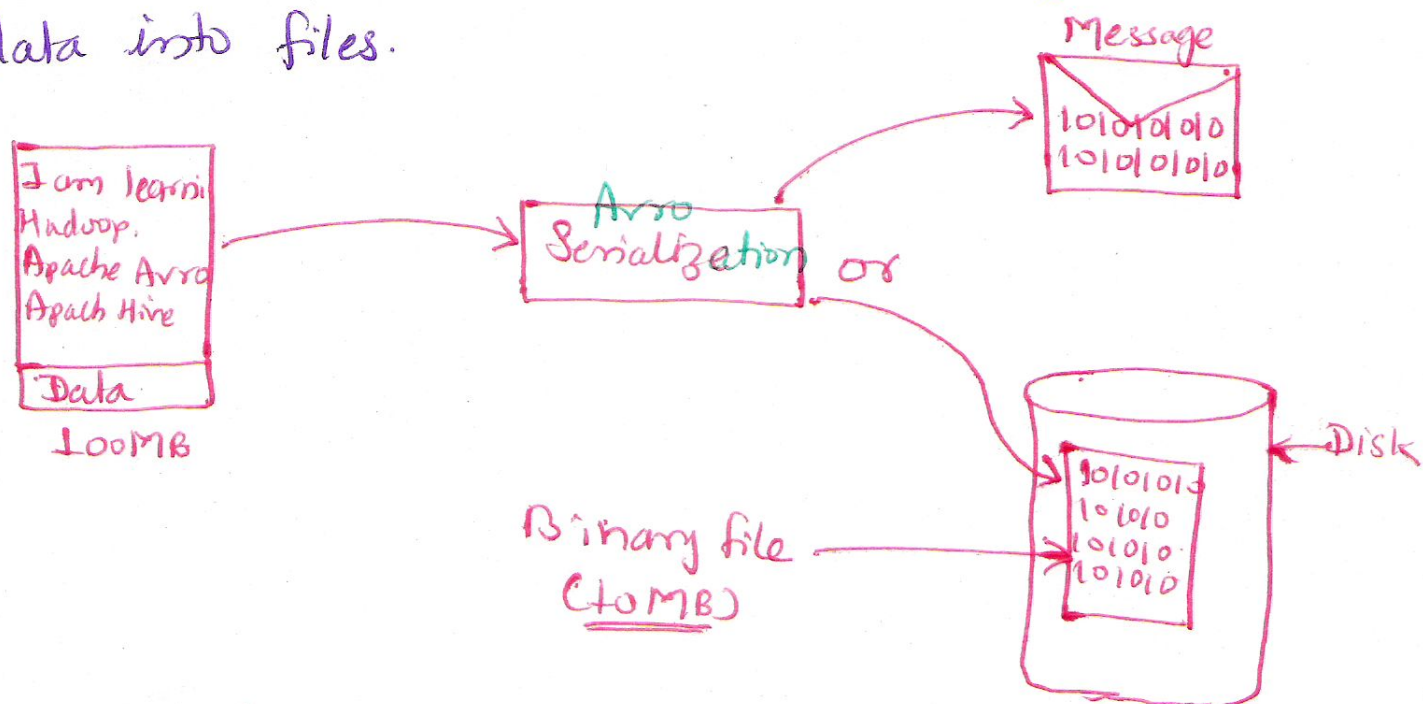
① Data Serialization

② Data Exchange (RPC)

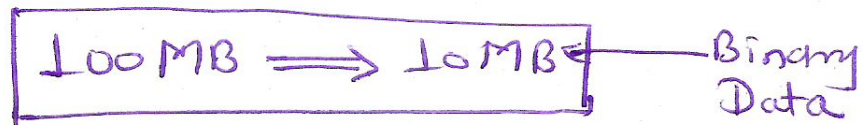
⇒ Avro files are language neutral.



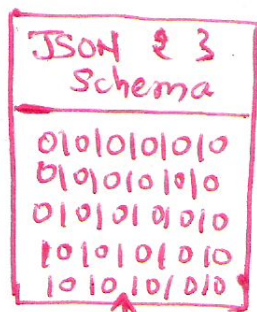
Avro * Serialization: Program can efficiently serialize data into files.



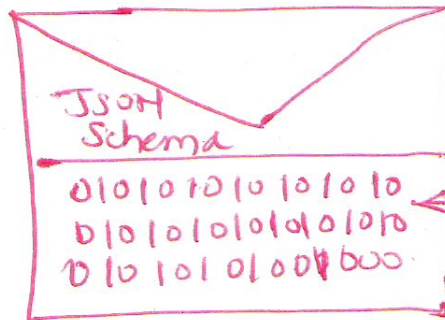
⇒ Data Storage is compact & efficient.



⇒ Avro stores both the data definition and data together in one message or one file.

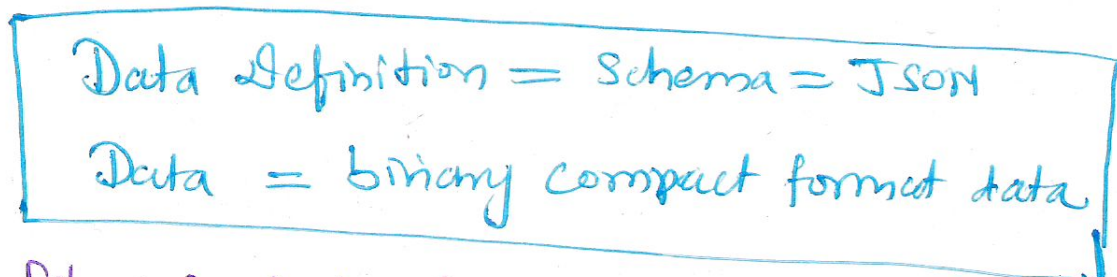


Actual Data



Message
Payload
(Binary)

⇒ Avro stores data and its definition (schema) together. Which helps program to understand the information stored in an Avro file or message.



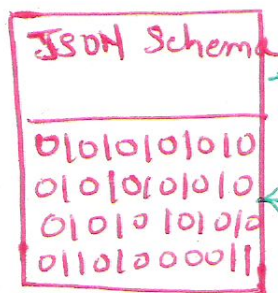
⇒ Avro files include markers that can be used to splitting large datasets into subset (HDFS) blocks which make it suitable for MapReduce processing.

Schema Evolution

⇒ It's a key feature of Avro in this case Schema can change over time. for instance.

→ Adding a field

→ Removing a field



← This can change over time.

← Data may not change

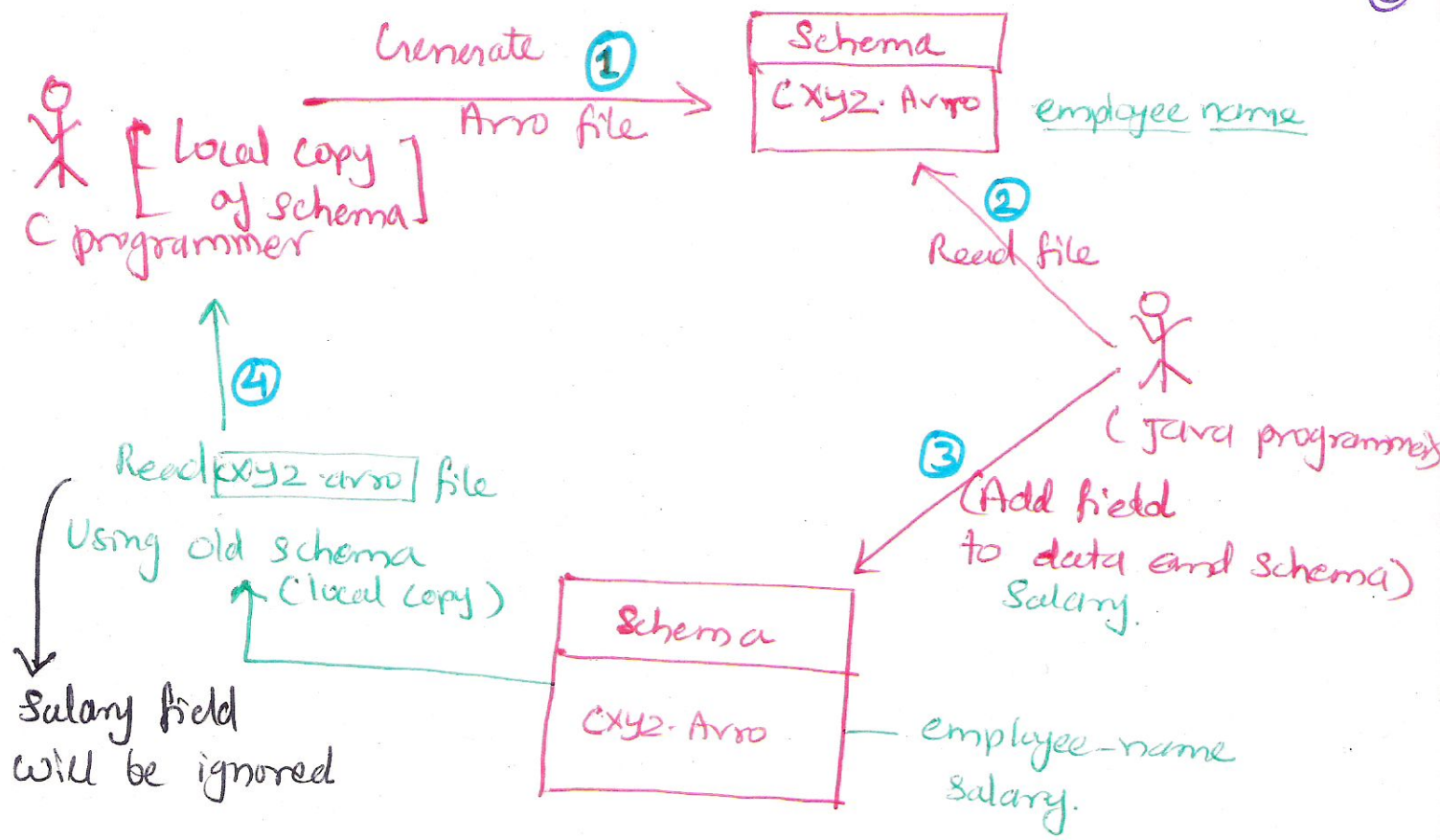
Still there is no problem for reading and writing data with changed schema or old (unchanged) schema.

⇒ Avro cleanly handles Schema changes like

- ① Missing fields
- ② Newly added fields
- ③ changed fields.

Because of this

- ① Old program can read ~~new~~ data [Based on new schema]
- ② New program can read ~~old~~ data [Based on old schema]



⇒ Summary :-

- ① Apache Avro is a framework that allows you to serialize data in a format that has Schema built in
- ② It serialize data in a compact binary format that does not require proxy objects or code generation.
- ③ Instead of using generated proxy libraries and strong typing, Avro relies heavily on the schemas that are sent along with the serialized data.

- ④ Including Schemas with the Avro message allows any application to deserialize the data.
- ⑤ Avro Schemas describe the format of the message and are defined using JavaScript Object Notation (JSON)

Visit WWW.HadoopExam.com for all the lectures (40 Modules completed) Including Hands on Session as well.

Become BigData / Hadoop Expert in few days

— * —

HadoopExam Learning Resource provides the following material for the Advanced Technologies.
Please visit www.HadoopExam.com for more detail this is just a few products from portfolio.

Price start for training with Just \$79/3500INR



Apache Spark
Professional
Training with
HandsOn Session

+ Certification
Material



Hadoop Professional
Training with
HandsOn Session

+ Certification
Material



HBase Professional
Training with
HandsOn Session

+ Certification
Material



Certification
Material



Certification
Material



Certification
Material



Certification
Material



Certification
Material



Microsoft Azure

Certification
Material



Certification
Material