Working With ADF At Scale learnings from the last 4+ years

April 9, 2019

Allotrope has been working for perhaps 4 years on developing and deploying Allotrope Data Format capability

- We have learned that there are approximately 3 aspects to effective deployment of Allotrope ADF technology
    1. Establishing how to model (instrument/measurement)
    2. Establishing how to encode the (instruments/measurement) in ADF format
    3. Establishing how to manage a corpus of ADF files

Modeling Approaches: three main forms

- (instrument/measurement) specific modeling
  - Typically creating a fully BFO aligned model
- Functional specific pattern modeling
  - A pattern for samples
  - A pattern for peaks
  - Etc.
- Generalized meta-pattern modeling
  - Leaf nodes
  - Generalized data cube models

Once modeled, we need an approach for encoding data in ADF files

- Again typically three approaches
  1. Hand encoding using text editors
  2. Specific purpose programming to encode the model
  3. Automation to take simple data structures and automate encoding of the models

Managing a corpus of ADF files

- Again, approximately three approaches are being explored
    1. Traditional file systems
    2. Big Data approaches (e.g. Hadoop with schema on read designs)
    3. Large graph data bases

The challenge is to be able to do this at scale

- How do we model new (instruments/measurements) at scale?

- How do we generate ADF files at scale?

- How do we manage ADF files at scale?

Observations on scale

- Hand crafting does not scale well

- To scale at the ADF file we need some level of automation/abstraction:
  - Modeling of (instruments/measurements) at scale can be accelerated with meta-patterns
  - Meta-patterns are also highly amenable to automation

- To scale at ADF data/file management the storage environment needs to comprehend the ADF structure
  - E.g. ADF files are based on HDF5, and HDF5 does not work well in some sharded file environments

Hypothesis: modeling of and creation of ADF files

- Meta-patterns allow us to:
  - have a generalized modeling approach for large numbers of (instruments/measurements)
  - Develop generalize code for writing ADF files

- Example meta-patterns:
  - Leaf node model for encoding name-value pairs
  - Data cube structure for OLAP cubes/graphs/tables

This leads to what is called "ADF files as a service"

- A generalized service which can:
  - Using simple data structures encode ADF files
  - Decode ADF files back into simple data stuctures

- This prototyping of ADF as a service was first reviewed At Waldbronn meeting. A number of teams:
  - Allotrope ontology working group
  - Amgen
  - Tetrascience has been piloting on this

- Tetrascience will now discuss their pilot efforts at ADF as a service

TetraScience discussion of ADF as a service

Vincent Chan

Previously...

TETRASCIENCE

ADF File Converter

ADF is the Allotrope Data Format. Learn more about the ADF and Allotrope Foundation here. To understand more about the Allotrope framework, you can read our guide Introduction to Allotrope. This ADF converter will follow the leaf node approach described in the guide.

Drag and drop your file here or
Browse your local device

Instructions:

✎ Format your excel file according to our template here.

👆 Drag and drop your selected file above or browse your local computer.

⬇ Download the final ADF file, data description file and SHACL file.

ⓘ Terms & Conditions

JSON/Excel to ADF

Allotrope Conversion
as a Service

JSON/Excel to ADF

ADF to JSON

ADF to JSON

Allotrope Conversion
as a Service

LeafNode.ttl
LeafNode.shacl.ttl
LeafNode.adf

REST API   **POST** LeafNode.xlsx –OR– **POST** LeafNode.adf
GET /workflows/:workflowId/LeafNode.json

Demo

https://allotrope.tetrascience.com
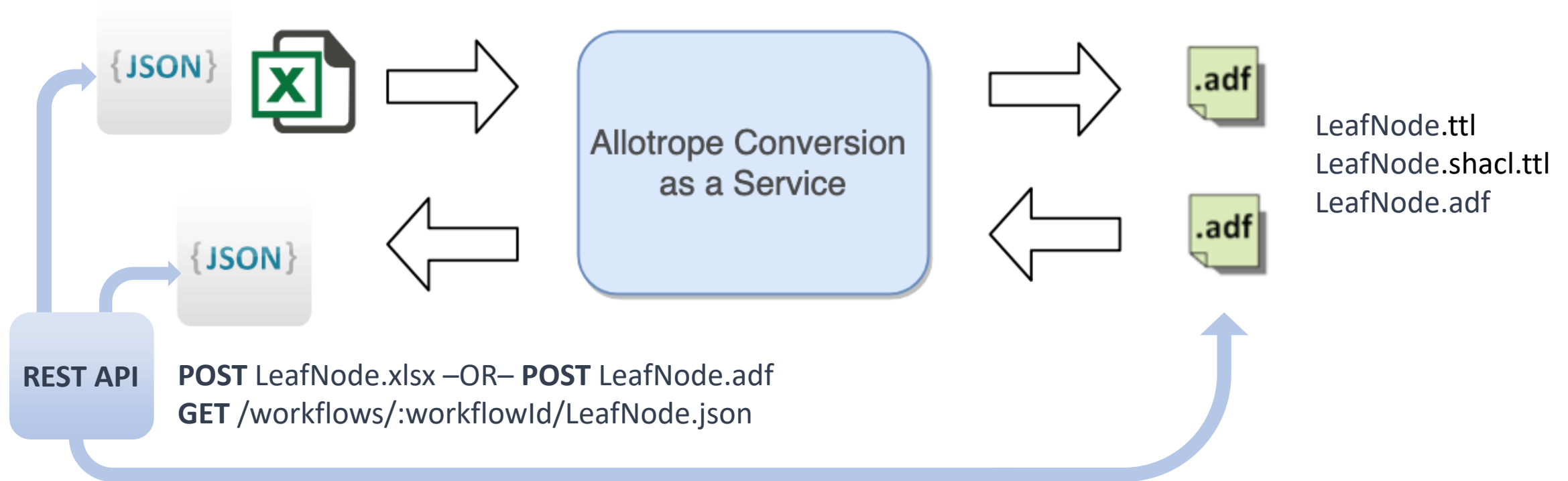
Configurable

Leaf Node
or
Full Graph

HDFS cluster

Azure

AWS

Amazon Neptune

Azure Cosmos DB

W3C SPARQL

Convert

Conversion Module

Store

Storage Module

Merge

Graph DB Module

Query

Query Management Module

Allotrope Conversion
as a Service (ACAS)

Next Steps

- Test and exploit ADF as a service to accelerate creating ADF files at scale

- Learn how to manage ADF files at scale
  - For next meeting?

- Start getting value out of using ADF files at scale