

2024 GTS Intern Final Presentations



xLPT S3 Data Migration

Srinidhi Jayaprakash

Network Data Lake Team - AI&D

Manager: Anil Padala

Mentor: Freddy Rodriguez, Aliza Rockstead,

Stacy Dalton

VP: Michael Raj

SVP: Kalyani Sekar



**Class of 2026
Computer Science**



Problem: xLPT Data is currently retrieved using **legacy web service** calls. About 1200 calls/hr are made by NDL. This means:

- Data must be **manually requested and pulled**
- Slow response time and bad performance
- Data quality issues for downstream systems

Solution: Migrate xLPT Data to **Openshift S3** ingestion pipelines.

- Data is **automatically pushed**
- Better data quality, speed, and reliability of data



Completed migration of 2 xLPT datasets

- Created Apache NiFi flows to route S3 files to the Google Cloud Platform
- Executed Jenkins pipeline to create table schemas in BigQuery for data feeds
- Inserted data into BigQuery tables.

DataLake Architecture





ListS3-JSON-enb_sec_carrier_hrly

ListS3 1.23.1.2.1.6.0-323

org.apache.nifi - nifi-aws-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	1 / 00:02:49.783	5 min

Name **success**

Queued 0 (0 bytes)



RouteOnAttribute-FilterJSON

RouteOnAttribute 1.23.1.2.1.6.0-323

org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name **JSON_filter**

Queued 0 (0 bytes)

How can we make sure all files are **loaded consistently**?

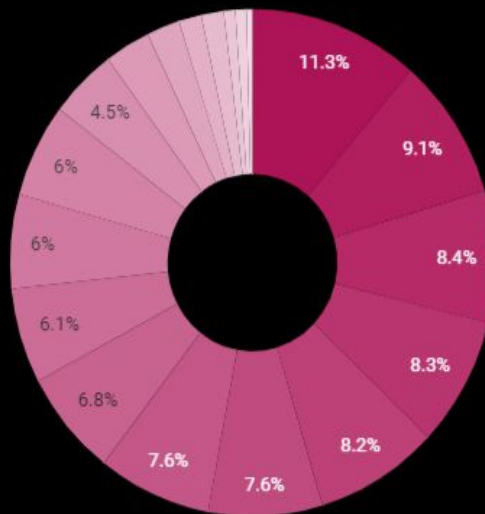
xLPT Data Tracking Dashboard

File Count per Market Location for enb_sec_carrier_hrly_preproc_v2 Dataset

mme_location / file_count								
trans_dt	MTPL	Mid...	Mic...	Car...	S	Gre...	Geo...	FI Grand total
Jun 7, 2024	360	288	264	264	264	240	240	3,180
Jun 6, 2024	360	288	264	264	264	240	240	3,184
Jun 5, 2024	360	288	264	264	264	239	240	3,185
Jun 4, 2024	360	288	264	264	264	240	240	3,181
Jun 3, 2024	360	285	264	264	264	240	240	3,177
Jun 2, 2024	360	288	264	264	264	240	240	3,170
Jun 1, 2024	360	288	264	264	264	240	240	3,170
May 31, 2024	360	286	264	264	264	240	240	3,180
May 30, 2024	360	288	264	264	264	240	240	3,183
May 29, 2024	360	288	264	264	264	240	240	3,184
May 28, 2024	360	286	264	264	264	240	240	3,182

Select date range

mme_location



- MTPL
- MidWest_IL_WI_LRA
- Michigan_Indiana_Kentucky
- Carolinas_Tennessee
- S
- Great_Plains
- Georgia_Alabama
- Florida
- Illinois_Wisconsin
- Southwest
- N
- Pacific_Northwest
- Ind_Kent_LRA
- Pacific_Northwest_LRA
- Great_Plains_LRA
- Car_Tenn_LRA
- Alaska
- Hawaii
- Houston_Gulf_Coast

Savings:

Web service calls

daily_nokia_fsmkpsall_carrier_raw_v1



daily_nokia_fsmkpsall_carrier_preproc_v1

Json migration

Saved 10 Web service calls daily.

Web service calls

enb_sec_carrier_hrly_raw_v2



enb_sec_carrier_hrly_preproc_v2

Json migration

Saved 500 Web service calls daily.

Including reruns: Additional 30% of Web service calls saved. In total...

About 660 Web service calls to 0 Web service calls.

Benefits include: Reduced computational needs, reduced operational issues, and improved data quality.

Next Steps:

- **Expand the dashboard by generalizing it to many datasets.**
- **Create on demand alerts, so the Data monitoring team can directly use the dashboard to go back and fix the data where there are holes/mistakes.**
- **Complete migration of the rest of the xLPT Datasets, include these loads in dashboard.**

Internship Takeaways



Google Cloud Platform



Google
Big Query



kafka



Jenkins



Presenting

Project planning,
managing tasks



Communicating

Problem Solving



Thank You!

Kalyani Sekar

AI&D Leadership Team

Anil Padala

Stacy Dalton

Samantha Vega

Katie White

Aliza Rockstead

Caroline Osmani

Christine Martinez

Alexander Lee

Kayla Chao

Freddy Rodriguez

Michael Raj

Alan Myers

Susan Kelly

Q&A

