

ENVS 5726 Guest Lecture

*Building Data Pipelines to Support
Mineral Exploration Activity*

Abdel Alfahham
alfahham@sas.upenn.edu

Founding Data Engineer

TABLE OF CONTENTS

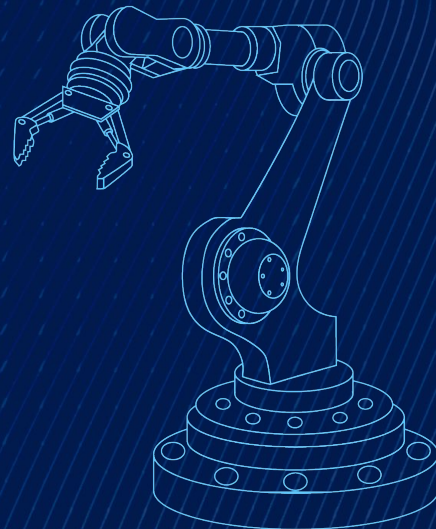
01 Introduction

02 Geophysical Surveys Fundamentals

03 Building Data Pipelines

04 Demo

05 Conclusion





01

Introduction

Summary and Objective

Career Timeline

ENVS 5726

- BS Environmental Sciences [2009 - 2013]
- Environmental Consulting [2013 - 2018]
- Penn MSAG [2016 - 2019]
- Data Engineer [2019 - Current]

- Self Taught Python (coding in general)
- Intro to SQL in Environmental Consulting
- Intro to GIS/QGIS/CAD in Environmental Consulting



Founding Data Engineer
ecue ai · Full-time

Jul 2024 - Present · 1 yr 6 mos
New York, New York, United States · Remote



Software Engineer
Grata · Full-time

Sep 2022 - Aug 2024 · 2 yrs
New York, New York, United States · Remote



KoBold Metals
Full-time · 2 yrs 11 mos

San Francisco Bay Area · Remote



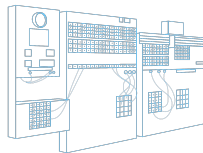
Data Engineer
Nov 2019 - Sep 2022 · 2 yrs 11 mos



Lecture Objective

ENVS 5726

1. Provide a *detailed but practical overview* on the various components of an end-to-end *data platform*.
2. Introduce critical *data engineering, data infrastructure and software development concepts*.
3. Demonstrate the *impact of data engineering in earth science*.



- During this lecture I will be discussing and demonstrating a hypothetical pipeline that I created to share the knowledge I have learned from the past 7 years or so as a data engineer.
- The main objective is to share the fundamentals and give jumping off points for individuals that are interested in pursuing data engineering or data science roles.



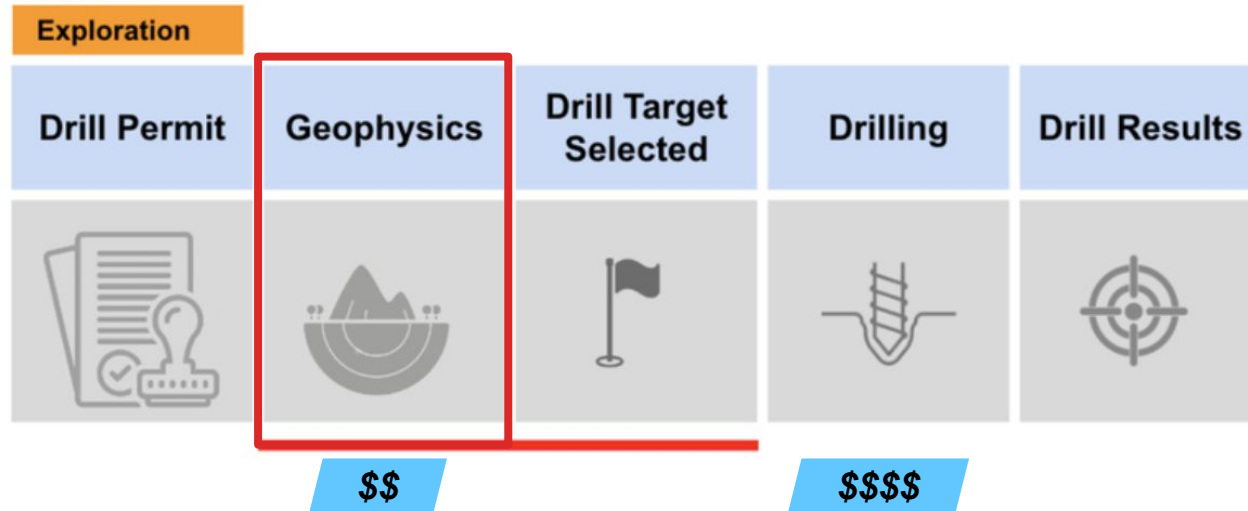
02

Geophysical Surveys

Ground EM Fundamentals

Geophysics in Mineral Exploration

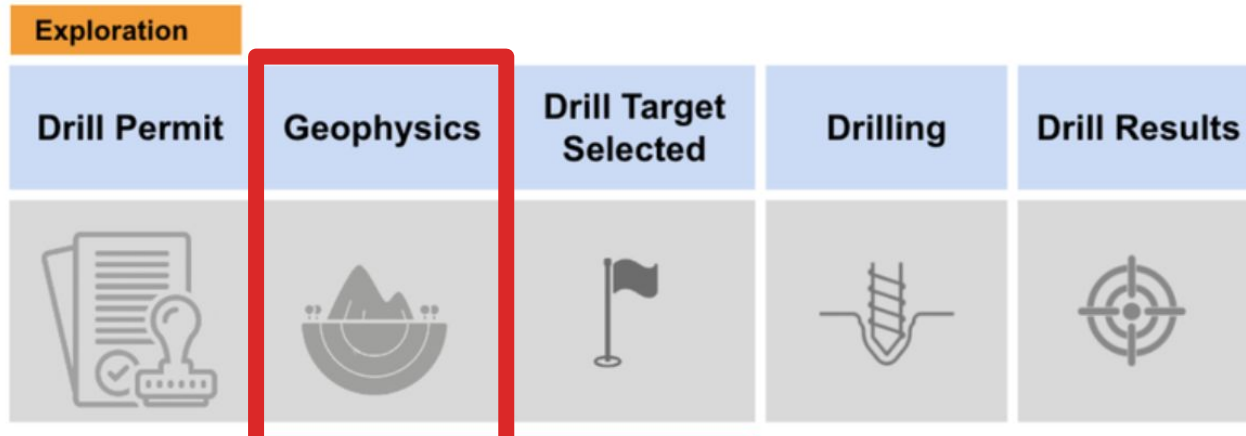
ENVS 5726



Geophysical surveys allows exploration teams to identify promising areas without having to dig or drill, saving time and money in the early stages of mineral exploration.

Types of Geophysical Surveys

ENVS 5726



**Electromagnetics
(EM)**

Gravity

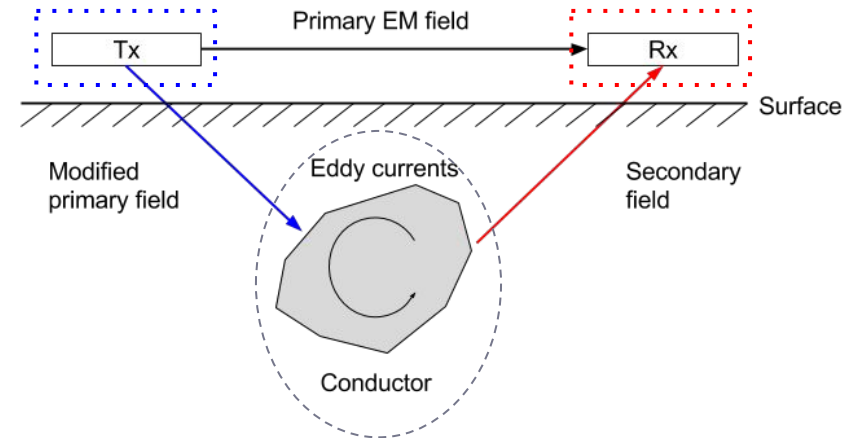
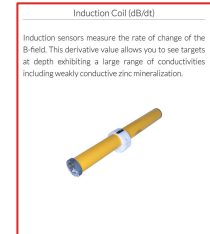
**Induced
Polarization (IP)**

Magnetics

Electromagnetic (EM) Surveys

ENVS 5726

1. An artificial electromagnetic signal from the **transmitter (Tx)** is generated and it penetrates deep into the earth.
2. When the signal hits **conductors** (conductive layers of rock or minerals like nickel, copper, and other base metals) it changes.
3. A **receiver (Rx)** picks up these changed signals and the information is saved.



Electromagnetic (EM) Surveys

ENVS 5726

What it measures:

- Ground EM surveys detect how well electrical current flows through subsurface materials.
- Highly conductive materials (like ore deposits) conduct electricity easily, while resistive materials don't

How it works:

- An artificial electromagnetic signal is transmitted into the ground, and the response is measured; conductive materials create stronger secondary signals that the receiver detects

Why it matters:

- Conductivity patterns reveal the location, size, and type of subsurface targets; anomalies indicate potential mineral deposits, geological structures.

Revisiting Electromagnetics (EM)

ENVS 5726

- **Strong Conductors and Anomalies:** Conductors and anomalies are prominent features on an EM survey, these are zones that conduct electricity particularly well. *Note: depending on the type of mineral the company is searching for, they may be looking for low conductivity.*
- **Shape and Depth of Anomalies:** The shape and depth of a conductive anomaly can help with modelling structures underground.
- **Context of Geological Features:** EM survey results are often used in conjunction with other geophysical data such as rock chip sampling, this allows an explorer to be further informed about the potential. For example, an anomaly along a well-known deposit belt or occurrence can serve to increase the confidence of prospectivity in a project.
- **Comparison with previous data:** If there is historical data and older survey results for the region, it can be compared with the new EM data to give the company a better understanding of its project.

Electromagnetic (EM) Survey Timeline

ENVS 5726

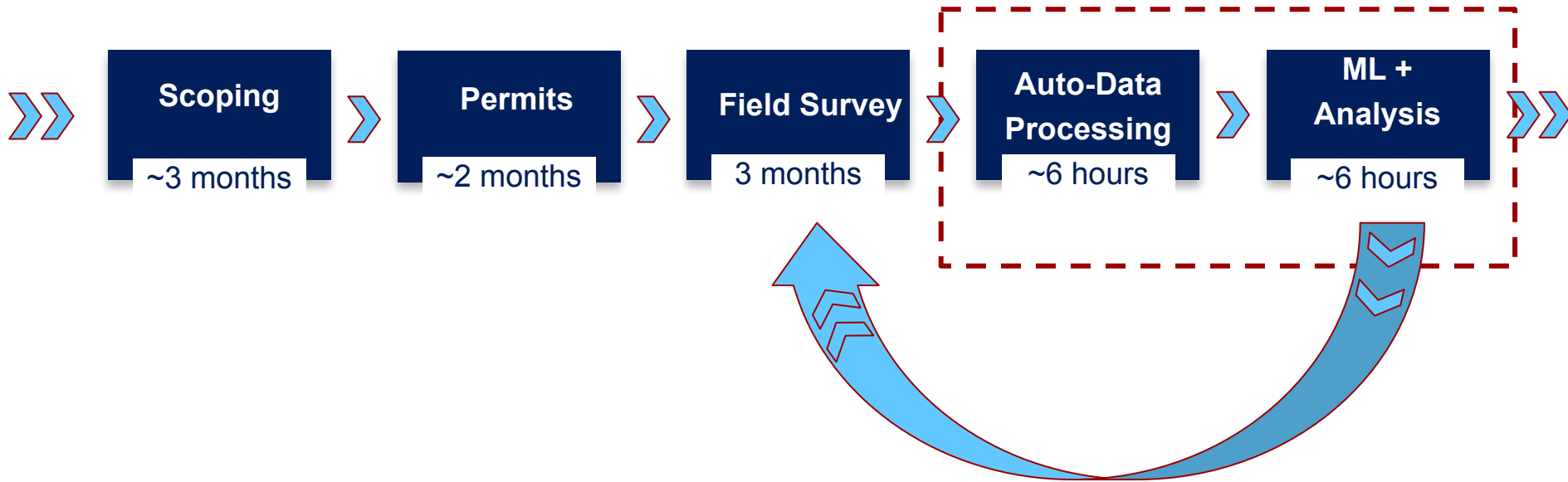
Conventional Survey Analysis Pace (12 Months)



Electromagnetic (EM) Survey Timeline

ENVS 5726

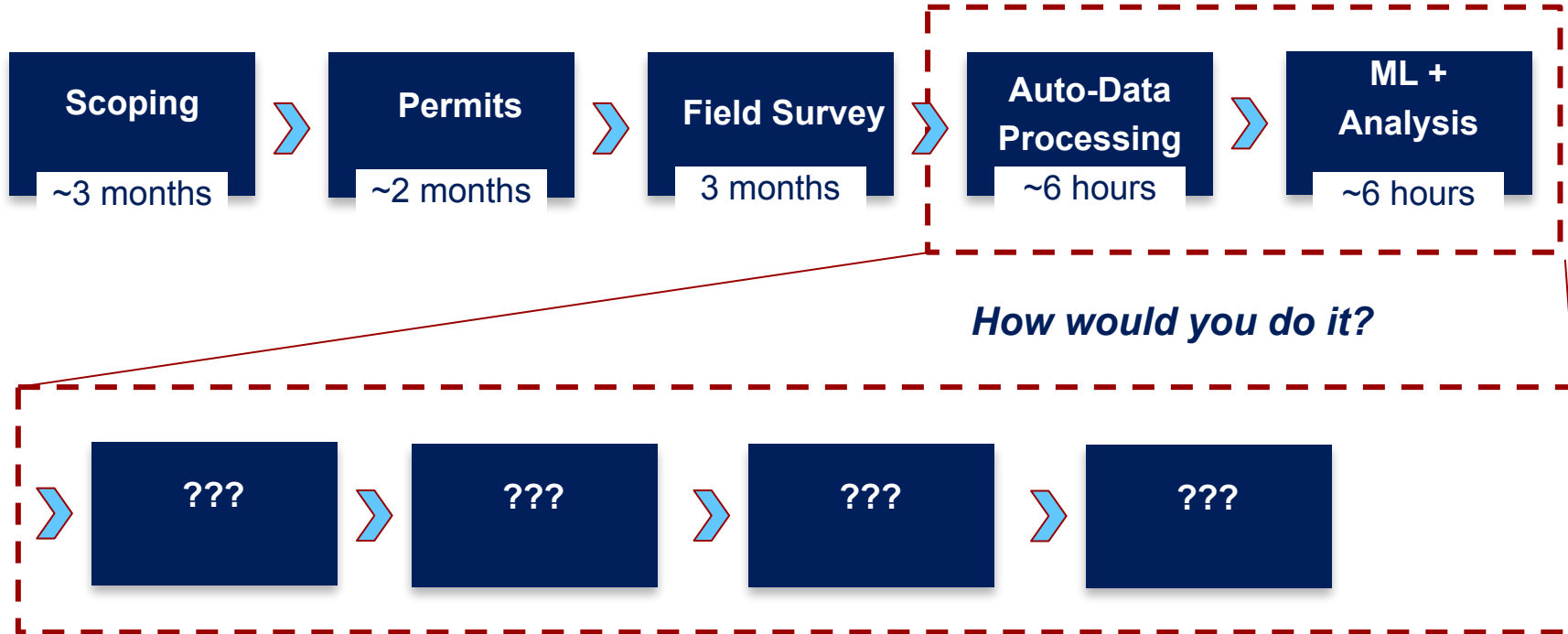
Proposed Survey Analysis Timeline (Fast Iteration)



Electromagnetic (EM) Survey Timeline

ENVS 5726

Proposed Survey Analysis Timeline (Fast Iteration)





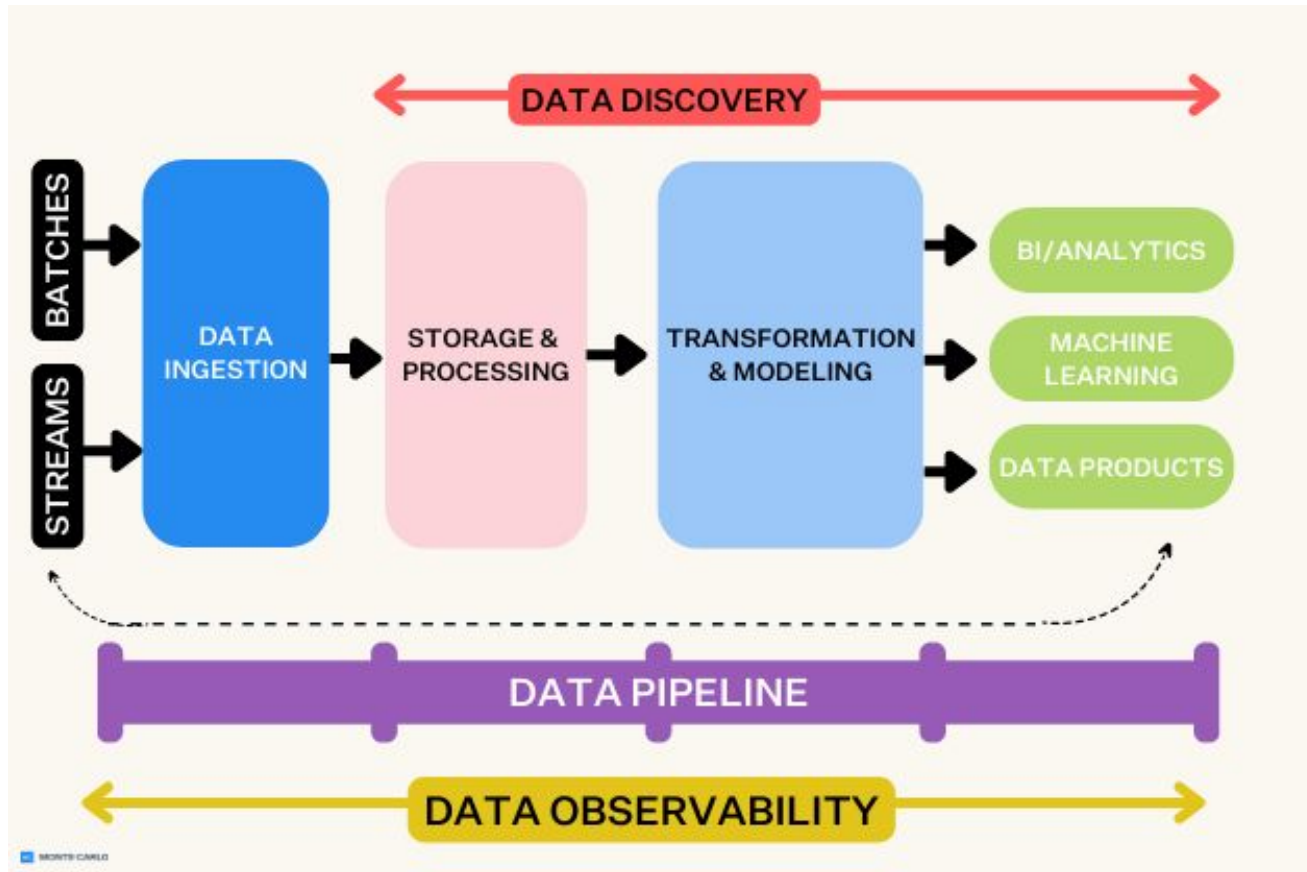
03

Data Pipelines

Automation, Scaling, Design

Basic Components of Data Platform

ENVS 5726



Let's Start the Demo Now!

ENVS 5726

```
<FMT> 230 ~ data format
1 <FMT> 230 ~ data format
2 <DIR> nanoTesla/sec ~ data units
3 <OPR> Josh Way ~ operator's name
4 <XYP> 0 0 0 ~ X-Y probe #, S0A, tool #, tool id
5 <CUR> 20.0 ~ peak current in loop (2 X meter reading)
6 <TXS> 000.0 700.0 ~ loop size (x y units) (units: 0 = m, 1 = ft)
7 ~ Transmitter Loop Co-ordinates:
8 <L00> 477180.2 7245172.5 666.6 0 ~ loop coord. | 4 numbers - (x,y,z,units)
9 <L01> 477098.4 7245207.5 667.0 ~ loop coord. | values for units:
10 <L02> 477180.1 7245222.5 672.6 0 ~ loop coord. | 0 = metres; 1 = feet
11 <L03> 477181.2 7245250.5 676.4 0
12 <L04> 477181.2 7245294.5 684.9 0
13 <L05> 477099.4 7245338.0 686.5 0
14 ~ Hole/Profile Co-ordinates:
15 <P00> 476497.0 7245021.5 672.3 0 0 ~ 0=Collar/Start| 5 numbers: (x,y,z,units,sten)
16 <P01> 476525.1 7245024.5 668.3 0 25.0 ~ [1-99 = (x,y,z)] values for 'units':
17 <P02> 476549.0 7245024.0 669.2 0 50.0 ~ [or segment's] 0 = metres; 1 = feet
18 <P03> 476575.0 7245024.0 665.8 0 75.0 ~ [grid azimuth,] 2=az(deg) dip(deg) length(m)
19 <P04> 476681.0 7245023.5 666.0 0 100.0 ~ [ve down dip,] 3=az(deg) dip(deg) length(ft)
20 <P05> 476625.9 7245025.0 665.6 0 125.0 ~ [segm't length] 'sten': surface PDM (WSS -ve)
21 <P06> 476656.1 7245024.5 666.8 0 150.0
22 <HE3> Data Scaled by Coil Area of 6500/4300
23 ~Tags for comment list: <GBN>, <STA0><Stn>, <STX><Stn>, <STY><Stn>, <STZ><Stn>
24 ~ <R00><Rdgp>, <R00>
25 ~Tags for List & plot comments: <SPAb><Stn>, <SPC><Stn>, <SPY><Stn>, <SPZ><Stn>
26 ~Tags for headings: <HE1>, <HE2>, <HE3>, <HE4>
27 North American Nickel
28 P-008
29 S0N
30 P-008
31 August 19, 2016
32 S-COIL Metric Crystal-Master 16.66 1500 10 42
33 #117 9.20 May19,2015Z L50N N 1000 4300 +
34
35 0.000e+00 0.000e+00 1.597e-05 3.193e-05 6.387e-05 1.277e-04 2.555e-04
36 5.109e-04 1.022e-03 2.844e-03 4.087e-03 8.175e-03 1.635e-02
37
38 0N XR8 0 A 1940.0 90 256 10 12
39 D4 49.074 -2310.1 41264 41328.7
40 49.07 0.2773 0.2773 0.004442 -0.003059 -0.004113 -0.002263
41 -0.002603 -0.001756 -0.00107 -0.03265
42
43 25N XR8 0 A 1940.0 90 256 10 32
44 D4 1270.99 -1865.06 35762.5 35833.7
45 1271 -0.1581 -0.1577 -0.05358 -0.008725 -0.003266 -0.001537
46 -0.0002606 0.0005181 0.0006375 -0.2959
47
48 S0N XR6 0 A 1940.0 90 192 10 52
49 D4 1433.13 -1711.27 32900 32975.7
50 1433 -0.2749 -0.2248 -0.05942 -0.0003389 0.003292 0.002578
51 0.004293 0.003807 0.002182 -1.143
52
53 75N XR9 0 A 1940.0 90 288 10 72
54 D4 1536.21 -1682.48 38739 38826.4
55 1536 -0.2921 -0.2921 -0.05601 0.0003726 0.002262 0.001559
56 0.002201 0.001282 0.0005714 -1.661
```



id	survey_id	line_id	survey_date	datum	projection	utm_zone	data_format	data_units	
7	P-246	NULL	2025-12-03 00:26:51.834163	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Morgan Glynn
8	P-142	NULL	2025-12-03 00:26:33.314972	WGS 1984	UTM	22N	230	picoTesla	~ data units Keith Falardeau
5	P-141	NULL	2025-12-03 00:26:44.344029	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Morgan Glynn
4	P-094	NULL	2025-12-03 00:26:38.871304	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Henry
3	P-040	NULL	2025-12-03 00:24:56.508166	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Morgan Glynn
2	P-022	NULL	2025-12-03 00:24:39.799894	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Morgan Glynn
1	P-008	NULL	2025-12-03 00:24:06.839264	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Josh

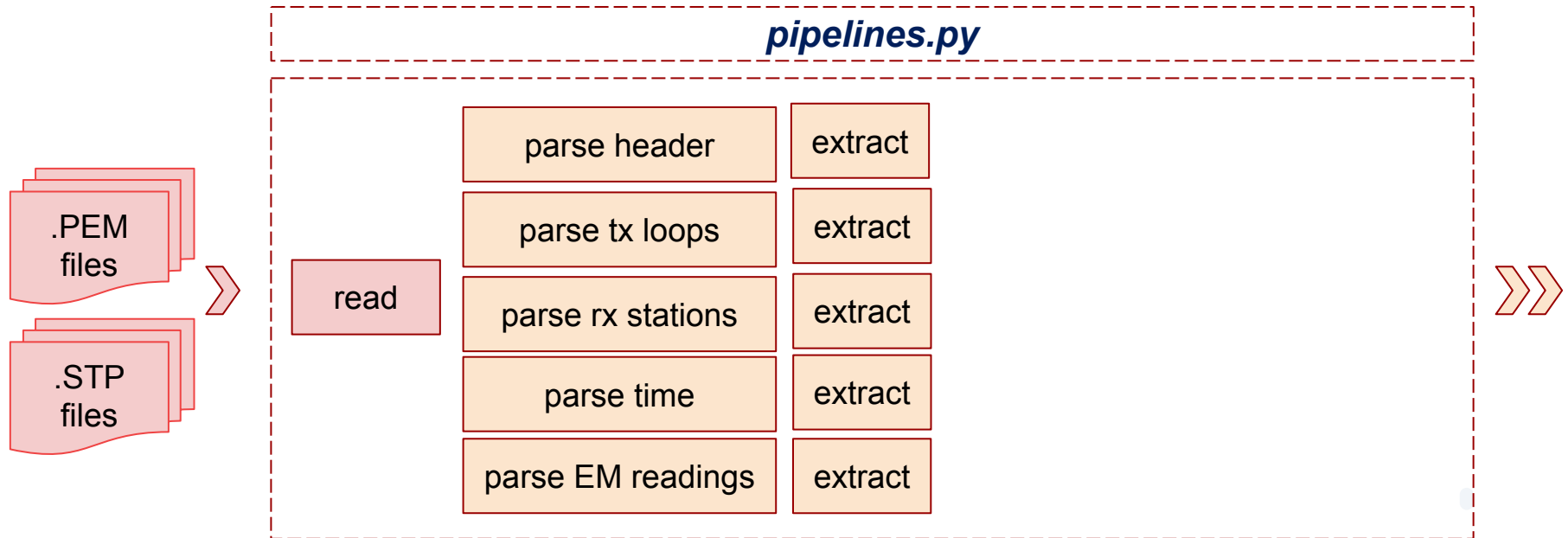
id	survey_id	receiver_station_id	component	source_file	station_label	receiver_code	primary_pulse_nt_per_sec	receiver_number	angle_degrees	num_s
1816	4+	53 + Z	L308AV_STP	0n	Z018			175300	NULL	NULL
918	4+	63 + Z	L1308AV_STP	250n	Z06			178300	NULL	NULL
1090	4+	53 + Z	L408AV_STP	0n	Z09			168000	NULL	NULL
1038	4+	53 + Z	L308AV_PDM	0n	Z018			168235.90693	NULL	NULL
997	4+	53 + Z	L1008AV_PDM	0n	Z07			162556.48356	NULL	NULL
981	4+	53 + Z	L1008AV_STP	0n	Z07			165100	NULL	NULL
940	4+	63 + Z	L1308AV_PDM	250n	Z06			165822.90106	NULL	NULL
1117	4+	53 + Z	L408AV_PDM	0n	Z09			162931.83268	NULL	NULL

id	survey_id	loop_point_number	loop_label	easting	northing	elevation	loop_size_x_units	loop_size_y_units	geometry
1	1+	0	NULL	477180.2	7245172.5	666.6	600	700	SRID=32622:POINT(477180.2 7245172.5)
2	1+	1	NULL	477098.4	7245207.5	667	600	700	SRID=32622:POINT(477098.4 7245207.5)
3	1+	2	NULL	477180.1	7245222.5	672.6	600	700	SRID=32622:POINT(477180.1 7245222.5)
4	1+	3	NULL	477181.2	7245250.5	678.4	600	700	SRID=32622:POINT(477181.2 7245250.5)
5	1+	4	NULL	477181.2	7245294.5	684.9	600	700	SRID=32622:POINT(477181.2 7245294.5)
6	1+	5	NULL	477099.4	7245338.0	686.5	600	700	SRID=32622:POINT(477099.4 7245338.0)
7	1+	6	NULL	477180.8	7245366.0	684.5	600	700	SRID=32622:POINT(477180.8 7245366.0)
8	1+	7	NULL	477181.8	7245392.0	676.5	600	700	SRID=32622:POINT(477181.8 7245392.0)

id	survey_id	station_number	station_label	line_direction	line_distance	easting	northing	elevation	distance_along_profile_m	geometry
1	1+	0	0m	N	0	476582.9	7245076.5	682.7	0	SRID=32622:POINT(476582.9 7245076.5)
2	1+	1	1.25n	N	25	476523.1	7245076.0	685.8	25	SRID=32622:POINT(476523.1 7245076.0)
3	1+	2	2.50n	N	58	476547.7	7245076.0	682.1	58	SRID=32622:POINT(476547.7 7245076.0)
4	1+	3	3.75n	N	75	476570.4	7245076.0	677.6	75	SRID=32622:POINT(476570.4 7245076.0)
5	1+	4	4.00n	N	100	476597.2	7245076.0	672.7	100	SRID=32622:POINT(476597.2 7245076.0)
6	1+	5	5.125n	N	125	476622.5	7245076.5	671.5	125	SRID=32622:POINT(476622.5 7245076.5)
7	1+	6	6.150n	N	158	476647.4	7245076.5	667	158	SRID=32622:POINT(476647.4 7245076.5)
8	1+	7	7.175n	N	175	476675.0	7245076.5	666.8	175	SRID=32622:POINT(476675.0 7245076.5)

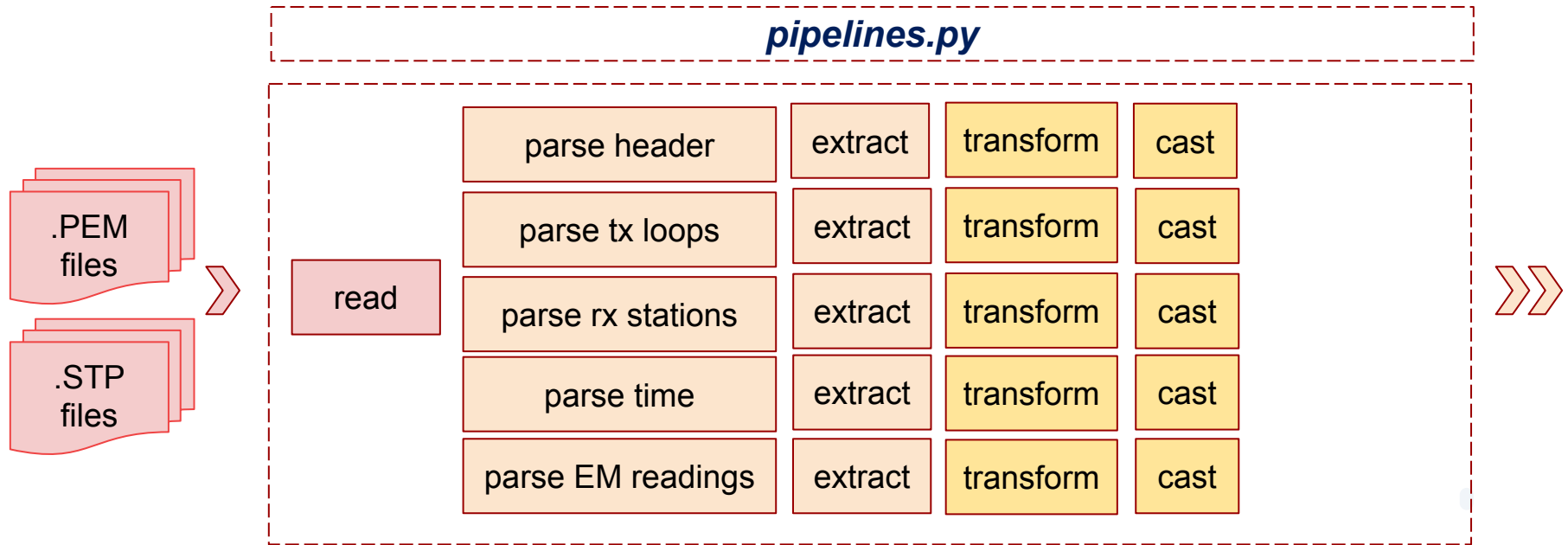
Basic Components of Data Pipeline

ENVS 5726



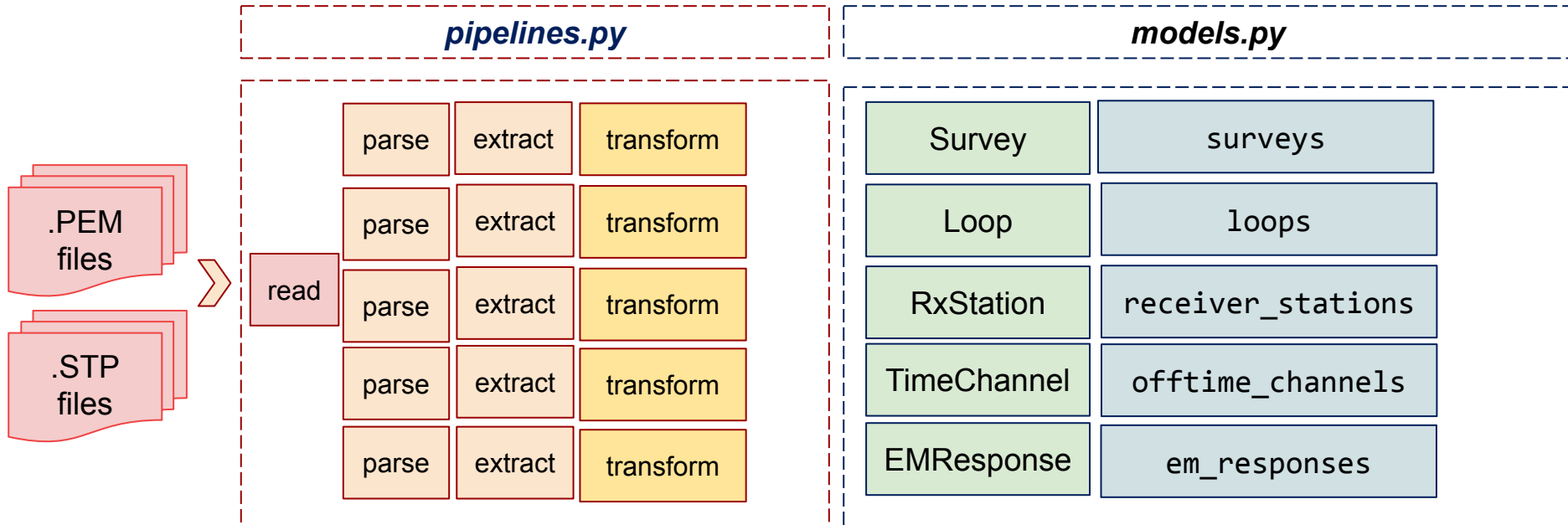
Basic Components of Data Pipeline

ENVS 5726



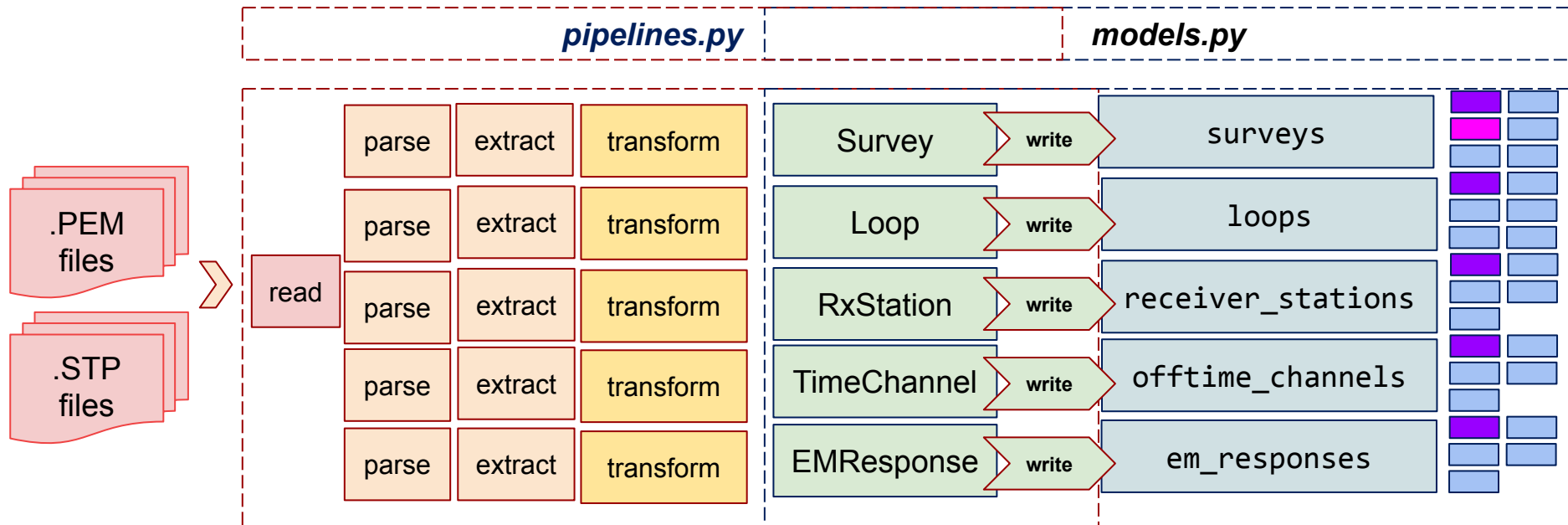
Basic Components of Data Pipeline

ENVS 5726



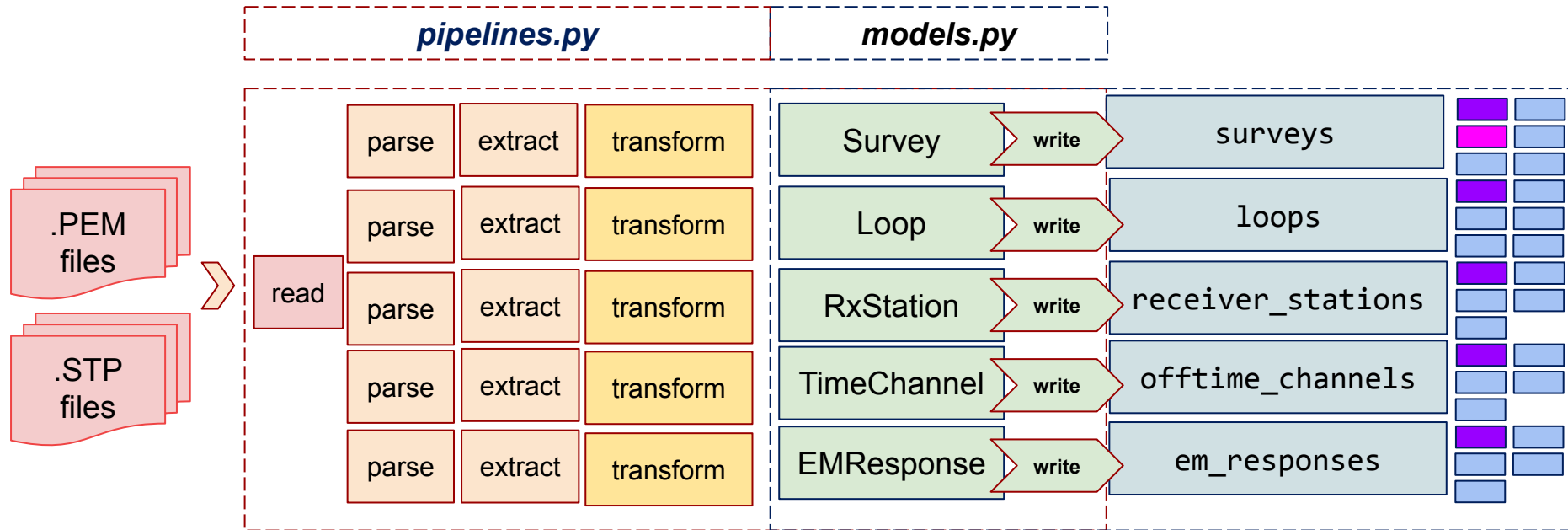
Basic Components of Data Pipeline

ENVS 5726



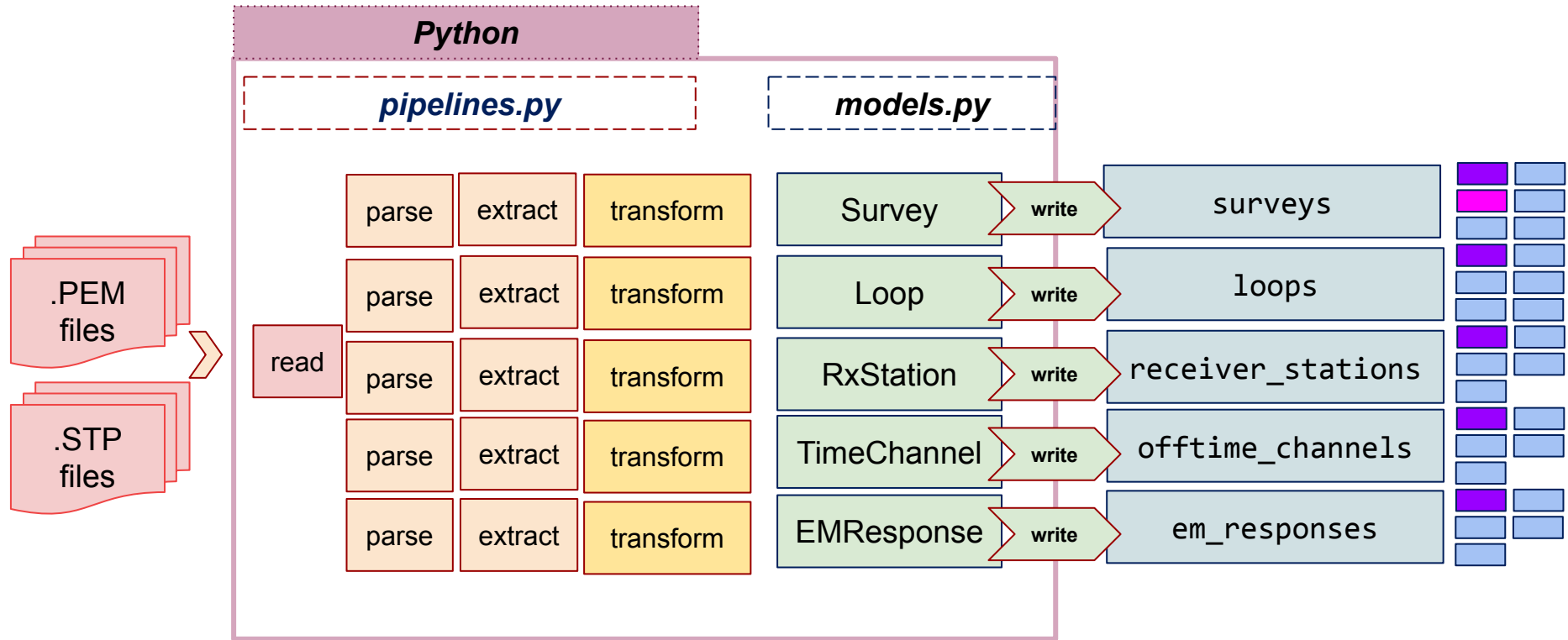
Basic Components of Data Pipeline

ENVS 5726



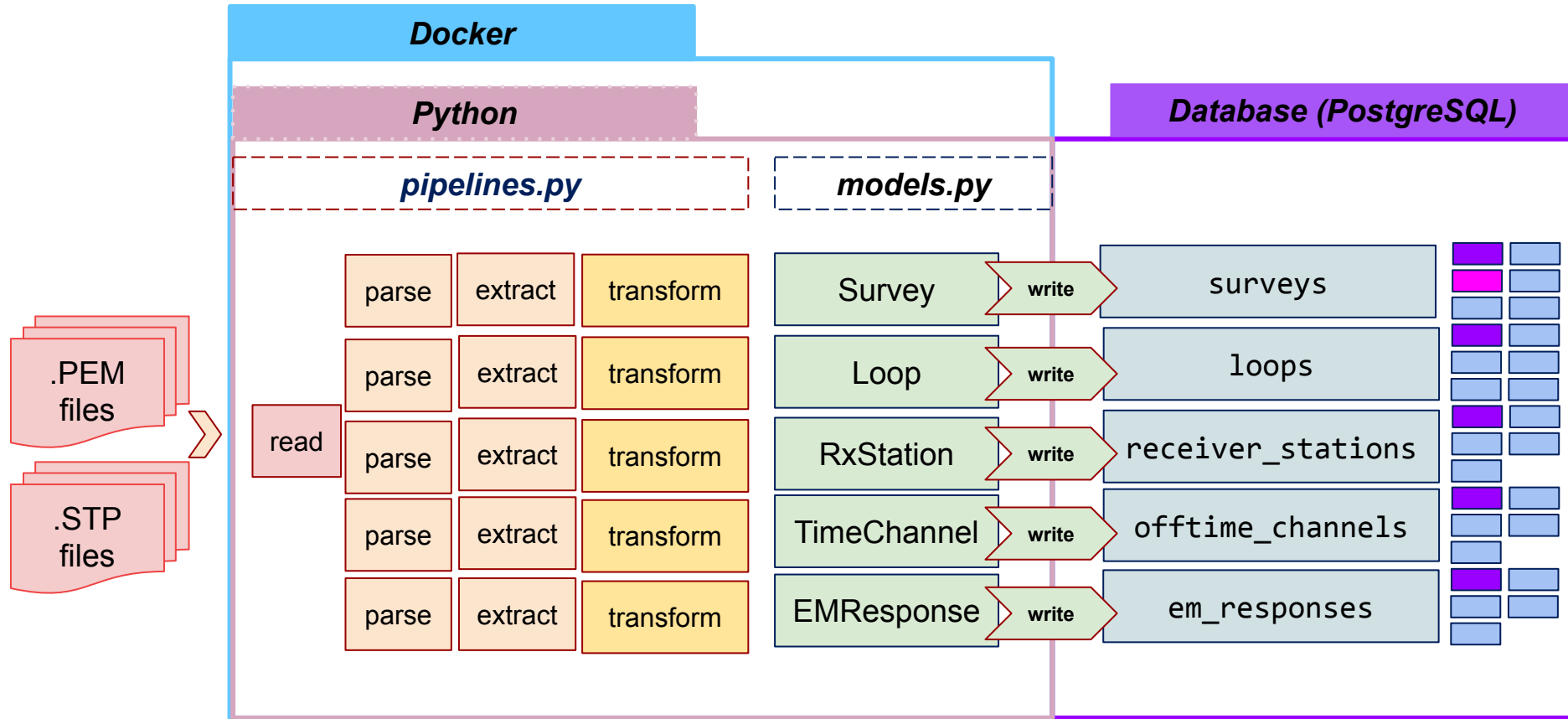
Basic Components of Data Pipeline

ENVS 5726



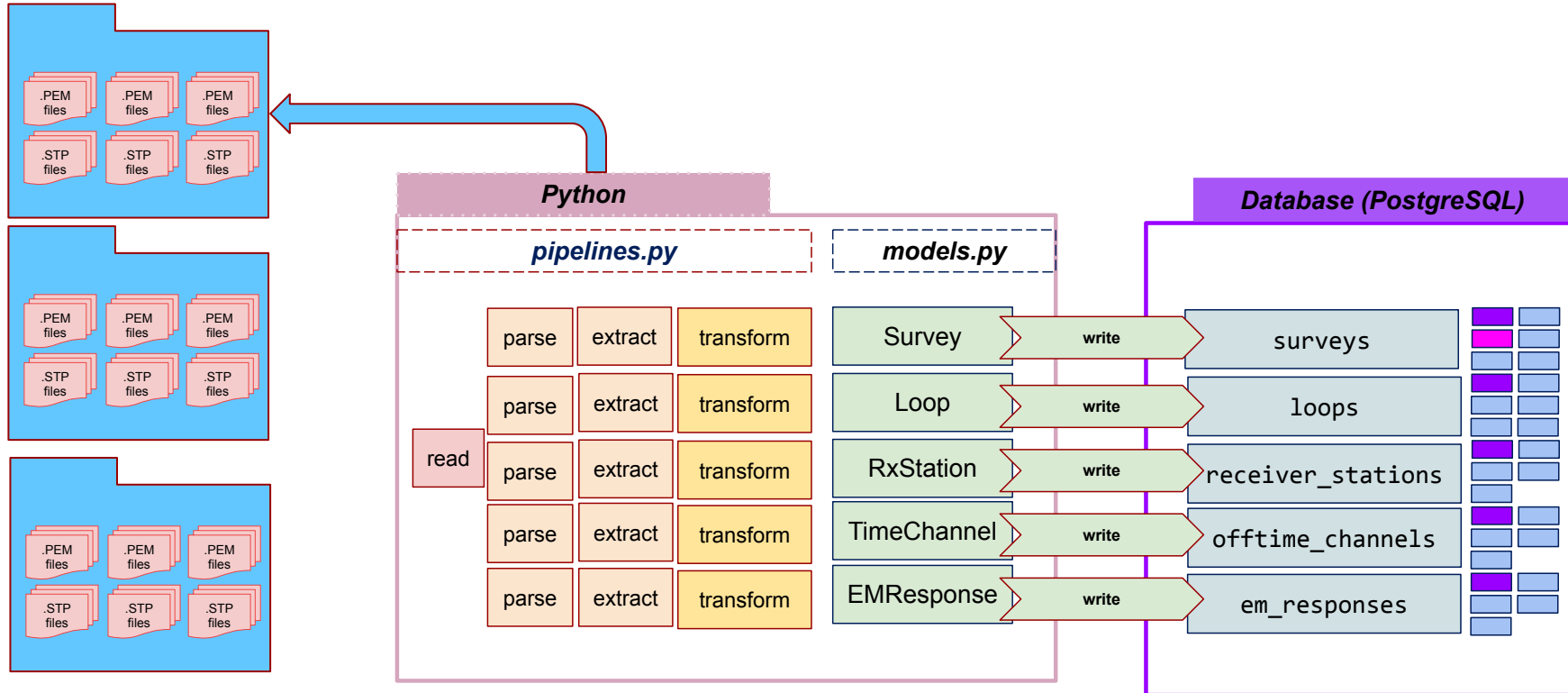
Basic Components of Data Pipeline

ENVS 5726



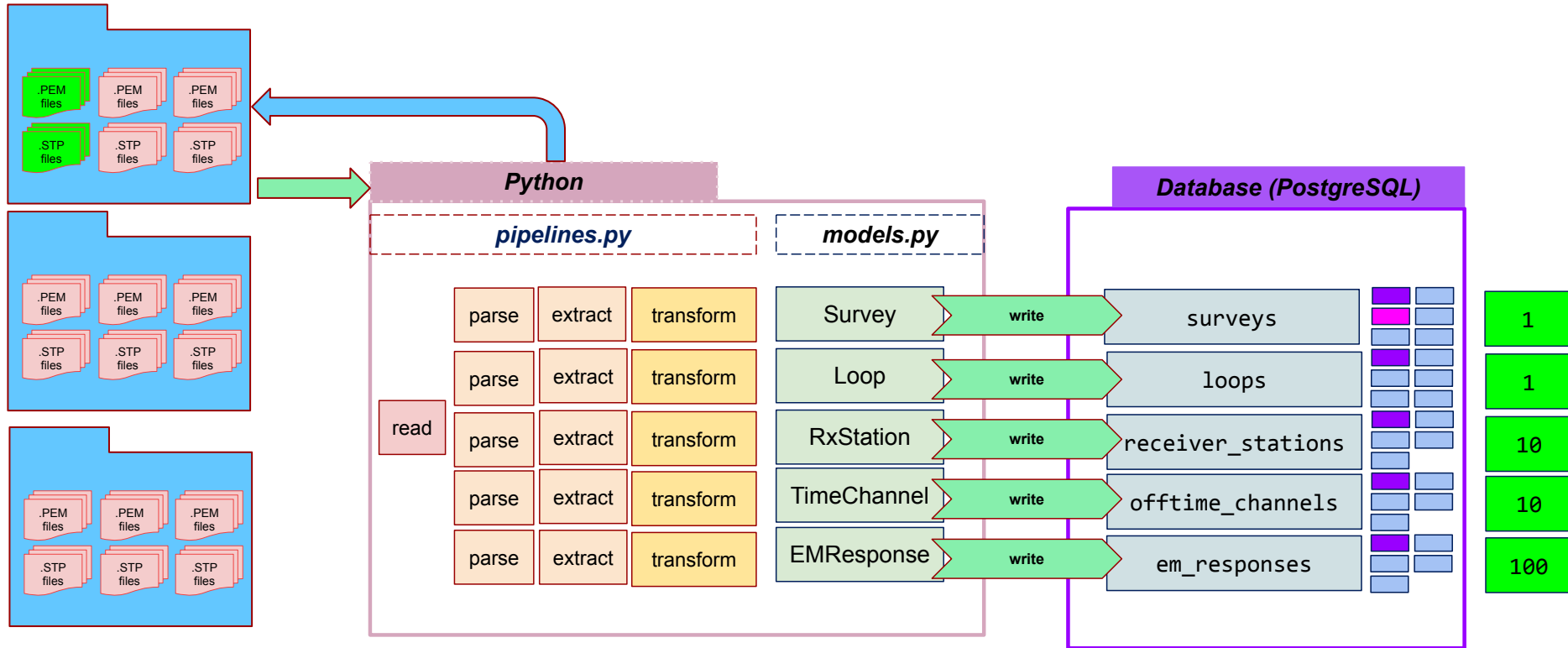
Scaling Data Pipeline

ENVS 5726



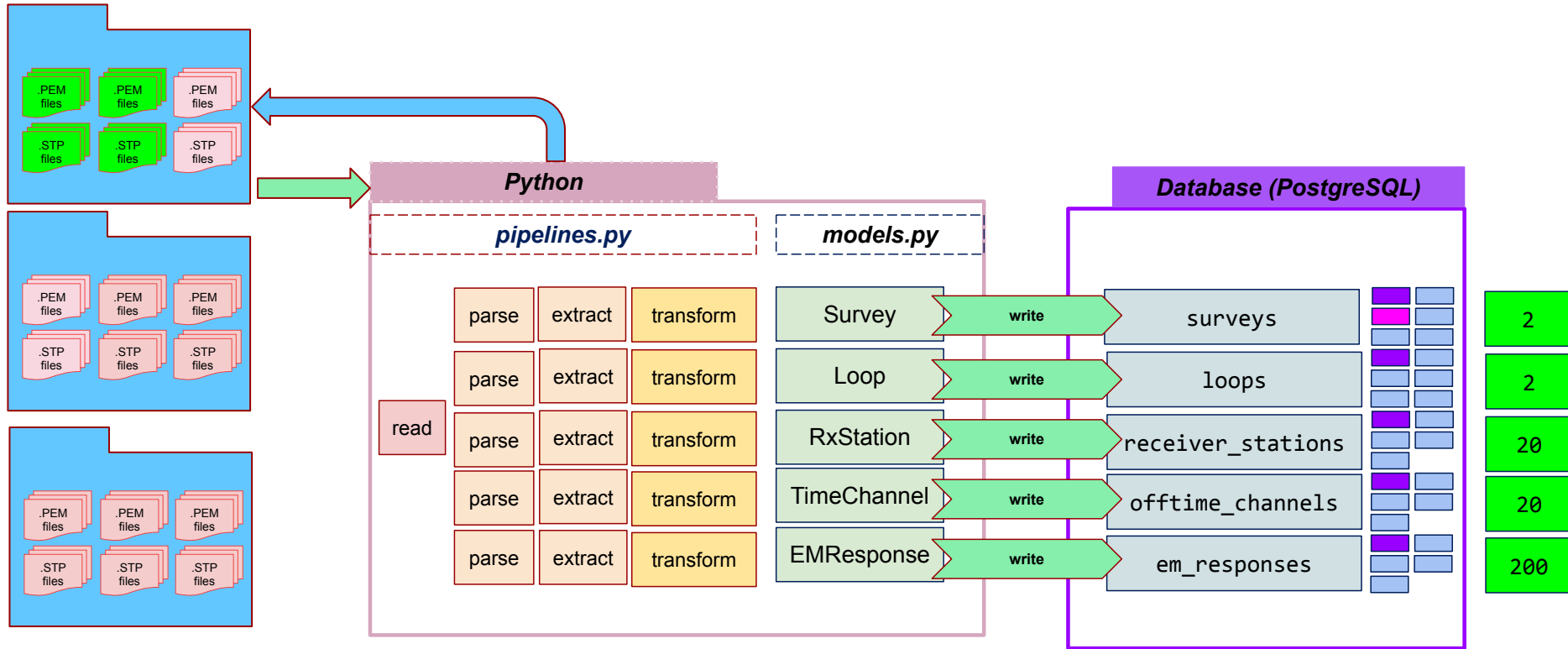
Scaling Data Pipeline

ENVS 5726



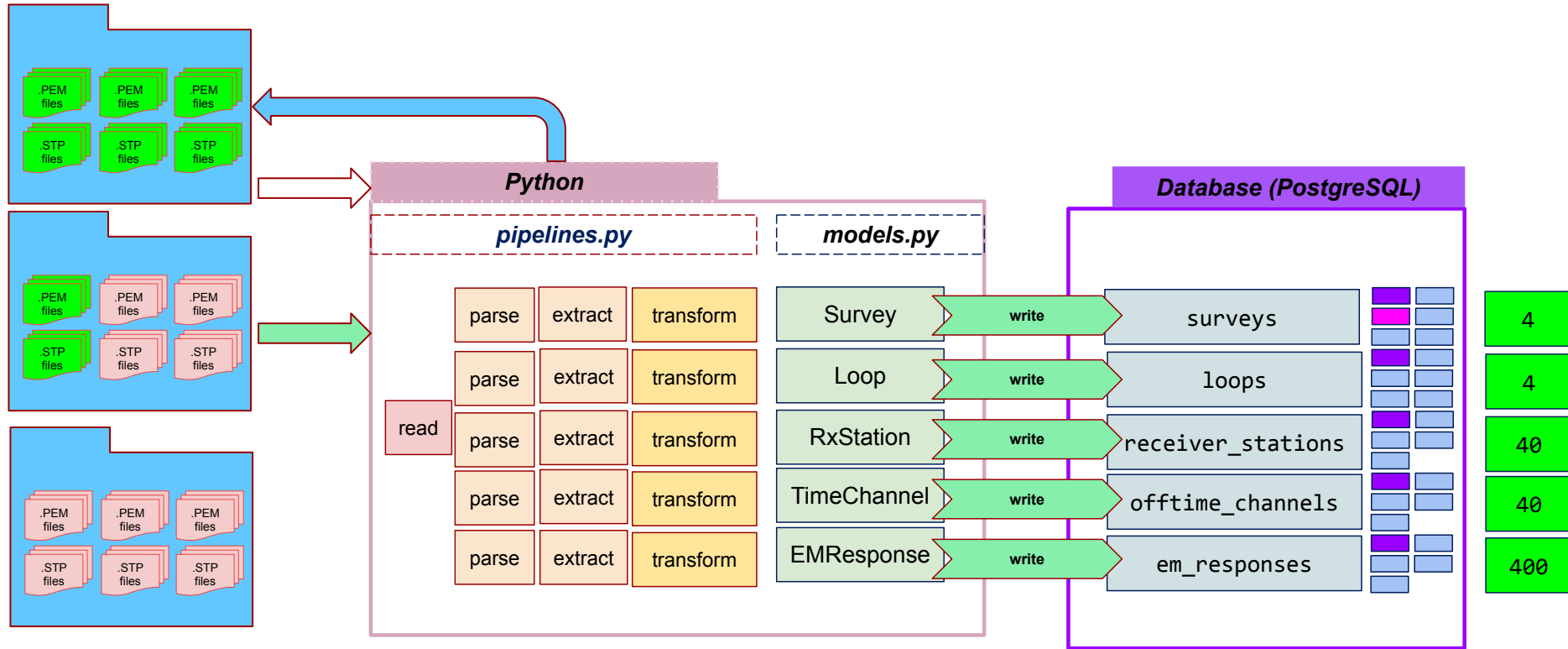
Scaling Data Pipeline

ENVS 5726



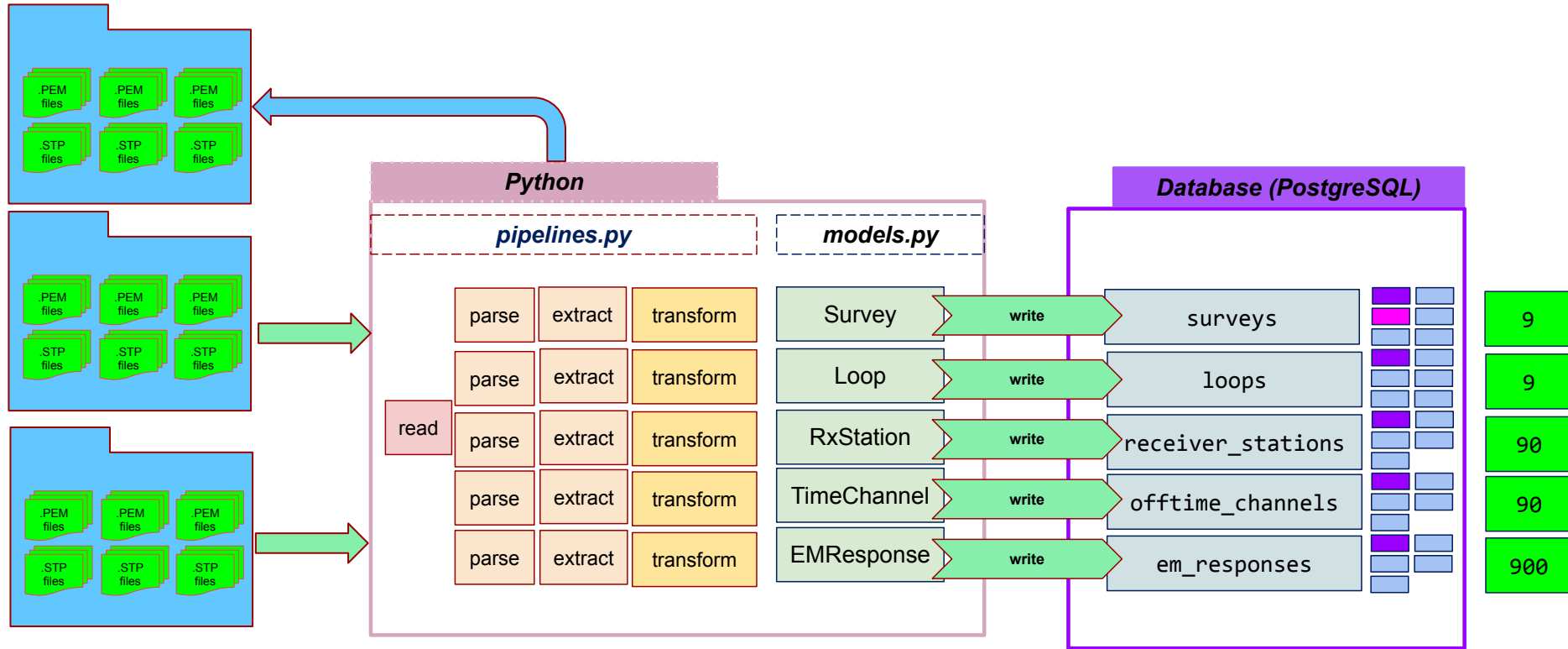
Scaling Data Pipeline

ENVS 5726



Scaling Data Pipeline

ENVS 5726



04

Demo Check-In

Demo time!

ENVS 5726

```
<FMT> 230 ~ data format
1 <FMT> 230 ~ data format
2 <DIR> nanoTesla/sec ~ data units
3 <OPR> Josh Way ~ operator's name
4 <XYP> 0 0 0 ~ X-Y probe #, S0A, tool #, tool id
5 <CUR> 20.0 ~ peak current in loop (2 X meter reading)
6 <TXS> 000.0 700.0 0 ~ loop size (x y units) (units: 0 = m, 1 = ft)
7 ~ Transmitter Loop Co-ordinates:
8 <L00> 477180.2 7245172.5 666.6 0 ~ loop coord. | 4 numbers - (x,y,z,units)
9 <L01> 477098.4 7245207.5 667.0 0 ~ loop coord. | values for units:
10 <L02> 477180.1 7245222.5 672.6 0 ~ loop coord. | 0 = metres; 1 = feet
11 <L03> 477181.2 7245250.5 676.4 0
12 <L04> 477181.2 7245294.5 684.9 0
13 <L05> 477099.4 7245338.0 686.5 0
14 ~ Hole/Profile Co-ordinates:
15 <P00> 476497.0 7245021.5 672.3 0 0 ~ 0=Collar/Start| 5 numbers: (x,y,z,units,sten)
16 <P01> 476525.1 7245024.5 668.3 0 25.0 ~ [1-99 = (x,y,z)] values for units:
17 <P02> 476549.0 7245024.0 669.2 0 50.0 ~ [or segment's] 0 = metres; 1 = feet
18 <P03> 476575.0 7245024.0 665.8 0 75.0 ~ [grid azimuth,] 2=az(deg) dip(deg) length(m)
19 <P04> 476681.0 7245023.5 666.8 0 100.0 ~ [-ve down dip,] 3=az(deg) dip(deg) length(ft)
20 <P05> 476625.9 7245025.0 665.6 0 125.0 ~ [segm't length] 'sten': surface PDM (WGS -ve)
21 <P06> 476656.1 7245024.5 666.8 0 150.0
22 <HE3> Data Scaled by Coil Area of 6500/4300
23 ~Tags for comment list: <GBN>, <STA><Stn>, <STX><Stn>, <STY><Stn>, <STZ><Stn>
24 ~ <R0G><Rdgp>, <R00>
25 ~Tags for List & plot comments: <SPAb><Stn>, <SPQ><Stn>, <SPY><Stn>, <SPZ><Stn>
26 ~Tags for headings: <HE1>, <HE2>, <HE3>, <HE4>
27 North American Nickel
28 P-008
29 S0N
30 P-008
31 August 19, 2016
32 S-COIL Metric Crystal-Master 16.66 1500 10 42
33 #117 9.20 May19,2015Z L50N N 1000 4300 +
34
35 0.000e+00 0.000e+00 1.597e-05 3.193e-05 6.387e-05 1.277e-04 2.555e-04
36 5.109e-04 1.022e-03 2.844e-03 4.087e-03 8.175e-03 1.635e-02
37
38 0N XR8 0 A 1940.0 90 256 10 12
39 D4 49.074 -2310.1 41264 41328.7
40 49.07 -0.2773 0.2773 0.004442 -0.003059 -0.004113 -0.002263
41 -0.002603 -0.001756 -0.00107 -0.03265
42
43 25N XR8 0 A 1940.0 90 256 10 32
44 D4 1270.99 -1865.06 35762.5 35833.7
45 1271 -0.1581 -0.1577 -0.05358 -0.008725 -0.003266 -0.001537
46 -0.000266 0.0005181 0.0006375 -0.2959
47
48 50N XR6 0 A 1940.0 90 192 10 52
49 D4 1433.13 -1711.27 32900 32975.7
50 1433 -0.2748 -0.2248 -0.05942 -0.0003389 0.003292 0.002578
51 0.004293 0.003807 0.002182 -1.143
52
53 75N XR9 0 A 1940.0 90 288 10 72
54 D4 1636.21 -1682.48 38739 38826.4
55 1636 -0.2921 -0.2921 -0.05601 0.0003726 0.002262 0.001559
56 0.002201 0.001282 0.0005714 -1.661
```



id	survey_id	line_id	survey_date	datum	projection	utm_zone	data_format	data_units	op
7	P-246	NULL	2025-12-03 00:26:51.834163	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Morgan Glynn
8	P-142	NULL	2025-12-03 00:26:33.314972	WGS 1984	UTM	22N	230	picoTesla	~ data units Keith Falardeau
9	P-141	NULL	2025-12-03 00:26:44.344029	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Morgan Glynn
4	P-484	NULL	2025-12-03 00:20:38.673104	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Henry
3	P-849	NULL	2025-12-03 00:24:56.580166	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Morgan Glynn
2	P-822	NULL	2025-12-03 00:24:39.799894	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Morgan Glynn
1	P-808	NULL	2025-12-03 00:24:06.839264	WGS 1984	UTM	22N	230	nanoTesla/sec	~ data units Josh

id	survey_id	receiver_station_id	component	source_file	station_label	receiver_code	primary_pulse_nt_per_sec	receiver_number	angle_degrees	num_s
1816	4 +	53 + Z	L308AV_STP	0n	Z018n			175300	NULL	NULL
918	4 +	63 + Z	L1308AV_STP	25n	Z06			178300	NULL	NULL
1090	4 +	53 + Z	L408AV_STP	0n	Z09			168000	NULL	NULL
1038	4 +	53 + Z	L308AV_PDM	0n	Z018			168235.90693	NULL	NULL
997	4 +	53 + Z	L1408AV_PDM	0n	Z07			162556.48356	NULL	NULL
981	4 +	53 + Z	L1408AV_STP	0n	Z07			165100	NULL	NULL
940	4 +	63 + Z	L1308AV_PDM	25n	Z06			165822.90106	NULL	NULL
1117	4 +	53 + Z	L408AV_PDM	0n	Z09			162931.81068	NULL	NULL

id	survey_id	loop_point_number	loop_label	easting	northing	elevation	loop_size_x_units	loop_size_y_units	geometry
1	1 +	0	NULL	477180.2	7245172.5	666.6	700	SRID=32622:POINT(477180.2 7245172.5)	SRID=32622:POINT(477180.2 7245172.5)
2	1 +	1	NULL	477098.4	7245207.5	667	600	700 SRID=32622:POINT(477098.4 7245207.5)	SRID=32622:POINT(477098.4 7245207.5)
3	1 +	2	NULL	477180.1	7245222.5	672.6	600	700 SRID=32622:POINT(477180.1 7245222.5)	SRID=32622:POINT(477180.1 7245222.5)
4	1 +	3	NULL	477181.2	7245250.5	678.4	600	700 SRID=32622:POINT(477181.2 7245250.5)	SRID=32622:POINT(477181.2 7245250.5)
5	1 +	4	NULL	477181.2	7245294.5	684.9	600	700 SRID=32622:POINT(477181.2 7245294.5)	SRID=32622:POINT(477181.2 7245294.5)
6	1 +	5	NULL	477099.4	7245338.0	686.5	600	700 SRID=32622:POINT(477099.4 7245338.0)	SRID=32622:POINT(477099.4 7245338.0)
7	1 +	6	NULL	477180.8	7245366.5	684.5	600	700 SRID=32622:POINT(477180.8 7245366.5)	SRID=32622:POINT(477180.8 7245366.5)
8	1 +	7	NULL	477181.8	7245392.5	676.5	600	700 SRID=32622:POINT(477181.8 7245392.5)	SRID=32622:POINT(477181.8 7245392.5)

id	survey_id	station_number	station_label	line_direction	line_distance	easting	northing	elevation	distance_along_profile_m	geometry
1	1 +	0	0n	N	0	476582.9	7245076.5	682.7	0	SRID=32622:POINT(476582.9 7245076.5)
2	1 +	1	125n	N	25	476523.1	7245076.5	685.8	25	SRID=32622:POINT(476523.1 7245076.5)
3	1 +	2	250n	N	58	476547.7	7245076.5	682.1	58	SRID=32622:POINT(476547.7 7245076.5)
4	1 +	3	375n	N	75	476570.4	7245076.5	677.6	75	SRID=32622:POINT(476570.4 7245076.5)
5	1 +	4	410n	N	100	476597.2	7245076.5	672.7	100	SRID=32622:POINT(476597.2 7245076.5)
6	1 +	5	5125n	N	125	476622.5	7245076.5	671.5	125	SRID=32622:POINT(476622.5 7245076.5)
7	1 +	6	6150n	N	148	476647.4	7245076.5	667	148	SRID=32622:POINT(476647.4 7245076.5)
8	1 +	7	7175n	N	175	476675.7	7245076.5	666.8	175	SRID=32622:POINT(476675.7 7245076.5)

- **Scalable Task Design with Deduplication Checks:** Each save function (save_survey, save_loops, save_receiver_stations) includes upfront checks that query for existing records before inserting. This prevents duplicate data ingestion if the pipeline re-runs, returning early with counts of existing records which is a critical pattern for reliable ETL pipelines.
- **Batch Processing with Scalable Looping Patterns:** The code processes large datasets efficiently by iterating through lists of parsed records (loops, stations, measurements) and committing them in controlled batches. Rather than individual commits per record, functions like save_receiver_stations loop through all stations, add them to the session, and commit once thus reducing database round-trips and enabling the pipeline to scale to thousands of measurements per file.
- **Rich Data Extraction with Structured Parsing:** The parsing tasks (e.g., parse_measurements, parse_stations) use regex patterns and line-by-line iteration to extract complex, nested data structures from unstructured file content. Each function returns well-typed dictionaries with metadata context (line names, component types), ensuring downstream storage functions receive clean, validated data ready for database insertion.

Data Engineering

Models and Schema

ENVS 5726

- **Robust Data Integrity with Cascades:** The models explicitly define parent-child relationships using *cascade="all, delete-orphan"*. This ensures that when a parent record (like a survey) is deleted, all associated children (loops, stations, measurements) are automatically cleaned up, preventing orphaned records in the database.
- **Explicit Database Constraints and Indexing:** proactively enforces data quality and performance through `__table_args__`. It defines `UniqueConstraint` to prevent logical duplicates (e.g., ensuring a specific `loop_point_number` only appears once per survey) and creates specific Index entries (e.g., on easting and northing) to speed up lookups.
- **Intentional Typing and Spatial Awareness:** The models go beyond basic types by integrating specialized fields like Enum for fixed categories (e.g., `ComponentEnum`) and PostGIS Geometry columns with specific SRIDs (Spatial Reference System Identifiers). This ensures the database enforces valid values and correctly handles complex spatial data rather than relying solely on application logic.

05

Questions and Conclusions

Conclusion

ENVS 5726

Knowledge about underlying data and user objective is critical to design a resilient and effective data pipeline.

A thoughtful schema design can go a long way (quality, resilience, and reliability).

Off-the-shelf orchestration tools are critical for pipeline observability.

Learning and applying the fundamentals (earth science, data engineering, software development) is critical despite LLMs.

Get feedback from users and stakeholders, and incorporate that feedback into the logic.

Discussion

ENVS 5726

Questions

ENVS 5726

