# Intended Question to Answer

- How do different weather conditions affect NFL player's stats?

Introduction     Dataset 1     Dataset 2     Cross-Dataset     AirFLow DAG     Conclusion
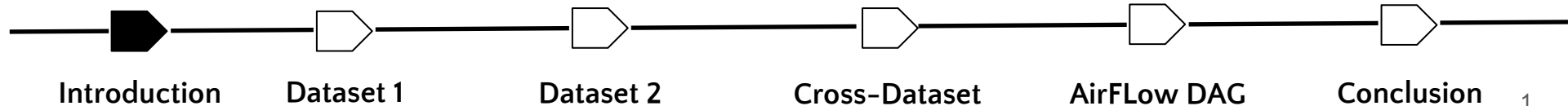
# Dataset 1 (NFL player stats)

- Basic Stats
  - 1 Table
  - Basic player information
- Game Logs
  - 2 Tables (Runningback and Quarterback Game Logs)
  - Displays player stats in a single game
- Career Stats
  - 4 Tables (Defensive, Passing, Rushing, and Receiving Career Stats
  - Displays player stats over player's entire NFL career

**Introduction**   **Dataset 1**   **Dataset 2**   **Cross-Dataset**   **AirFLow DAG**   **Conclusion**
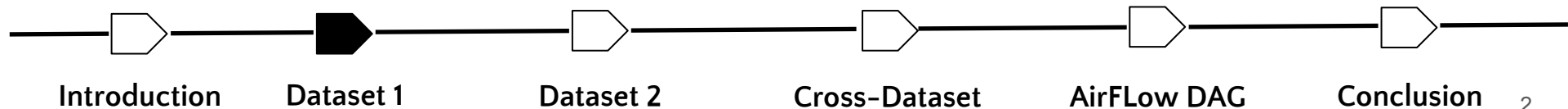
# Dataset 1 ERD Modeled

Primary Changes Included:

- Attempt to cast fields from STRING to INT/FLOAT (unsuccessful due to "--" entries)
- Attempt to cast game_date from STRING to DATE (unsuccessful due to incorrect format)
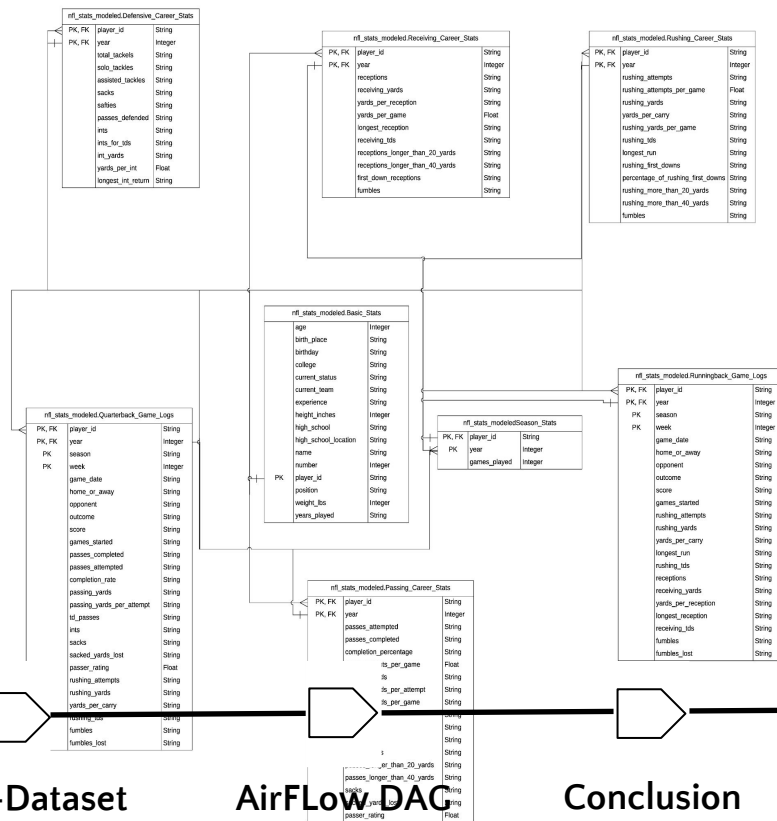
# Dataset 1 Beam

Changes made:

- Game Logs and Career Stats Tables
  - Replaced "--" entries with 0 and converted from STRING to INT/FLOAT
  - Combined game_date and year fields and converted from STRING to DATE
- Basic Stats
  - Made experience field uniform (Rookie → 1 Season(s), 2nd → 2 Season(s))
  - Split years_played into year_started and year_ended and converted from STRING to INT

Introduction    Dataset 1    Dataset 2    Cross-Dataset    AirFLow DAG    Conclusion
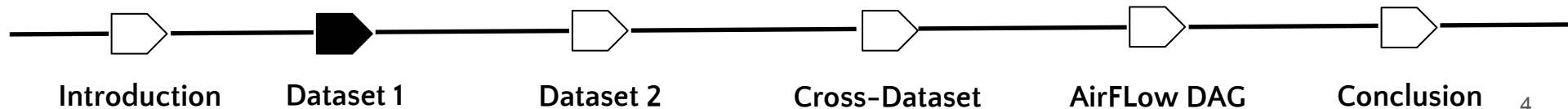
# Dataset 2 - Addition of Weather Data

- 3 Tables:
    - Nfl_stadiums - a list of NFL stadiums with size, weather, and geographic information
    - Nfl_teams - a list of NFL teams
    - Spreadspoke_scores - game day information regarding teams, score, and location

**Introduction    Dataset 1    Dataset 2    Cross–Dataset    AirFLow DAG    Conclusion**
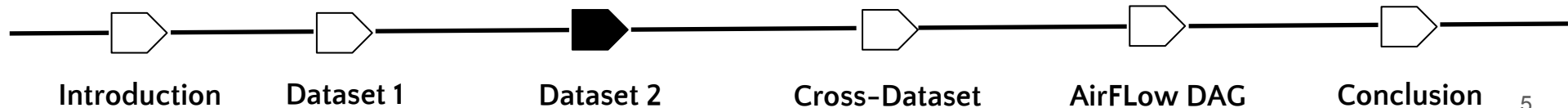
# Dataset 2 ERD Modeled

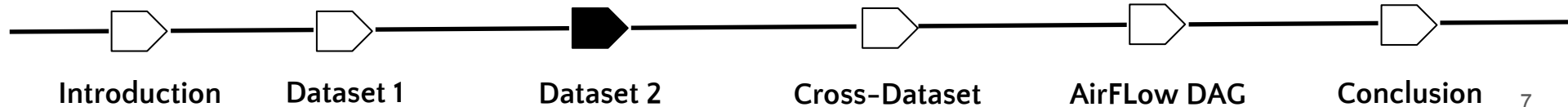No changes were required of the dataset

- Transferred data directly from NFL_weather staging tables
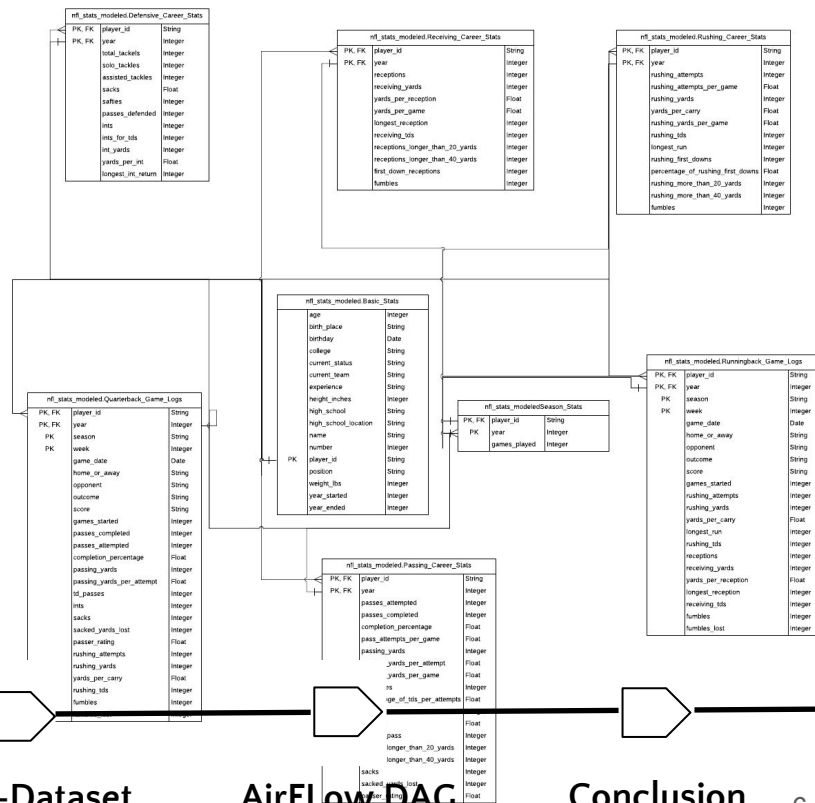
# Dataset 2 Beam

Changes made:

- Spreadspoke_scores Table
    - Broke up the stadium location field into two fields: city and state

**Introduction**   **Dataset 1**   **Dataset 2**   **Cross-Dataset**   **AirFLow DAG**   **Conclusion**

# Cross-Dataset Queries

```sql
SELECT a.player_id, avg(a.completion_percentage) as avg_completion_percentage
FROM `nfl_stats_modeled.Quarterback_Game_Logs_Beam_DF` a
join `nfl_weather_modeled.spreadspoke_scores` b
on a.game_date = b.schedule_date
WHERE b.weather_wind_mph > 20
GROUP BY a.player_id
ORDER BY avg_completion_percentage DESC
```

```sql
select a.player_id, round(avg(a.rushing_yards), 2) as avg_rushing_yards
from nfl_stats_modeled.Quarterback_Game_Logs_Beam_DF a
join nfl_weather_modeled.spreadspoke_scores b
on a.game_date = b.schedule_date
where a.year = b.schedule_season
and   b.weather_temperature < 20.0
group by a.player_id
```

```sql
select q.player_id, avg(passing_yards) as avg_passing_yards
from nfl_stats_modeled.Quarterback_Game_Logs_Beam_DF q
join nfl_weather_modeled.spreadspoke_scores s
on q.game_date = s.schedule_date
where s.weather_detail = "DOME"
group by q.player_id
order by avg_passing_yards desc
```



**Introduction**   **Dataset 1**   **Dataset 2**   **Cross-Dataset**   **AirFLow DAG**   **Conclusion**

# Airflow DAG

```
create_staging_dataset >> create_modeled_dataset >> split
split >> load_nfl_stadiums >> create_nfl_stadiums >> nfl_stadiums_beam >> nfl_stadiums_dataflow
split >> load_nfl_teams >> create_nfl_teams
split >> load_spreadspoke_scores >> create_spreadspoke_scores
```

**Introduction**     **Dataset 1**     **Dataset 2**     **Cross-Dataset**     **AirFLow DAG**     **Conclusion**
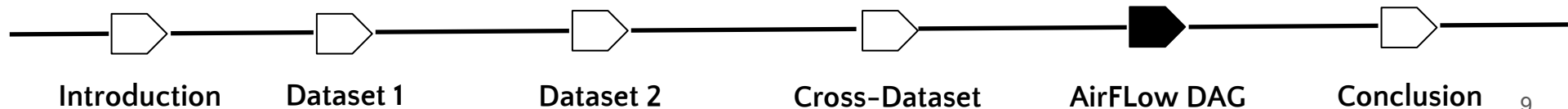
# Future Improvements

- More comprehensive data
  - Specifically looking into the type of weather - details about rain and snow ect
  - Our current dataset has mostly null values for the weather_detail field

**Introduction**     **Dataset 1**     **Dataset 2**     **Cross-Dataset**     **AirFLow DAG**     **Conclusion**