

# Automated Data Pipeline for Hockey Team Analytics

Technologies: Python (BeautifulSoup, pandas), PySpark, Snowflake, Web Scraping

- Developed an end-to-end automated data pipeline to scrape, process, and analyze hockey team data, leveraging Python for web scraping (BeautifulSoup) and data manipulation (pandas).
- Engineered a scalable storage solution in Snowflake, utilizing PySpark for efficient handling and loading of large datasets, enhancing the data's accessibility for real-time analytics.
- Implemented dynamic database schema creation, automating the mapping of data to Snowflake table structures, significantly reducing manual schema alignment efforts.
- Facilitated advanced data analytics by providing a structured, query-ready dataset in Snowflake, supporting stakeholders in making informed decisions based on up-to-date hockey team insights.
- Achieved a streamlined workflow from data collection to analysis, enabling timely and insightful analytics that support strategy formulation for hockey teams and enthusiasts.

## Source Data And Stored Data

Team Name	Year	Wins	Losses	OT Losses	Win %	Goals For (GF)	Goals Against (GA)	Pct.
Boston Bruins	1990	44	24	0	0.65	284	204	88
Buffalo Sabres	1990	31	50	0	0.38	282	275	14
Calgary Flames	1990	45	26	0	0.63	344	205	81
Chicago Blackhawks	1990	43	23	0	0.65	294	211	73
Detroit Red Wings	1990	34	36	0	0.49	275	298	-25
Edmonton Oilers	1990	37	37	0	0.50	272	272	0
Hartford Whalers	1990	31	56	0	0.36	239	276	-36
Los Angeles Kings	1990	48	24	0	0.67	340	234	88
Minnesota North Stars	1990	27	59	0	0.31	256	266	-10
Montreal Canadiens	1990	39	30	0	0.57	275	240	24
New Jersey Devils	1990	32	30	0	0.52	272	234	9
New York Islanders	1990	25	45	0	0.36	223	260	-47
New York Rangers	1990	36	31	0	0.54	267	205	32
Philadelphia Flyers	1990	33	37	0	0.47	252	257	-15
Pittsburgh Penguins	1990	41	35	0	0.54	342	245	37
Quebec Nordiques	1990	16	60	0	0.21	238	354	-118
St. Louis Blues	1990	47	22	0	0.68	310	205	65

Team Name	Year	Wins	Losses	OT Losses	Win %	Goals For (GF)	Goals Against (GA)	Pct.
Boston Bruins	1990	44	24	0	0.65	284	204	88
Buffalo Sabres	1990	31	50	0	0.38	282	275	14
Calgary Flames	1990	45	26	0	0.63	344	205	81
Chicago Blackhawks	1990	43	23	0	0.65	294	211	73
Detroit Red Wings	1990	34	36	0	0.49	275	298	-25
Edmonton Oilers	1990	37	37	0	0.50	272	272	0
Hartford Whalers	1990	31	56	0	0.36	239	276	-36
Los Angeles Kings	1990	48	24	0	0.67	340	234	88
Minnesota North Stars	1990	27	59	0	0.31	256	266	-10
Montreal Canadiens	1990	39	30	0	0.57	275	240	24
New Jersey Devils	1990	32	30	0	0.52	272	234	9
New York Islanders	1990	25	45	0	0.36	223	260	-47
New York Rangers	1990	36	31	0	0.54	267	205	32
Philadelphia Flyers	1990	33	37	0	0.47	252	257	-15
Pittsburgh Penguins	1990	41	35	0	0.54	342	245	37
Quebec Nordiques	1990	16	60	0	0.21	238	354	-118
St. Louis Blues	1990	47	22	0	0.68	310	205	65

## Web Scrape ETL Process Design

