Kurzvortrag pandas

Franko Jolic, Janne Buhr

2. November 2022

Inhaltsverzeichnis

1 Timeline

2008: Development of pandas started
2009: pandas becomes open source
2012: First edition of Python for Data Analysis is published
2015: pandas becomes a NumFOCUS sponsored project
2018: First in-person core developer sprint

2 Library Highlights

- A fast and efficient DataFrame object for data manipulation with integrated indexing
- Tools for reading and writing data between in-memory data structures and different formats: CSV and text files, Microsoft Excel, SQL databases, and the fast HDF5 format
- Intelligent data alignment and integrated handling of missing data: gain automatic label-based alignment in computations and easily manipulate messy data into an orderly form
- Flexible reshaping and pivoting of data sets
- Intelligent label-based slicing, fancy indexing, and subsetting of large data sets
- Columns can be inserted and deleted from data structures for size mutability
- Aggregating or transforming data with a powerful group by engine allowing split-apply-combine operations on data sets
- High performance merging and joining of data sets
- Hierarchical axis indexing provides an intuitive way of working with high-dimensional data in a lower-dimensional data structure
- Time series-functionality: date range generation and frequency conversion, moving window statistics, date shifting and lagging. Even create domain-specific time offsets and join time series without losing data
- Highly optimized for performance, with critical code paths written in Cython or C
- Python with pandas is in use in a wide variety of academic and commercial domains, including Finance, Neuroscience, Economics, Statistics, Advertising, Web Analytics, and more

3 Mission

pandas aims to be the fundamental high-level building block for doing practical, real world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open source data analysis / manipulation tool available in any language.

4 Vision

A world where data analytics and manipulation software is:

- Accessible to everyone
- Free for users to use and modify
- Flexible
- Powerful
- Easy to use
- Fast

5 Quellen

https://pandas.pydata.org