I have been working as a software engineer at Two Sigma for the last 4.5 years (with the last two on Smooth Storage). I was an engineer on the AWS S3 team before that. I received my masters in Computer Science from the University of Pittsburgh, and bachelors from the Indian Institute of Technology, Varanasi.

Smooth is a distributed storage system for managing structured time series data at Two Sigma. Smooth's design emphasizes scale, both in terms of size and aggregate request bandwidth, reliability and storage efficiency. It is optimized for large parallel streaming read/write accesses over provided time ranges. Smooth has a clear separation between the metadata and data layers, and supports multiple pluggable object stores for storing data files. Data can be replicated or moved between different stores and data centers to support availability, performance and storage tiering objectives.

Smooth is widely used at Two Sigma by various applications including modelling research workflows, data pipelines and various data analysis jobs. Smooth has been in development for about 5 years, currently stores multiple PBs of compressed data, and serves peak aggregate throughput in excess of 100 GB/s.

In this talk I will discuss the design and implementation of Smooth, our experience running it over the past two years, ongoing challenges and future directions.