Speedrunning a lakehouse: a composable FaaS over object storage (Keynote Talk)

Jacopo Tagliabue
CTO, Bauplan Labs
jt1339@nyu.edu

ABSTRACT

The lakehouse is emerging as the foundational design for data and AI workloads. Heterogeneous runtimes, requirements, integrations are suited for the composable stack, but flexibility may come at the expense of simplicity: how can we reason about such complex systems, both as final users and system developers? In this talk, we discuss the opportunities and challenges of building a "Function-as-a-Service" (FaaS) lakehouse: every workload is served by one or more functions, which unifies the mental model for the user (e.g. pipelines are "chained functions"), as well as developers.

Dissatisfied with existing models, we built a composable system on top of object storage, which required innovation across the entire stack: from function scheduling to containerization, from "Git for Data" to differential Arrow caching. Today, companies run tens of thousands of ephemeral functions as part of Bauplan pipelines: we conclude by discussing the importance of API design for viral enterprise adoption, and why composable user abstractions should hide the composability in the stack.

Speaker Biography: Jacopo Tagliabue is the co-founder and CTO of Bauplan. Educated in several acronyms across the globe (UNISR, SFI, MIT), Jacopo was co-founder and CTO of Tooso, an AI startup acquired by TSX:CVO in 2019. He led Coveo's AI from scale-up to IPO, and built out Coveo Labs, a prolific R&D practice whose libraries, models and datasets have garnered tens of millions of downloads. Throughout his career, he has been fortunate enough to collaborate with incredible folks in industry and academia (e.g. Netflix, NVIDIA, Stanford, Univ. of Wisconsin-Madison), and publish contributions in a variety of fields: Information Retrieval (RecSys, SIGIR), Data Science (KDD), Artificial Intelligence and NLP (ICML, NAACL), Data Management (SIGMOD, VLDB), Computer Systems (Middleware). While building his new company, he is teaching ML Systems at NYU, which is mostly notable because it is the only job he ever had that his parents understand.