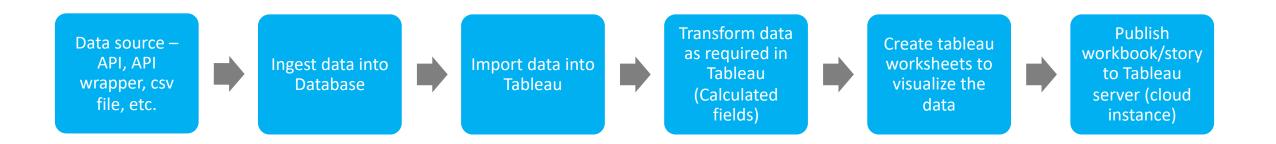
Designing a Data pipeline for visualization

Akanxa Padhi

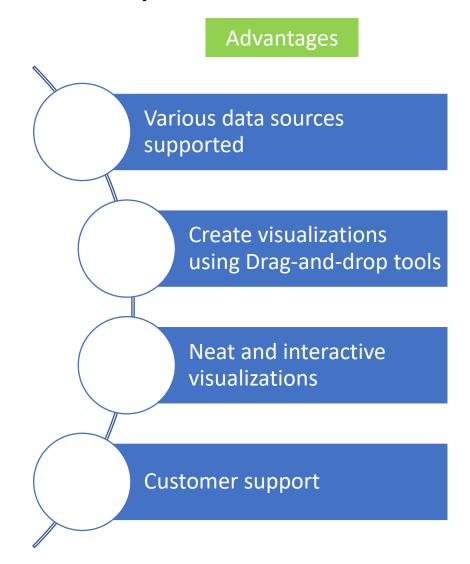
STATW5702 – EDAV Fall 2020 CC

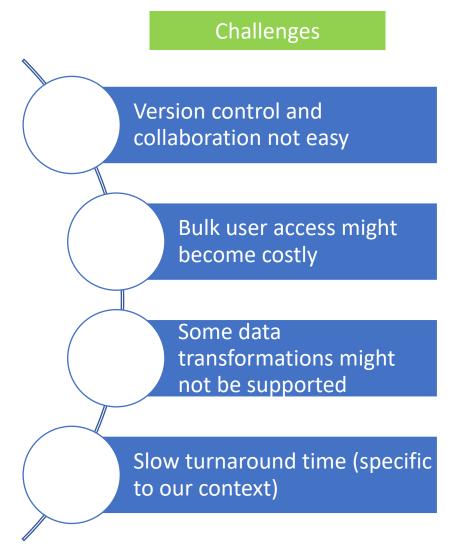
11/18/2020

Current Pipeline - Tableau



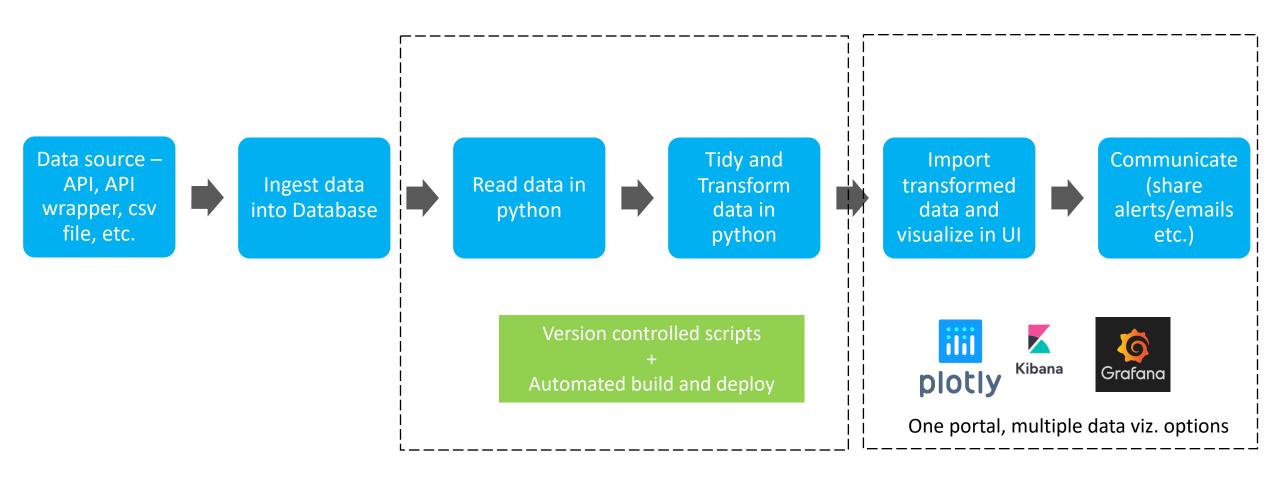
Current Pipeline - Tableau







Future pipeline – Python + UI of choice



Future pipeline – Python + UI of choice

Challenges we hope to fix

Better version control

Bulk user access

Complex data transformations

Self-serviced data pipeline; Fast turnaround time Lookahead opportunities

Simple, reusable code

Platform availability

Keep up to date with data and programming skills

Maintenance (cost) of selfserviced pipeline

Key Takeaways for designing a data pipeline



Build vs Buy - Licensing cost for tool vs cost of self-hosted platform

Team skillset

Data processing and modeling vs Data visualization

Simple reusable architecture

Easy and robust connection with multiple tools

Extensible and customizable

Fast scaling and quick turnaround for varied use cases