WhereToFly

a Spark-Hibernate Project

Silvia Fracalossi

List of content

- Data description
- Project Architecture and Technologies
  - Spark Component
  - Data Loader
  - Prediction Component
- Conclusion and Lesson Learned

List of content

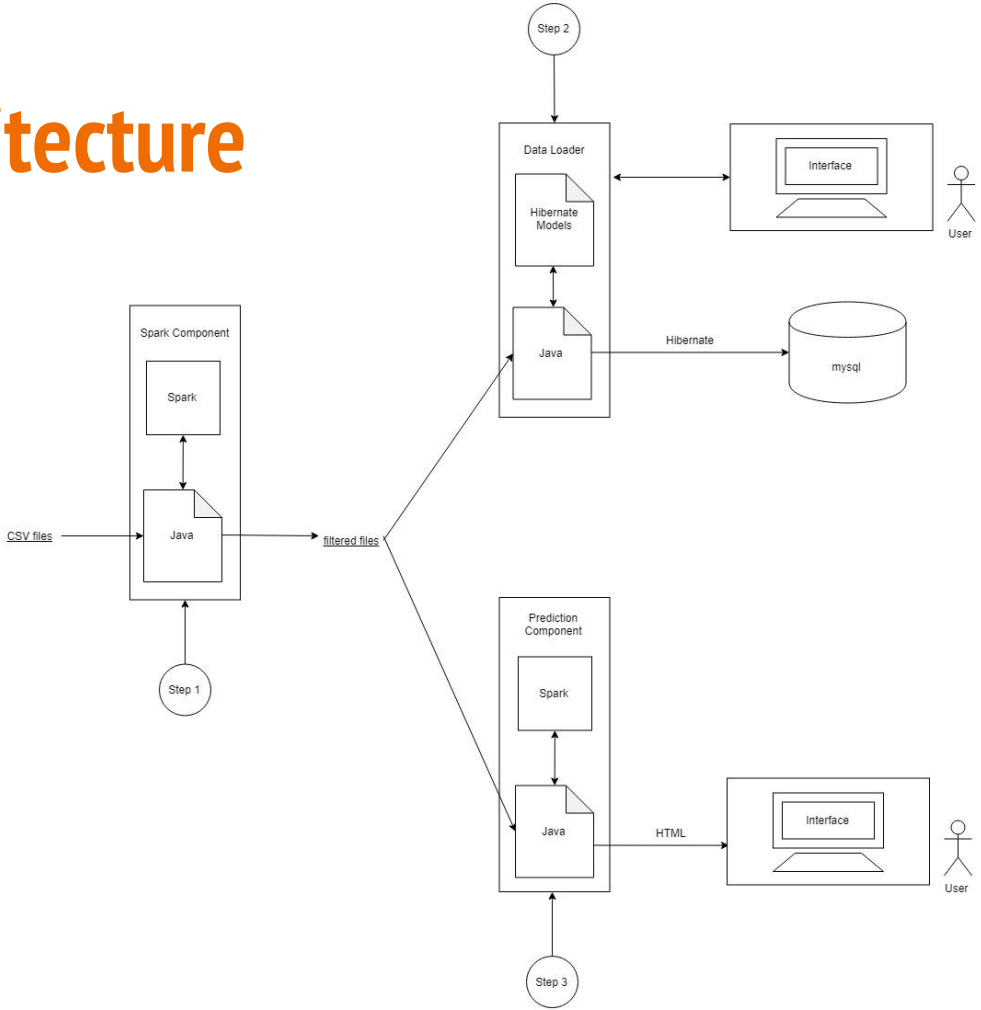
Data Description

- Airlines, from Kaggle
- Airports, from Kaggle
- Flights, from Kaggle
- Transportation_airfare, from transportation.gov

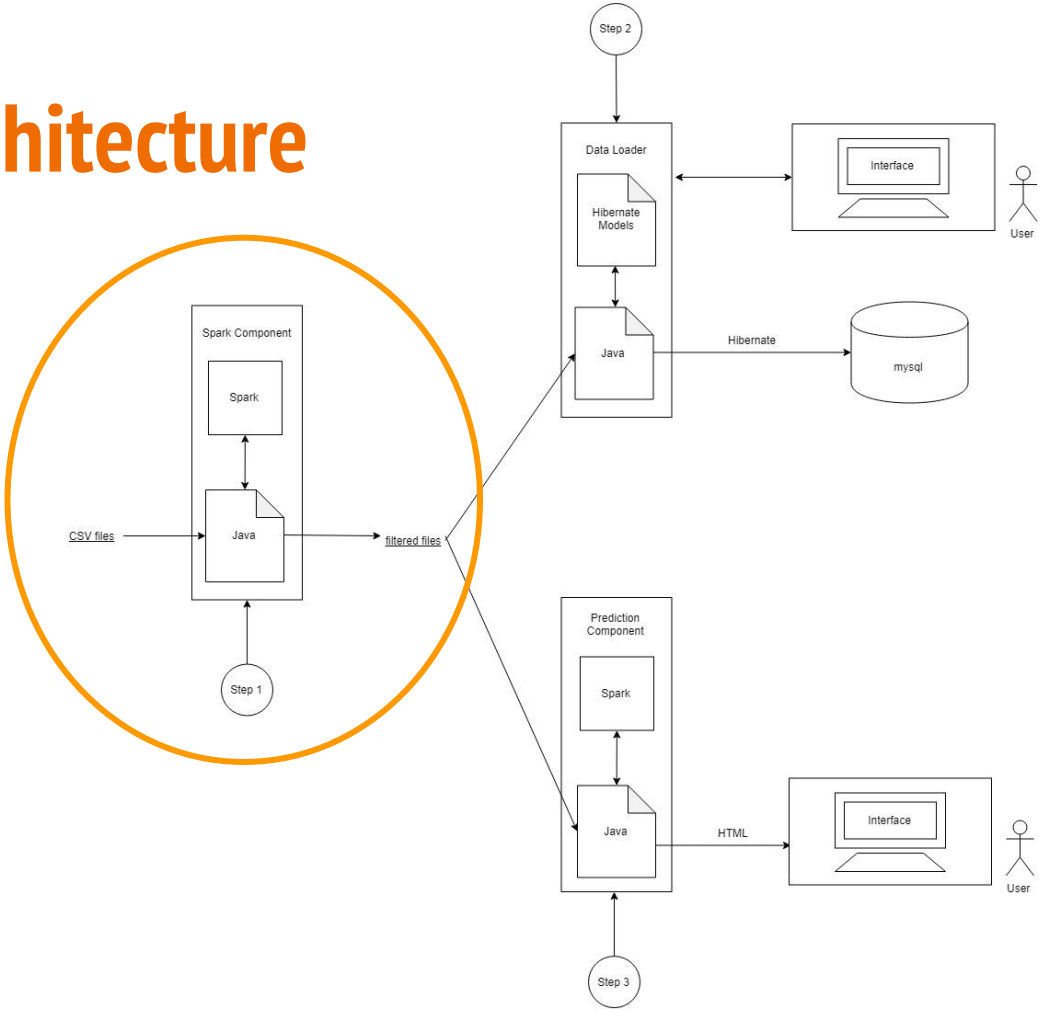
List of content

- Data description
- **Project Architecture and Technologies**
  - Spark Component
  - Data Loader
  - Prediction Component
- Conclusion and Lesson Learned

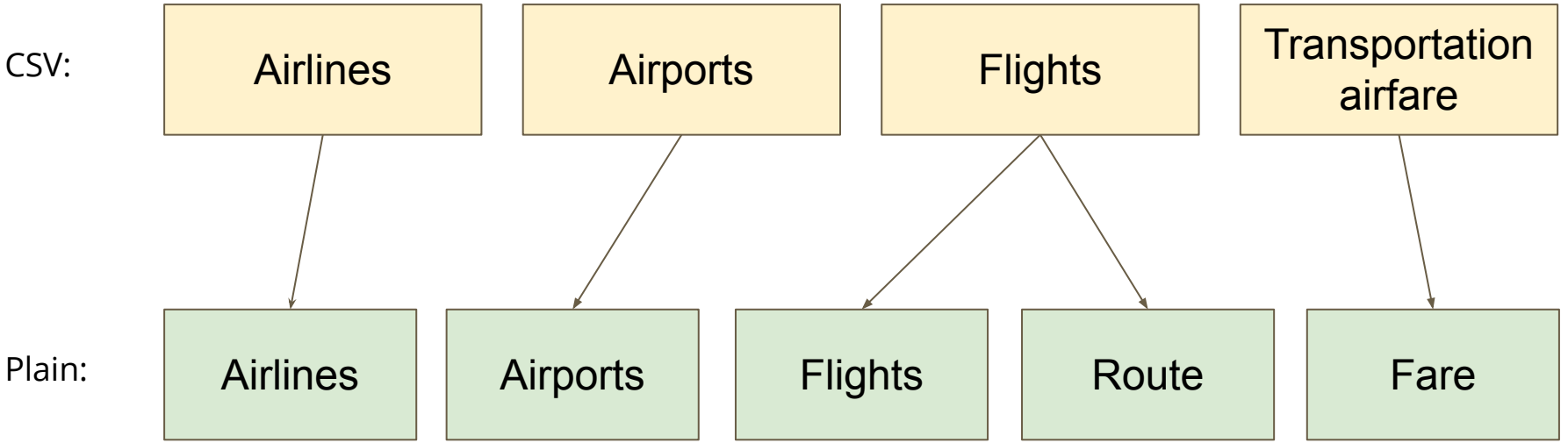
Project Architecture



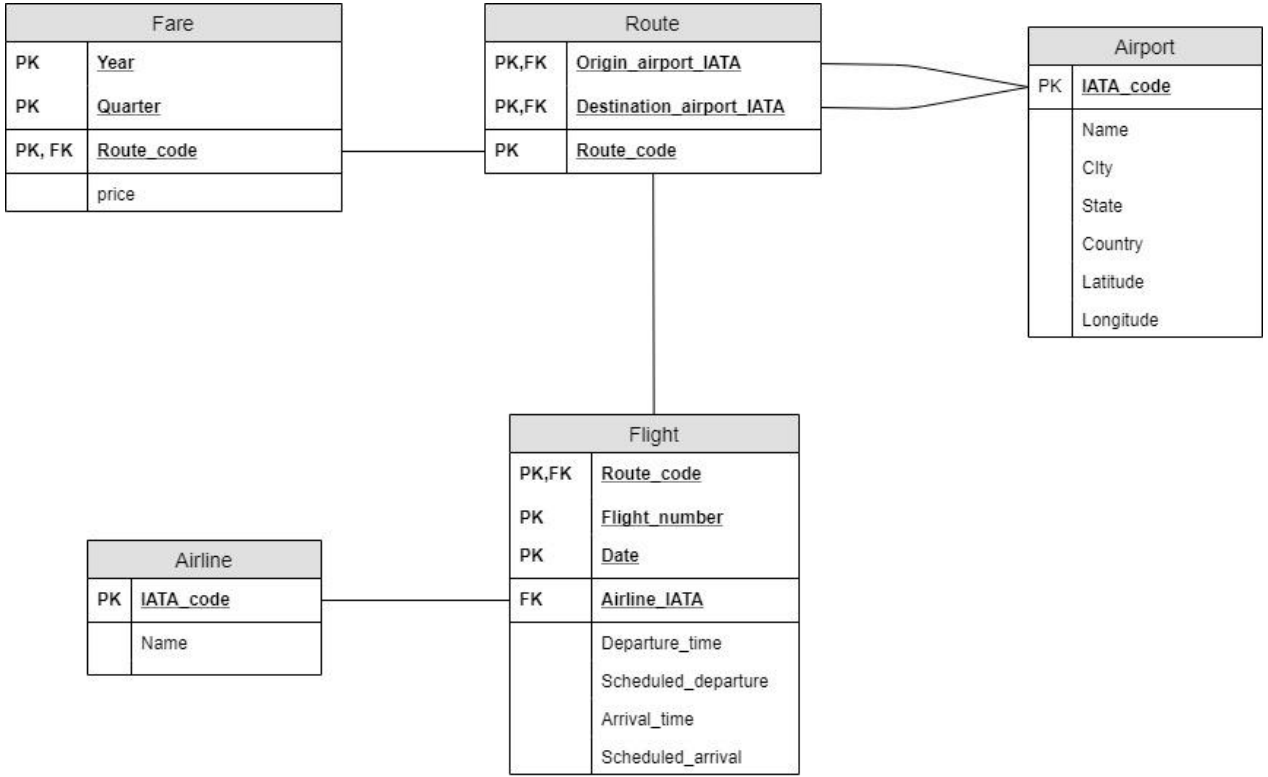
Project Architecture



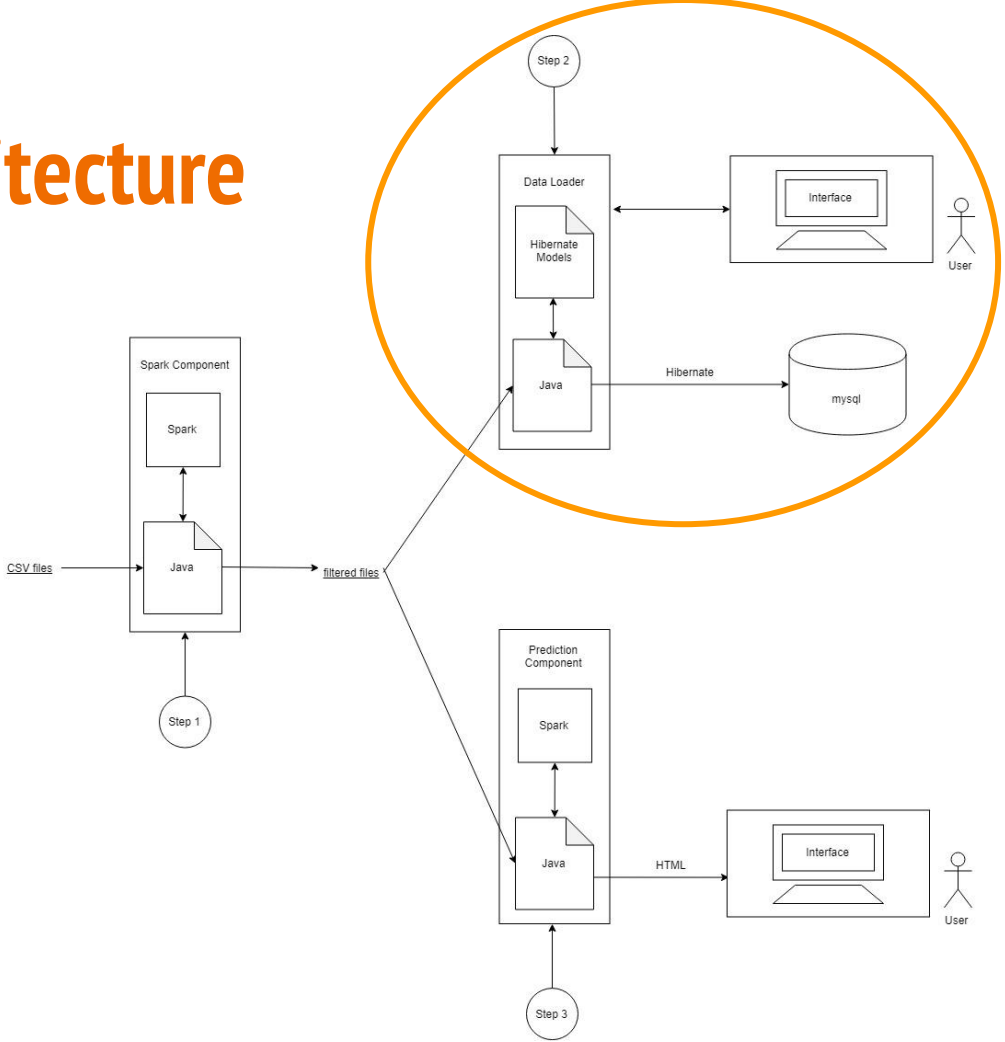
Step 1: Spark Component



Step 1: Spark Component



Project Architecture



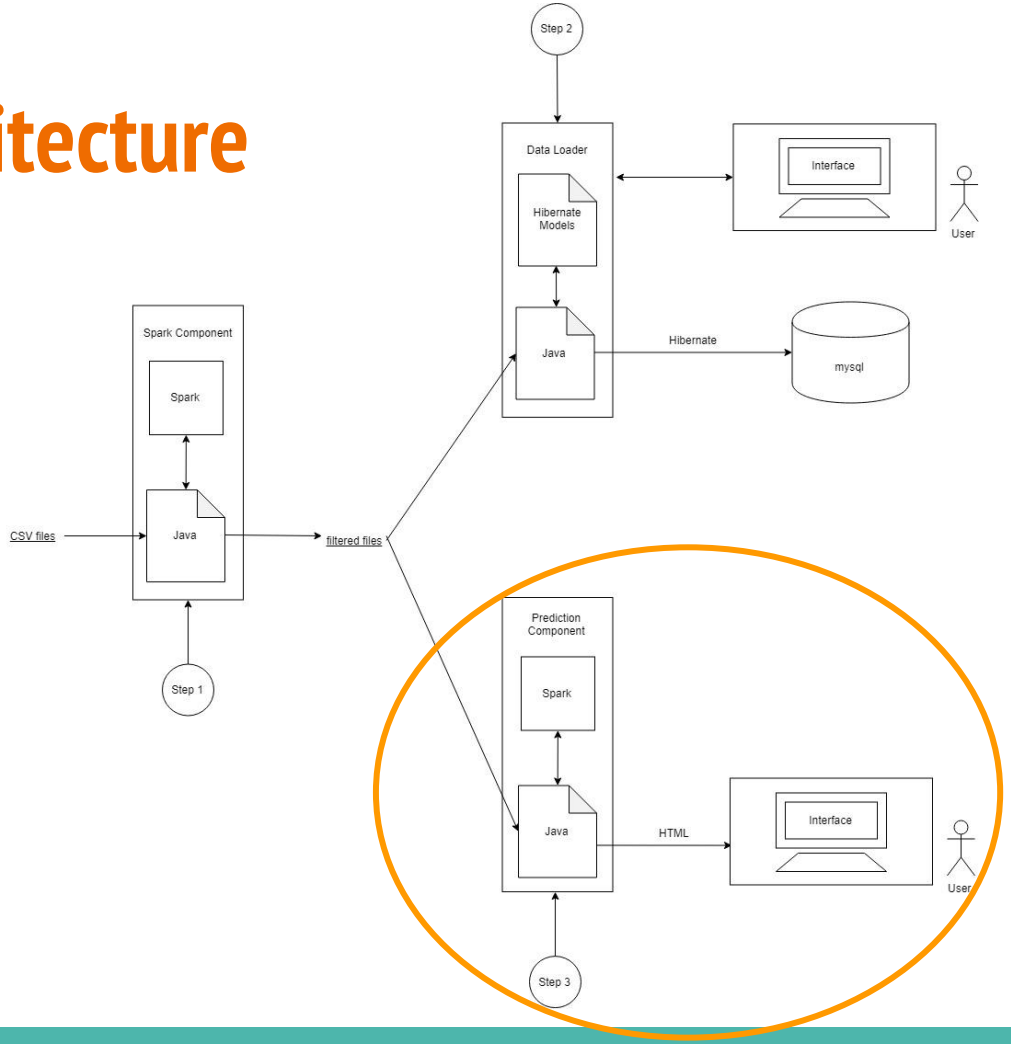
Step 2: Data Loader

- Interface
  - Flight Explorer
- Analytics
  1. What are the most expensive routes?
  2. What are the airports with the highest number of flights?
  3. What are the two most connected airports?
  4. What are the airlines which serve the lowest number of flights?
  5. What are the airports causing the greatest departure delay per flight?

Step 2: Data Loader

Demo

Project Architecture



Step 3: Prediction Component

- Prediction on the flights delay
- Analysis on “ANC-SEA” route
- Features:
  - Flight number;
  - Day of flight;
  - Departure Time;
  - Arrival Time

Step 3: Prediction Component

Demo

List of content

- Data description
- Project Architecture and Technologies
  - Spark Component
  - Data Loader
  - Prediction Component
- **Conclusion and Lesson Learned**

Conclusion and Lesson Learned

- Java Project, Spark + Hibernate as main technologies, MySQL as database, HTML+JavaScript and Java windows as interface
- Hibernate: fast queries but slow upload
- Improvement: interface enhancement and control panel creation

Thank you