Data Processing Pipeline

Pre-Processing MySQL

- Measurements every 15 min
- MySQL prefilters invalid data
- Generation of 3 CSV files:
  - Faults
  - Customer
  - DSM (Service and Maintenance Department)

Scala / Spark

- Reads 3 CSV files
- Filters invalid locations
- Joins Customers with DSMs
- Joins (Customer, DSM) with Faults
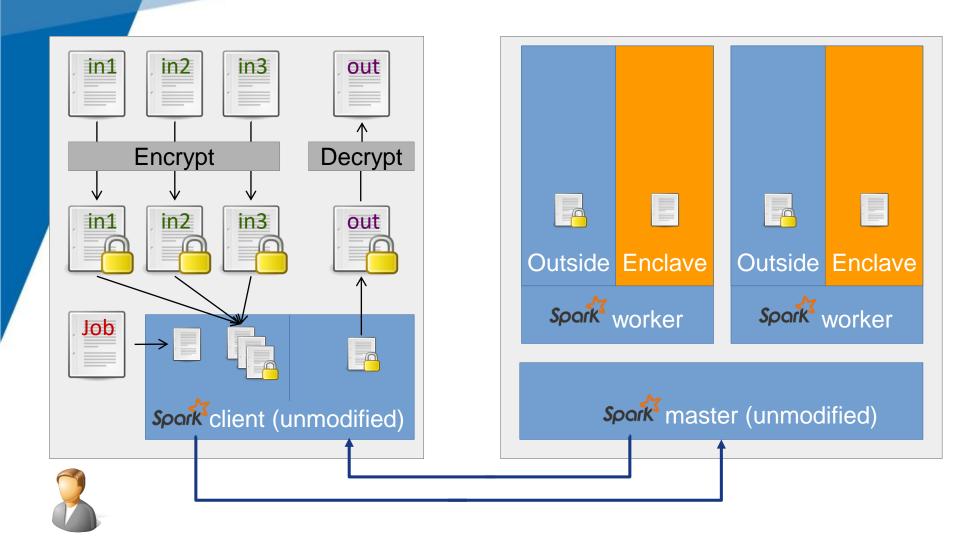- Computes total number of faults and total fault duration for each customer

Output

- Number of faults per consumer with respective service interruption times

Customer _id	Customer Name	Latitude	Longitude	Contract number	Medidor Serial	Service area abreviation	Service area name	Number of faults	Total fault duration in seconds
5467	USICAP	-23.27197	-51.05277	39633896	31606197	LNA	Londrina	2	1800

Data Processing Pipeline