



Apache Beam Overview

Israel Herraiz
@herraiz



What is Apache Beam?

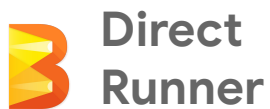
Apache Beam is an open source, **unified model** for defining both **batch** and **streaming** data-parallel processing pipelines. Using the open source Beam SDKs, you build a program defining the pipeline, that then is executed by one of Beam's supported distributed processing backends.



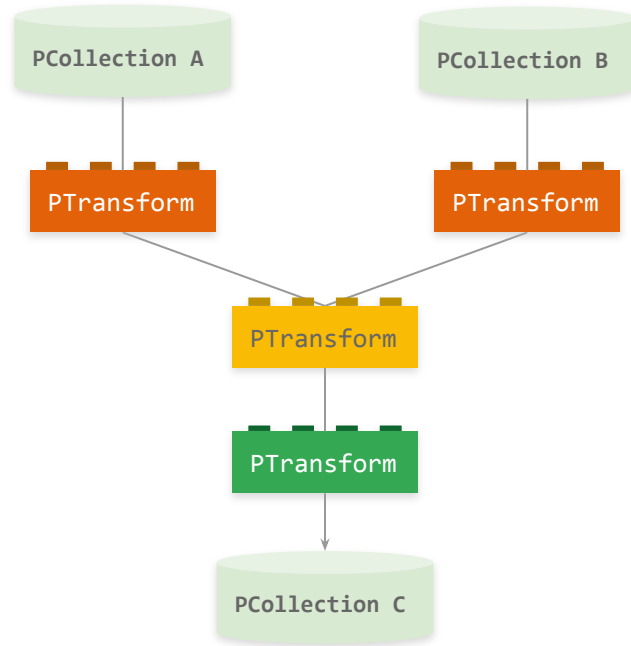
SDKs



Pipeline Runners



Pipeline



Overview of the workshop

Input: text of Don Quixote

Decía él, que el Cid Ruy Díaz había sido muy buen caballero; pero que no tenía que ver con el caballero de la ardiente espada, que de sólo un revés había partido por medio dos fieros y descomunales gigantes. Mejor estaba con Bernardo del Carpio, porque en Roncesvalle había muerto a Roldán el encantado, valiéndose de la industria de Hércules, cuando ahogó a Anteo, el hijo de la Tierra, entre los brazos. Decía mucho bien del gigante Morgante, porque con ser de aquella generación gigantesca, que todos son soberbios y descomedidos, él solo era afable y bien criado; pero sobre todos estaba bien con Reinaldos de Montalbán, y más cuando le veía salir de su castillo y robar cuantos topaba, y cuando en Allende ...

Output: most frequent words

que, 20590
de, 18193
y, 18138
la, 10357
a, 9849
en, 8200
el, 8199
no, 6192
los, 4744
se, 4690
con, 4183
por, 3889
las, 3464
lo, 3455
le, 3398
...

Hands on

Get the code from the Beam College repo

<https://github.com/Beam-College/season-2022>

Go to day-1

