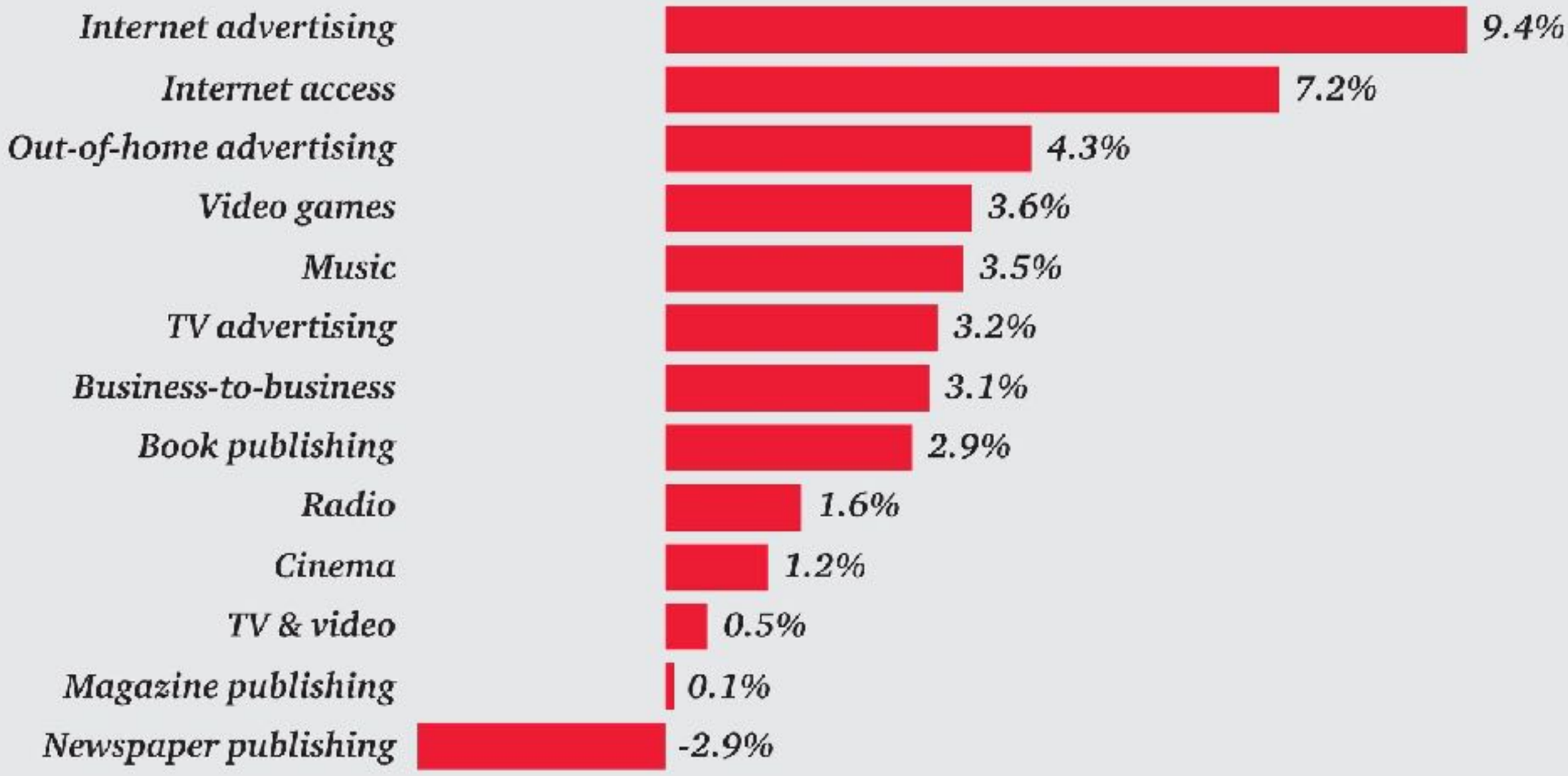
In-Game Recommender Engine

PwC: US Outlook for Entertainment & Media Segments

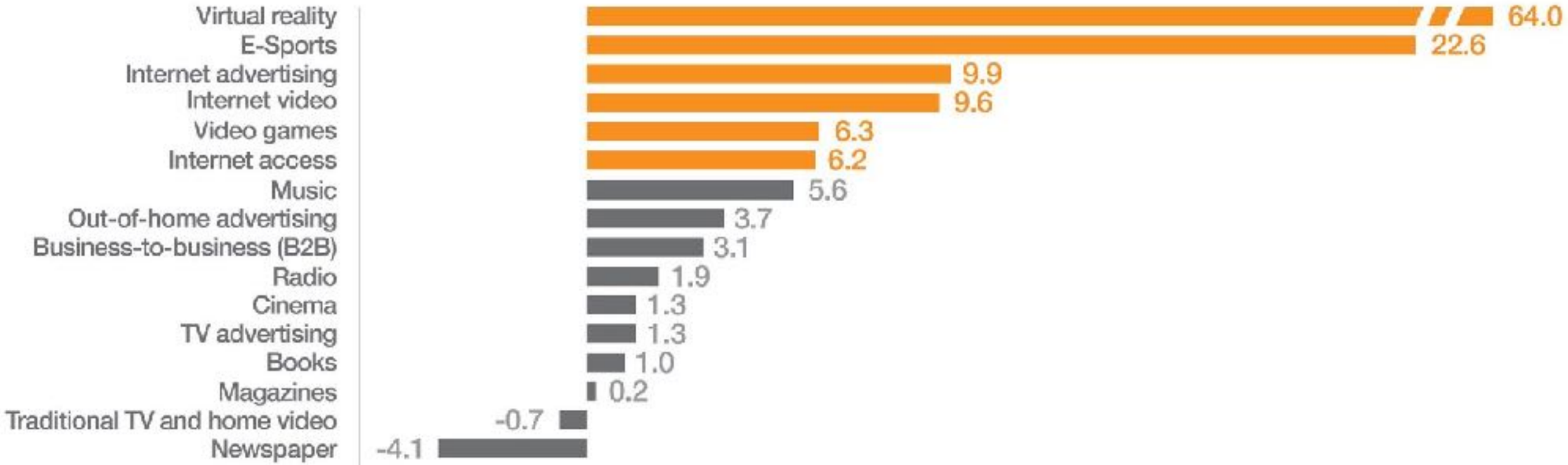
US growth in aggregate spending (2016 –2020 CAGR)



pwc

US segment CAGRs

2016 – 2021 (%)



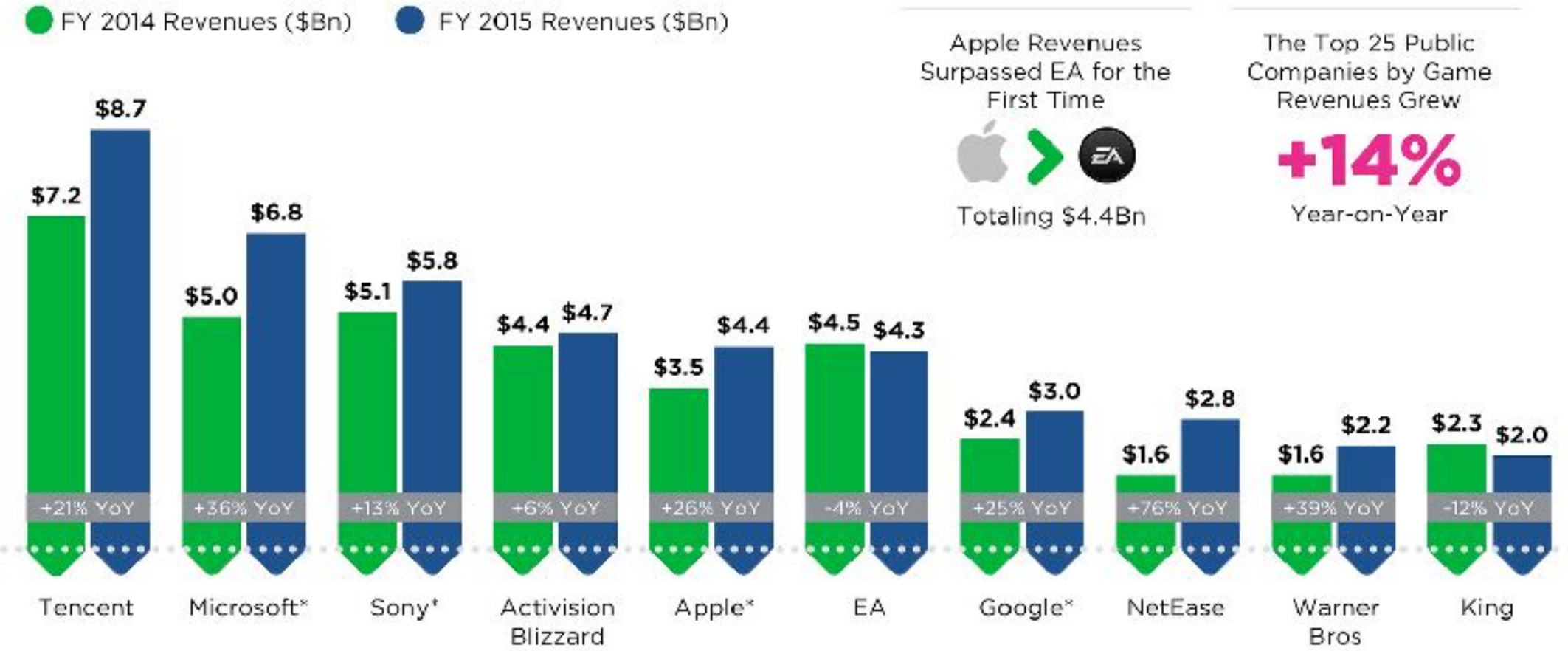
pwc

Background


Development
Publishing
Manufacturing
Distribution
Retail

TOP 10 PUBLIC COMPANIES BY GAME REVENUES

COMPARISON OF FY 2015 AND FY 2014 REVENUES ($BN)



Apple Revenues Surpassed EA for the First Time

Totaling $4.4Bn

The Top 25 Public Companies by Game Revenues Grew

+14%

Year-on-Year

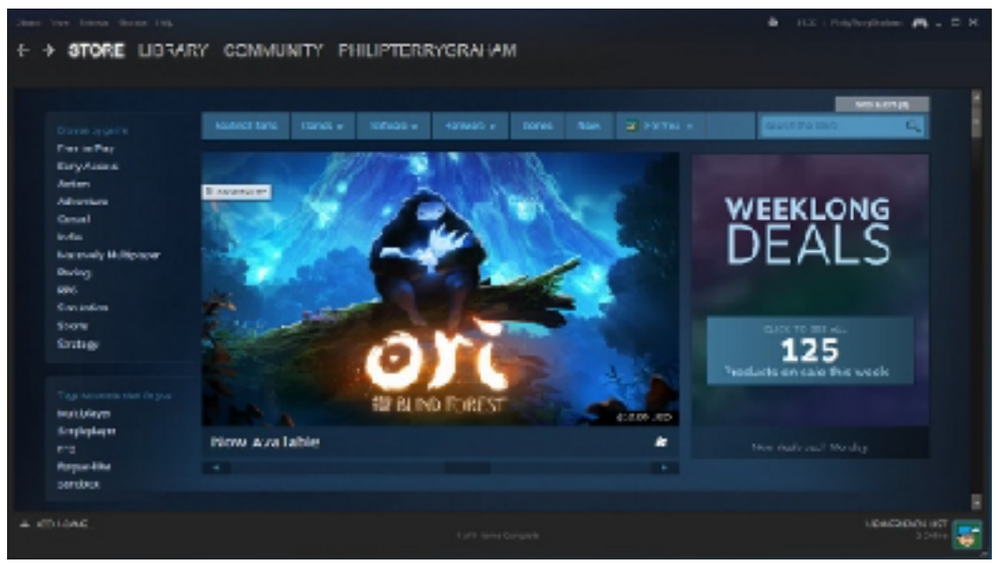
*Estimates of public game revenues using quarterly earning reports in which these are not specifically segmented out

Source: @Newzoo | Global Games Market Report Premium | newzoo.com/companyrankings/

Steam

- The largest digital distribution platform for PC gaming.

- Over 7,000 games available on Steam, with over 125 million active users.



store.steampowered.com/app/352460/

dead realm - Google Search | Dead Realm on Steam

Install Steam | login | language

STEAM

STORE COMMUNITY ABOUT SUPPORT

Featured Items | Games | Software | Hardware | Videos | News | For You | search the store

All Games > Action Games > Dead Realm

Dead Realm

Community Hub

Ghosts prey on the living inside the haunted mansion of a long dead electricity tycoon. Dead Realm is a creepy, multiplayer action game with beautiful, immersive environments. You can play as either a Ghost or a Human character and work with your friends to run, hide, survive... or die.

User reviews: Very Positive (2,692 reviews)

Release Date: Aug 3, 2015

Popular user-defined tags for this product:

Early Access | Horror | Multiplayer | Survival Horror | +

Dead Realm

60

DEAD REALM

PLAY

OPTIONS

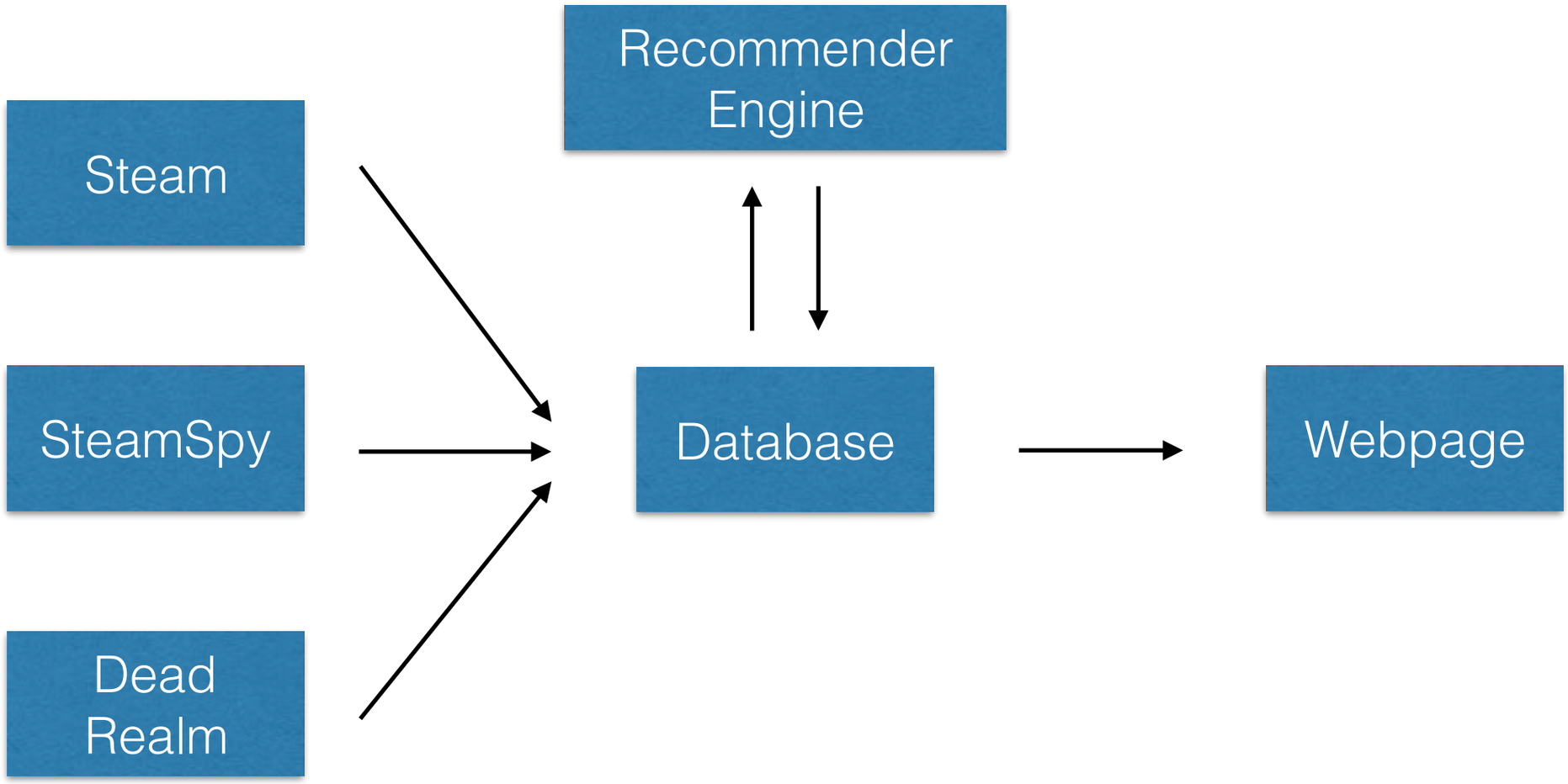
NEWS

QUIT

Submit Your Character

GO!

20


Recommender Engine
Steam
SteamSpy
Dead Realm
Database
Webpage

Schedule

WK1: Understand API and GIT; write crawler

WK2: Process raw data and set up a database

WK3: Build a recommender engine in Spark

Version Control

- Git
- Bitbucket
- Track and manage scripts

API

- Application Programming Interface
- Mostly, Web API
- Base URL + endpoints + token

e.g. YouTube API

- YouTube data API - public data
- Create a developer account and register your app @ google api console
- Read documents, look for base url, endpoints, and auth requirement

Crawler

- Basic: Requests, BeautifulSoup
- Good to have: Multiprocessing

Requests

- HTTP library for Python, safe for human consumption (a.k.a. readable)
- r = requests.get(url); r.json()
- Cons: can’t deal with ajax and javascript
- Alternatives: Selenium, Scrapy, etc.

BeautifulSoup

- A powerful HTML parser
- When API is not available, you have to collect data from webpages
- soup = BeautifulSoup(r.text, ‘lxml’)
- https://www.crummy.com/software/BeautifulSoup/bs4/doc/

e.g. requests + soup

- import requests

- from bs4 import BeautifulSoup

Our Project

- Recommender engine: (user_id, product_id, rating)

- Steam Web API: GetOwnedGames

Our Project

- App id not readable, we need app info as well
- Steam doesn’t offer official API for this
- Workaround: use 3rd party API - steamspy
  - step1: get all app id
  - step 2: get app detail for each id
  - be careful with quota and rate

Assignment

- Create a steam developer account and get game inventory for a sample of 5000 users (file attached)
- Get the latest app details using steam spy api
- Save into .txt files and each line should be a json string.

Hint

```python
path_file = 'id_to_crawl.txt'
with open(path_file, 'rb') as f:
    lst_user_id = [i for i in f.readlines()]
    for user_id in lst_user_id:
        # do something ...
        # do something ...
```

read file

```python
path_file = 'data.txt'
with open(path_file,'wb') as f:
    for json_data in lst_data:
        f.write(json.dumps(json_data))
        f.write('\n')
```

save file

- Optional: try multiprocessing to improve efficiency