CS 660: Big Data

CS 660: BIG DATA

Project Document - Results

Objective: Mine Google Play Store Repository

Name: Govind Yatnalkar

MUID: 901875614

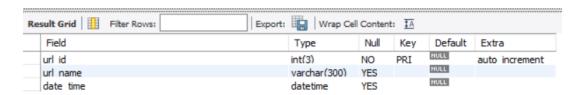
Date: 10th December 2018

Motive of the document:

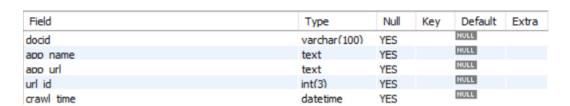
This document provides the structure of the tables and the results of the project.

Structure of tables: -

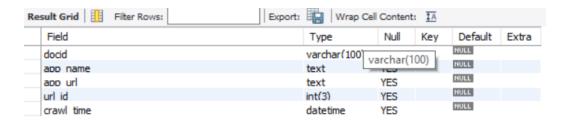
1. Table: urls - Table for storing an url like https://play.google.com/store/apps/top with an id and a timestamp.



2. appdata - Table for storing Google Store Crawled App Data



3. **bestappdata** – Table for storing Best Apps Site Crawled Data



Python Script Output for Play Store Link:

https://play.google.com/store/apps/category/BUSINESS

```
Data Was Inserted in MySQL Database for docid: com.ics.creditcardreader
Data Was Inserted in MySQL Database for docid: com.squareup
Data Was Inserted in MySQL Database for docid: com.twansoftware.invoicemakerpro
Data Was Inserted in MySQL Database for docid: com.invoiceasap.nativeapp
Data Was Inserted in MySQL Database for docid: com.concur.breeze
Data Was Inserted in MySQL Database for docid: com.paypal.here
This App Was Found in This Crawl Event. Only Updating Time for Docid: com.intuit.quickbooks
Data Was Inserted in MySQL Database for docid: com.intuit.turbotax.mobile
Data Was Inserted in MySQL Database for docid: com.hrblock.blockmobile
Data Was Inserted in MySQL Database for docid: com.taxact.express2013
Data Was Inserted in MySQL Database for docid: com.iqboxyinc.iqboxy
Data Was Inserted in MySQL Database for docid: com.taxfyle.TaxfyleMobile
Data Was Inserted in MySQL Database for docid: com.waveaccounting.receipts
Data Was Inserted in MySQL Database for docid: com.intuit.qbse
This App Was Found in This Crawl Event. Only Updating Time for Docid: com.intuit.quickbooks
Data Was Inserted in MySQL Database for docid: com.microsoft.rdc.android
```

Entertainment Play Store Link Apps Crawled

```
Data Was Inserted in MySQL Database for docid: com.harkinstheatres.android.showtimeapp
This App Was Found in This Crawl Event. Only Updating Time for Docid: com.fandango
Data Was Inserted in MySQL Database for docid: com.fandango.regal
Data Was Inserted in MySQL Database for docid: com.imdb.mobile
This App Was Found in This Crawl Event. Only Updating Time for Docid: com.crunchyroll.crunchyroid
This App Was Found in This Crawl Event. Only Updating Time for Docid: com.aetn.history.watch
Data Was Inserted in MySQL Database for docid: com.showtime.showtimeanytime
Data Was Inserted in MySQL Database for docid: com.HBO
Data Was Inserted in MySQL Database for docid: com.cw.seed.android
Data Was Inserted in MySQL Database for docid: com.nbcuni.nbc
This App Was Found in This Crawl Event. Only Updating Time for Docid: com.netflix.mediaclient
This App Was Found in This Crawl Event. Only Updating Time for Docid: com.hulu.plus
This App Was Found in This Crawl Event. Only Updating Time for Docid: com.tubitv
This App Was Found in This Crawl Event. Only Updating Time for Docid: com.gotv.crackle.handset
Data Was Inserted in MySQL Database for docid: com.google.android.apps.youtube.kids
Data Was Inserted in MySQL Database for docid: com.hmusic.memory7080
```

An Example of Cross URL Apps

(Netflix in Top Charts & Entertainment Crawl URL)

docid	app_name	app_url	url_id	crawl_time
com.netflix.mediaclient	Netflix	/store/apps/details?id=com.netflix.mediaclient	1	2018-12-11 00:48:58
com.hulu.plus	Hulu: Stream TV. Movies & more	/store/apps/details?id=com.hulu.plus	3	2018-12-1100:48:58
com.nick.android.nickir	Nick Jr Shows & Games	/store/apps/details?id=com.nick.android.nickir	13	2018-12-1100:48:58
com.fandango	Fandango Movie Tickets & Times	/store/apps/details?id=com.fandango	13	2018-12-1100:48:58
com.viki.android	Viki: Korean Drama. Movies & Asian TV	/store/apps/details?id=com.viki.android	13	2018-12-1100:48:58
com.nickonline.android.nickapp	Nick	/store/apps/details?id=com.nickonline.android	13	2018-12-1100:48:58
com.laughoutloud.laughoutloud	Laugh Out Loud by Kevin Hart	/store/apps/details?id=com.laughoutloud.laugh	13	2018-12-1100:48:58
com.hamilton.app	Hamilton — The Official App	/store/apps/details?id=com.hamilton.app	13	2018-12-1100:48:58
com.tubitv	Tubi - Free Movies & TV Shows	/store/apps/details?id=com.tubitv	13	2018-12-1100:48:58
com.gotv.crackle.handset	Sonv Crackle - Free TV & Movies	/store/apps/details?id=com.gotv.crackle.handset	13	2018-12-11 00:48:58
1 1 1 1	^ ! "	7 - 7 - 71 - 4 × 1 - 1 - 1 - 1	**	2012 12 11 22 12 12

If the link is present and if apps are found, only time is updated for url and for apps

A screenshot of MySQL after crawling 32 URLS 2 times:

docid	app_name	app_url	url_id	crawl_time
com.ietstartoames.chess	Chess	/store/apps/details?id=com.ietstartgames.chess	32	2018-12-11 00:58:44
com.LoopGames.Domino	Dominoes	/store/apps/details?id=com.LoopGames.Domino	32	2018-12-11 00:58:44
com.cardoame.solitaire	Solitaire	/store/apps/details?id=com.cardgame.solitaire	31	2018-12-11 00:57:57
com.roblox.dient	ROBLOX	/store/apps/details?id=com.roblox.client	31	2018-12-11 00:57:57
com.king.candvcrushsaga	Candy Crush Saga	/store/apps/details?id=com.king.candvcrushsaga	31	2018-12-11 00:57:5
com.superbox.aos.iewelsiunale	Jewels Junale : Match 3 Puzzle	/store/apps/details?id=com.superbox.aos.iewel	31	2018-12-11 00:57:5
com.king.candvcrushsodasaga	Candy Crush Soda Saga	/store/apps/details?id=com.king.candvcrushsod	31	2018-12-11 00:57:5
com.minidip.eightballpool	8 Ball Pool	/store/apps/details?id=com.miniclip.eightballpool	31	2018-12-11 00:57:5
com.linkdesks.bubblegames.bubbleshooter	Bubble Shooter Genies	/store/apps/details?id=com.linkdesks.bubblega	31	2018-12-11 00:57:5
com.hectopixel.match3	Match 3 Jewels	/store/apps/details?id=com.hectopixel.match3	31	2018-12-11 00:57:5
com.finaersoft.hcr2	Hill Climb Racing 2	/store/apps/details?id=com.fingersoft.hcr2	31	2018-12-11 00:57:5
com.ticketmaster.mobile.android.na	Ticketmaster Event Tickets	/store/apps/details?id=com.ticketmaster.mobile	14	2018-12-11 00:57:2
com.stubhub	StubHub - Tickets to Sports	/store/apps/details?id=com.stubhub	14	2018-12-11 00:57:2
com.seatoeek.android	SeatGeek â€" Tickets to Sport	/store/apps/details?id=com.seatgeek.android	14	2018-12-11 00:57:2
com.espn.score center	ESPN	/store/apps/details?id=com.espn.score center	30	2018-12-11 00:57:2
com.gotv.nflgamecenter.us.lite	NFL	/store/apps/details?id=com.gotv.nflgamecenter	30	2018-12-11 00:57:2
com.nbadigital.gametimelite	NBA	/store/apps/details?id=com.nbadigital.gametim	30	2018-12-11 00:57:2
com.livescore	LiveScore: Live Sport Updates	/store/apps/details?id=com.livescore	30	2018-12-11 00:57:2
com.handmark.sportcaster	CBS Sports App - Scores, Ne	/store/apps/details?id=com.handmark.sportcaster	30	2018-12-11 00:57:2
com.vahoo.mobile.client.android.sportacular	Yahoo Sports - scores, stats,	/store/apps/details?id=com.vahoo.mobile.client	30	2018-12-11 00:57:2
com, fivemobile, thescore	theScore: Live Sports Scores	/store/apps/details?id=com, fivemobile, thescore	30	2018-12-11 00:57:2
com.hole 19aolf.hole 19.beta	Hole 19: Golf GPS App. Range	/store/apps/details?id=com.hole19golf.hole19	30	2018-12-11 00:57:2
com.manutd.mutv	MUTV - Manchester United TV	/store/apps/details?id=com.manutd.mutv	30	2018-12-11 00:57:2
com.bleacherreport.android.teamstream	Bleacher Report: sports news	/store/apps/details?id=com.bleacherreport.and	30	2018-12-11 00:57:2
com.espn.fantasv.lm.football	ESPN Fantasy Sports	/store/apps/details?id=com.espn.fantasv.lm.fo	30	2018-12-11 00:57:2
com.vahoo.mobile.client.android.fantasvf	Yahoo Fantasy Sports - #1 R	/store/apps/details?id=com.vahoo.mobile.client	30	2018-12-11 00:57:2
com.cbs.sports.fantasv	CBS Sports Fantasy	/store/apps/details?id=com.cbs.sports.fantasv	30	2018-12-11 00:57:2
com.vividseats.android	Vivid Seats - Concert Tickets	/store/apps/details?id=com.vividseats.android	30	2018-12-11 00:57:2
com gametime gametime	Gametime - Tickets to Sports	/store/anns/details?id=com gametime gametime	30	2018-12-11 00-57-2

TO RUN THE CODE:

There are about 30-32 Google Play URLS. These have been already crawled but still can be crawled again. For Play Store, select any URL from urls notepad file and execute the program.

For best app website, use the following page: https://www.androidpit.com/best-android-apps. The working is similar to both the tasks. URL's are saved in the database and are updated according to the crawl events.