The Blue Alliance - 3 Apr 2012 Data Loss Postmortem

- On April 3rd, 2012, The Blue Alliance experienced data loss due to unprotected data deletion mechanisms designed for local development and debugging. Efforts to restore data failed because my.usfirst.org was down.
- On April 4th, it was discovered that The Blue Alliance's scrapers high-velocity attempt to restore data had brought my.usfirst.org down, and were now being blocked by FIRST.
- On April 5th, 2012, appropriate scraping rates were established with FIRST, The Blue Alliance's scrapers were unblocked, and all data was restored.

Timeline

All times Pacific

- 3 Apr 2012, ??:?? - Data is lost in prod
- 3 Apr 2012, 18:38 - Greg contacted by user reporting security hole where methods designed to be used for test server data purging are improperly protected, and can be called without authentication
- 3 Apr 2012, 18:51 - Greg replies
- 3 Apr 2012, 21:27 - Reporting user replies that methods were unknowingly called during probe of security hole. Thus, data was lost in prod.
- 3 Apr 2012, 23:30 - Greg patches security hole, pushes new version TBA to prod
  - https://github.com/gregmarra/the-blue-alliance/commit/d8c87fe92f6ddfd914b064fe7c4b9e765c0ed25e
- 3 Apr 2012, 23:35 - Greg partially restores data from test_data/events.csv and test_data/matches.csv, which contained a TBA.net export current as of 2010. Cron jobs appear to have scraped 2012 data again. Notably, 2011 is missing.
- 3 Apr 2012, 23:40 - USFIRST's data pages appear to be down. http://www.usfirst.org/whatsgoingon is not loading, so TBA scrapers can't restore data from USFIRST.
- 4 Apr 2012, 00:06 - Most data that can be restored has been restored. Notably missing is 2011.
- 4 Apr 2012, 9:39 - More data restored while my.usfirst.org was temporarily back up, but went back down
- 4 Apr 2012, 20:00 - Updated Event scrapers to be able to scrape historic years, but for some reason prod is behaving differently than local dev and not fetching events, including for 2012.
- 4 Apr 2012, 23:00 - FIRST is blocking scraping attempts from TBA
- 5 Apr 2012, 14:00 - Resolved issues with FIRST web team
- 5 Apr 2012, 19:46 - New task queue deployed, rescraping FIRST
- 5 Apr 2012, 20:01 - Events from 2011, 2012 completely restored

- 5 Apr 2012, 20:16 - Matches from 2011, 2012 completely restored

Data Restoration Done

- Cron jobs appear to have rescraped 2012 Events, Teams, Matches
- Events 2010 and back have been restored
- Matches for 2010 and back have been restored
- Reimport all YouTube videos from [spreadsheet](#)
- Rebuild EventTeams
- Rebuild tba_videos
- Repaired invalid data from csv imports
- Rerun team scrapers for 2009, 2010, 2011, 2012
- Rerun event scrapers for 2011, 2012
- Rerun match scrapers for 2011, 2012
- Rerun tbavideo scrapers for 2011, 2012