



# Wikimedia Content API:

A Cassandra Use-case

Eric Evans <[eevans@wikimedia.org](mailto:eevans@wikimedia.org)>  
[@jericevans](https://twitter.com/jericevans)

Berlin Buzzwords | June 6, 2016



# Our Vision:

A world in which every single human can freely share in the sum of all knowledge.

# About:

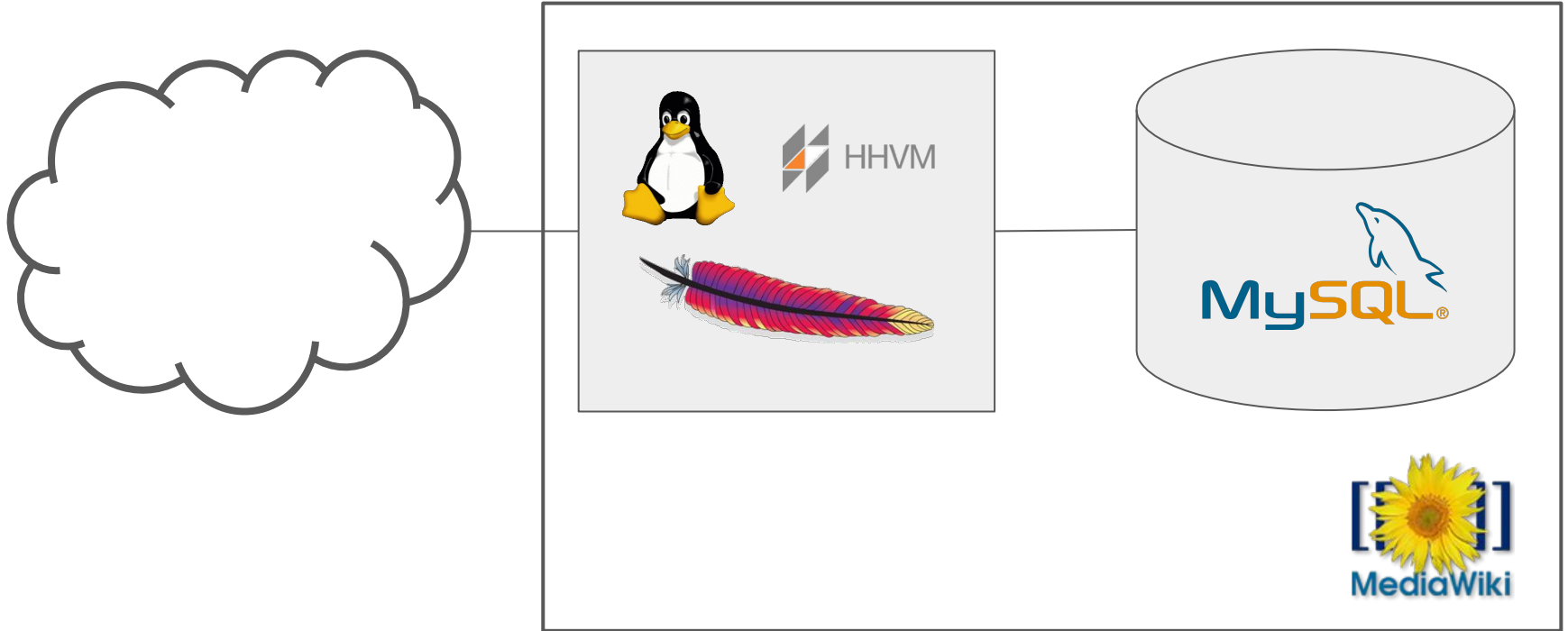
- Global movement
- Largest collection of free, collaborative knowledge in human history
- 16 projects
- 16.5 billion total page views per month
- 58.9 million unique devices per day
- More than 13k new editors each month
- More than 75k active editors month-to-month

# About: Wikipedia

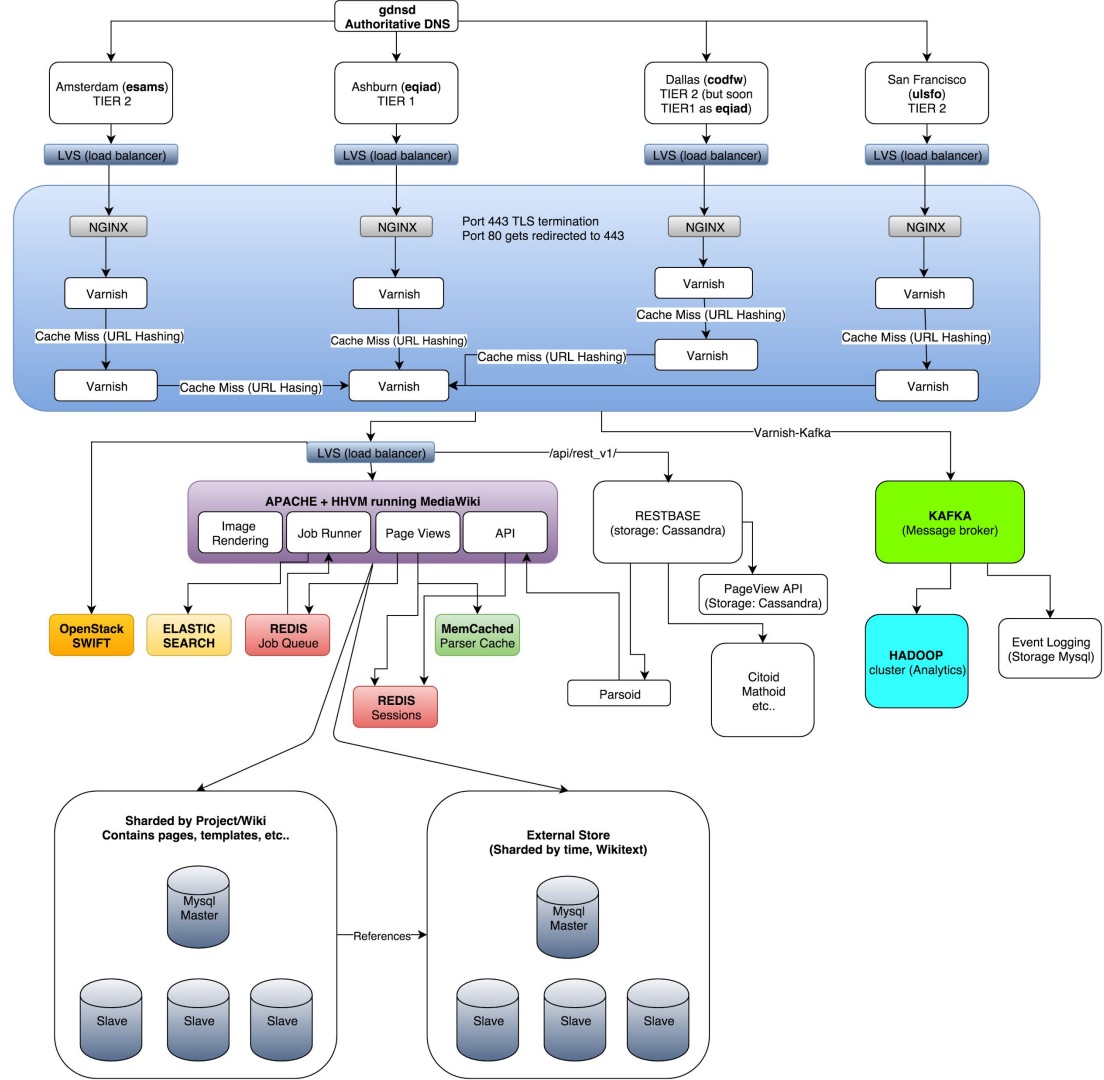
- More than 38 million articles in 290 languages
- Over 10k new articles added per day
- 13 million edits per month
- Ranked #6 globally in web traffic

# Wikimedia Architecture

# LAMP



# THE ARCHITECTURE



ALL OF IT





visual editor

# Wikitext

= Star Wars: The Force Awakens =

Star Wars: The Force Awakens is a 2015 American epic space opera film directed, co-produced, and co-written by [[J. J. Abrams]].

# HTML

```
<h1>
```

```
  Star Wars: The Force Awakens
```

```
</h1>
```

```
<p>
```

```
  Star Wars: The Force Awakens is a 2015 American epic space opera  
  film directed, co-produced, and co-written by
```

```
  <a href="/wiki/J._J._Abrams" title="J. J. Abrams">
```

```
    J. J. Abrams
```

```
  </a>
```

```
</p>
```



# HTML

Star Wars: The Force Awakens – Wikipedia, the free encyclopedia – Google Chrome

W Star Wars: The Force x

https://en.wikipedia.org/wiki/Star\_Wars:\_The\_Force\_Awakens

Jericevans 0 0 Talk Sandbox Preferences Beta Watchlist Contributions Log out

Article Talk Read Edit Edit source View history Search

WIKIPEDIA  
The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
Donate to Wikipedia  
Wikipedia store

Interaction  
Help  
About Wikipedia  
Community portal  
Recent changes  
Contact page

Tools  
What links here  
Related changes  
Upload file  
Special pages

**Star Wars: The Force Awakens**


From Wikipedia, the free encyclopedia

**Star Wars: The Force Awakens** (also known as **Star Wars: Episode VII – The Force Awakens**) is a 2015 American [epic space opera](#) film directed, co-produced, and co-written by [J. J. Abrams](#). The seventh installment in the main *Star Wars* film series, it stars [Harrison Ford](#), [Mark Hamill](#), [Carrie Fisher](#), [Adam Driver](#), [Daisy Ridley](#), [John Boyega](#), [Oscar Isaac](#), [Lupita Nyong'o](#), [Andy Serkis](#), [Domhnall Gleeson](#), [Anthony Daniels](#), [Peter Mayhew](#), and [Max von Sydow](#). Produced by [Lucasfilm](#) and Abrams' [Bad Robot Productions](#) and distributed worldwide by [Walt Disney Studios Motion Pictures](#), *The Force Awakens* is set 30 years after *Return of the Jedi*; it follows [Rey](#), [Finn](#), and [Poe Dameron](#)'s search for [Luke Skywalker](#) and their fight alongside the [Resistance](#), led by veterans of the [Rebel Alliance](#), against [Kylo Ren](#) and the [First Order](#), a successor group to the [Galactic Empire](#).

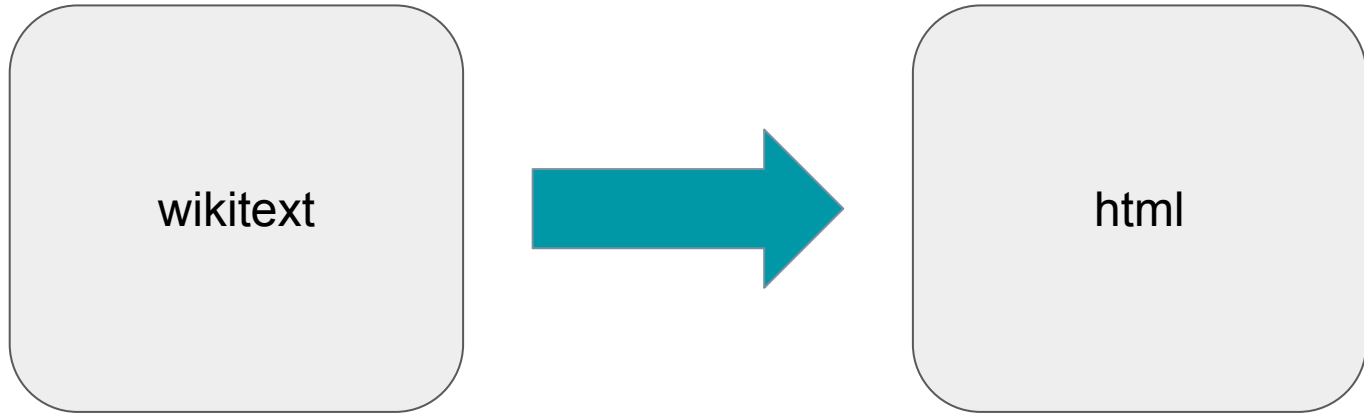
*The Force Awakens* is the first film in the [Star Wars sequel trilogy](#) announced after Disney's acquisition of Lucasfilm in October 2012. It was produced by Abrams, his longtime collaborator [Bryan Burk](#), and Lucasfilm president [Kathleen Kennedy](#). Abrams and [Lawrence Kasdan](#), co-writer of the [original trilogy](#) films *The Empire Strikes Back* (1980) and *Return of the Jedi* (1983), rewrote an initial script by [Michael Arndt](#). [John Williams](#), composer for the previous six films, returned to compose its score. *Star Wars* creator [George Lucas](#) served as creative consultant

Accepted (latest)

**Star Wars: The Force Awakens**



# Conversion



# WYSIWYG

Editing Star Wars: The Force Awakens – Wikipedia – Google Chrome

W Editing Star Wars: The Force Awakens – Wikipedia – Google Chrome

https://en.wikipedia.org/wiki/Star\_Wars:\_The\_Force\_Awakens?veaction=edit


Jericevans 0 0 Talk Sandbox Preferences Beta Watchlist Contributions Log out

Article Talk Read Edit Edit source View history Search

Paragraph A Cite Insert Ω ? ! ≡ || Save page


**Star Wars: The Force Awakens** Accepted (latest)

From Wikipedia, the free encyclopedia



**Star Wars: The Force Awakens** (also known as **Star Wars: Episode VII – The Force Awakens**) ⓘ is a 2015 American *epic space opera* film directed, co-produced, and co-written by [J. J. Abrams](#). The seventh installment in the main *Star Wars* film series, it stars [Harrison Ford](#), [Mark Hamill](#), [Carrie Fisher](#), [Adam Driver](#), [Daisy Ridley](#), [John Boyega](#), [Oscar Isaac](#), [Lupita Nyong'o](#), [Andy Serkis](#), [Domhnall Gleeson](#), [Anthony Daniels](#), [Peter Mayhew](#), and [Max von Sydow](#). ⓘ Produced by [Lucasfilm](#) and Abrams' [Bad Robot Productions](#) and distributed worldwide by [Walt Disney Studios Motion Pictures](#), *The Force Awakens* is set 30 years after

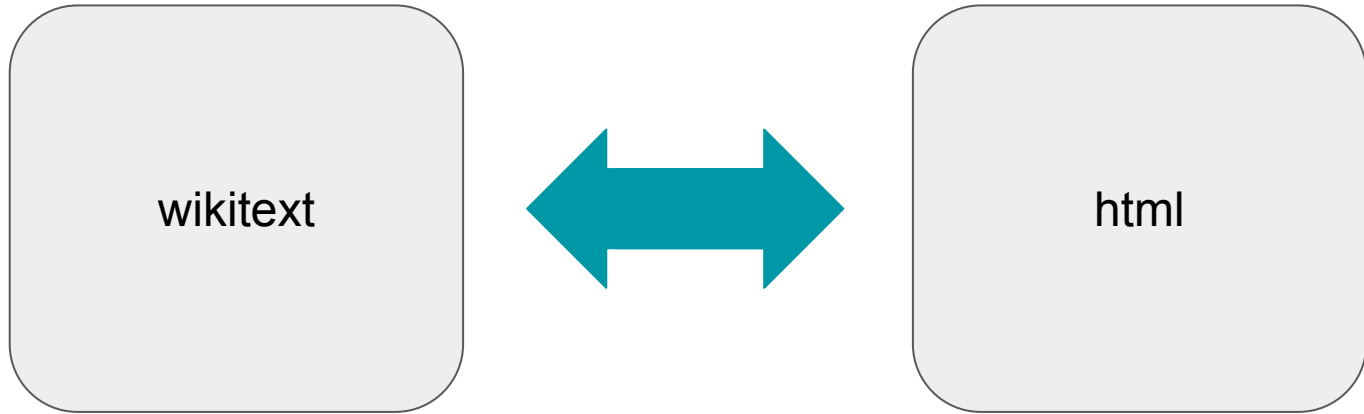
**Star Wars: The Force Awakens**



https://en.wikipedia.org/w/index.php?title=Star\_Wars:\_The\_Force\_Awakens



# Conversion



# Character-based diffs

Star Wars: The Force Awakens: Difference between revisions - Wikipedia, the free encyclopedia - Google Chrome

W Star Wars: The Force x

https://en.wikipedia.org/w/index.php?title=Star\_Wars:\_The\_Force\_Awakens&diff=prev&oldid=716626073

Jericevans 0 0 Talk Sandbox Preferences Beta Watchlist Contributions Log out

Article Talk Read Edit Edit source View history Search

## Star Wars: The Force Awakens: Difference between revisions

From Wikipedia, the free encyclopedia

**Revision as of 20:24, 22 April 2016 (edit)**      **Revision as of 20:30, 22 April 2016 (edit) (undo) (thank)**

**Line 50:**

<pre>- * \$245-259 million (net)&lt;ref name="BOM" /&gt; - &lt;ref&gt;http://deadline.com/2016/03/star-wars-the-force-awakens-movie-profit-2015-lucasfilm-disney-1201726142/&lt;/ref&gt; - }} -   gross = \$2.066 billion&lt;ref name="BOM"&gt;{{cite web -  url=http://www.boxofficemojo.com/movies/?id=starwars7.htm  title=Star Wars: The Force Awakens (2015) publisher=[[Box Office Mojo]] accessdate=April 18, 2016}}&lt;/ref&gt; - }} - </pre>	<pre>+ * \$245 (net)&lt;ref name="BOM" /&gt; + }} +   gross = \$2.066 billion&lt;ref name="BOM"&gt;{{cite web +  url=http://www.boxofficemojo.com/movies/?id=starwars7.htm  title="Star Wars: The Force Awakens" (2015) publisher=[[Box Office Mojo]] accessdate=April 18, 2016}}&lt;/ref&gt; + }} + </pre>
---	---

**Line 141:**

# Metadata

```
[[Foo|bar]]
```

```
<a rel="mw:WikiLink" href="./Foo">bar</a>
```

# Metadata

```
[[Foo|{{echo|bar}}]]
```

```
<a rel="mw:WikiLink" href="./Foo">  
  <span about="#mwt1" typeof="mw:Object/Template"  
    data-parsoid="{...}" >bar</span>  
</a>
```

# Parsoid

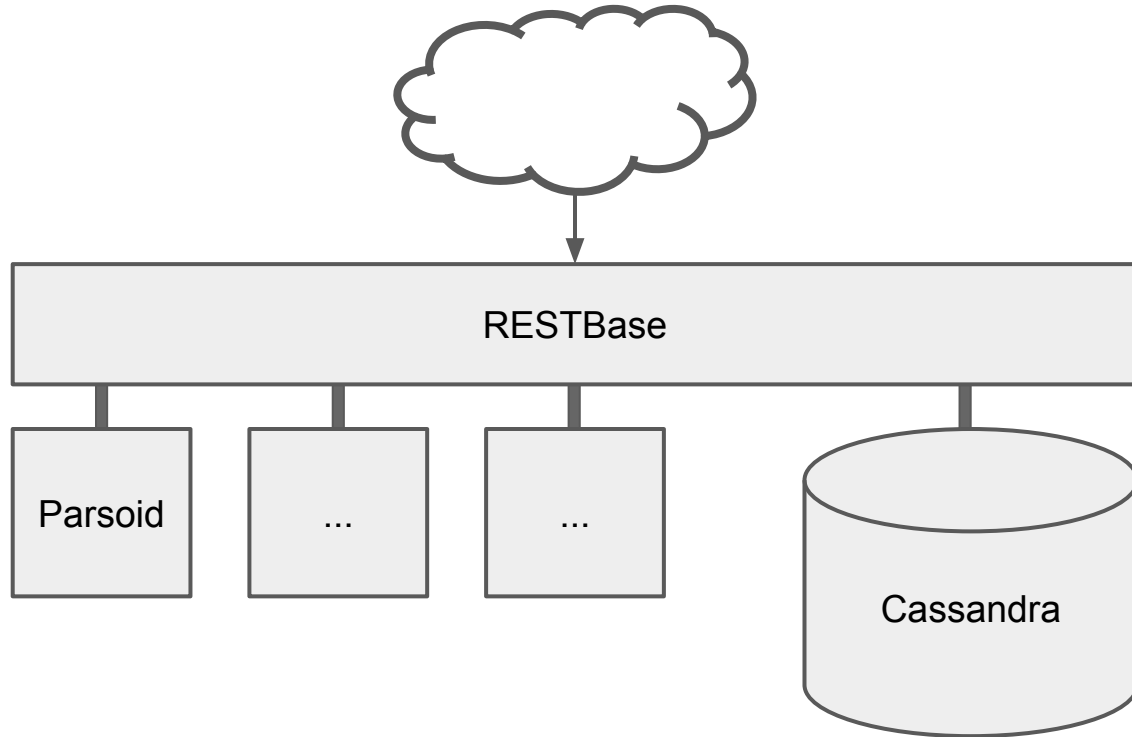
- Node.js service
- Converts wikitext to HTML/RDFa
- Converts HTML/RDFa to wikitext
- Semantics, *and* syntax (avoid dirty diffs)!
- Expensive (slow)
- Resulting output is large



# RESTBase

- Services aggregator / proxy (REST)
- Durable cache (Cassandra)
- Wikimedia's content API (e.g. [https://en.wikipedia.org/api/rest\\_v1?doc](https://en.wikipedia.org/api/rest_v1?doc))

# RESTBase



# Other use-cases

- Mobile content service
- Math formula rendering service
- Dumps
- ...



Cassandra

# Environment

- 2 datacenters
- 3 racks per datacenter
- 18 hosts (16 core, 128G, SSDs)
- 54 nodes
- Deflate compression (~14-18%)
- 31T storage (~206T uncompressed)
- Cassandra 2.1.13 (moving to 2.2.6)
- Read-heavy workload (5:1)

Data model

# Data model

```
CREATE TABLE data (  
    domain text,  
    title text,  
    rev int,  
    tid timeuuid,  
    value blob,  
    PRIMARY KEY ((domain, title), rev, tid)  
) WITH CLUSTERING ORDER BY (rev DESC, tid DESC)
```

# Data model

en.wikipedia.org + Star\_Wars:\_The\_Force\_Awakens

717862573

717873822

...

1f2dd66c...7c7a913d3d8a

97466b12...7c7a913d3d8a

...

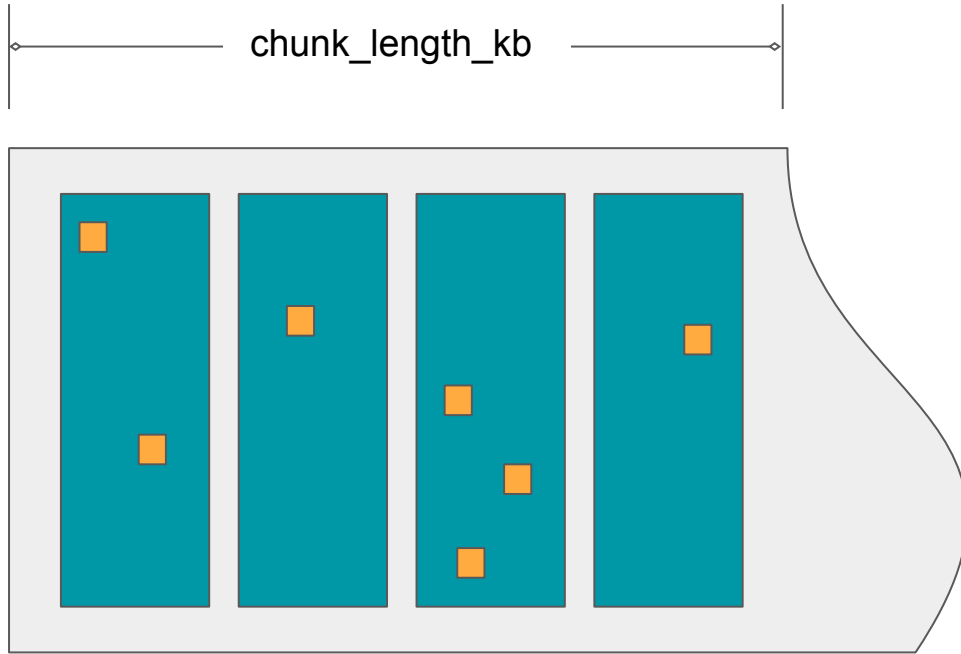
827e2ec2...7c7a913d3d8a

bdebc9a6...7c7a913d3d8a

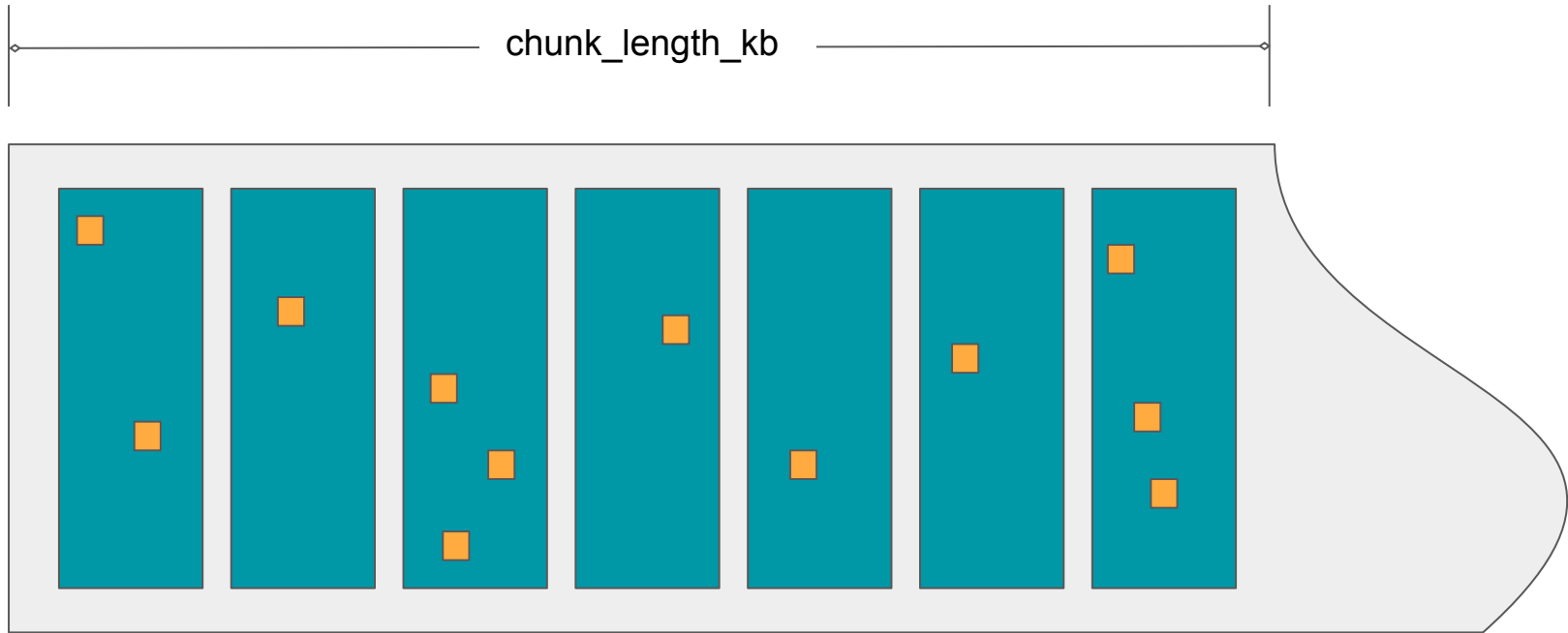
09877568...7c7a913d3d8a

# Compression

# Compression



# Compression





# Brotli compression

- Brought to you by the folks at Google; Successor to deflate
- Cassandra implementation (<https://github.com/eevans/cassandra-brotli>)
- Initial results very promising
- Better compression, lower cost (apples-apples)
- And, wider windows are possible (apples-oranges)
  - GC/memory permitting
  - Example: level=1, lgblock=4096, chunk\_length\_kb=4096, yields 1.73% compressed size!
  - <https://phabricator.wikimedia.org/T122028>
- Stay tuned!

# Compaction

# Compaction

- The cost of having log-structured storage
- Asynchronously (post-write) optimize data on disk for reads
- At a minimum, reorganize into fewer files
  - Dropping what is obsolete
  - Expiring TTLs
  - Removing deleted (aka tombstoned) data (after a fashion)
- Reorganize data so results are nearer each other

# Compaction strategies

- **Size-tiered**
  - Combines tables of similar size
  - Oblivious to column distribution; Works best for workloads with no overwrites/deletes
  - Minimal IO
- **Leveled**
  - Small, fixed size files in levels of exponentially increasing size
  - Files have non-overlapping ranges within a level
  - Very efficient reads, but also quite IO intensive
- **Date-tiered**
  - For append only, total ordered data
  - Avoids mixing old data with new
  - Cold data eventually ceases to be compacted

# Compaction strategies

- **Size-tiered**
  - Combines tables of similar size
  - Oblivious to column distribution; Works best for workloads with no overwrites/deletes
  - Minimal IO
- **Leveled**
  - Small, fixed size files in levels of exponentially increasing size
  - Files have non-overlapping ranges within a level
  - Very efficient reads, but also quite IO intensive
- **Date-tiered**
  - For append only, total ordered data
  - Avoids mixing old data with new
  - Cold data eventually ceases to be compacted



# DTCS: Well...no, actually

- Hard to reason about
- Optimizations easily defeated
- See: <https://phabricator.wikimedia.org/T126221>



# DTCS: So now what?

- Size-tiered compaction? Might as well.
- TimeWindowCompactionStrategy (<https://github.com/jeffjirsa/twcs>)? Maybe...
- Reduce node density?

# Garbage Collection



# G1GC

- Early adopters of G1 (aka “Garbage 1st”)
- Successor to Concurrent Mark-sweep (CMS)
- Incremental parallel compacting collector
- More predictable performance than CMS

# Humongous objects



- Anything  $\geq \frac{1}{2}$  region size is classified as *Humongous*
- Humongous objects are allocated into *Humongous Regions*
- Only one object for a region (wastes space, creates fragmentation)
- Until 1.8u40, humongous regions collected only during full collections (Bad)
- Since 1.8u40, end of the marking cycle, during the cleanup phase (Better)
- Treated as exceptions, so should be exceptional
  - For us, that means 8MB regions
- Enable GC logging and have a look!

Node density

*“Many smaller-sized Cassandra nodes is always better than fewer, dense ones.”*

— Everyone

# Motivation

- Compaction
- GC
- ...

# What we do

- Processes (yup)
- Puppetized configuration
  - `/etc/cassandra-a/`
  - `/etc/cassandra-b/`
  - systemd units
  - Etc
- Shared RAID-0



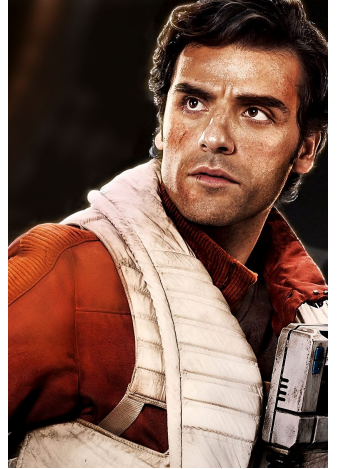
# What we should have done

- Virtualization
- Containers
- Blades
- Not processes



# Cassandra: The Good

- Fault-tolerance
- Availability
- Datacenter / rack awareness
- Visibility
- Ubiquity
- Nice, helpful people (tickets, IRC, etc)





# Cassandra: The Bad

- Usability
  - Compaction
  - Streaming
  - JMX
  - etc
- Vertical scaling
- JVM



# Cassandra: The Ugly

- Upgrading
- Release process



