

DataScience for Development and Social Change, 2015

Big Data

What's special about it, how to handle it...

“The Firehose”



Volume

Mobile Stats in East Africa



Source:

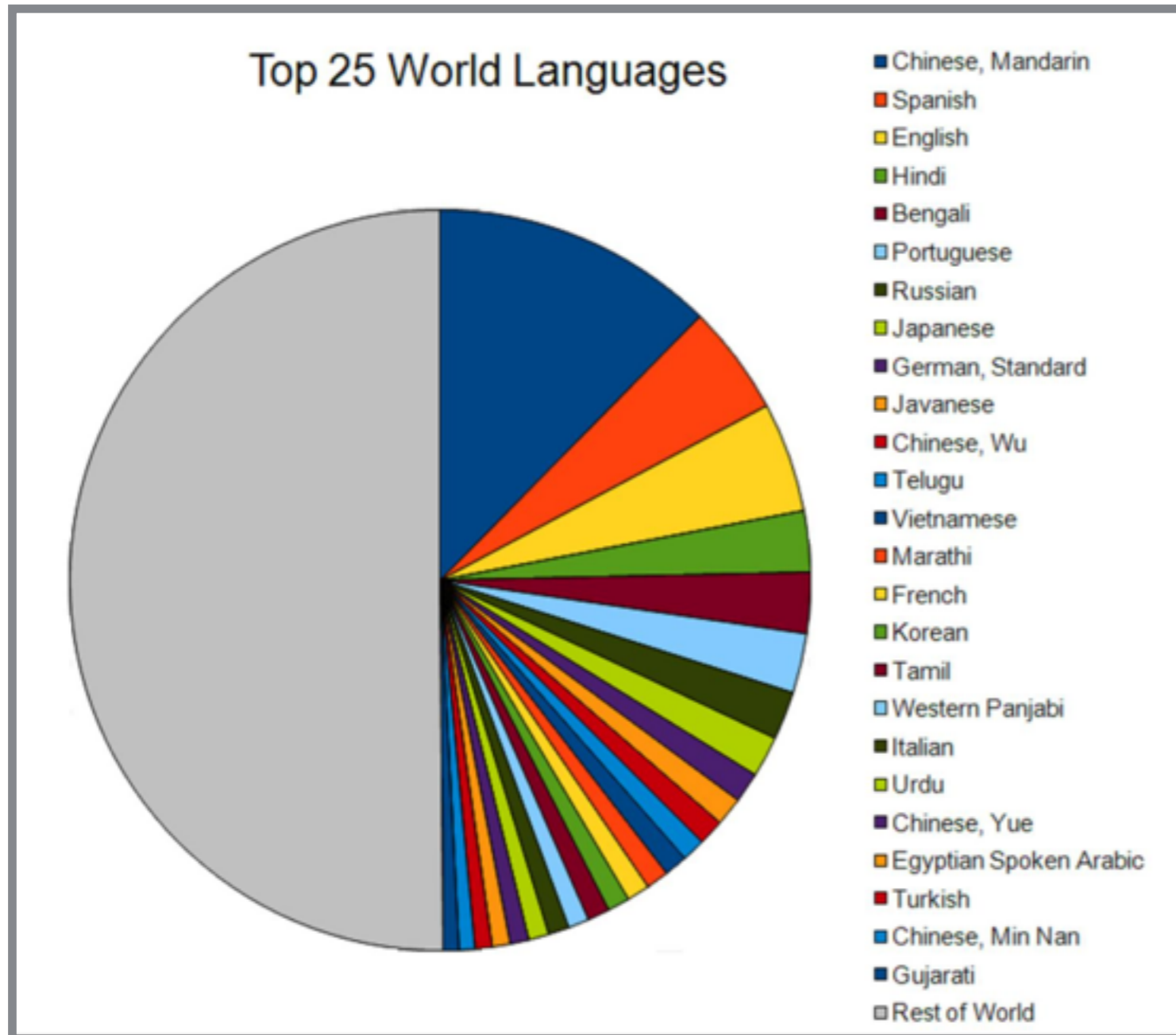
<http://www.itel.com/press-releases/2013/03/2013-03-18-ethiopia-52-percent-teledensity/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-kenya-75-8-percent-teledensity/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-uganda-57-3-percent-teledensity/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-tanzania-61-percent-teledensity/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-kenya-29-8-million-mobile-subscribers/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-uganda-6-0-million-mobile-subscribers/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-tanzania-27-6-million-mobile-subscribers/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-ethiopia-18-3-million-mobile-subscribers/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-kenya-221-287-fixed-lines/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-uganda-42-523-fixed-lines/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-tanzania-169-786-fixed-lines/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-ethiopia-464-849-fixed-lines/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-kenya-16-444-061-internet-users/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-uganda-935-125-internet-users/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-tanzania-2-995-000-internet-users/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-ethiopia-4-8-million-internet-users/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-kenya-7-261-862-384-voice-traffic-in-minutes/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-uganda-1-470-290-060-voice-traffic-in-minutes/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-tanzania-7-186-127-614-voice-traffic-in-minutes/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-ethiopia-520-million-outgoing-and-incoming-sms/>

<http://www.itel.com/press-releases/2013/03/2013-03-18-kenya-29-8-million-mobile-subscribers/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-uganda-6-0-million-mobile-subscribers/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-tanzania-27-6-million-mobile-subscribers/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-ethiopia-18-3-million-mobile-subscribers/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-kenya-221-287-fixed-lines/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-uganda-42-523-fixed-lines/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-tanzania-169-786-fixed-lines/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-ethiopia-464-849-fixed-lines/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-kenya-16-444-061-internet-users/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-uganda-935-125-internet-users/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-tanzania-2-995-000-internet-users/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-ethiopia-4-8-million-internet-users/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-kenya-7-261-862-384-voice-traffic-in-minutes/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-uganda-1-470-290-060-voice-traffic-in-minutes/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-tanzania-7-186-127-614-voice-traffic-in-minutes/>
<http://www.itel.com/press-releases/2013/03/2013-03-18-ethiopia-520-million-outgoing-and-incoming-sms/>

Velocity



Variety



CSV, json, xml,
excel, pdf,
text,
webpages, rss,
scanned pages,
images, videos,
audiofiles,
maps,
proprietary
formats etc.

Veracity



The “3 other Vs” are Viability, Validation, Verification.

Validation, checking that inputs are real, is a big deal for development data.

Map/Reduce

- ❖ Get Hadoop for VirtualBox: http://www.cloudera.com/content/cloudera/en/downloads/quickstart_vms/cdh-5-3-x.html
- ❖ Go play!