

Creating Data Pipelines in Python:

Build Frameworks not Pipelines

Deep Data Intelligence

Gatis Seja

What did John Bere have to do to sell his product?



Source:

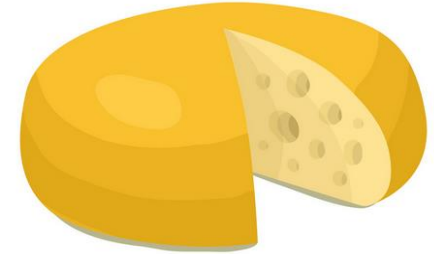
- Measurements
- Quality
- Vendors
- Price
- Language/Dialects
- Currencies
- Storage
- Laws & Religious Customs

Transport:

- Roads
- Security
- Packaging

Destination:

- Measurements
- Quality
- Price
- Language/Dialects
- Currencies
- Levies
- Storage
- Laws & Religious Customs





ANNO QUINTO

GEORGH IV. REGIS.

C A P. LXXIV.

An Act for ascertaining and establishing Uniformity of Weights and Measures. [17th June 1824.]

WHEREAS it is necessary for the Security of Commerce, and for the Good of the Community, that Weights and Measures should be just and uniform: And whereas notwithstanding it is provided by the Great Charter, that there shall be but one Measure and one Weight throughout the Realm, and by the Treaty of Union between England and Scotland, that the same Weights and Measures should be used throughout Great Britain as were then established in England, yet different Weights and Measures, some larger, and some less, are still in use in various Places throughout the United Kingdom of Great Britain and Ireland, and the true Measure of the present Standards is not verily known, which is the Cause of great Confusion and of manifest Frauds: For the Remedy and Prevention of these Evils for the future, and to the End that certain Standards of Weights and Measures should be established throughout the United Kingdom of Great Britain and Ireland; be it therefore enacted by the King's most Excellent Majesty, by and with the Advice and Consent of the Lords Spiritual and Temporal, and Commons, in this present Parliament assembled, and by the Authority of the same, That from and after the First Day of May One thousand eight hundred and twenty-five, the Straight Line or Distance between the Centres of the Two Points in the Gold Studs in the Straight Brass Rod, now in the Custody of the Clerk of the House of Commons, whereon the Words and Figures "Standard Length of Yard, 1760," are engraved, shall be and the same is hereby

7

The examination of medieval documents shows how unsuccessful were these efforts on the part of the central authority to establish uniform measures, when confronted by local customs so strong that many have survived to the present day; for there are still in use 25 local corn weights and measures, 12 different bushels, 13 different pounds, 10 different stones and 9 different tons⁶.

Magna Carta: (35) There shall be standard measures of wine, ale, and corn (the London quarter), throughout the kingdom. There shall also be a standard width of dyed cloth, russet, and haberject, namely two ells within the selvedges. Weights are to be standardised similarly.

959

A.D. 959-975. EDGAR. Ordinance. This is the ordinance that King Edgar, with the counsel of his witan, ordained, in praise of God, and in honour to himself, and for the behoof of all his people.

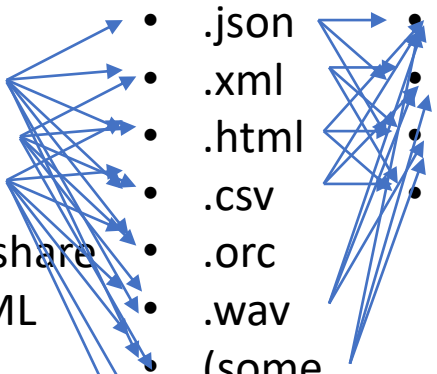
Cap. 8. And let one money pass throughout the king's dominion; and that let no man refuse; and let one measure and one weight pass, such as is observed at London and at Winchester.

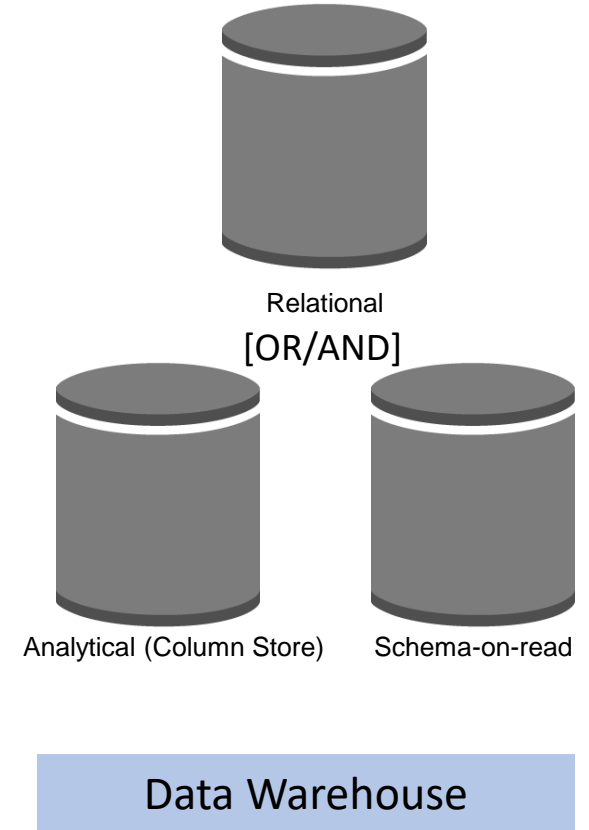
1215

1824

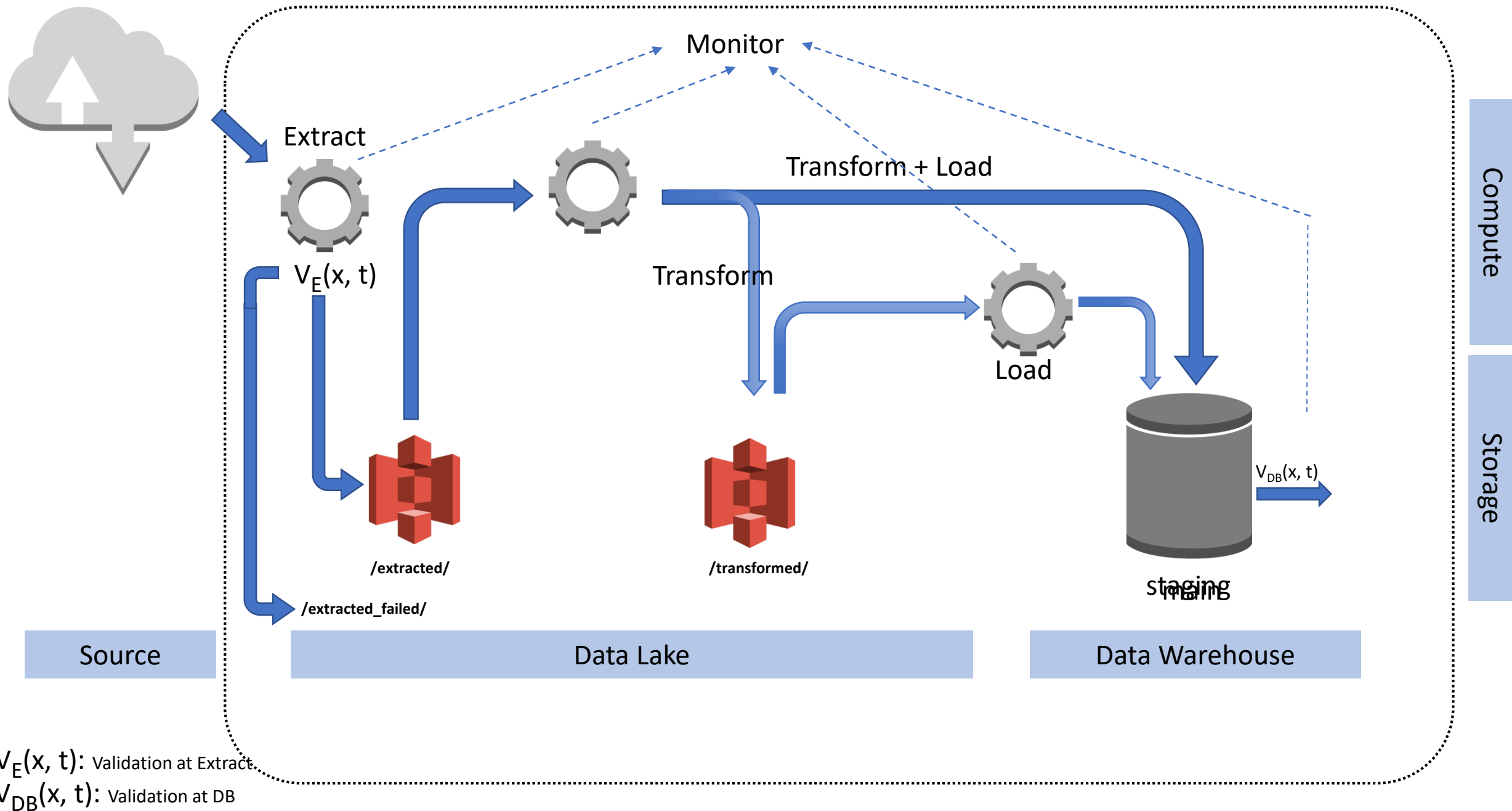
1924

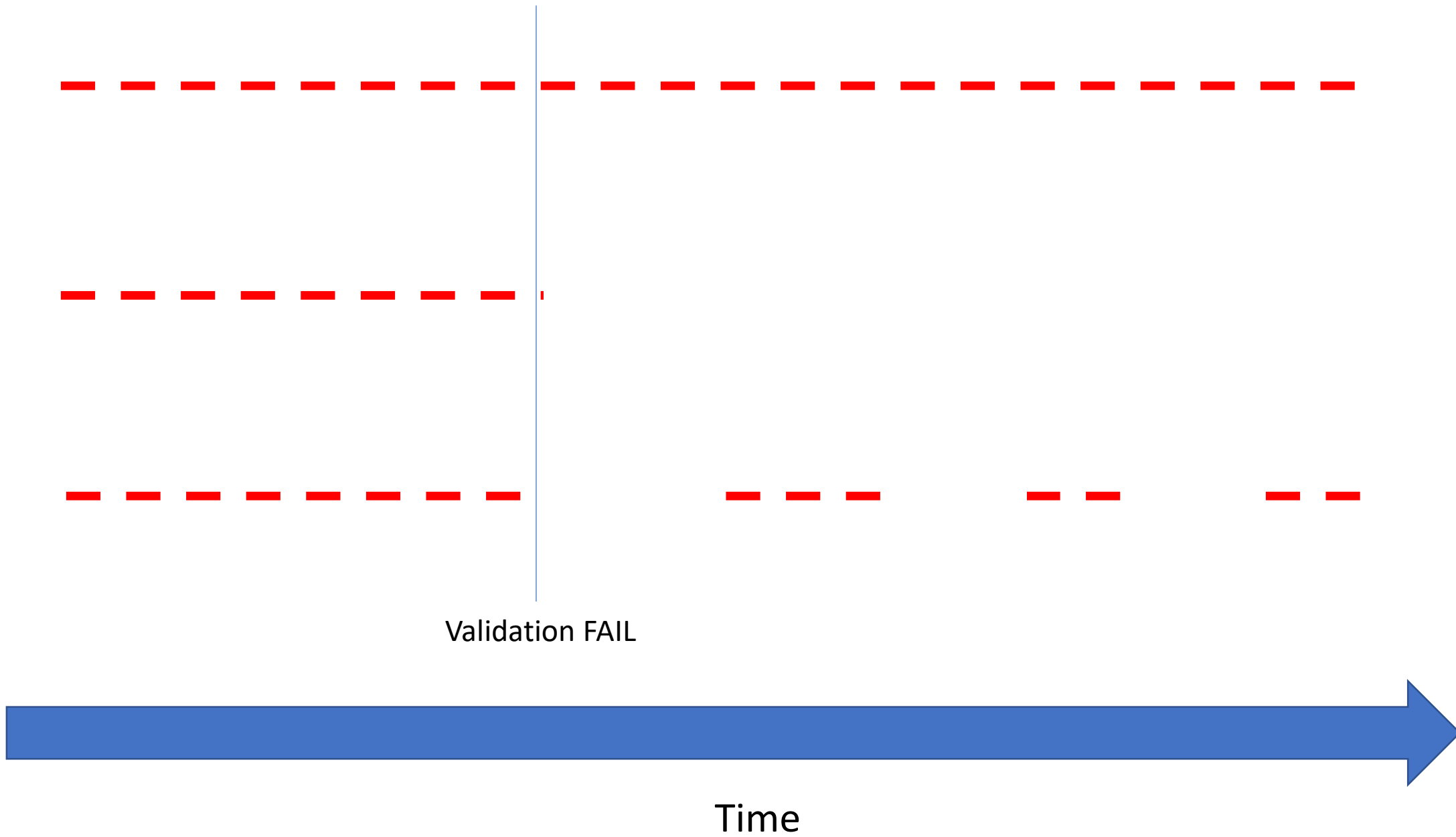
Sources: **Formats:** **Compression:**

- DB
 - FTP
 - API
 - S3
 - Fileshare
 - HTML
 - ...
- 
- The diagram shows a complex web of blue arrows connecting the sources to the formats and then to the compression methods. Each source is connected to multiple formats, and each format is connected to multiple compression methods.
- .json
 - .xml
 - .html
 - .csv
 - .orc
 - .wav
 - (some insane format that only the vendor created because they wanted to put that on their linkedin profile!!!)
 - ...
- .zip
 - .tar.gz
 - .gz
 - ...



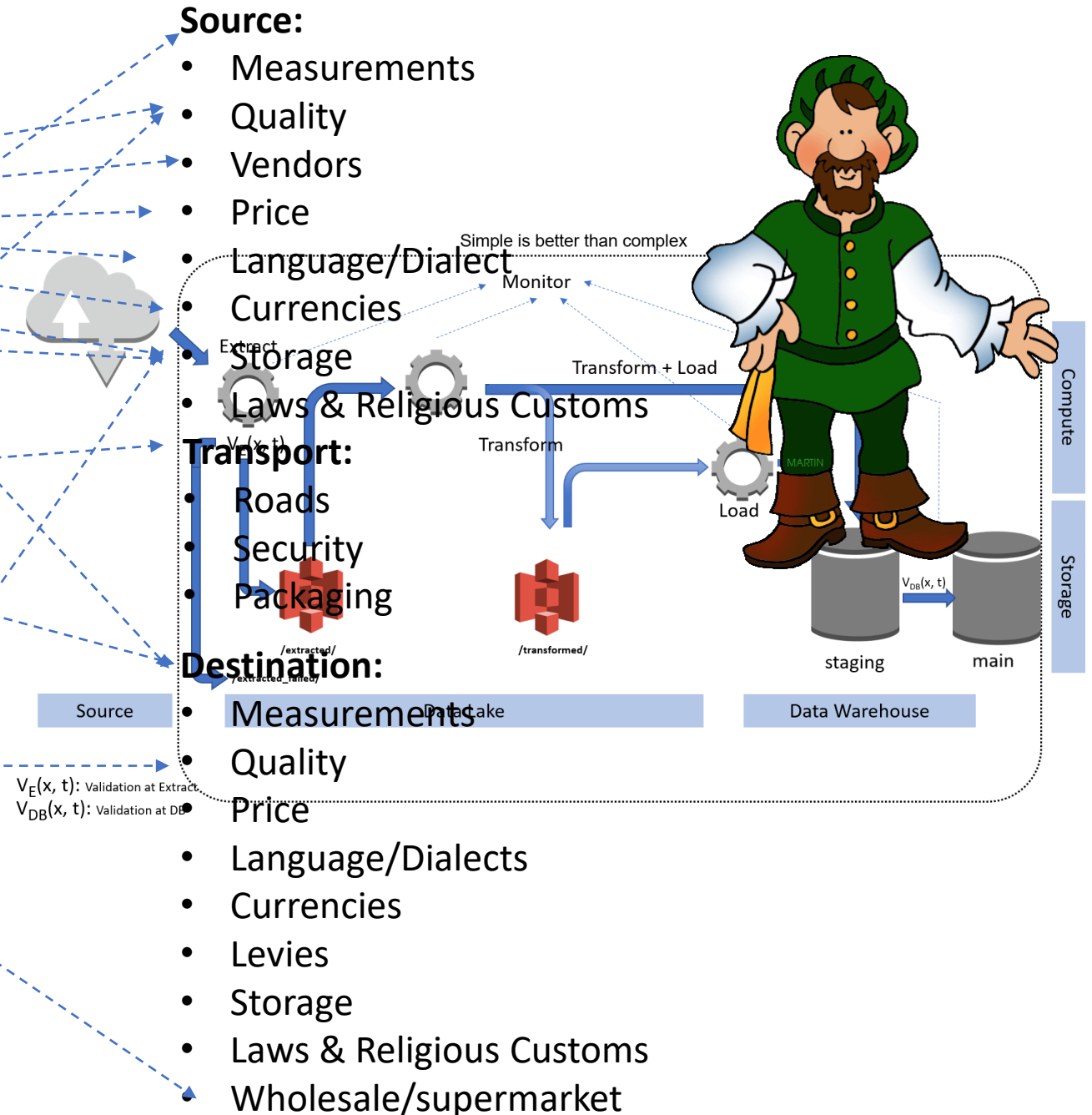
Simple is better than complex



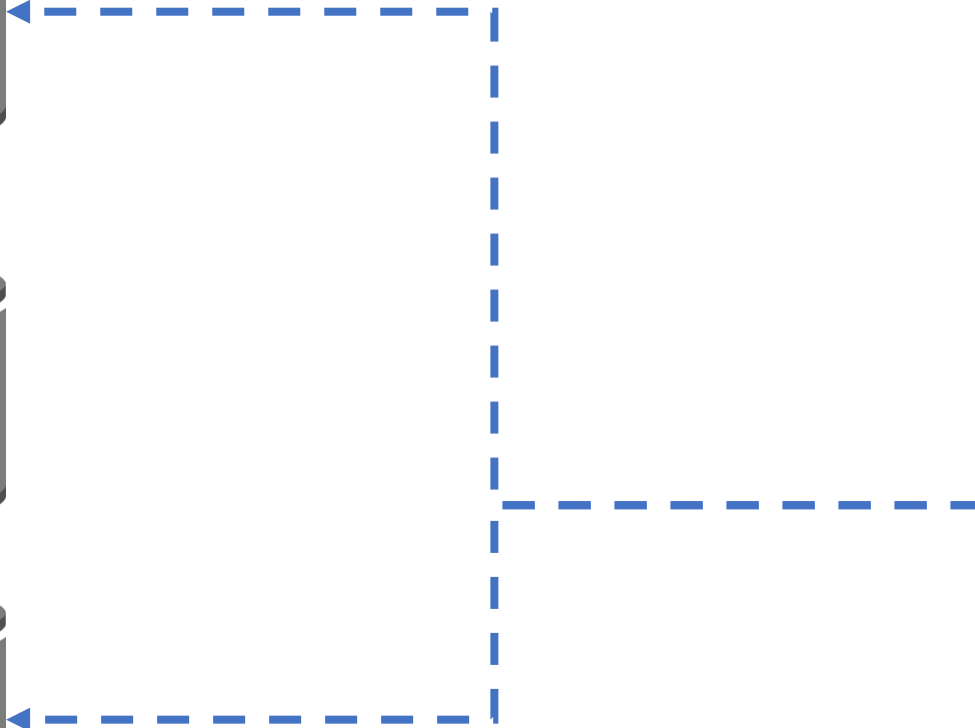
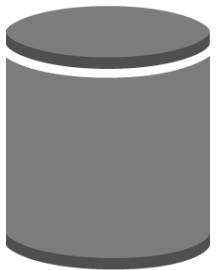
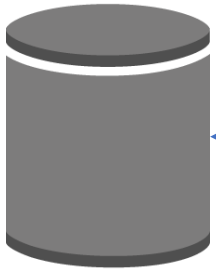
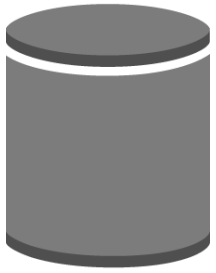
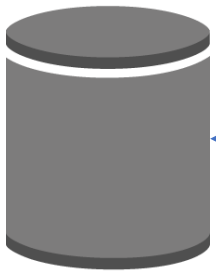


Principles

- Understand the data consumer
- Understand the data
- Keep data in its raw form
- Don't delete or move your raw data
- Validate your extracted data before saving
- Transform your data over all time: $T(x,t)$ not $T(x)$
- Separate your E-TL
- Minimize the number of data and compute nodes
- Store all your data if practically feasible to do so
- Make your ETL acyclical
- Validate your data before its given to consumers
- Join your data at the database level
- Monitor your data



Data as a Product



FTSE stocks|

ftse stocks
ftse stocks **to buy**
ftse stocks **and shares isa**
ftse stocks **2019**
ftse stocks **to watch**
ftse stocks **by dividend yield**
ftse stocks **with no stamp duty**
ftse stocks **to buy 2018**
ftse stocks **by market cap**
ftse stocks **at 52 week lows**

Deep Data Intelligence

www.ddintelligence.com

<https://www.linkedin.com/in/gatisseja/>

Primary Sources

Calendar of the Patent Rolls Preserved in the Public Record Office, Vol. III: Edward III, 1348–1350 (London, 1905)

Given-Wilson, C., Brand, P., Phillips, S., Ormrod, M., Martin, G., Curry, A. and Horrox, R. (eds.), *The Parliament Rolls of Medieval England* (Woodbridge, 2005)

Hall, H. and Nicholas, F. J. (eds.), ‘Select Tracts and Table Books Relating to English Weights and Measures (1100–1742)’, *Camden Third Series*, 41, pp. 1–68

Kitchin, G. W. (ed.), *Compotus Roll of the Obedientiaries of St. Swithun's Priory, Winchester* (London, 1892)

Online Primary Sources

<http://avalon.law.yale.edu/medieval/saxlaw.asp>

<https://www.bl.uk/magna-carta/articles/magna-carta-english-translation>

<http://www.overlandtrade.org/>

Secondary Sources

Britnell, R., ‘Commerce and markets’ in Crick, J. and Van Houts, E. (eds.), *A Social History of England, 900–1200* (Cambridge, 2011), pp. 179–87

Griffiths, D., ‘Towns and their hinterlands’ in *A Social History of England*, pp. 152–78

Prior, W. H., ‘Notes on the weights and measures of Medieval England’, *Bulletin du Cange* I (1924), pp. 77–97

West, C., ‘Urban populations and associations’ in *A Social History of England*, pp. 198–207

Zupko, R. E., *A Dictionary of Weights and Measures for the British Isles: the Middle Ages to the twentieth century* (Philadelphia, 1985)