



The importance of Open Source (in delivering large storage services)

Alberto Pace (alberto.pace@cern.ch)

CERN, Geneva, Switzerland



Why open source ?

- Many known reasons:
 - Control – in control of your business
 - Flexibility – license allows modify the code
 - Reliability / security – as development is shared, code is reviewed widely, fewer bugs
 - Cost – little entry cost, you pay the modifications
 - Longevity – you decide when drop it or when upgrade it

Why open source ?

- It has took many years to acknowledge the importance of software
 - Heard only few yeas ago: “Software is so flexible that can be done in the last minute ... by a student”
- Using proprietary software is outsourcing
 - You outsource something that is not important



When is it effective to outsource ?

- 3 requirements (apply to all fields, not just computing):
 - The activity is not strategic / core business
 - The activity has clear established standard (interfaces / protocols)
 - There are multiple independent vendors implementing these standards.
- If any of these 3 requirements is not satisfied, you are exposed to problems
 - Lock-in, business or service failure, blackmailing

Proprietary software, when you are 'big'

- Marketing arguments can be seriously misleading
- Traditional “procurement” rules may push you into disastrous strategy:
 1. Make a tender for the best commercial software available, and select the cheapest.
 2. Obtain an outstanding discount, which includes unlimited usage for xx years ...
 - Easily done when you are a 'big' customer. You can even get everything for free.
 3. Deploy rapidly, grow rapidly, ... for xx years, (with xx finite integer $< \infty$).
 4. Ensure that you are no longer in charge when you need to renegotiate the contract ...
- Does it make sense ?
 - Only if you have implemented a clear and tested exit strategy from the beginning

Today, things have changed for storage

- Software in storage
 - In many scenario software can be the most strategic component
 - When you are 'big', software should have a fixed-cost only
- Storage Hardware
 - If the "software" problem is correctly handled, the Hardware + Energy is where variable-costs are concentrated – scale out is possible at minimal 'marginal' cost.
- With this approach ...
 - the cost of adding a PB of storage is limited to the cost of a PB of HW
 - the cost of operating an additional PB of storage is limited to the cost of the required energy and hardware amortisation
 - If you do not have the critical mass cooperation is the solution

History repeats itself !

- Same story already heard:
 - 1990's Software: "is so flexible that can be done in the last minute ..."
 - 2010's Data: "Why care about data? Just put It in the cloud (for free)"
- This introduces the need for data sovereignty
 - Vendor lock-in has a huge impact: access to your data is at stake
 - Companies fails
 - Contract fails
 - Law changes
 - Subject to remote jurisdictions
 - Loss of your own data, not just access to software
 - Cannot be bought !

Conclusions

- The more critical a component is to your business, the more marketing pressure you will receive to outsource it (i.e. use proprietary sw)
- Outsource standard activities, well defined and interoperable
 - Do not outsource your own business
- Insource what is specific to you, and your critical activities
- The open source approach is the best practice to insource your critical activities at a minimum cost
 - This will guarantee fixed cost for software.
 - No license cost proportional to data volumes, or number of nodes, or cores, or disk, or data transferred.
 - This is why CERN has a Storage group

