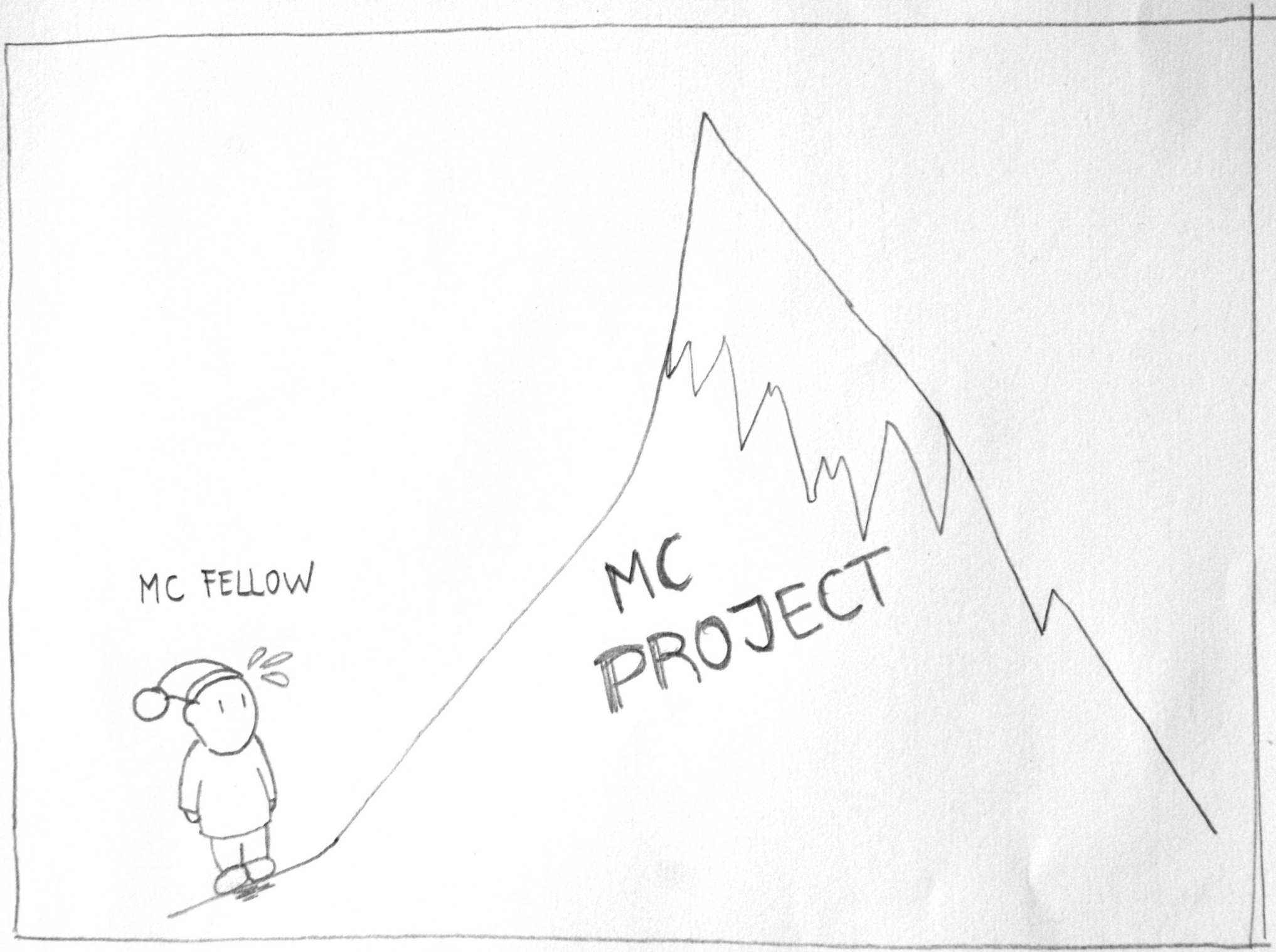
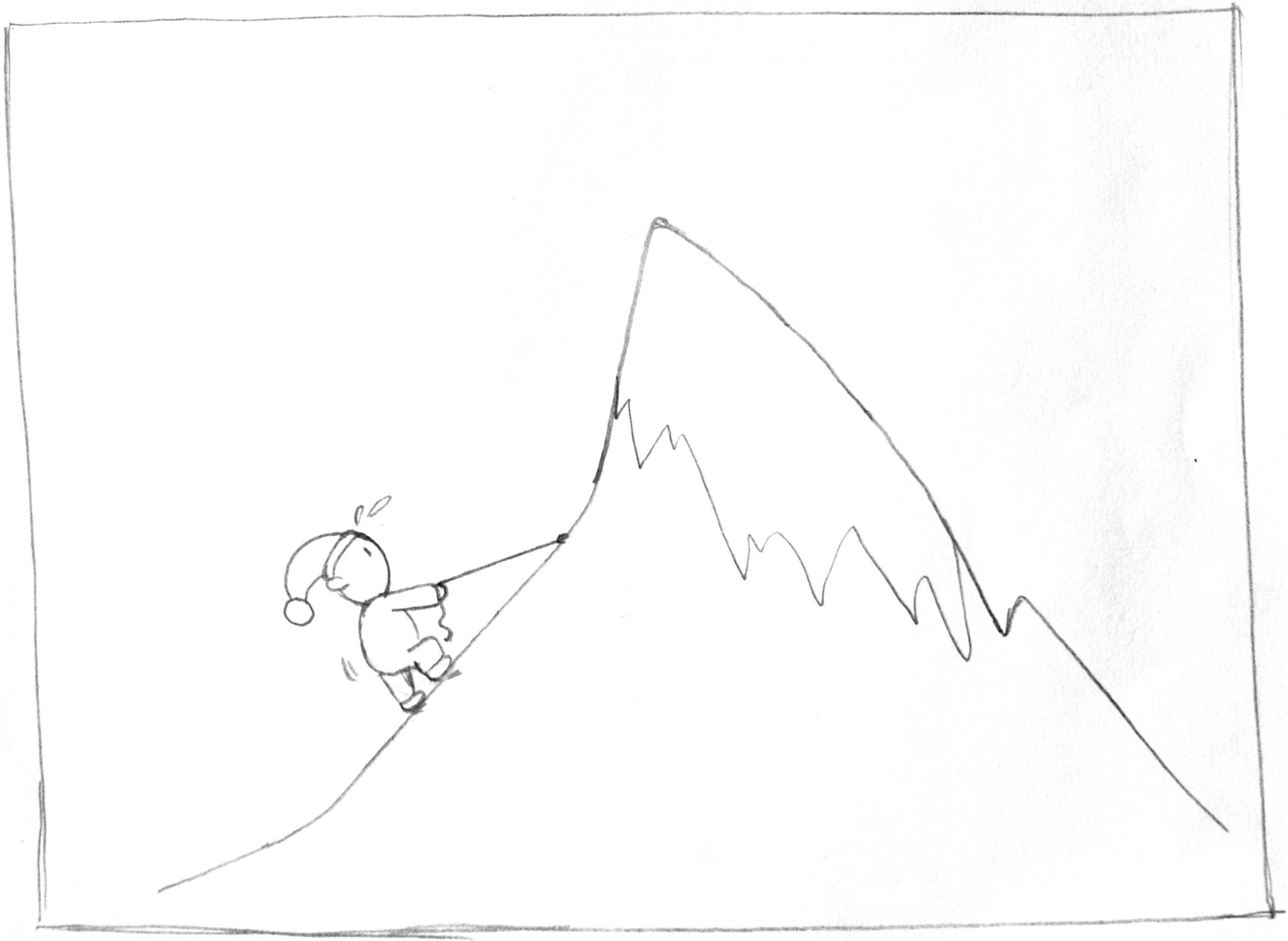
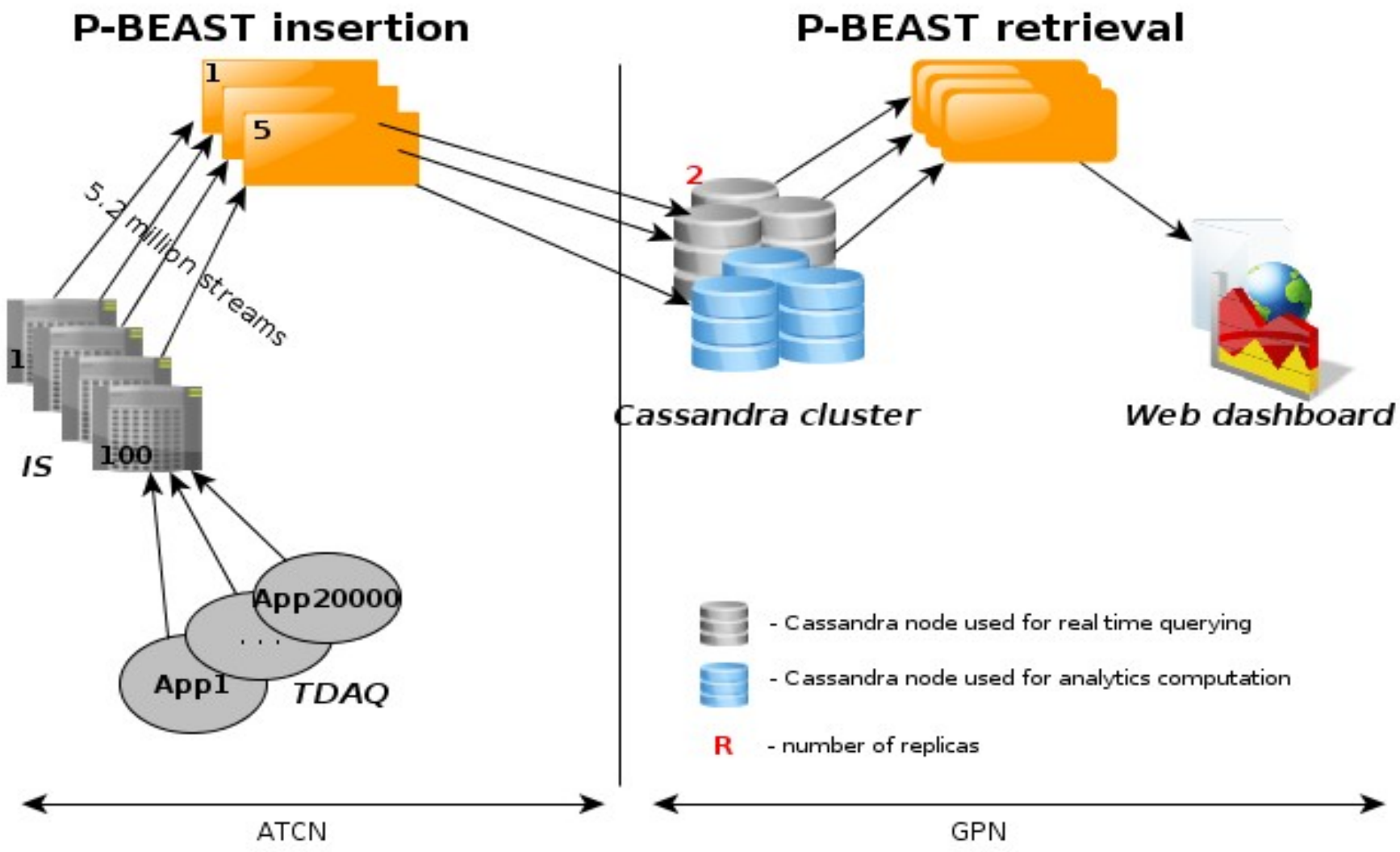

MC FELLOW
MC
PROJECT

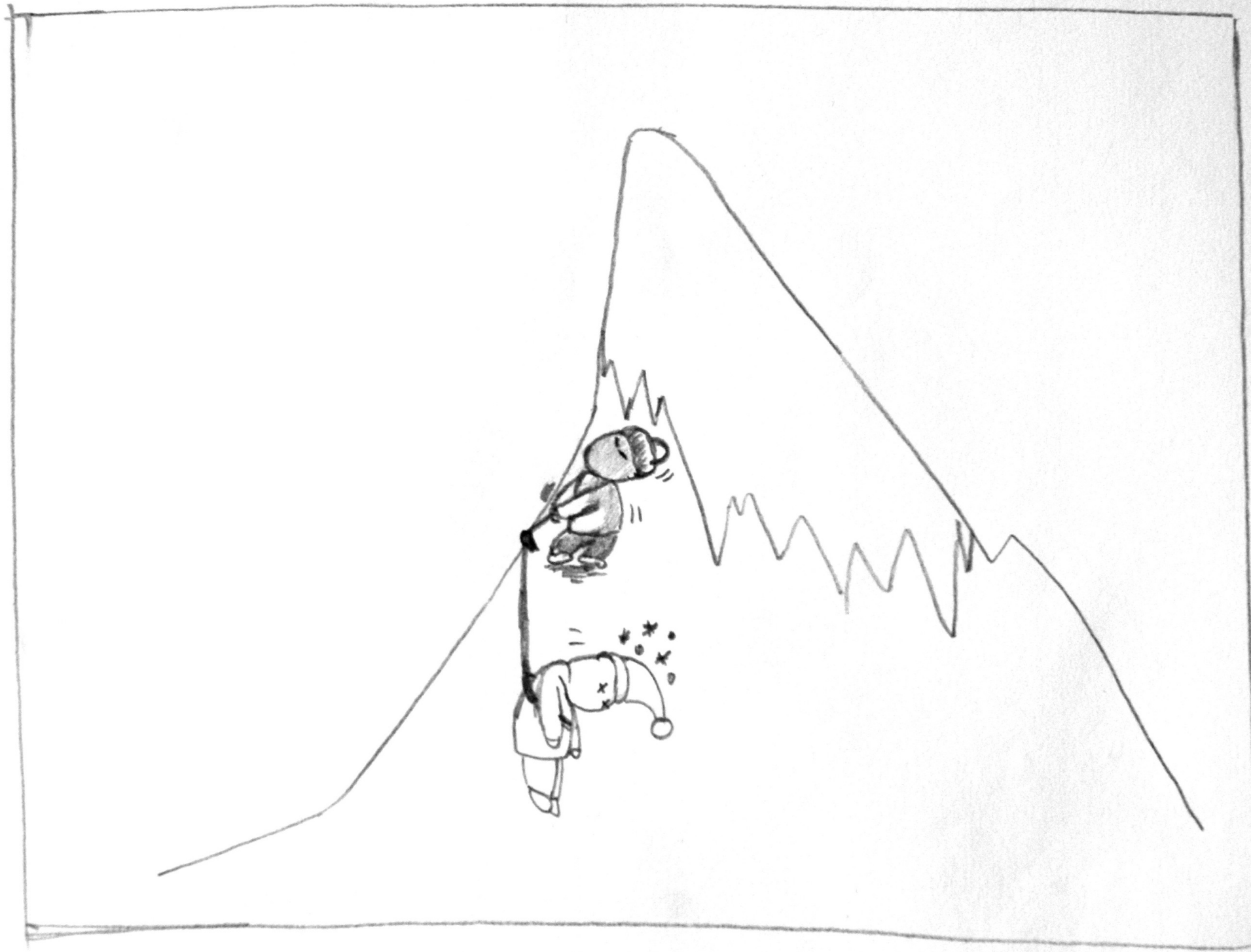
5 YEAR CAREER PLAN

Reach the level of technical, managerial and entrepreneurial experience that will confer the ability to form and manage teams of software engineers working on successful research or industrial projects within small or medium sized companies, particularly related to scalable back-end applications.



P-BEAST


P-BEAST insertion
P-BEAST retrieval
1
5
2
5.2 million streams
1
IS
100
App20000
. . .
App1
TDAQ
Cassandra cluster
Web dashboard
- Cassandra node used for real time querying
- Cassandra node used for analytics computation
R - number of replicas
ATCN
GPN

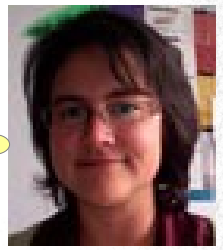


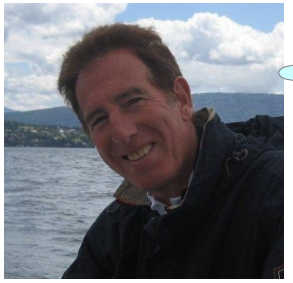
It's a long way from theory to practice

Design is the most important step of your work

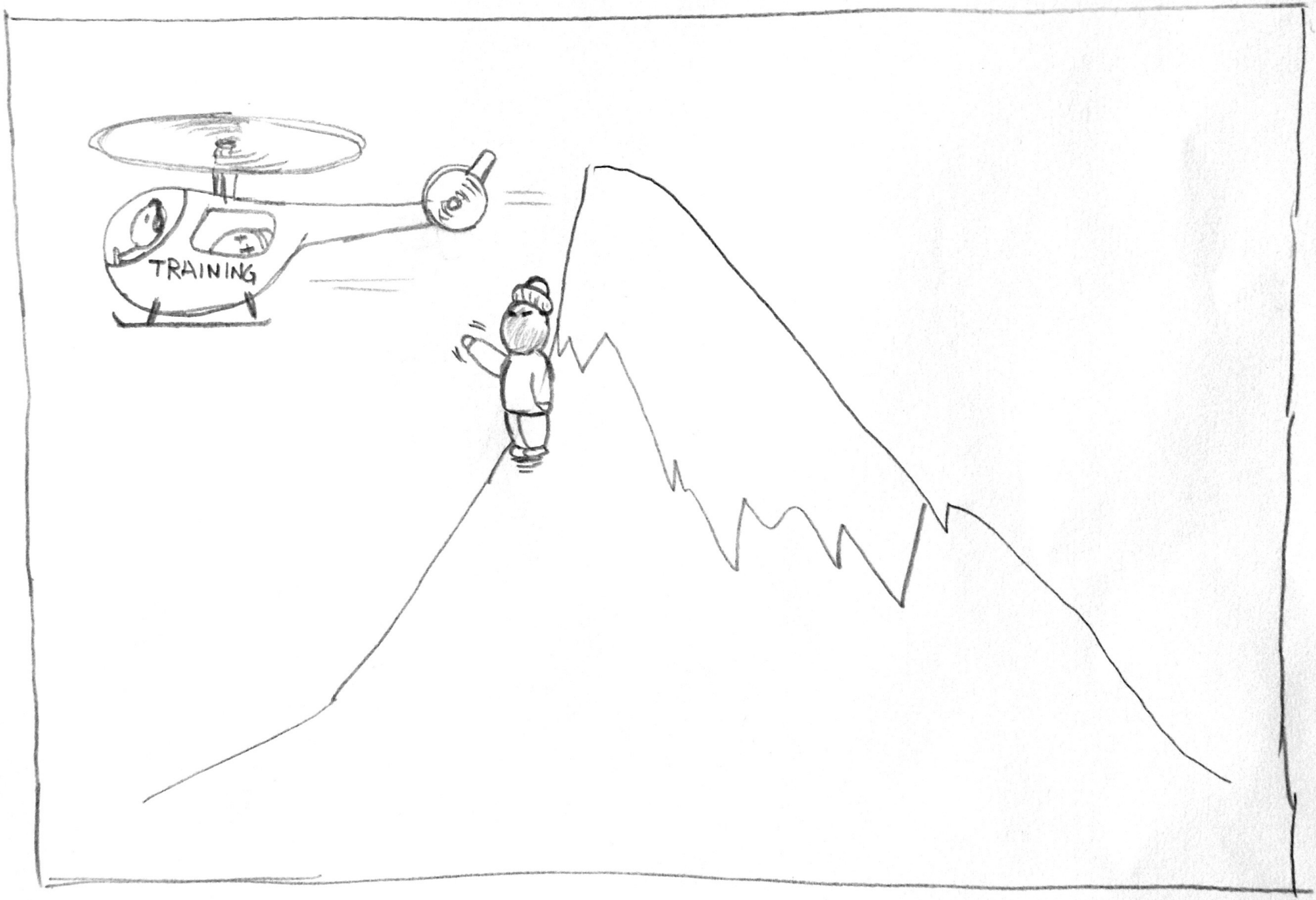
The idea is to solve it using a new technology

Have you considered Yoga?





Nothing is impossible!!!


TRAINING

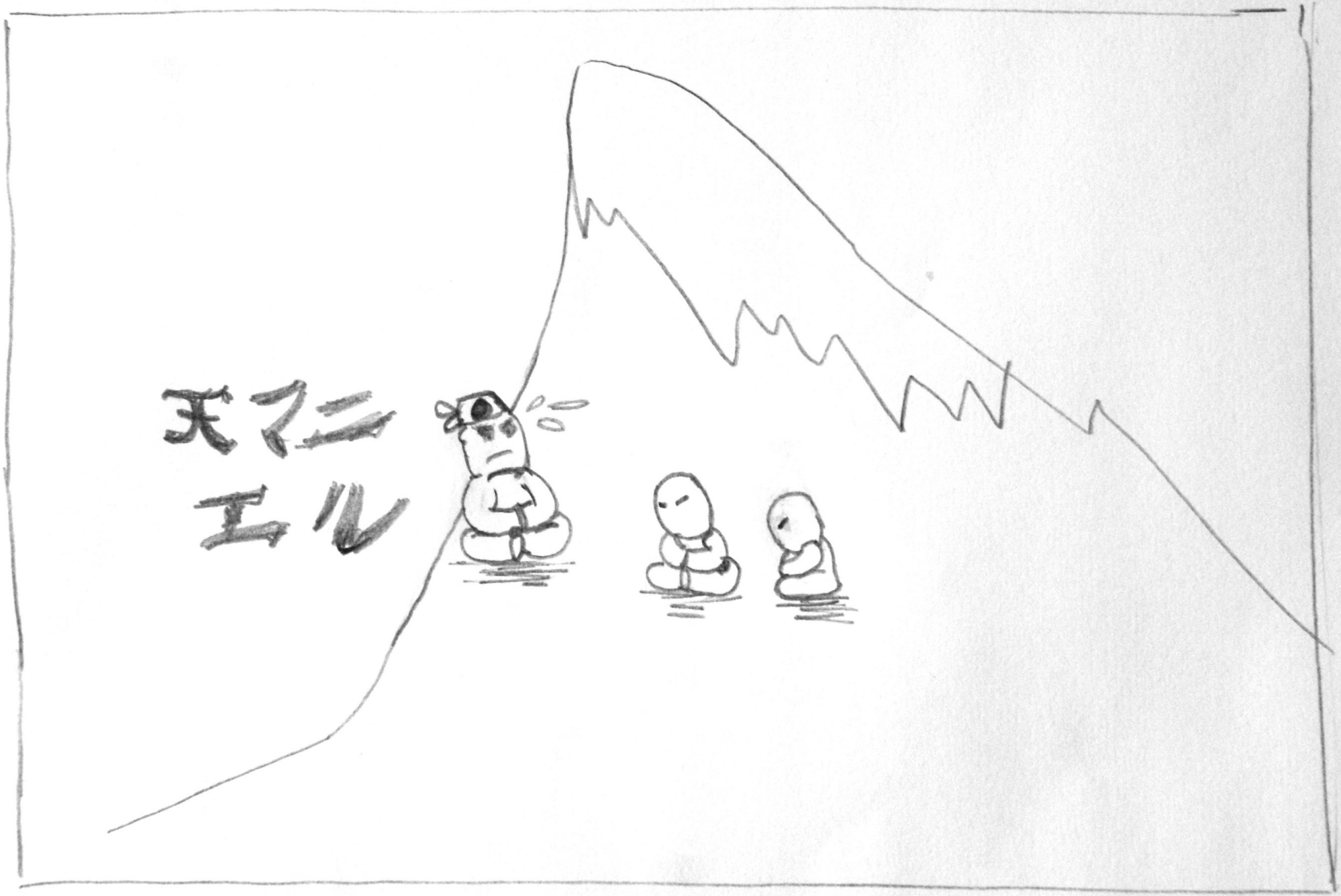
+ Cassandra EU Workshop (28 March 2012, ACUNU, London, UK)

+ Java Spring (14-17 May 2012, Springsource, New York City, USA).

+ Making Presentations (14, 15 December 2011 + 10 January 2012, Dodeca & CERN, Switzerland)

+ Entrepreneurship for researchers (3-4 November 2011, iOpener & CERN, Switzerland)

- Agile Project Management with Scrum (20-21 September 2012, CERN, Switzerland)


モニエル

Un jeu écologique remporte le Startup Weekend

My sollars.com a gagné la deuxième édition d'un concours entre jeunes entreprises

Compenser son empreinte carbone en jouant sur Facebook. C'est ce projet totalement dans l'air du temps qui a été récompensé lors de la seconde édition de Startup Weekend, un événement qui permet de créer une entreprise en 54 heures.

My sollars.com endosse le rôle d'une banque à CO_2. L'utilisateur, en jouant, calcule son empreinte écologique. Il la compense en versant de l'argent au moyen de cette banque à des entreprises ou des ons écologiques. Il peut t avec sa carte de crédit.

par des sociétés partenaires qui veulent améliorer leur responsabilité sociale.

Le fondateur, Miguel Molina, ingénieur de formation, a pu, durant ce week-end de dur labeur, rencontrer des spécialistes du marketing qui travailleront avec lui à ce projet. My sollars.com a remporté un accompagnement offert par la Fondetec, l'organisme de soutien aux entreprises de la ville de Genève, un pack juridique ainsi que du soutien pour une stratégie d'e-marketing.

Lors de cette seconde édition, 120 participants ont développé 14 projets qui ont été encadrés par 15 coaches. Pour Alexis Moeckli, l'organisateur de l'événement, la prochaine édition, en novembre 2012, gardera la même taille.

L'équipe de My Sollars.com, lauréate des Startup Weekend à la Haute paysage, d'ingénierie et d'architecture de Genève. PIERRE ALBOUY

A Persistent Back-End for the ATLAS Online Information Service (P-BEAST)

ATLAS EXPERIMENT http://atlas.ch

Author:
SICOE, Alexandru Dan
(Marie Curie ACEOLE Fellow, CERN, *asicoe@cern.ch*, tel: +41 2276 71113)

Contributors:
LEHMANN MIOTTO, Giovanni
KOLOS, Serguei (University of
MAGNONI, Luca (CERN)
SOLOVIEV, Igor (University of

1.Introduction

ATLAS is the largest of several detectors built along the Large Hadron Collider at CERN. Its aim is to measure particle production when protons collide at a very high center of mass energy, thus reproducing the behavior of matter a few instants after the Big Bang. The detecting techniques used for this purpose are very sophisticated and the amount of digitized data created by the sensing elements requires a very large trigger and data acquisition system (TDAQ[1]). This consists of approximately 30,000 applications running on 2000 interconnected computers.

There are several sub-systems responsible for facilitating information exchange between these applications and for monitoring their health. One of these is called the Information Service[1] (IS). It consists of a multitude of server applications running on dedicated machines. Any TDAQ application can be an IS client and can publish information objects of various types or it can subscribe to receive information objects from a specified source. The publishing rates vary widely and give a bursty nature to the traffic that IS is capable of generating.

During normal operation the rates have relatively steady levels. Peaks in the rates appear however when many applications publish data at the same time. This can happen when the state of the ATLAS infrastructure changes. For example, during a starting transition a lot of applications come alive and as soon as they do, they start publishing information about themselves.

2.What is P-BEAST?

P-BEAST wants to offer persistency to a large part of the information pu "Operational Information". IS already has a mechanism that buffers a ce but this is not sufficient for offline data analysis. What is needed is a sys data on disk such that it can be retrieved at any point by data flow exper help of specialized dashboards. Such functionality is useful for:

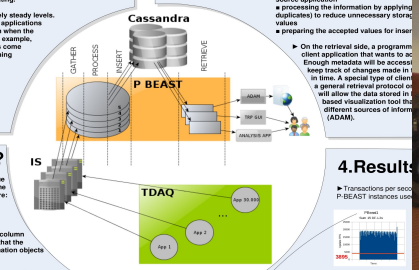
- understanding short/long term past behavior of different components
- comparing between physics data taking sessions of the detector
- investigating problems that occurred during a certain data taking sess

The project has thus two major parts which are reflected in its architectu

► The insertion path involves:

- gathering the required information by s callbacks whenever an information objec source application
- processing the information by applying duplicates) to reduce unnecessary storag values
- preparing the accepted values for inser

► On the retrieval side, a programm client application that wants to ac Enough metadata will be access keep track of changes made in in time. A special type of client a general retrieval protocol o will allow the data stored in based visualization tool tha different sources of inform (ADAM).



3.Why Cassandra?

The database technology of choice is a key-value distributed storage system called Cassandra. The main reasons for adoption of this technology are:

► built to sustain massive insertion data rates presented in an irregular fashion.

► within a top level logical partitioning of data (column family) Cassandra is schemaless which means that the stored data can follow the evolution of IS information objects over time in a seamless fashion

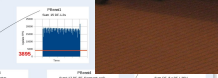
► easy to scale horizontally and configure a cluster to balance the load amongst its nodes

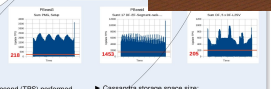
► data is arranged in rows of key-value pairs making it ideal to store time series data (timestamp as key).

► lots of sources of information: the Apache project homepage[2], the online community or the books[3][4] written about this technology

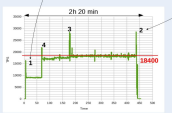
4.Results

► Transactions per seco
P-BEAST instances use





► Aggregate transactions per second (TPS) performed on the 3 Cassandra nodes in the cluster:



► Cassandra storage space size:

5.Conclusions

► the results are a good indication that P-BEAST can sustain the data rate generated by the ATLAS Online Information Service running within the TDAQ infrastructure

►measurements of the update rates confirm the varied behavior of different classes of IS servers with respect to the information rates they provide

►intermediate buffering in the P-BEAST gathering instances as well as Cassandra's insertion mechanism account for the spikes in the information rate

►the storage space required is significant due to the fact that the results shown were taken with only the mildest form of filtering applied to the incoming data (duplicates filtering). It is expected that further smoothing filters would further reduce the amount of stored data

►further work entails more testing for refining the insertion path and tuning filtering parameters, integration with TDAQ infrastructure and development of the retrieval mechanism

ACEOLE www.cern.ch/aceole

This research project has been supported by a Marie Curie Initial Training Network Fellowship of the European Community's 6 Seventh Framework Programme under contract number [FITN-GA-2008-211801-ACEOLE] and by the European Community's Seventh Framework Programme under the Grant Agreement no 212114 [ILHC-PP]

MARIE CURIE ACTIONS

References:

[1] ATLAS HLT/DAQ/DCS Group. (2003, Jul.). ATLAS high-level trigger, data acquisition and controls : Technical Design Report. CERN, SW [Online]. Available: http://cdsweb.cern.ch/record/616089

[2] The Apache Software Foundation. Apache Cassandra Home Page. [Online]. Viewed 2011 February. Available: http://cassandra.apache.org/

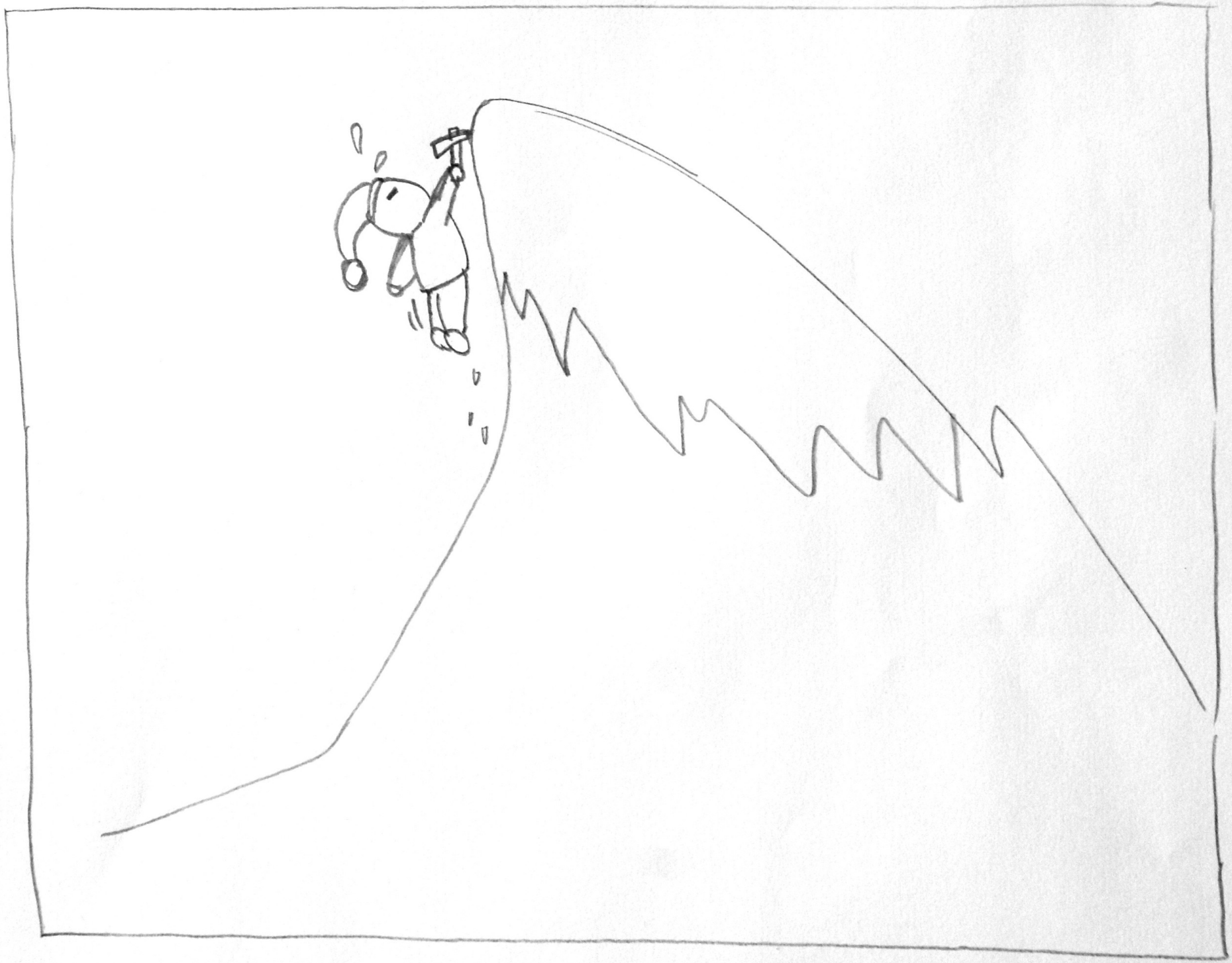
[3] E. Hewitt, Cassandra: The Definitive Guide, O'Reilly, ISBN: 978-1-449-39041-9

[4] E. Capriolo, Cassandra High Performance Cookbook, PACKT Publishing, ISBN: 978-1-849515-12-2

eldorajob

Job* module for CERN

1. Recruiters @CERN can define a job
2. Recruiters @CERN can involve Mentors to source valuable Candidates
3. Recruiters @CERN can see on the Map Candidates who match the best
4. Candidates are rated
5. Mentors are rated



ACHIEVEMENTS

A system to archive operational monitoring data from the ATLAS on-line software infrastructure using a modern distribute storage technology called Cassandra as a back-end.

CURRENT

Running in production archiving 1 GB/hour.

FUTURE

Long term data management.

Integrate with web front end.

Define an analytics framework.

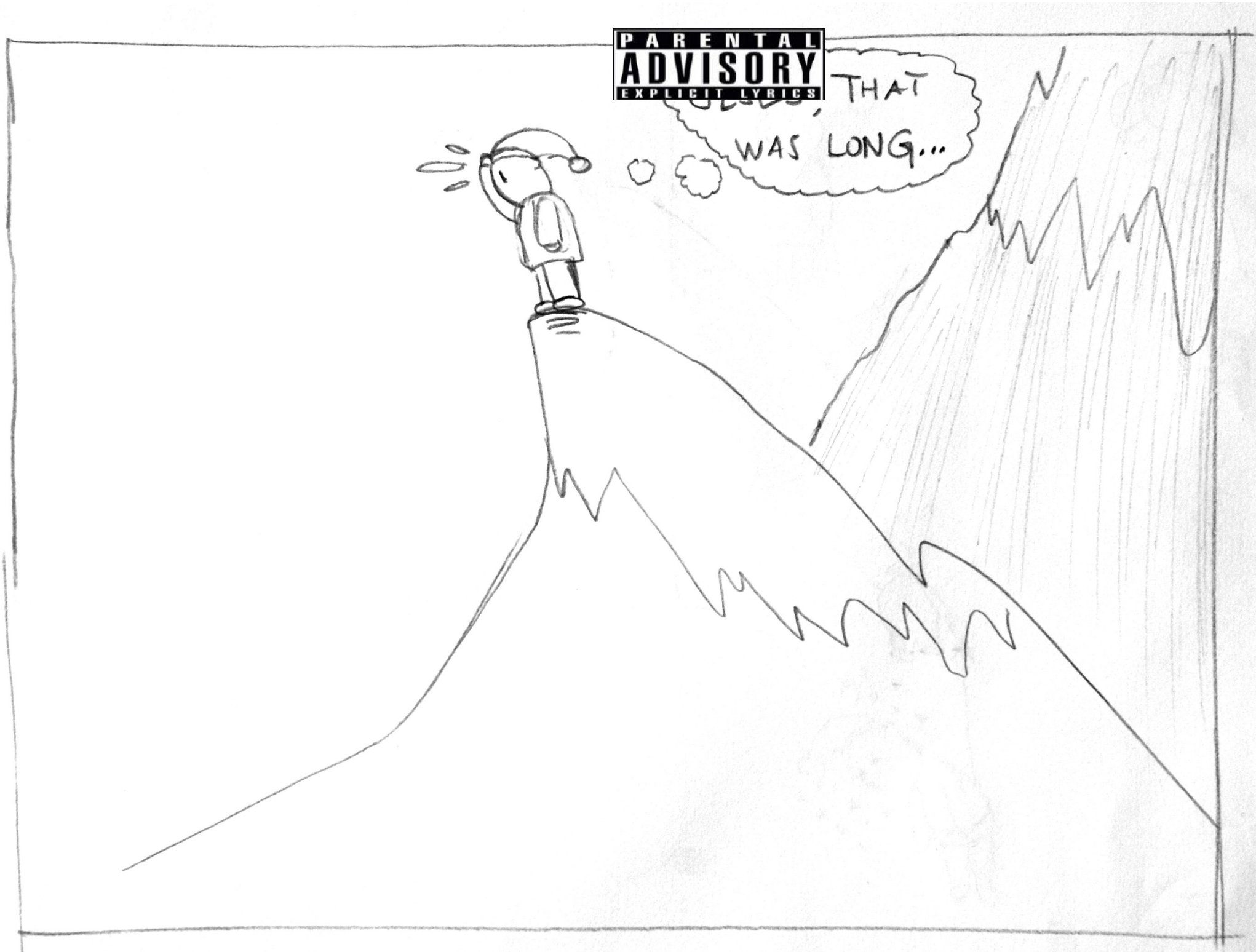
Job* module technical specification (with EldoraJob)

CURRENT

EldoraJob signed a contract to implement this with ATLAS.

FUTURE

Define a technical spec for another module called DreamTeam.


PARENTAL
ADVISORY
EXPLICIT LYRICS
THAT
WAS LONG...

WHAT I GOT FROM ACEOLE

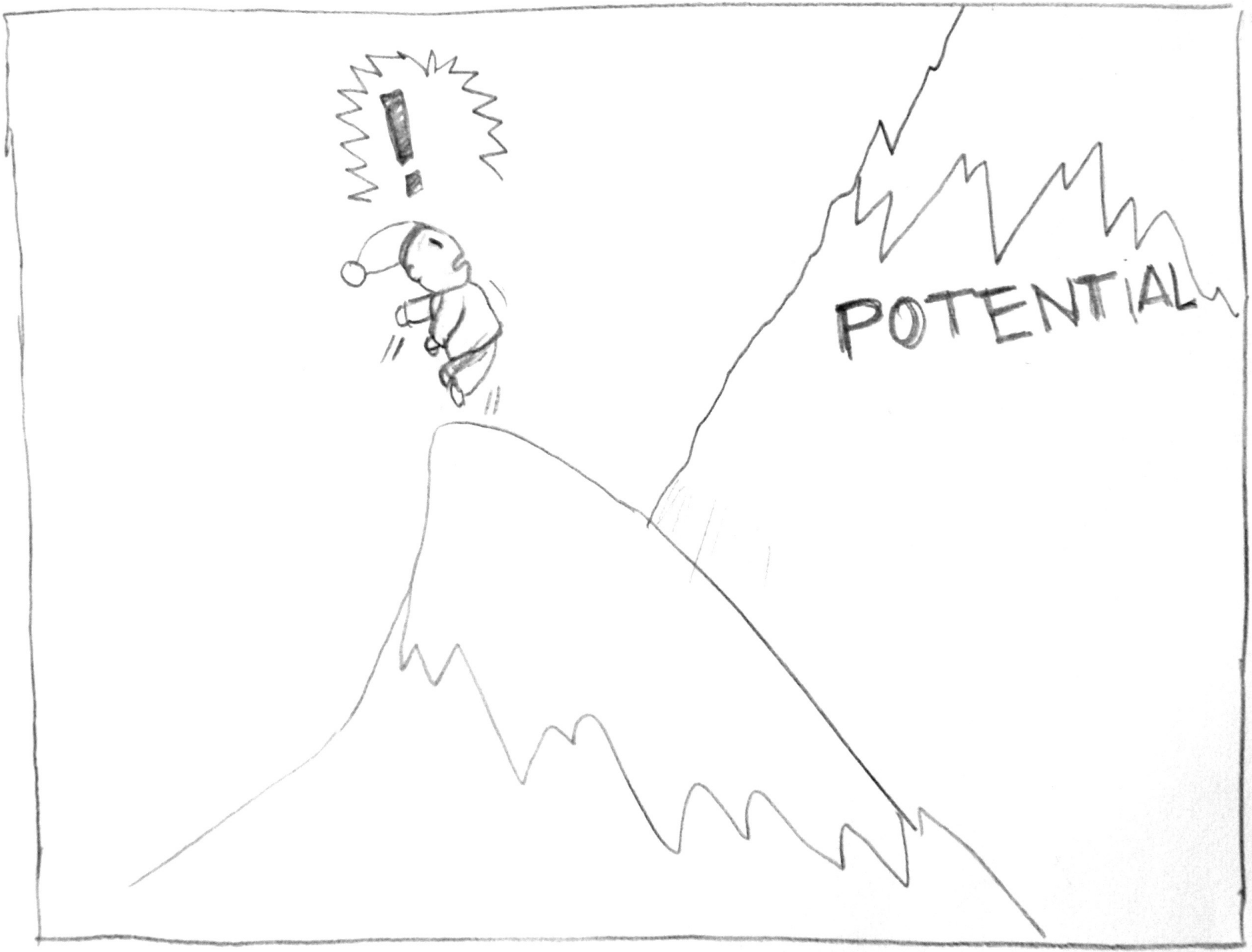
+ technical expertise/experience

+ self management

+ effective communication skills

+ presentation skills

+ self confidence


!
POTENTIAL

WHAT NEXT?

+ pursue a contract extension at CERN

+ continue collaborating with EldoraJob

+ join a small-medium sized company (start-up) operating in the big data arena