

Why I started my HPCC Systems

and ways to increase adoption.

By Fujio Turner

<http://www.meetup.com/hpcc-sv>

1995



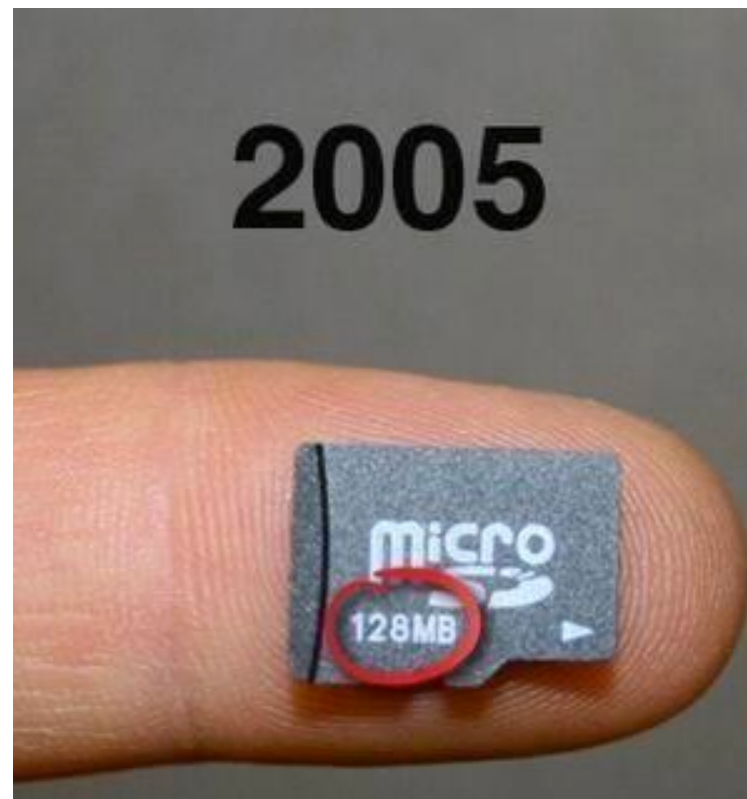
2014



1985



2005



2025





In school I would ask “Why is that?”



Data
Amount

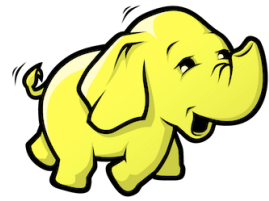


Query
Complexity

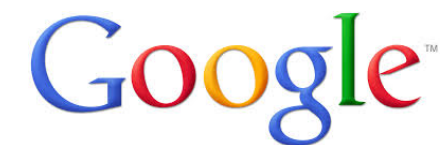
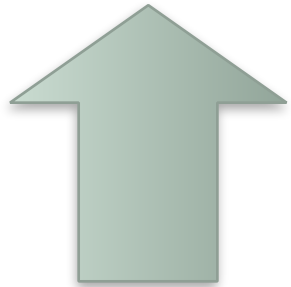


Query
Speed

3 Parts to Big Data



Hadoop = Airplane

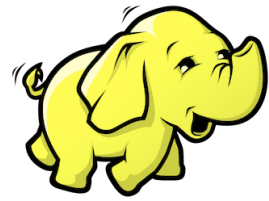


If Google made airplanes:

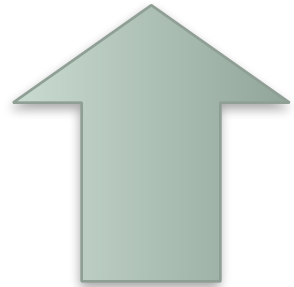
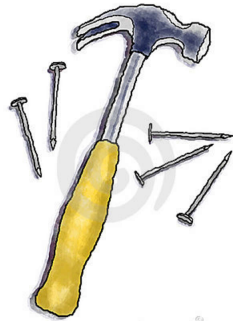
They released a white paper in 2004 on “how to fly”.

use:

- wings
- propeller

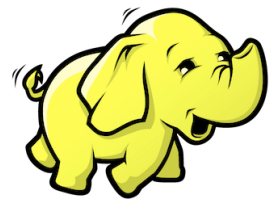


Hadoop = Airplane

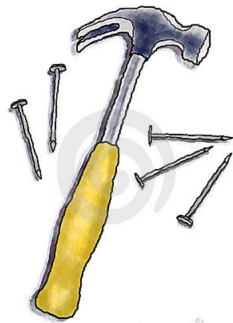


Home Made Airplane from White Paper:
Called "Hadoop" for free

Google™

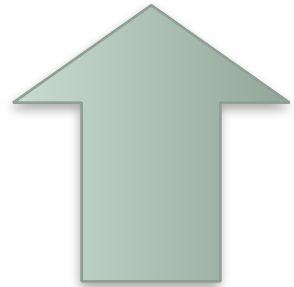


Hadoop = Airplane

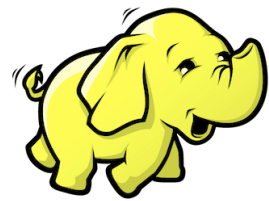


Business Popup to Sell Airplane:

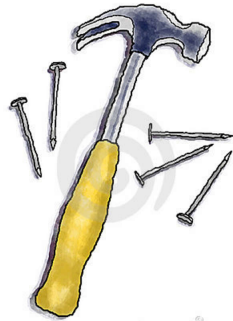
- Some airplanes are faster
- Some airplanes are easier to use
- Some airplanes are cheaper



Google™



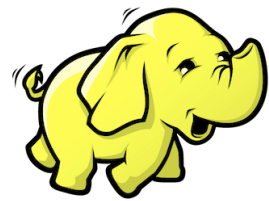
Hadoop = Airplane



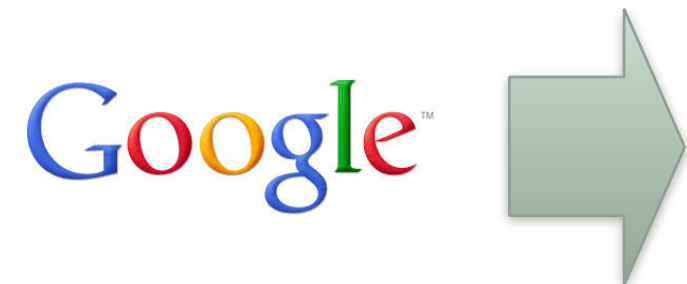
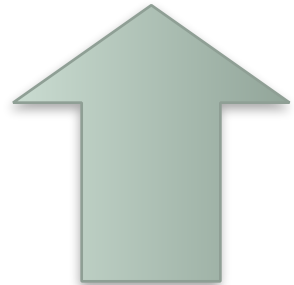
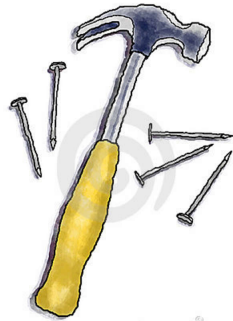
Why did Google release white paper?

Because all business have secrets.





Hadoop = Airplane



Answer:

They have a F-15 jet fighter and no longer use propeller airplanes.



"I do not need a jet fighter ,but can I have the jet engine?"





hadoop alternatives

About 348,000 results (0.14 seconds)

What Is Hadoop - Intel.com

Ad www.intel.com/BigData

Intel Delivers Technology Built to Scale Your Big Data Challenges.
Learn the fundamentals ... Big Data Planning Guide ... Big Data 101 Tutorial
Intel has 607,622 followers on Google+

[Big Data Planning Guide](#)

[Big Data Best Practices](#)

[Cloud-Based Big Data](#)

[Big Data White Paper](#)

What are the emerging alternatives to Hadoop? | ITworld

www.itworld.com/answers/.../what-are-emerging-alternatives-hadoop

Alternatives for Hadoop/MapReduce data storage and management ... The main advantages over other alternatives are the real-time delivery of data queries ...

What are some promising open-source alternatives to ...

www.quora.com/What-are-some-promising-open-source-alternativ... Quora

Answer 1 of 17: Hadoop consists of many subprojects: HDFS, MapReduce, Hive, Pig, HBase, and Avro. I believe this question refers to the MapReduce ...

Hadoop Alternative Hydra Re-Spawns as Open Source

www.datanami.com/.../hadoop_alternative_hydra_re-spawns_as_open_s...

Mar 12, 2014 - It may not have the name recognition or momentum of Hadoop. But Hydra, the distributed task processing system first developed six years ago ...

Hadoop Alternatives: When Your Data Isn't as Big as Yo...

java.dzone.com/articles/hadoop-alternatives-when-your

Sep 27, 2013 - This post from Chris Stuccio's blog takes a critical look at the use of Hadoop and Big Data as buzzwords by asking an interesting question: ...

Don't use Hadoop - your data isn't that big - Chris Stucchio

www.chrisstucchio.com/blog/2013/hadoop_hatred.html

Sep 16, 2013 - "So, how much experience do you have with Big Data and Hadoop? Hadoop will also nearly always be slower than the simpler alternatives.

Hadoop Fatigue — Alternatives to Hadoop « Byte Mining

www.bytemining.com/2011/08/hadoop-fatigue-alternatives-to-hadoop/

Aug 16, 2011 - Hadoop Fatigue — Alternatives to Hadoop. It's been a while since I have posted... in the midst of trying to plow through this dissertation while ...

Why HPCC is a superior alternative to Hadoop | HPCC ...

hpccsystems.com/Why-HPCC/HPCC-Vs-Hadoop HPCC

Read how the HPCC Systems platform is a superior alternative to Hadoop.

Ads

Google Hadoop Solution

cloud.google.com/Hadoop

Setup Hadoop And Powerful Big Data Analytics In Just A Few Minutes

Hadoop - Big Data eBook

www.pentaho.com/Big-Data

Free Guide to Real-Time Analytics
Learn the Future of Big Data Now

Hadoop & NoSQL

www.marklogic.com/

Get the best of both worlds.
MarkLogic: Try it free today!

Hadoop Apache

www.rackspace.com/CDN

Serve Your Media from 219 Akamai
CDN Edge locations from 10¢/GB

[See your ad here »](#)

7th



Open
Source ?
Market
Pricing



HPCC Systems

<http://hpccsystems.com>

Filters ▾

Find a repository...

HPCC-Platform

IMPORTANT - Issue tracking has been migrated to JIRA - use your github ID and password reset mechanism to log in

Updated 13 hours ago

C++



83



Size?
Future
Proof

[Any limitations on file size and number of files?](#)



Currently no single file can be greater than 4 exabytes (~~4~~ million terabytes). This is not a current concern as other limitations occur in real-world systems before this limitation is reached; such as total available disk space on your cluster. There are no theoretical limits to the number of files.



Talked To Dan Camper

who is currently using HPCC Systems in production.

Q: “How many jobs can you do in a day?”

Talked To Dan Camper

who is currently using HPCC Systems in production.

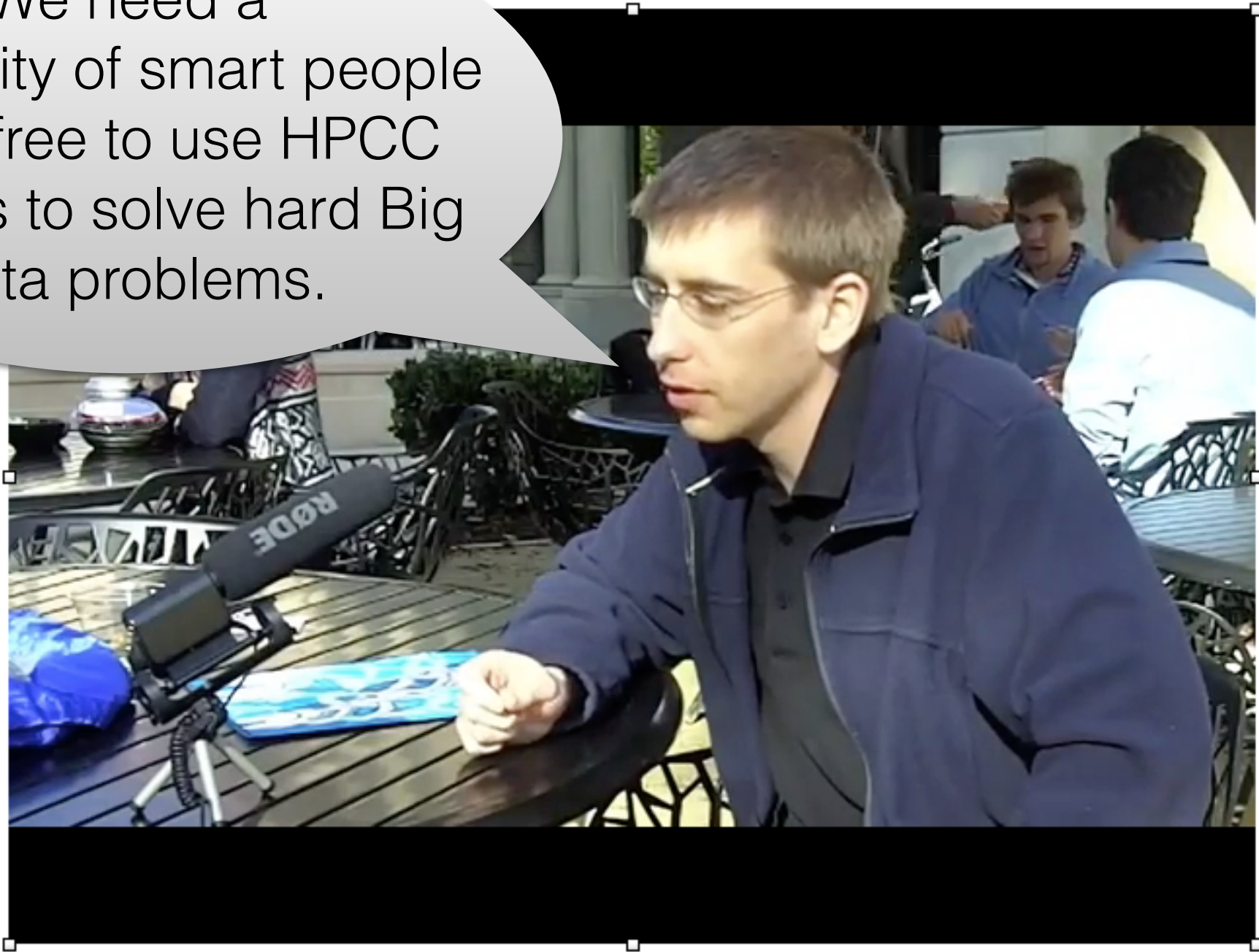
Q: “How many jobs can you do in a day?”

A: “Roxie can do 2,000 - 3,000 jobs/
second.”



Why did I start my MeetUp?

We need a community of smart people to feel free to use HPCC Systems to solve hard Big Data problems.



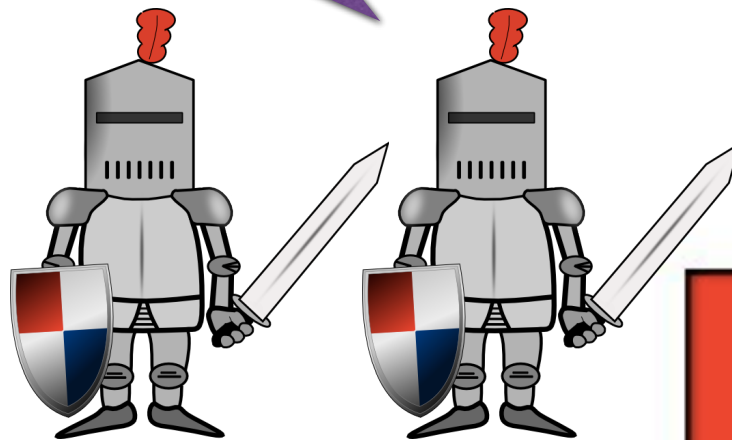
Charles Kaminski, Jr.

The three hurdles to adoption are to:

1. Understanding how business are organized
2. Understanding how business think
3. Understanding how business operate

Direction of the Company &
Makes Decisions

Carry Out Decisions &
Creates Stability



Do the Work

Senior and
Executive
Management

Middle Management

Frontline Management and Individual
Contributors

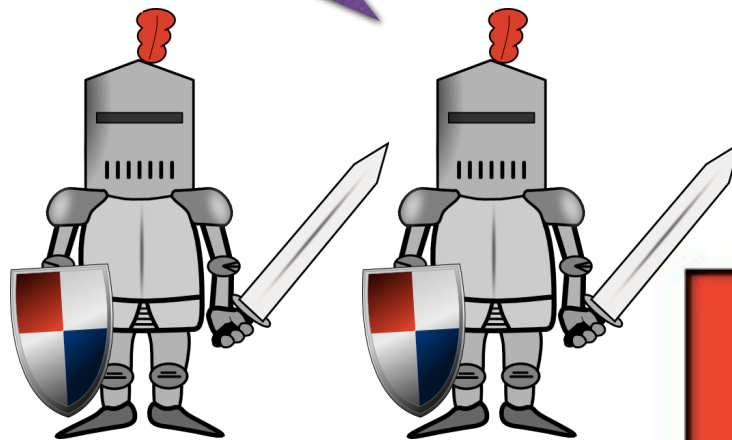


How business are organized

Direction of the Company &
Makes Decisions

Carry Out Decisions &
Creates Stability

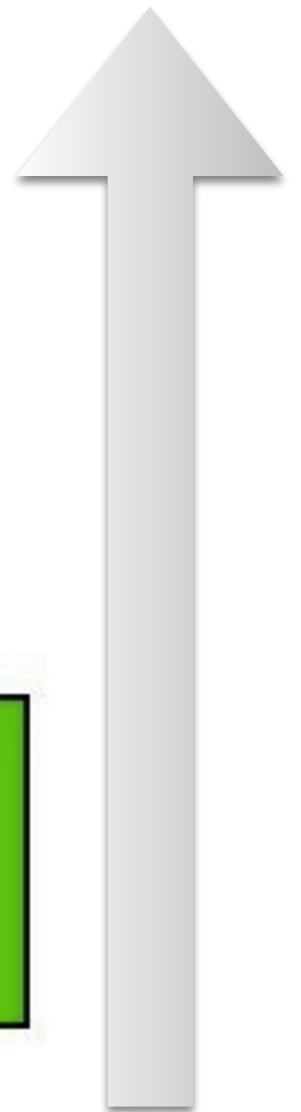
Do the Work



Senior and
Executive
Management

Middle Management

Frontline Management and Individual
Contributors

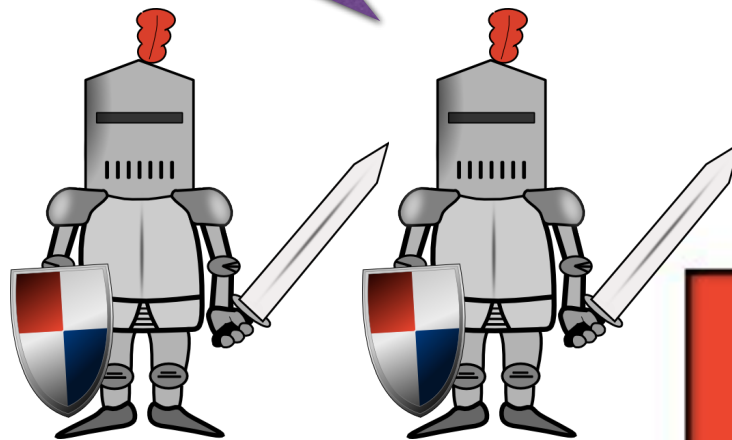


How business are organized

Less Technical
More Politics
Reduce Risk

Direction of the Company &
Makes Decisions

Carry Out Decisions &
Creates Stability



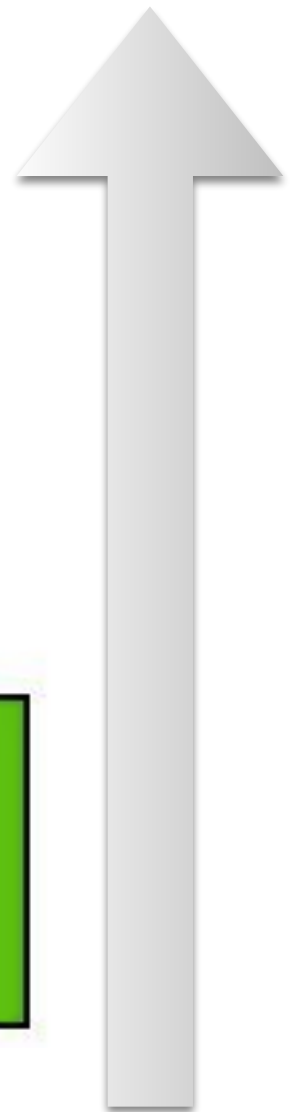
Do the Work



Senior and
Executive
Management

Middle Management

Frontline Management and Individual
Contributors

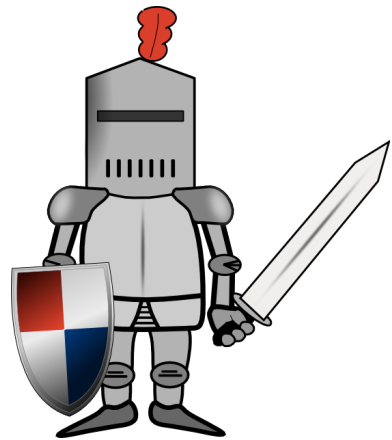


How business are organized

Most important to business ➔

Less Technical
More Politics
Reduce Risk

Why different parts of a business choose a Big Data technology

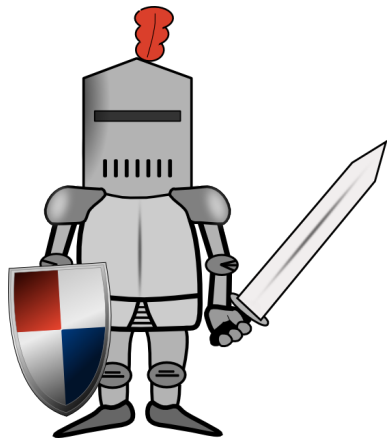


Why different parts of a business choose a Big Data technology

“I have the technology of Google.”



Google™ = 



Why different parts of a business choose a Big Data technology

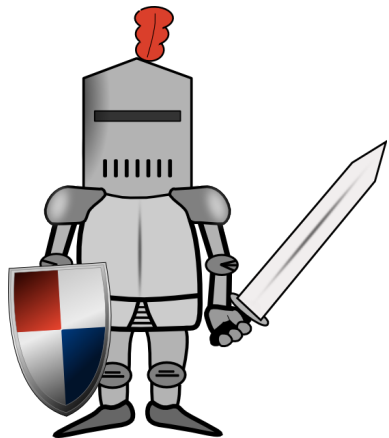
HPCC Systems can answer:

- More Questions Faster
- More Complex Questions



Google = 

“Makes me more money.”



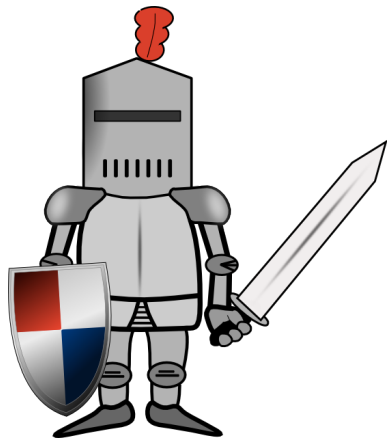
Why different parts of a business choose a Big Data technology

“It’s been open source since 2005.”



Google = 

“Makes me more money.”



“Already made the investment in Hadoop.”



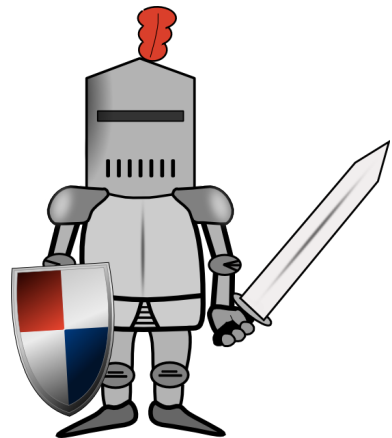
Why different parts of a business choose a Big Data technology

“Business requirement always change & increase and I’m ready for it.”



Google = 

“Makes me more money.”

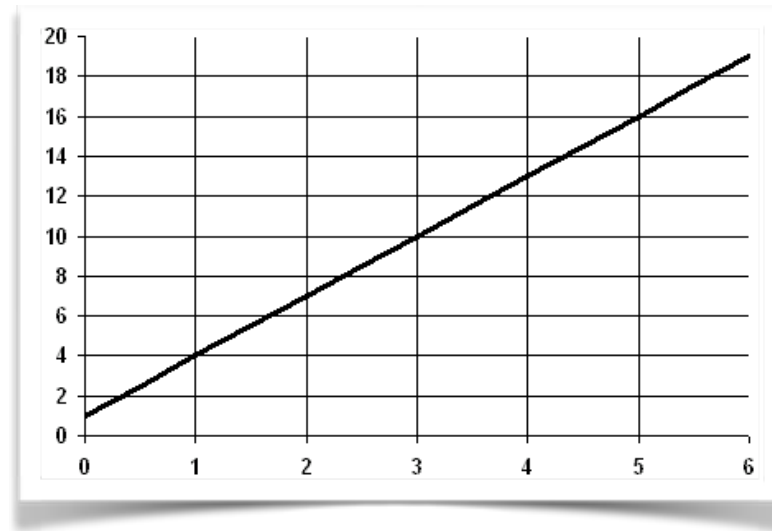


“Already made the investment in Hadoop.”

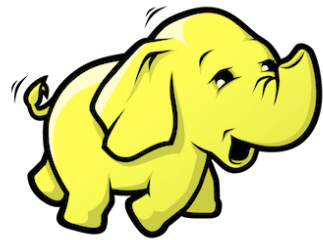
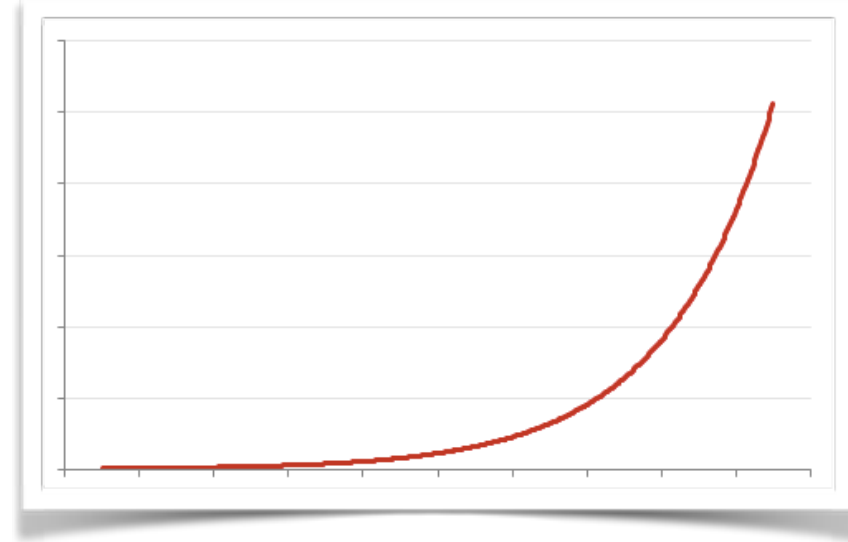
“Fits requirements plus more.”



What about the future?
Will you have more data?



VS

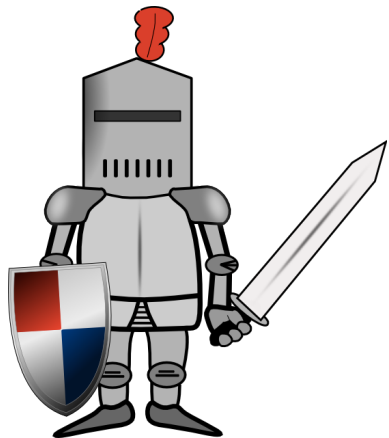


HPCC Systems



Google™ = 

“Makes me more money.”

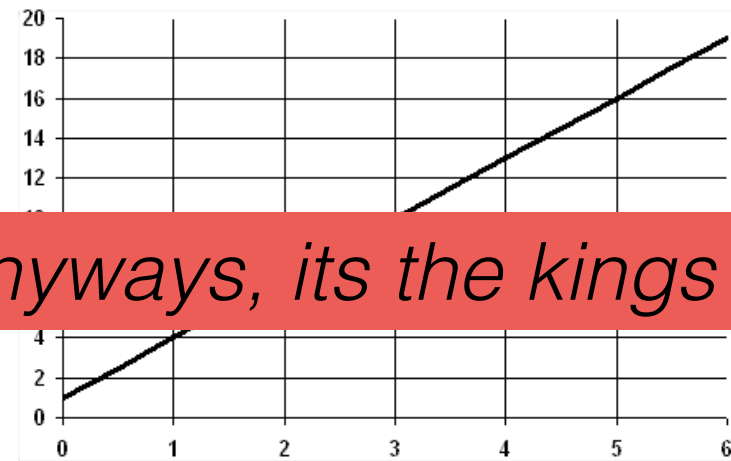


“Already made the investment in Hadoop.”

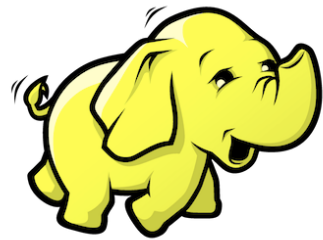
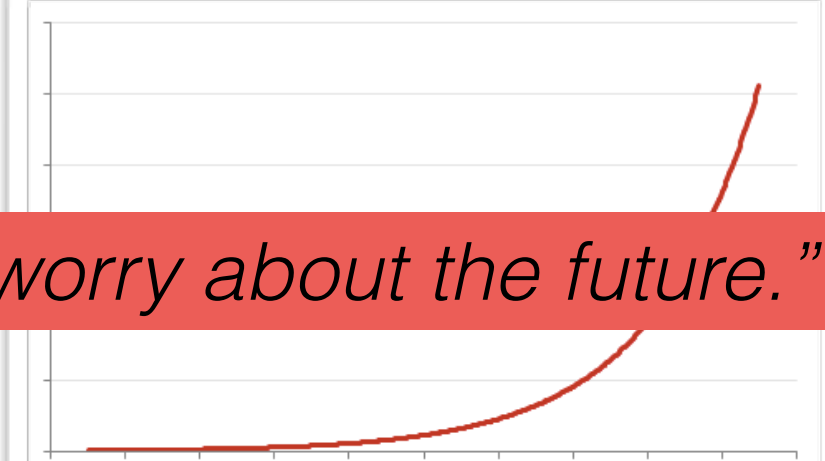
“Fits requirements plus more.”



What about the future?
Will you have more data?



“Anyways, its the kings job to worry about the future.”

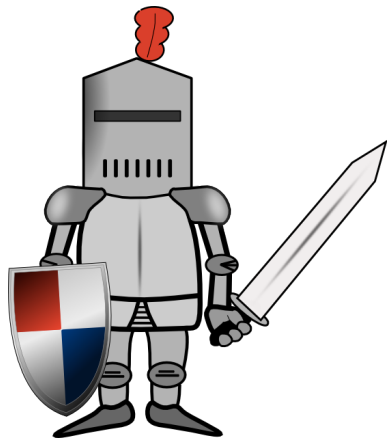


HPCC Systems



Google™ = 

“Makes me more money.”



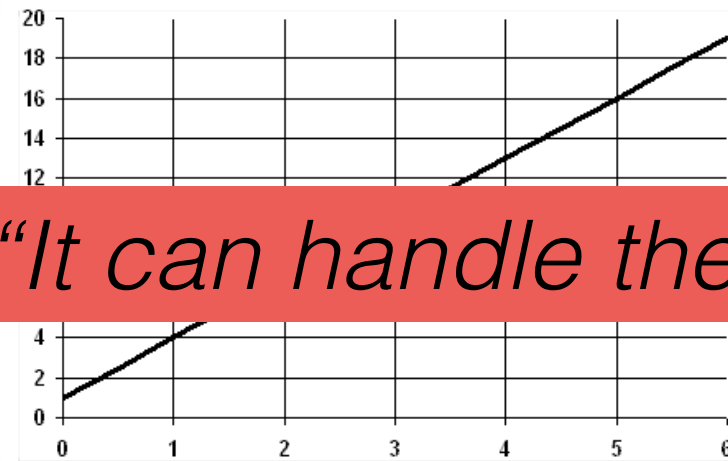
“Already made the investment in Hadoop.”

“Fits requirements plus more.”

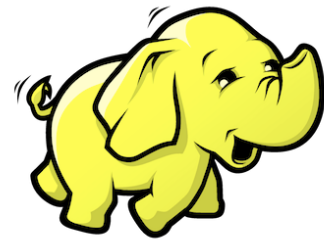
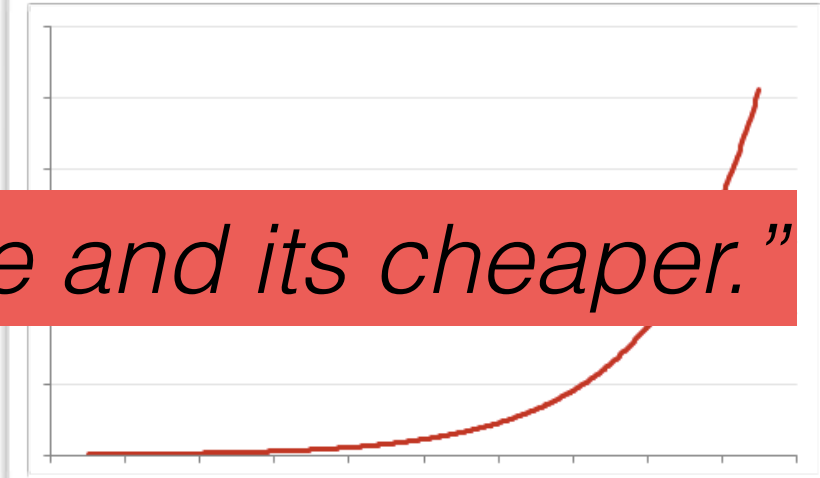
“I’ll worry about it later.”



What about the future?
Will you have more data?



"It can handle the future and its cheaper."

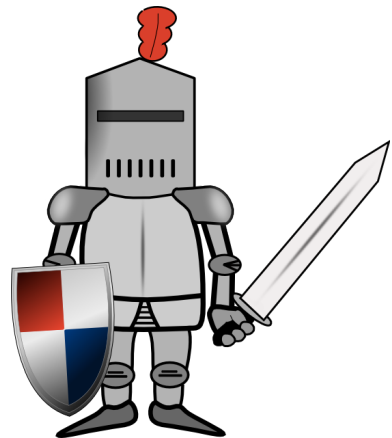


HPCC Systems



Google™ = 

"Makes me more money."



"Already made the investment in Hadoop."

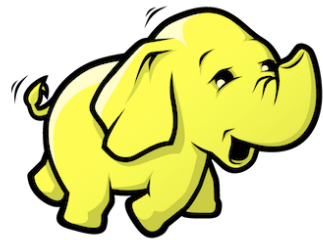
"Fits requirements plus more."

"I'll worry about it later."

"Reduced Costs"



Why different parts of business pick technology

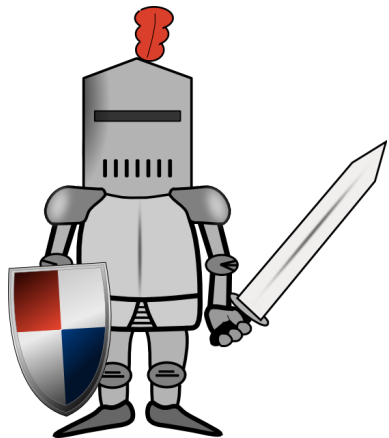


HPCC Systems



Google™ = 

“Makes me more money.”



“Already made the investment in Hadoop.”

“Fits requirements plus more.”

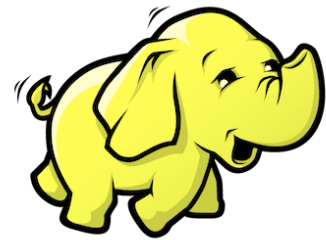
“I’ll worry about it later.”

“Reduced Costs”



“Companies pay a lot.”

Why different parts of business pick technology

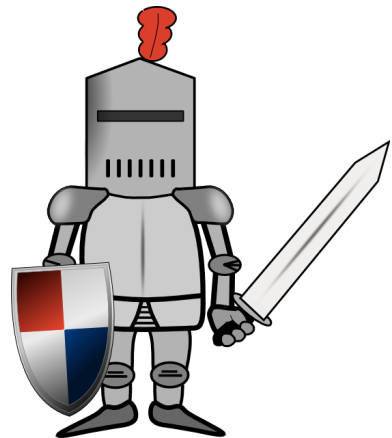


HPCC Systems



Google = 

“Makes me more money.”



“Already made the investment in Hadoop.”

“Fits requirements plus more.”

“I’ll worry about it later.”

“Reduced Costs”



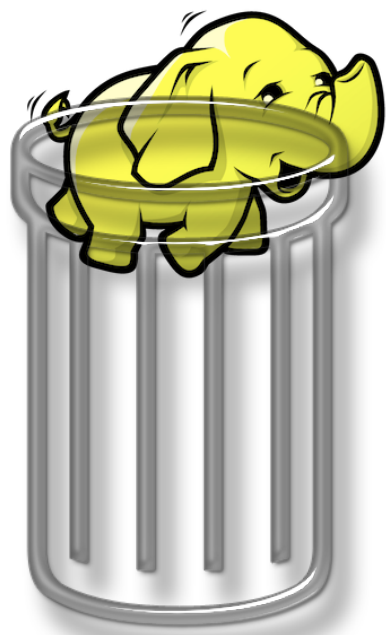
“Companies pay a lot.”

“Easy to manage & use.”

Analytics

Google Just Made Big Data Expertise Much Tougher to Fake

Well, Google (GOOG) officially put an end to the good times this week. During some talks at the company's annual developer conference, **Google executives declared that they're over Hadoop**. It's yesterday's buzzword. Anyone who wants to be a true Big Data jockey will now need to be conversant in Flume, MillWheel, Google Cloud Dataflow, and Spurch. (Okay, I made the last one up.)



Photograph by Jeff Chiu/AP Photo

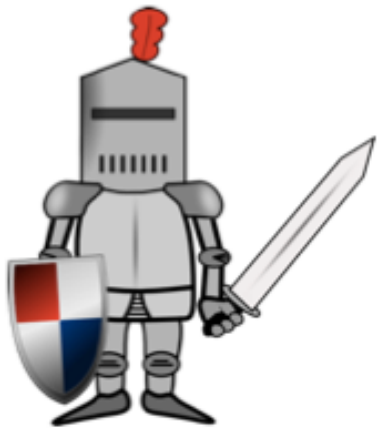


HPCC Systems



Google = 

“Makes me more money.”



“Already made the investment in Hadoop.”

“Fits requirements plus more.”

“I’ll worry about it later.”

“Reduced Costs”



“Companies pay a lot.”

“Easy to manage & use.”

How do companies think?

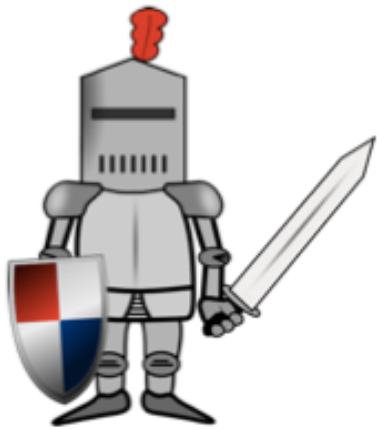


HPCC Systems



Google = 

“Makes me more money.”



“Already made the investment in Hadoop.”

“Fits requirements plus more.”

“I’ll worry about it later.”

“Reduced Costs”



“Companies pay a lot.”

“Easy to manage & use.”

“I want the reward ,without any risk.”

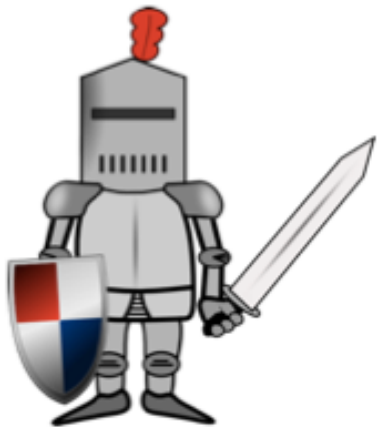


HPCC Systems



Google = 

“Makes me more money.”



“Already made the investment in Hadoop.”

“Fits requirements plus more.”

“I’ll worry about it later.”

“Reduced Costs”




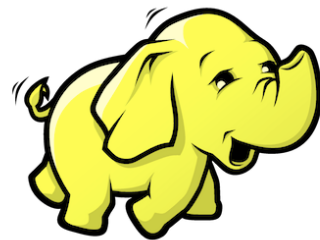
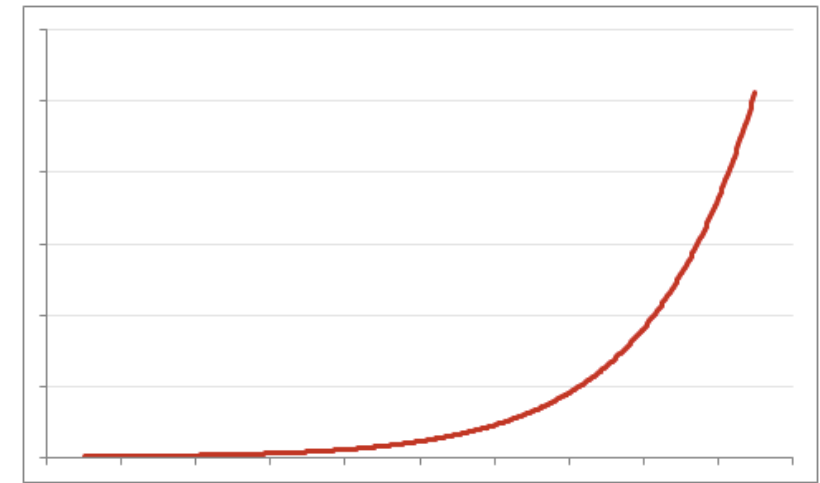
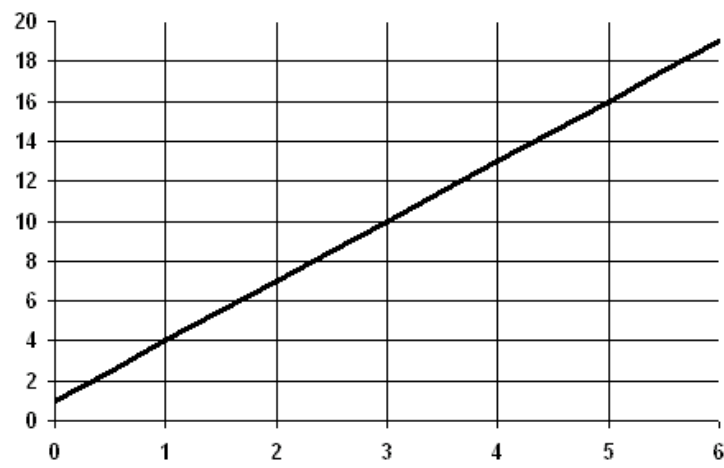
“Companies pay a lot.”

“Easy to manage & use.”

“I want the reward ,without any risk.”

“My company makes more money by being better then the competition.”





HPCC Systems

“Makes me money.”

“Makes me more money.”

“Competitive Advantage”

“Real Competitive Advantage”

“I need something off the shelf.”

“Reduced Costs”

“I don’t have that much data.”

“Companies pay a lot.”

“Easy to manage & use.”



“My main focus is my
business objectives.
Everything else is just noise.”

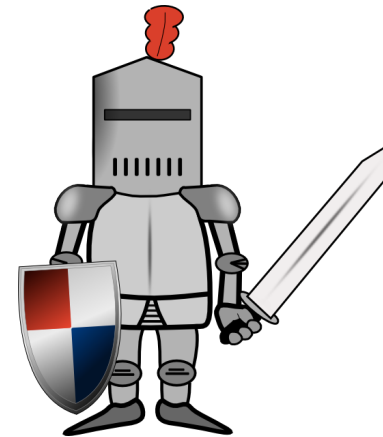
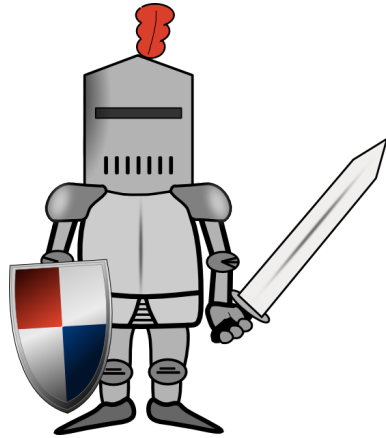




“I need more insight
into the business.
I need Big Data.”

Software

Hardware

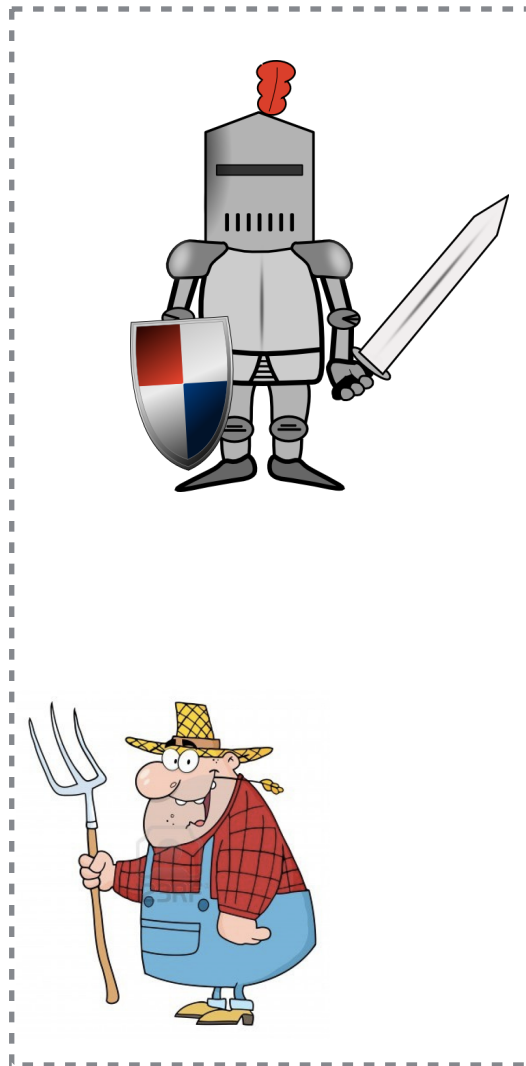


SQL

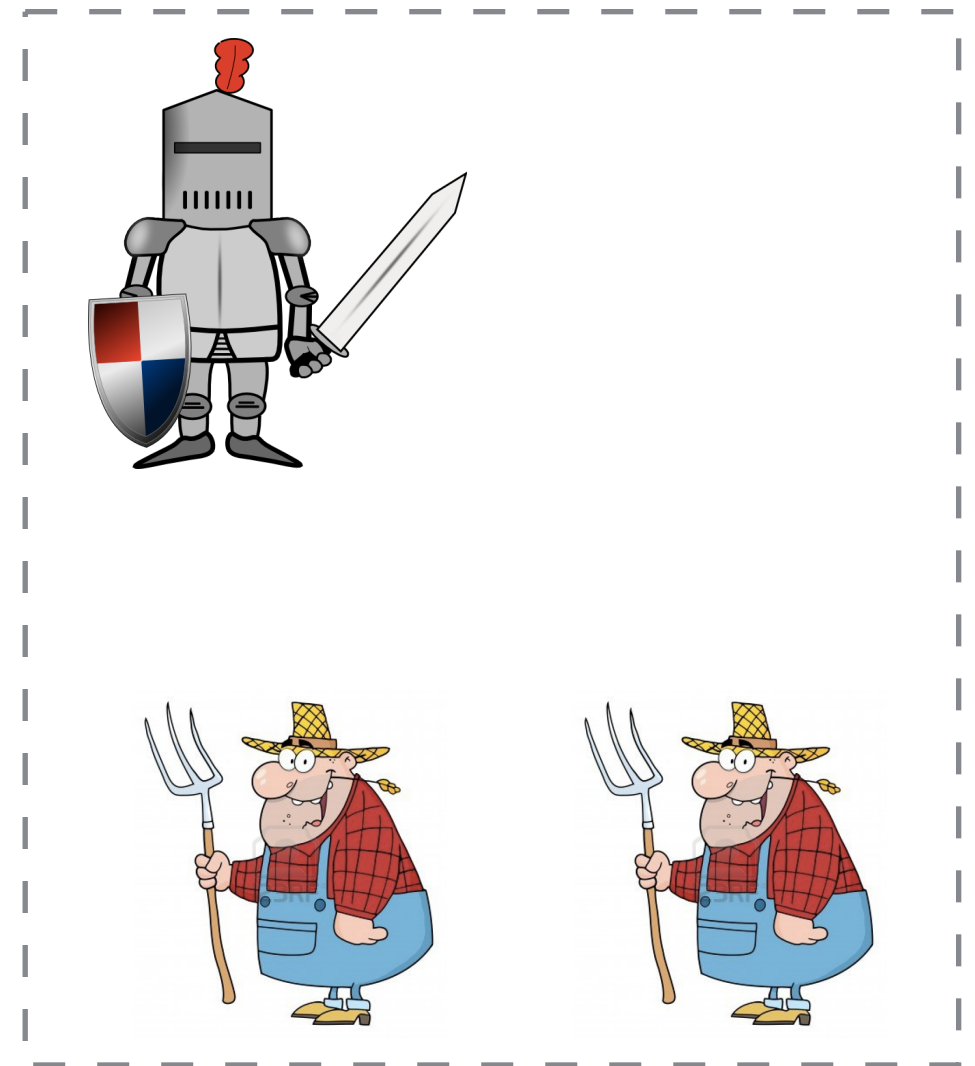


“HPCC Systems
does reduce costs
and is fast...”

Software



Hardware



SQL

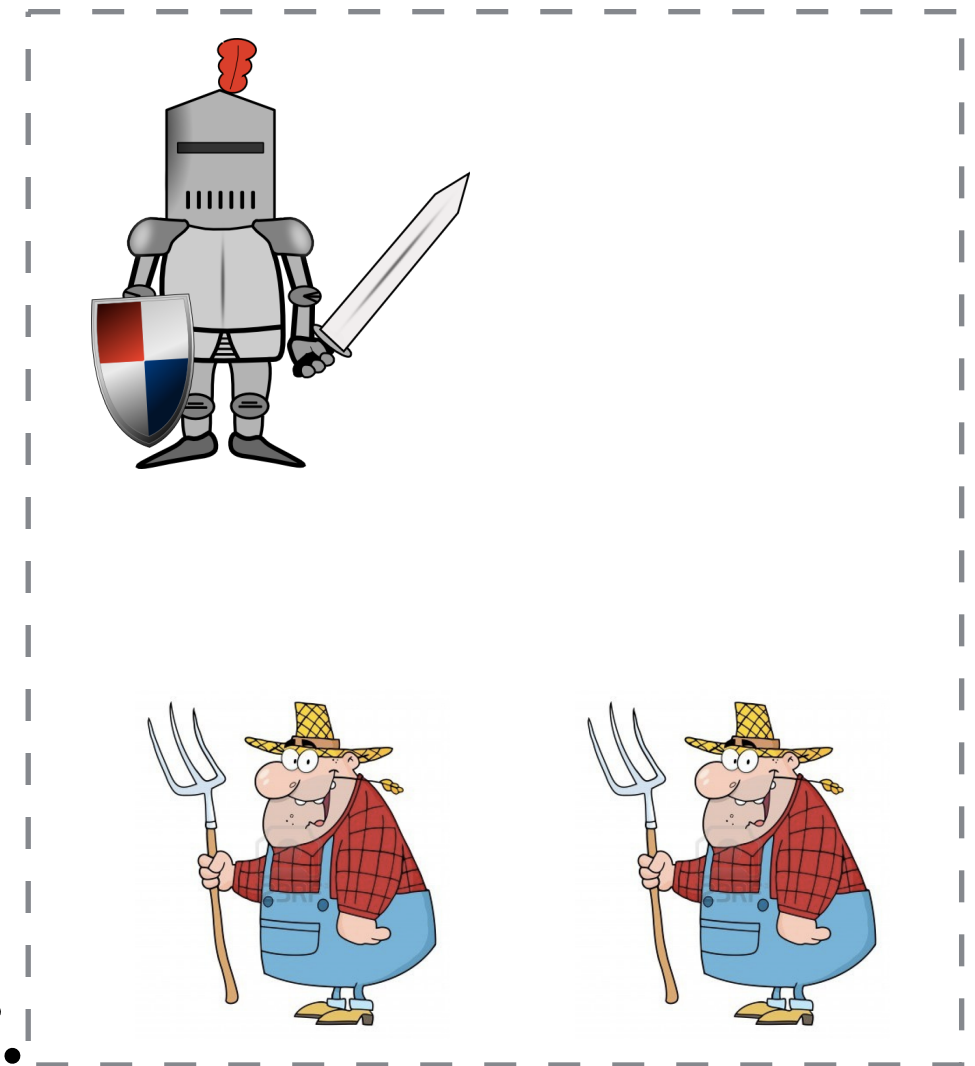
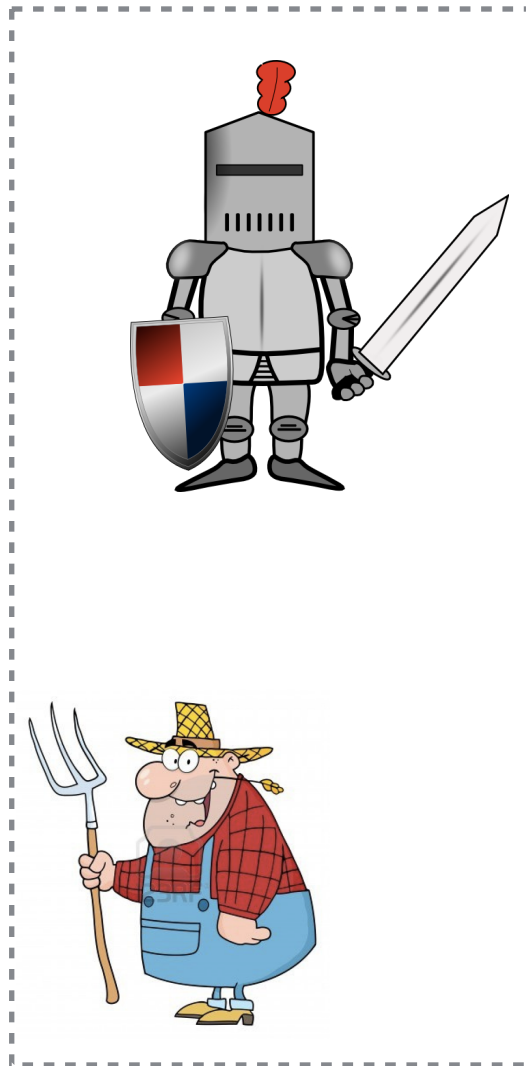




“.....,but if the king could ask the questions directly it would solve real business problems.”

Software *and/or*

Hardware



SQL

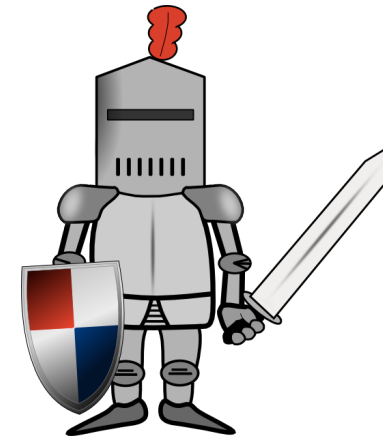
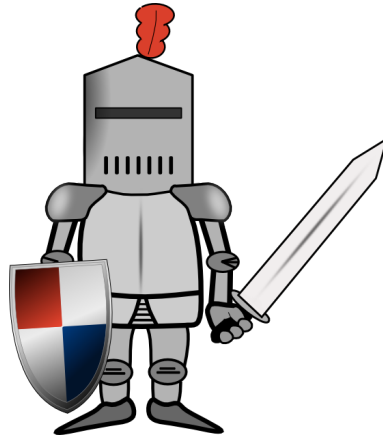


But, Don't Know Tech



Software *and/or*

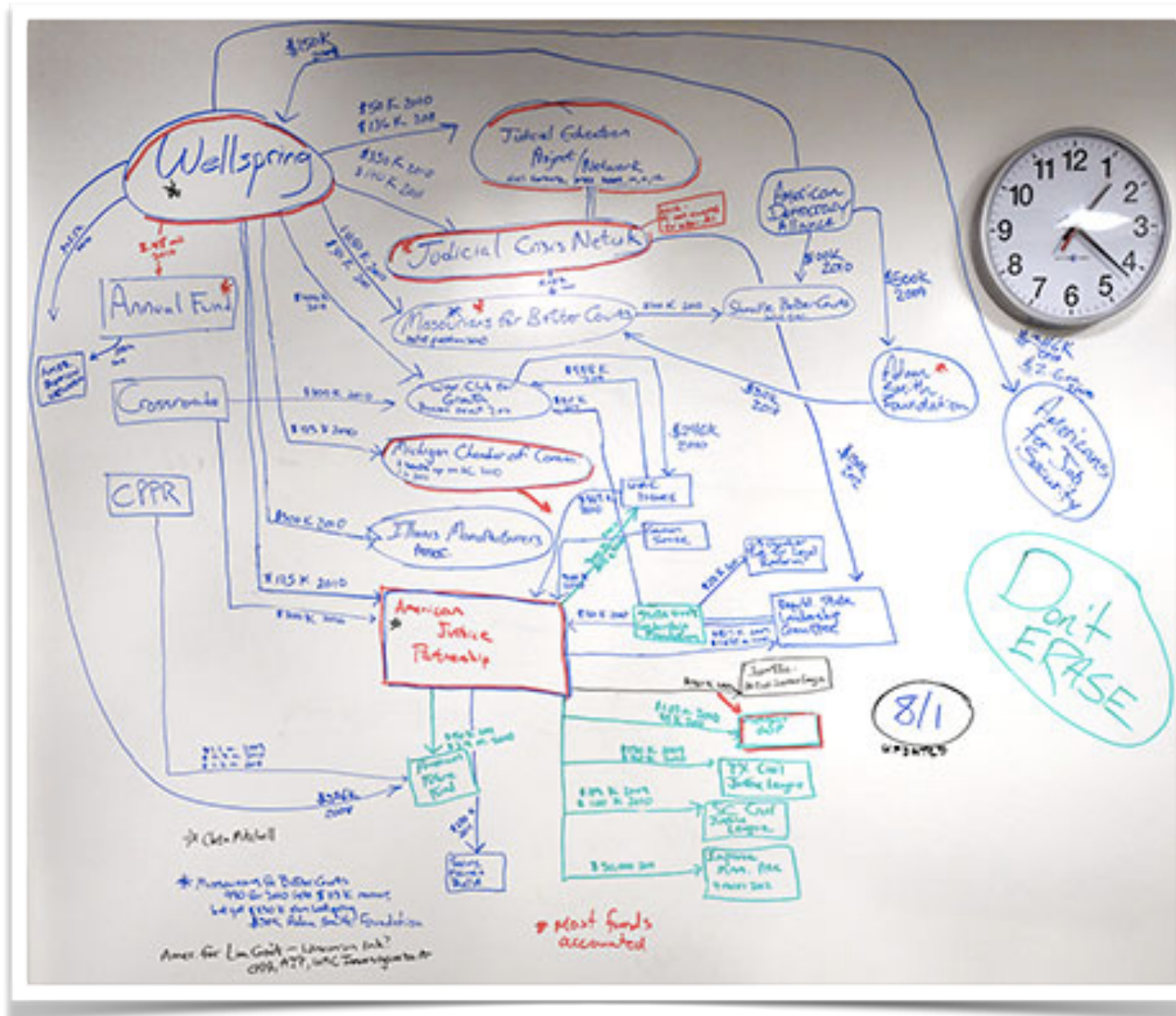
Hardware



SQL



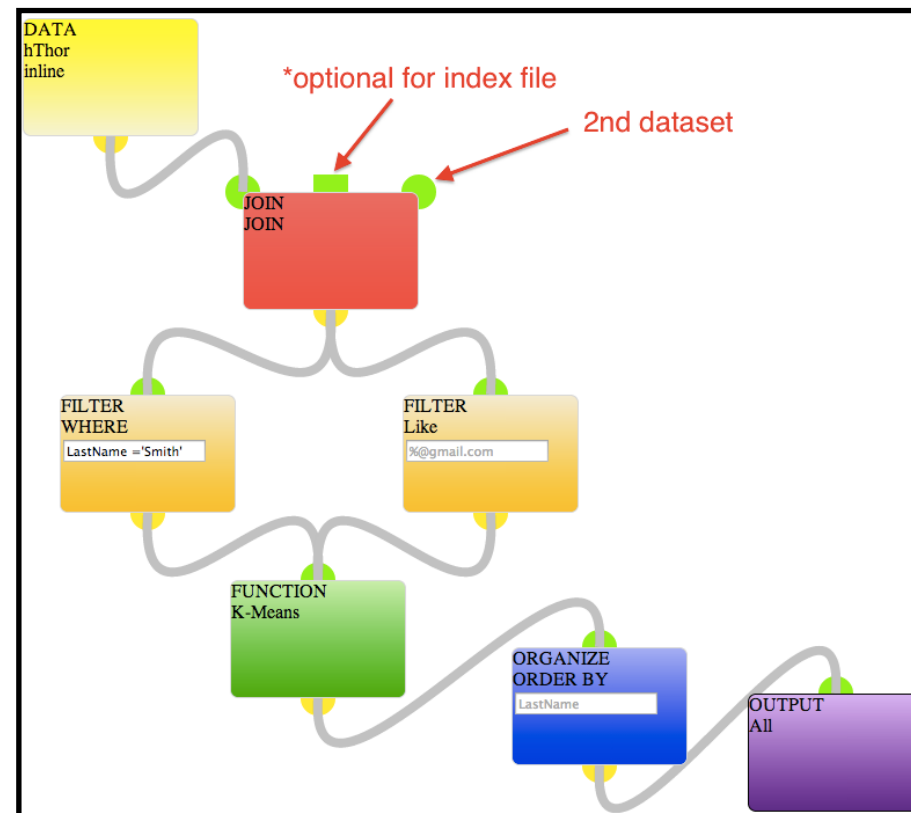
My Business Objective is



In meetings we work problems like this...



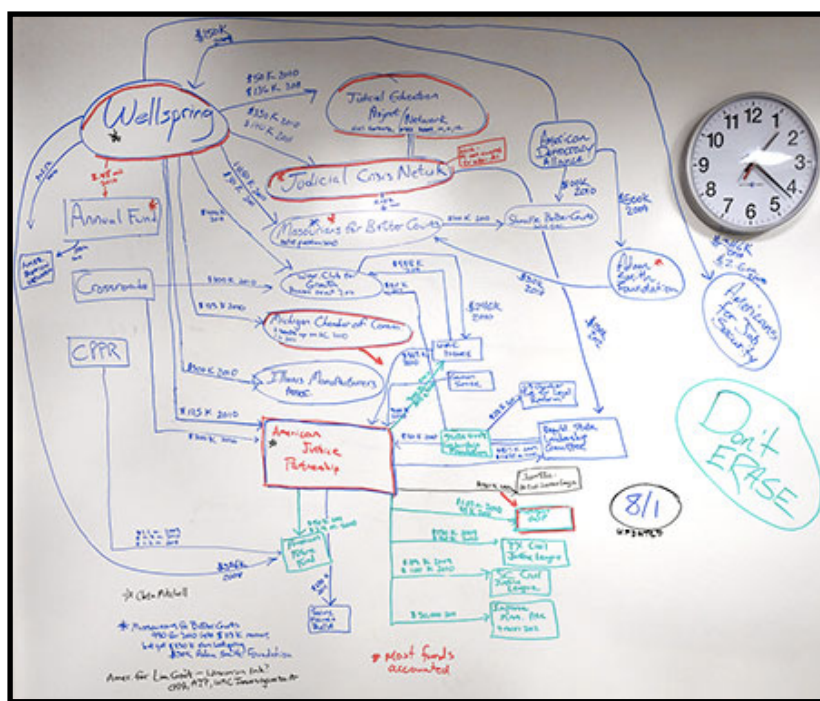
ECL Pipes combines the way business think and act and converts them to real action with Drag N' Drop to make ECL code.



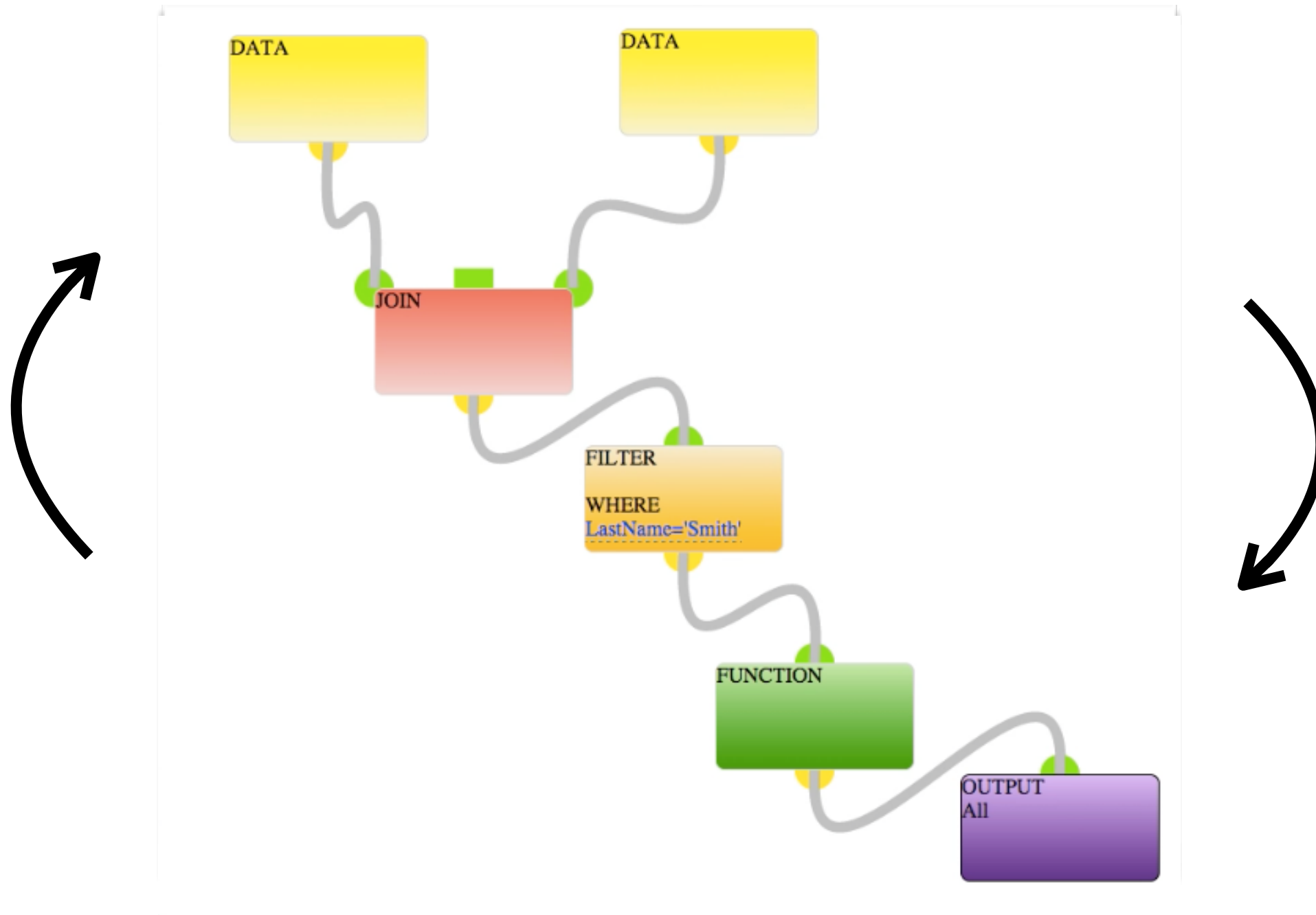
```

1 Layout_People := RECORD
2   STRING15 FirstName;
3   STRING25 LastName;
4   STRING15 MiddleName;
5   STRING5 Zip;
6   STRING42 Street;
7   STRING20 City;
8   STRING2 State;
9 END;
10
11 allPeople := DATASET('~test::originalperson',Layout_People,THOR);
12
13 Smiths := allPeople(LastName = 'Smith');
14
15 Smiths;

```



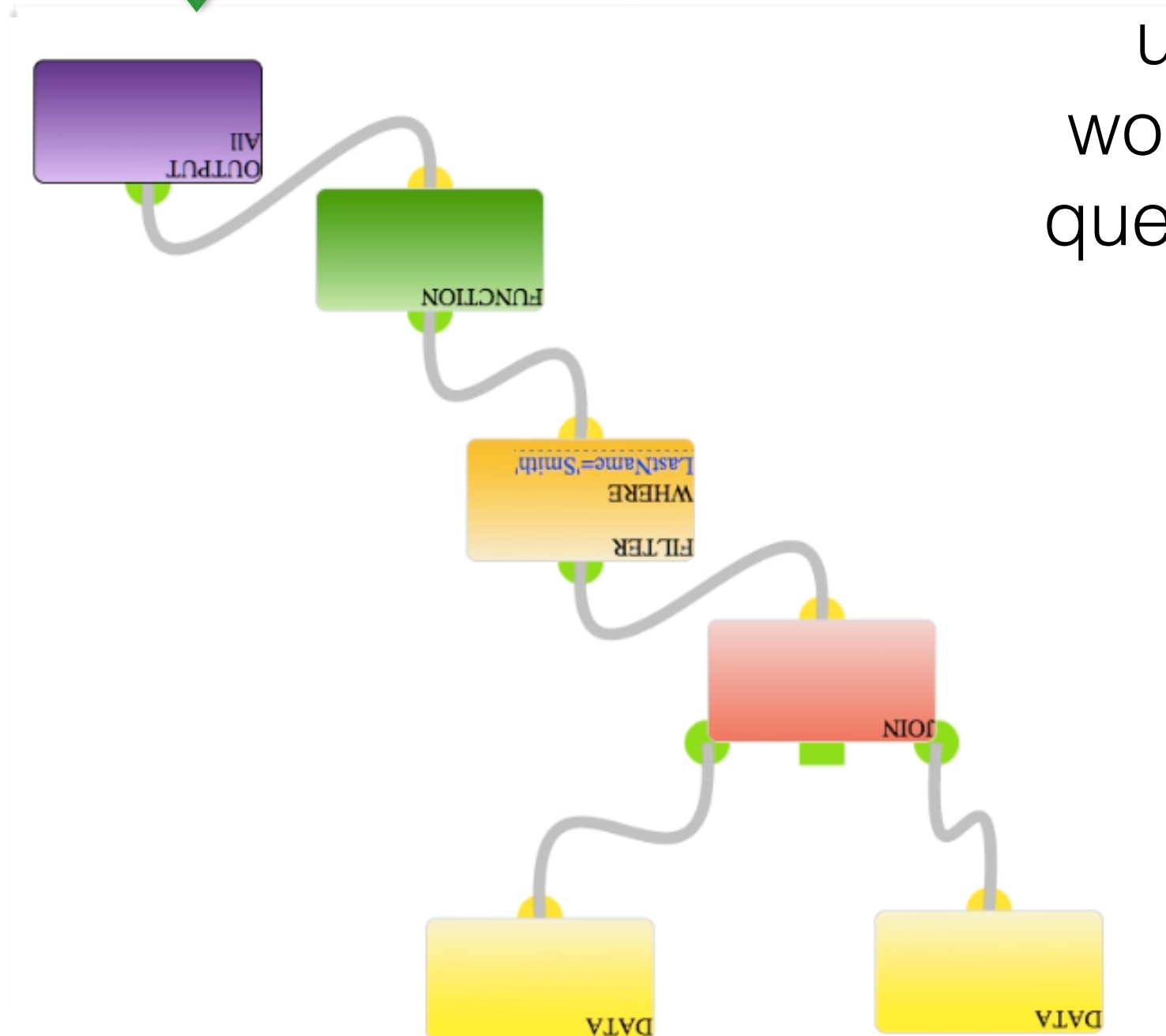
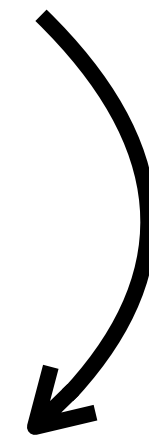
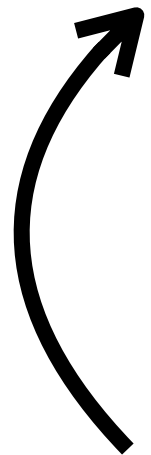
My Business Objective is?



My Business Objective is right here.



ECL Pipes helps
users create
workflows to ask
questions quickly.



Conclusion

