



Quantbot

Technologies, LP

Paul White
CEO, Co-founder

“Data About Data: Measuring ROI”

September 22, 2021

Neudata Data Insights & Scouting Summit - September 22, 2021



- Quantbot:
 - Registered Investment Adviser, servicing one large, institutional, client: Schonfeld
 - Global statistical arbitrage and quantitative futures
 - Offices in New York, London, Hong Kong and New Orleans
- Founding partners worked together for over 20 years
 - Morgan Stanley late 90s created Equity Trading Lab (ETL) business
 - Electronic and Algo trading - VWAP, SORT etc.
 - Proprietary trading - global statistical arbitrage
 - Moved to Merrill Lynch in early 2000s
 - Oversaw revamp of Electronic/Algo trading
 - Created new global proprietary trading business QSA
 - Created Quantbot in 2009 after the financial crisis hit

Data About Data : Measuring ROI



Quantbot
Technologies, LP

- ROI - hard to measure, let's define what we're talking about today
- Costs:
 - Data:
 - Licensing, storage and access
 - Processing:
 - Compute on-prem, cloud charges etc.
 - Other:
 - People, fixed costs etc.
- Return:
 - PnL after trading/investment costs

Data About Data : Lifecycle of Data



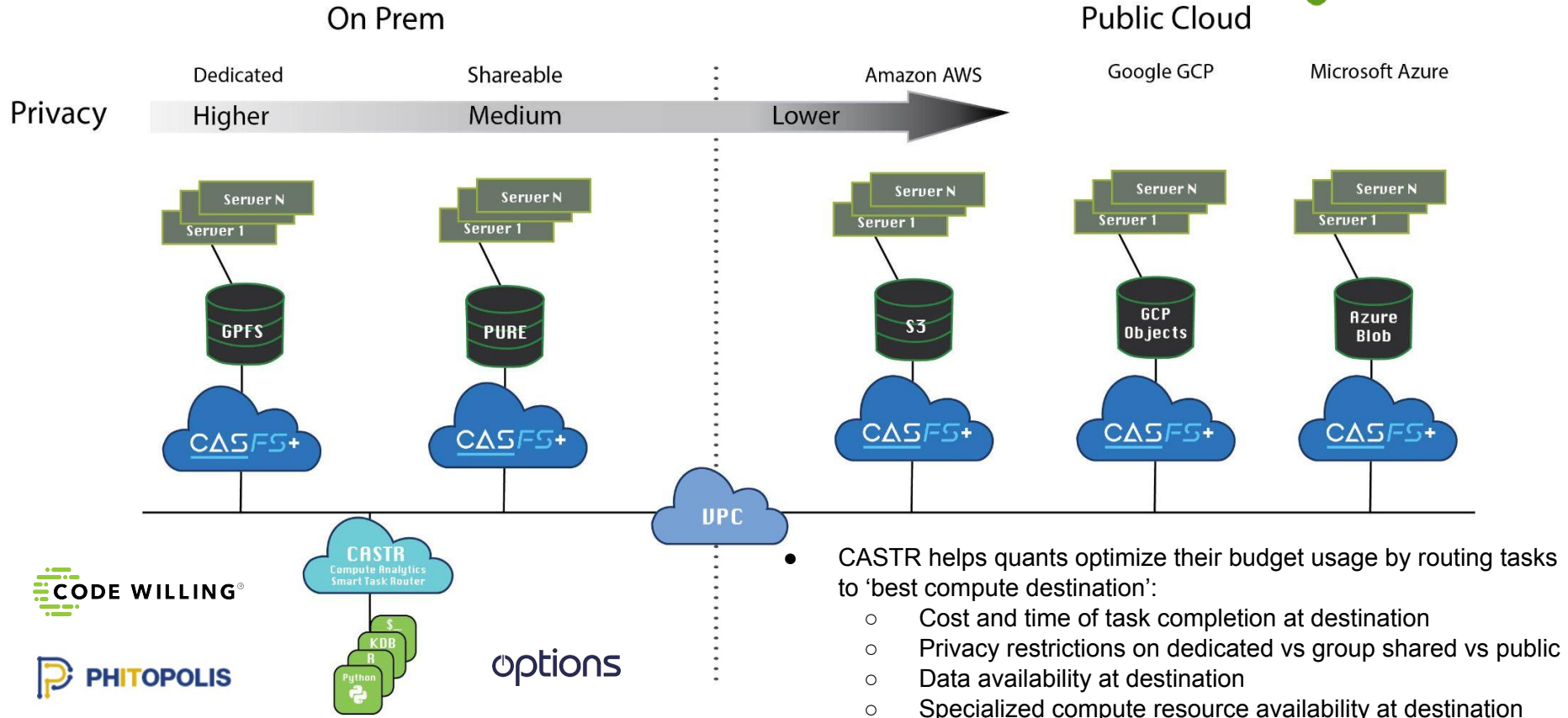
Quantbot
Technologies, LP

- Trial - most data go through a trial period before we subscribe
 - Some work needed in order to support initial trial research
 - Onboarding process for trial
 - Historical data integration
 - Dataset documentation
- Production (Our focus for ROI) :
 - Onboarding for full research and trading support
 - Production Ready - continuous mode of:
 - Research
 - Trading

Smart Task Router: CASTR (Where we get the Data)



Quantbot
Technologies, LP

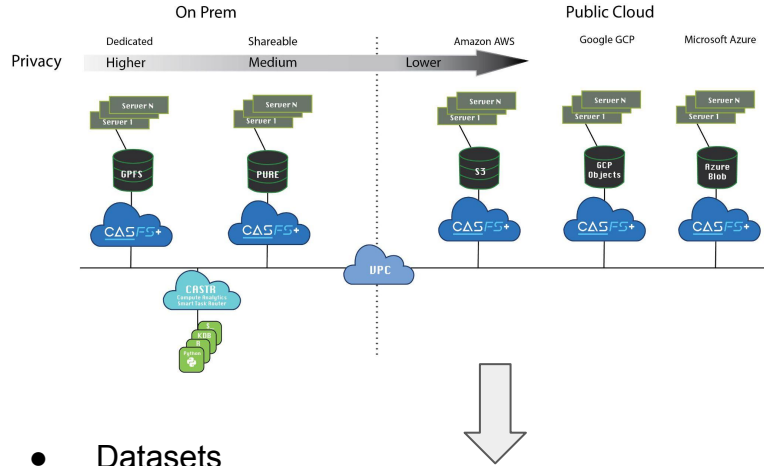


- CASTR helps quants optimize their budget usage by routing tasks to 'best compute destination':
 - Cost and time of task completion at destination
 - Privacy restrictions on dedicated vs group shared vs public
 - Data availability at destination
 - Specialized compute resource availability at destination

Data About Data : Metrics + Accounting



Quantbot
Technologies, LP



- Datasets
 - CRUD when, whom, how, what, stats
 - Cost
 - Processing
 - Storage
 - I/O ops
- Compute
 - Node usage: when, whom, how, what, stats
 - Cost



- Data costs
- PnL
- Miscellaneous costs

Metrics : What Can They Tell Us ?



Quantbot
Technologies, LP

What datasets are most/least accessed ?

Name	Last Access	Count
Data ?????	09/07/21 07:34:22	22.7K
Data ?????	09/07/21 07:25:08	21.3K
Data ?????	09/07/21 05:48:01	17.9K
Data ?????	09/07/21 04:55:45	13.2K
Data ?????	09/07/21 02:15:59	12.8K

Name	Last Access	Count
Data ?????	08/01/21 05:24:12	0
Data ?????	08/02/21 15:06:20	0
Data ?????	08/02/21 18:24:45	0
Data ?????	08/11/21 12:24:56	2
Data ?????	08/12/21 09:24:17	5

Metrics : What About Cost ?

Add costs and look at the least accessed again

Name	Last Access	Count	Cost
Data ?????	08/01/21 05:24:12	0	\$\$
Data ?????	08/02/21 15:06:20	0	\$\$
Data ?????	08/02/21 18:24:45	0	\$\$
Data ?????	08/11/21 12:24:56	2	\$
Data ?????	08/12/21 09:24:17	5	\$

CFO: I see a
chance to save
\$\$\$

Metrics : What Else Do We Have ?

Add in the user information to see who's using the data, let's select the most accessed dataset

Name	Quant Group	Last Access	Count
Data ?????	QG 11	09/01/21 05:24:12	6.4K
Data ?????	QG 2	09/01/21 04:22:36	4.2K
Data ?????	QG 9	08/31/21 22:59:00	3.4K
Data ?????	QG 14	08/31/21 18:07:59	3.0K
Data ?????	QG 22	08/30/21 11:45:27	2.9K

Metrics : What About Compute ?



Quantbot
Technologies, LP

Now add how much each group used for processing

Quant Group	Node Hours	Storage	I/Os	Cost
QG 11	2.7K	10.4T	3.9M	\$\$\$
QG 2	2K	8.2T	2.5M	\$\$\$
QG 22	1.7K	5.2T	1.8M	\$\$
QG 14	1.4K	1.2T	0.5M	\$\$
QG 7	0.5K	0.5T	0.2M	\$

ROI : Finally Lets View ROI by Group

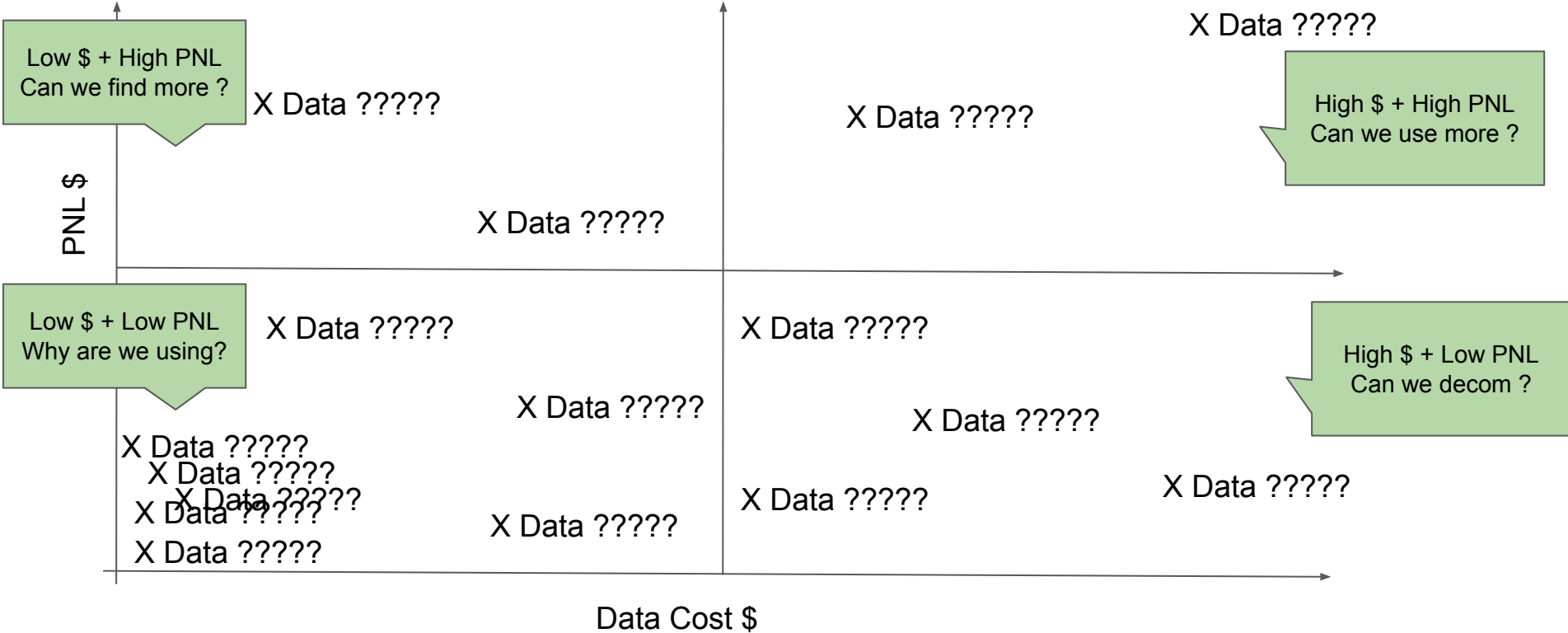


Quantbot
Technologies, LP

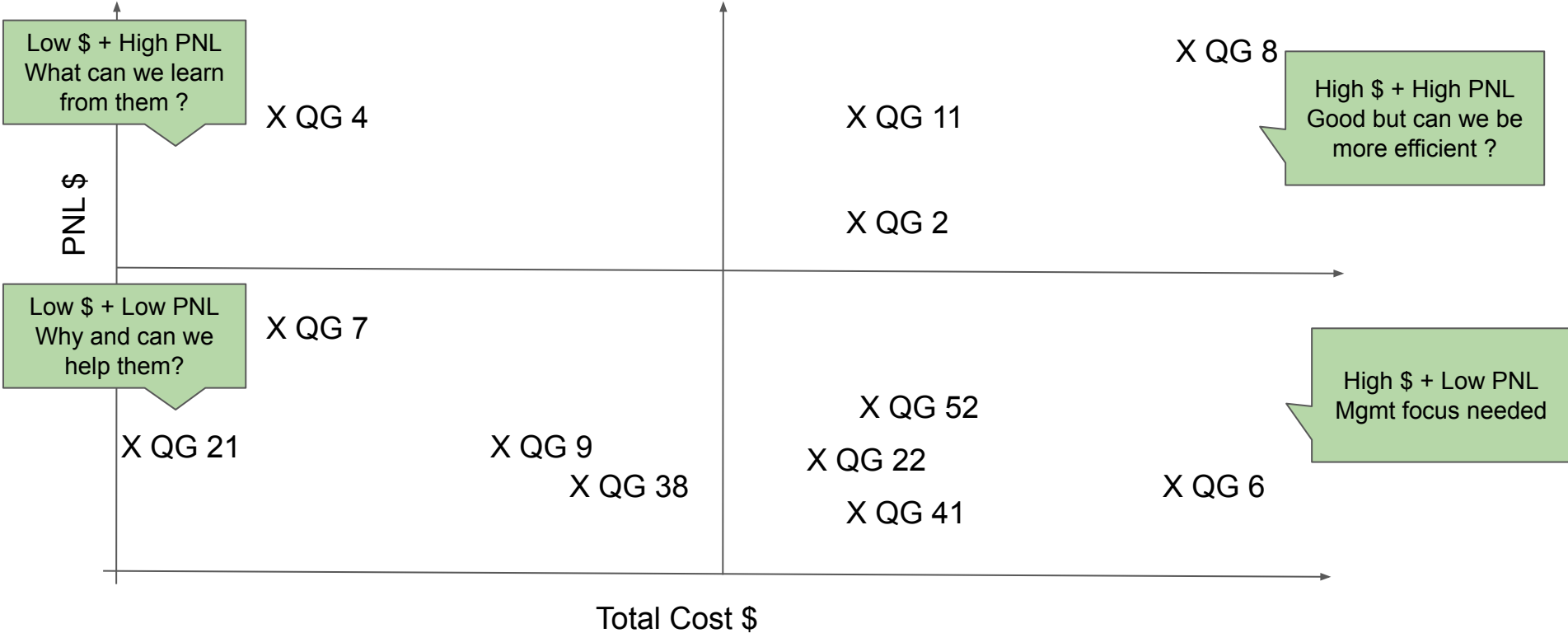
Now we have costs, let's add the return, i.e. the Net PnL after trading/investment costs

Quant Group	Data Cost	Compute Cost	PnL
QG 11	\$\$\$	\$\$\$	\$\$\$\$
QG 22	\$\$\$	\$\$	\$
QG 2	\$\$\$	\$\$\$	\$\$\$
QG 7	\$\$	\$	\$
QG 9	\$\$	\$\$\$	\$

ROI : Quadrant View of ROI by Dataset



ROI : Quadrant View of ROI by Quant Group





- Metrics:
 - Data about our data and research process gives us the ability to ask/answer lots of interesting questions in the ROI space
 - We have just scratched the surface on how to use these data to improve our efficiency
- Future interesting areas to investigate:
 - Expand ROI analysis
 - Team productivity/efficiency who's productive vs busy vs churning ?
 - Which Datasets are we sharing, with whom and can we leverage them more ?
 - Security reviews, who's allowed to access, why and how often ?
 - Budget planning/modeling - where should we spend our data/research \$
 - Trials - are we properly trying out the data within the window ?
 - etc.