Multi-Party Computation in Health Care

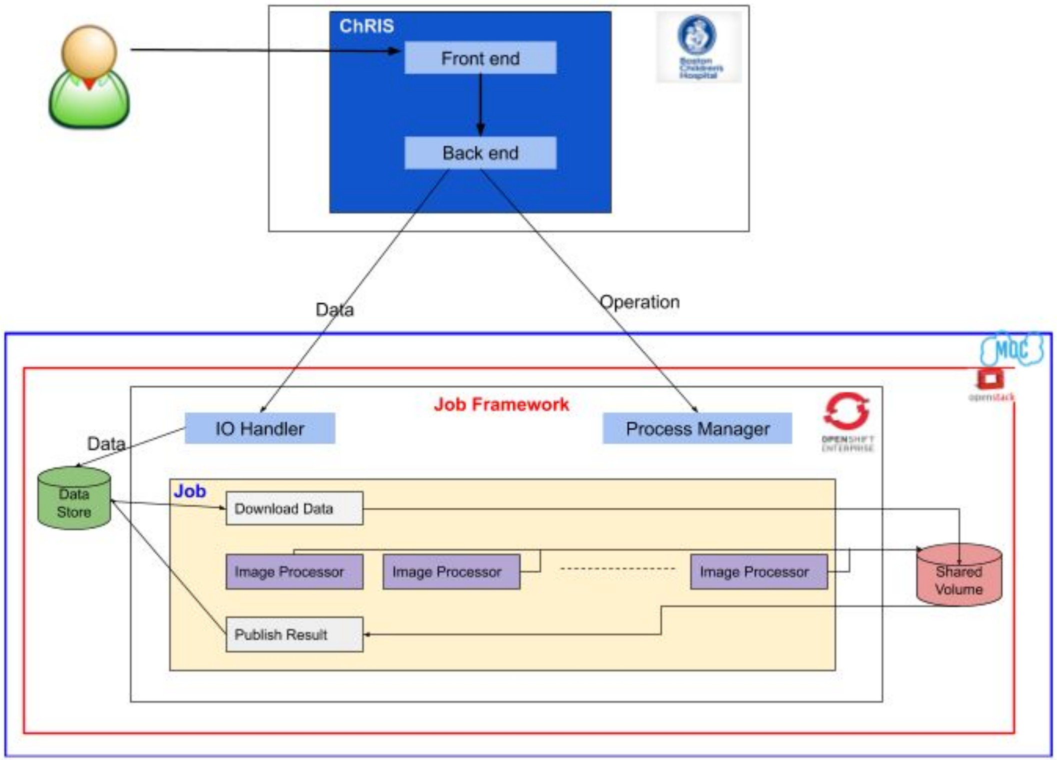
Parul Singh, Red Hat
Mayank Varia, Boston University

OPENSHIFT

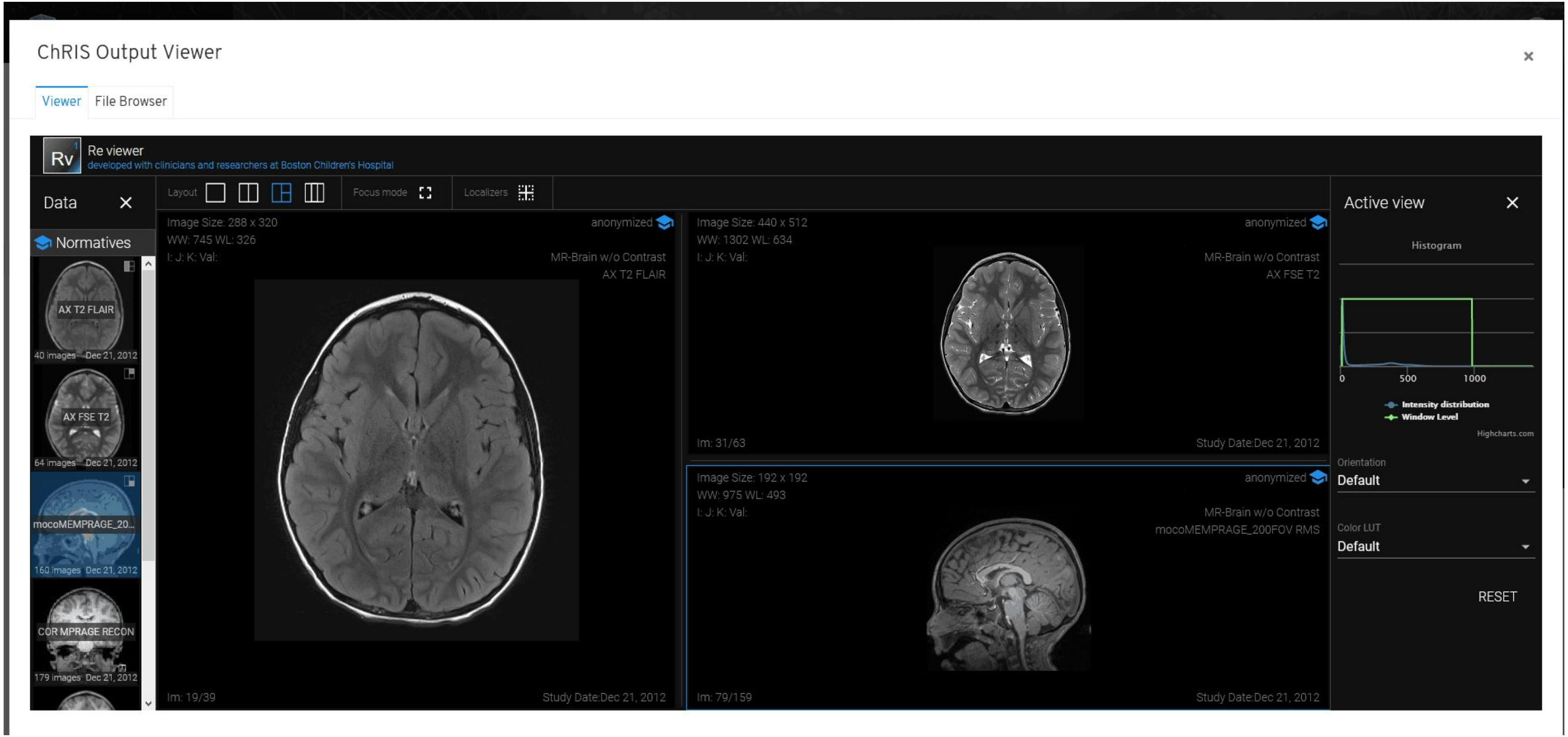
Secure Medical Image Processing

- At Red Hat we are involved in several research projects at universities and industries
  - Boston University & Boston Children's Hospital
- ChRIS Research Integration System (ChRIS)
  - Web-based medical image processing platform with quicker feedback
  - Democratize Medical Image Processing application development

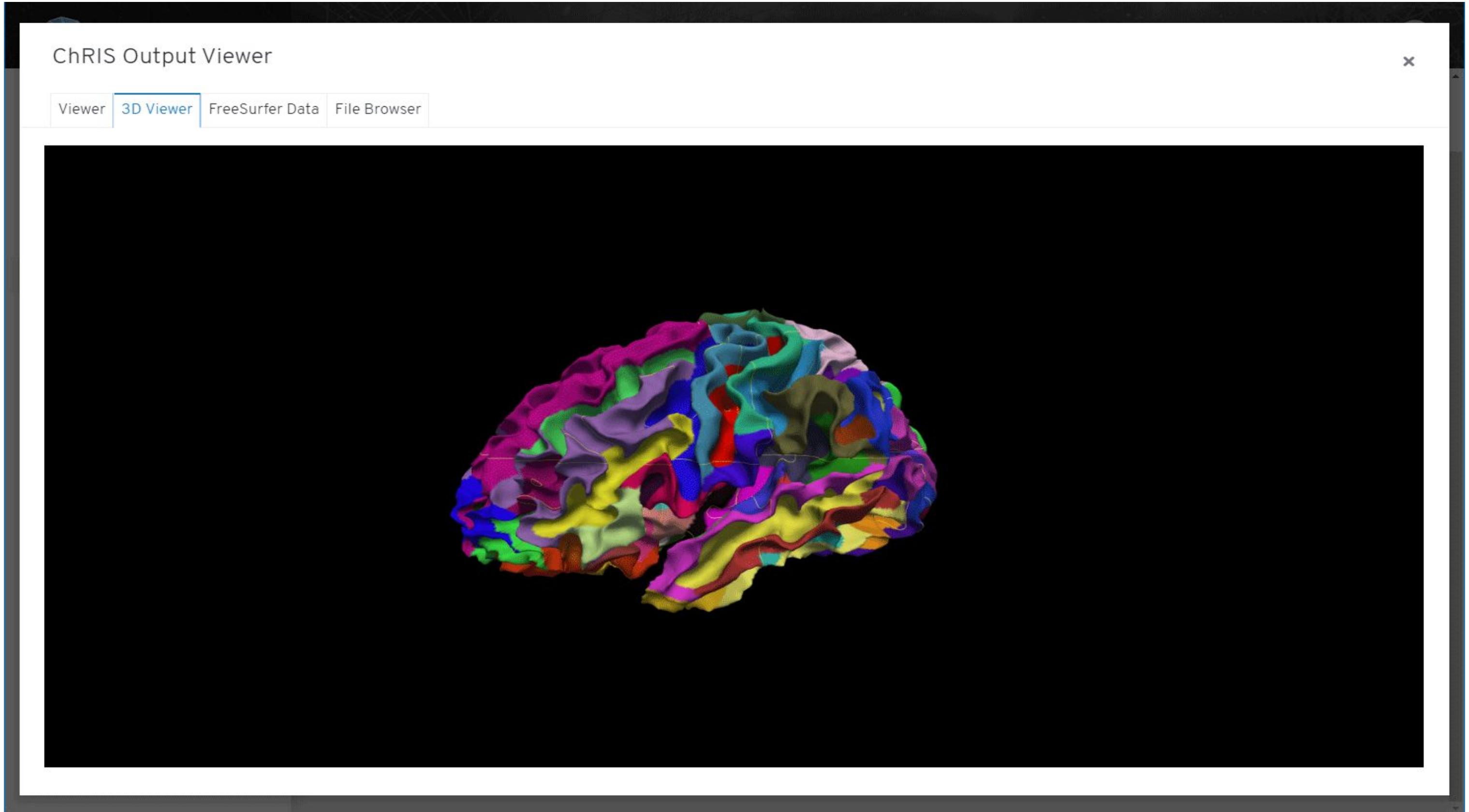
ChRIS Architecture



Medical Image Processing in ChRIS - Input



Medical Image Processing in ChRIS - Output



Conclave Cloud Dataverse (C2D)

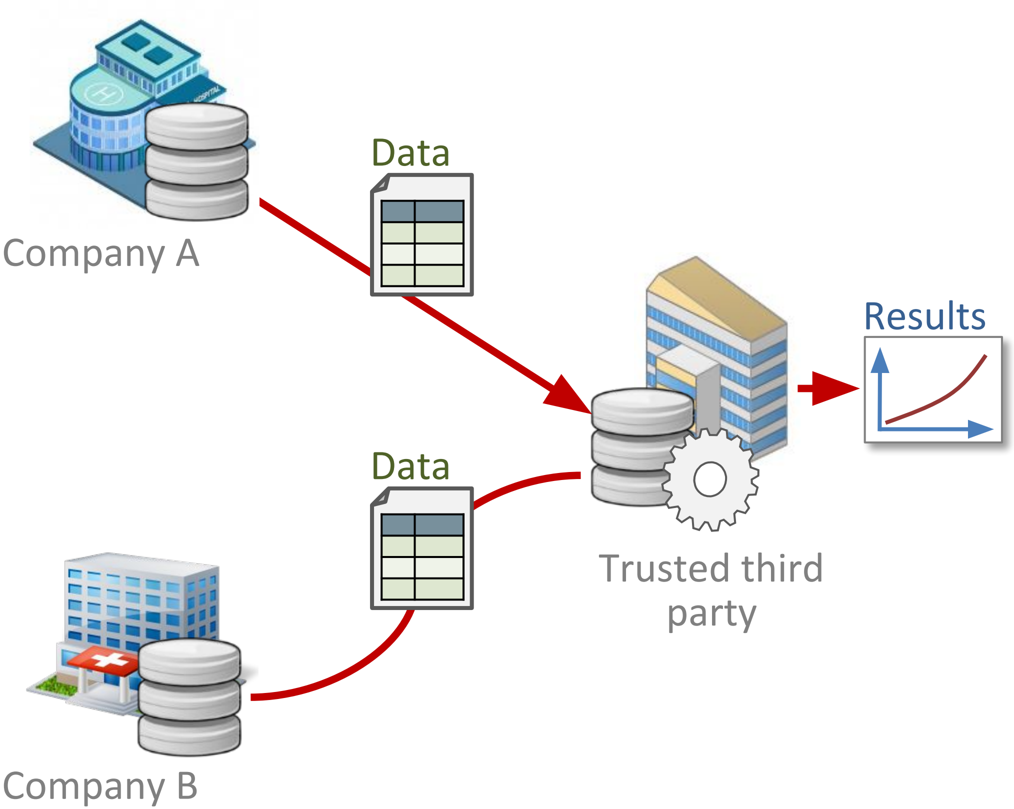
SAIL Software & Application Innovation Lab

BU

Red Hat

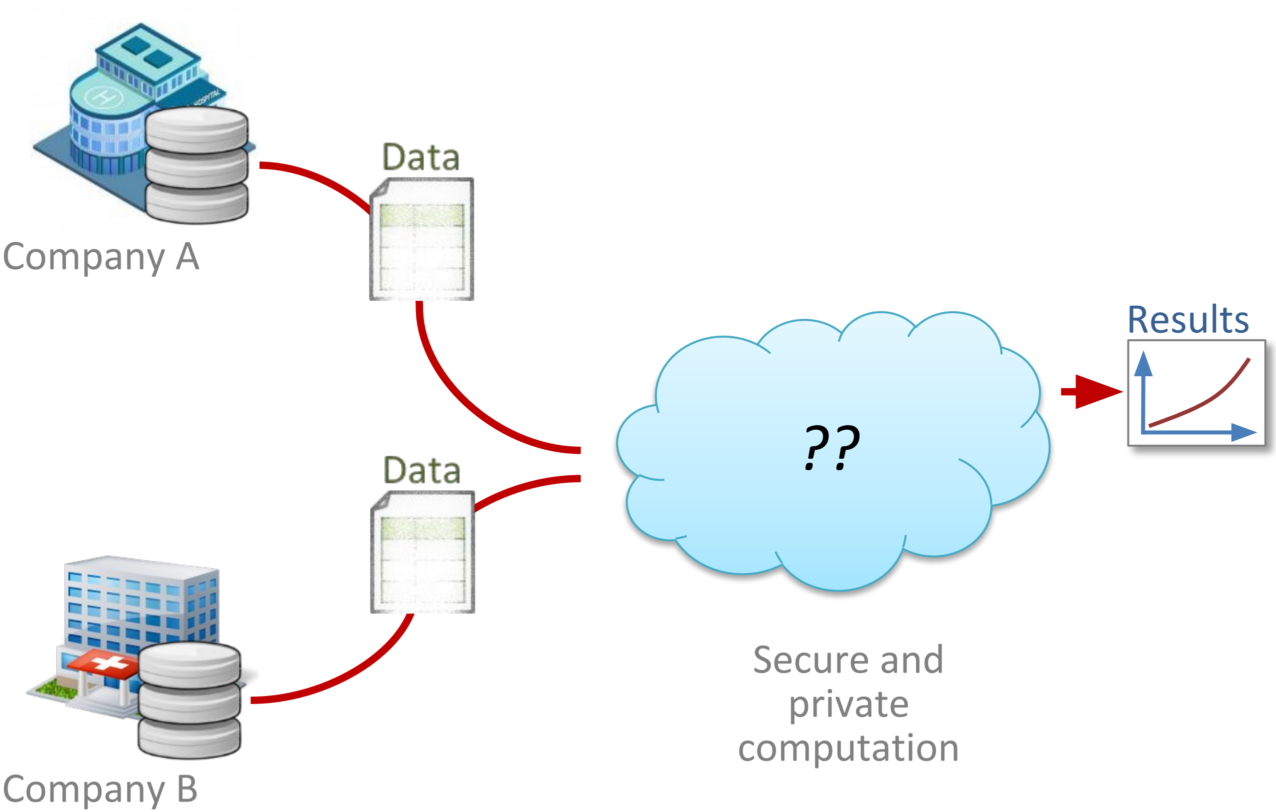
MOC

Typical Workflow: Centralize Data Storage & Analysis

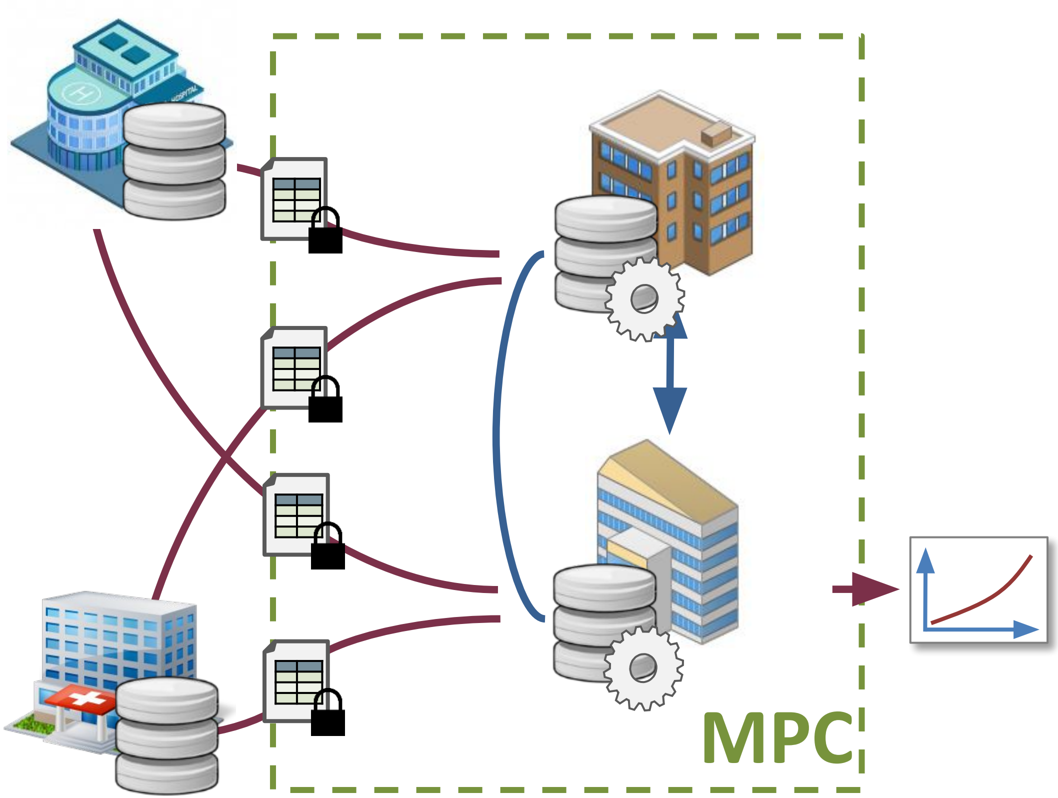


- Typically, transfer data to a common location
- All data contributors must **trust** the organization that performs the analysis
- Data may be vulnerable in transit, when stored, and during analysis

C2D Workflow: Compute Without Sharing Data

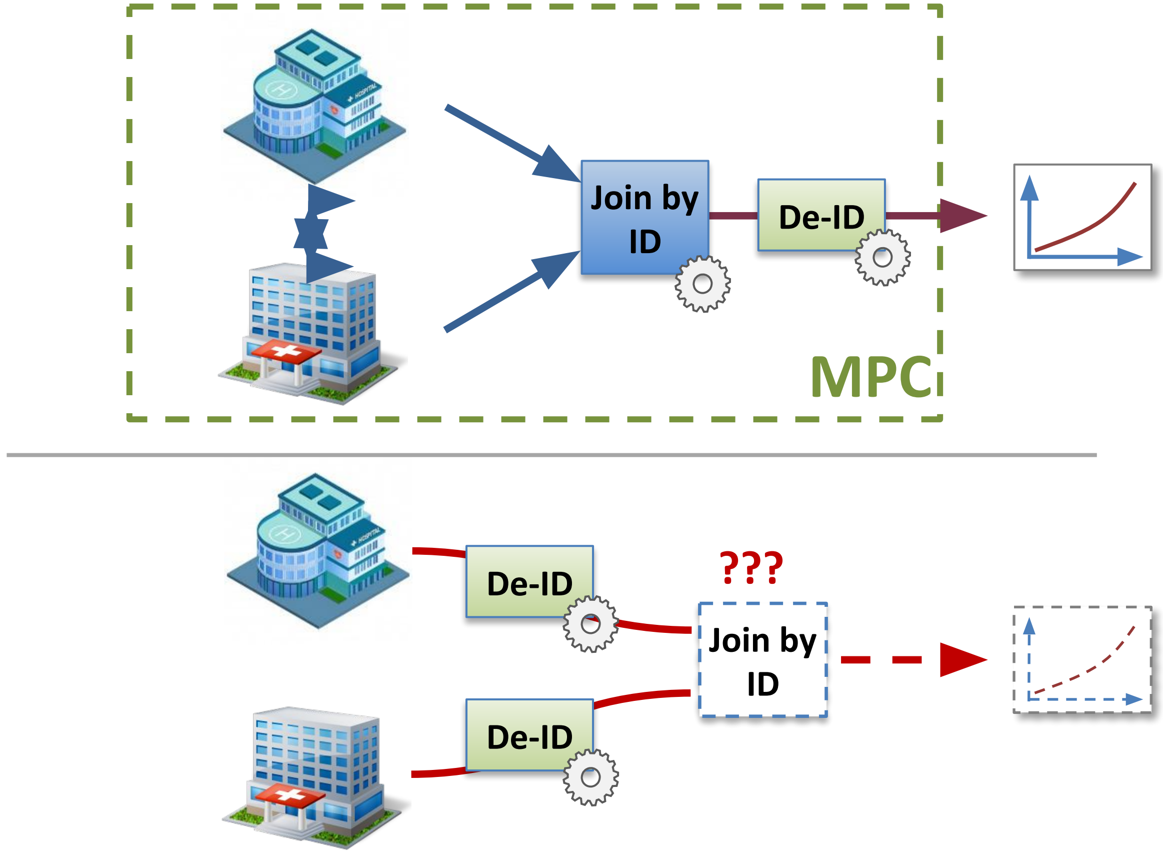


C2D Workflow: Federate Data Storage and Analysis



- Data **never leaves** any organization in the clear
- MPC protects data in use during the analysis
- Inputs stay private as long as **at least one** computing entity can be trusted to behave as specified

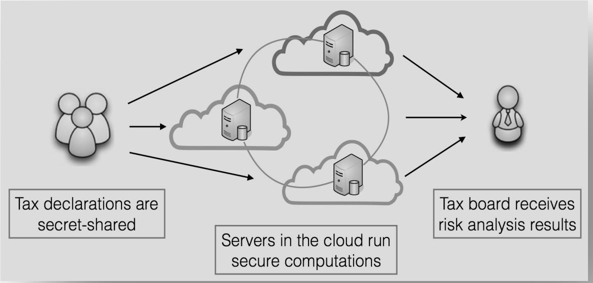
C2D Workflow: Valuable Analysis $\not\Rightarrow$ Share Data



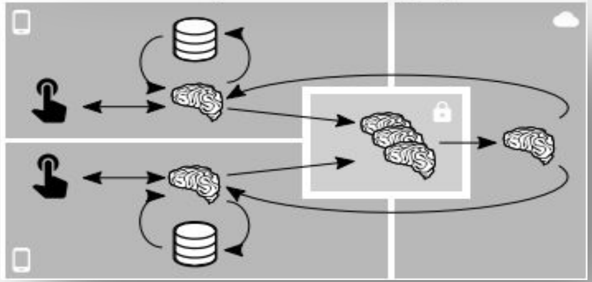
- New analyses can be privacy-preserving because data not shared prematurely
- Without MPC, would need data transfer, so protections applied "too early" in pipeline

Selected MPC deployments

Cybernetica: VAT tax audits

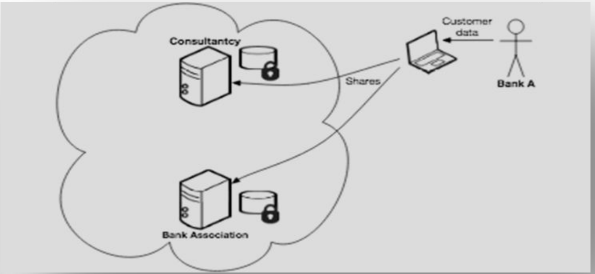


Google: Federated machine learning



BU: Pay equity in Boston

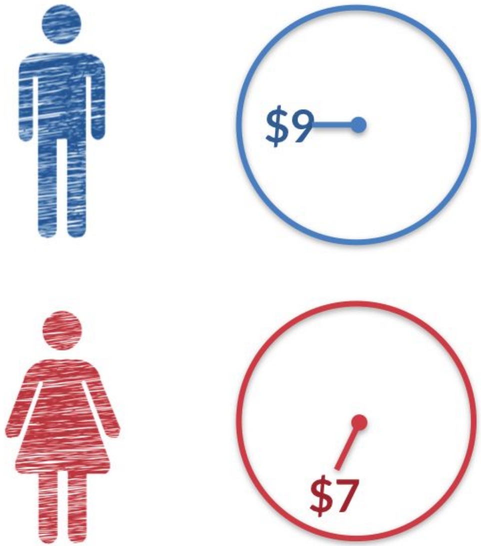
Partisia: Rate credit of farmers



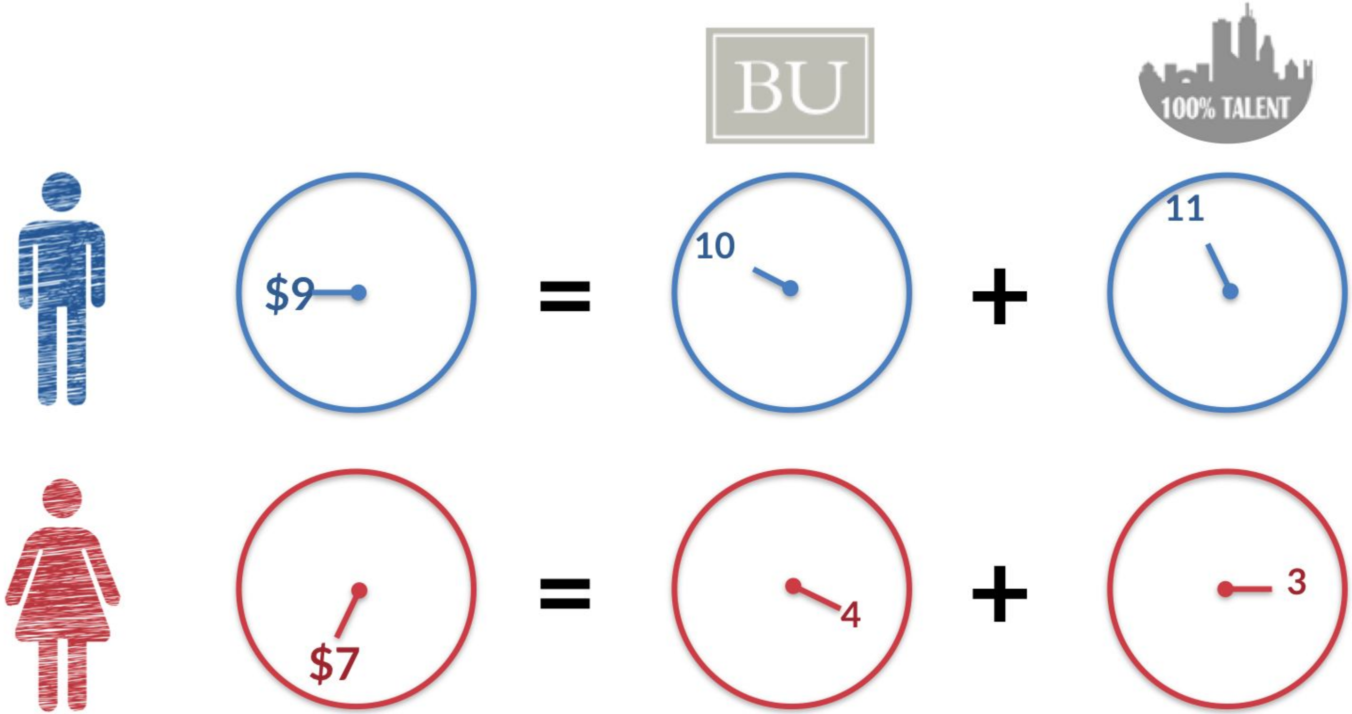
Unbound: Protect cryptographic keys

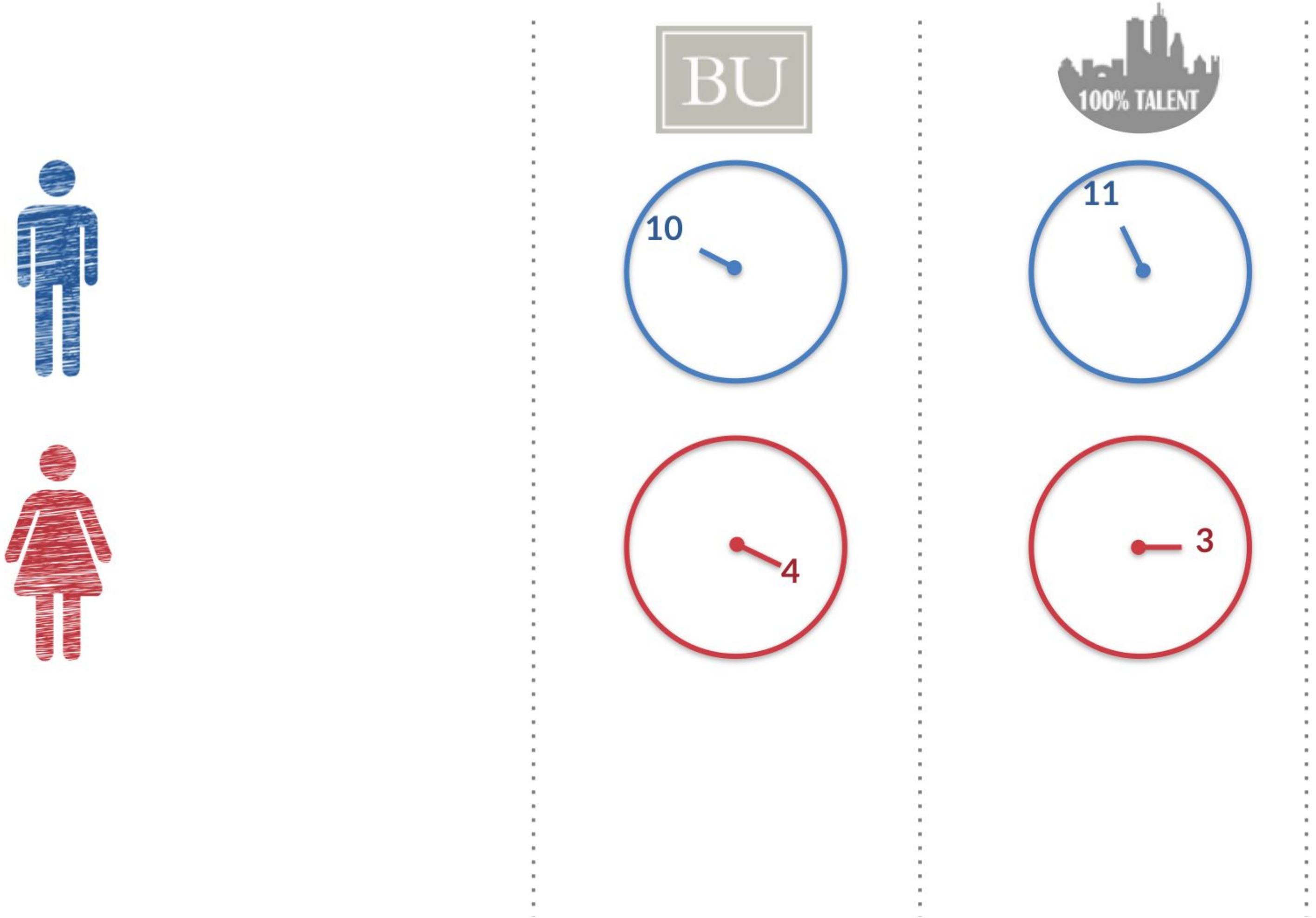
BU

100% TALENT



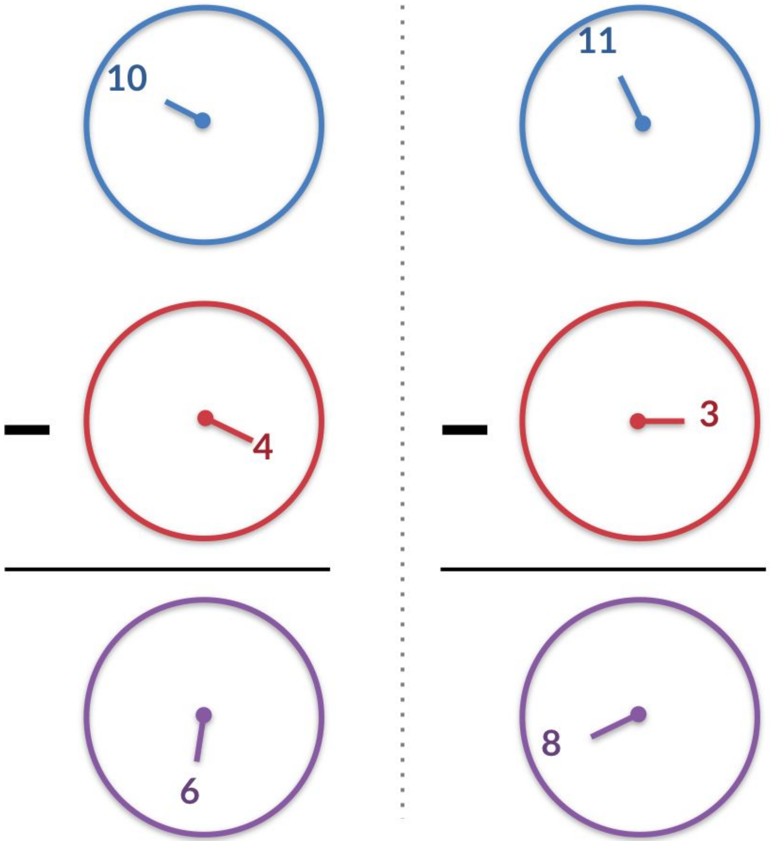
CITY of BOSTON


BU
100% TALENT
$9
=
10
+
11
$7
=
4
+
3

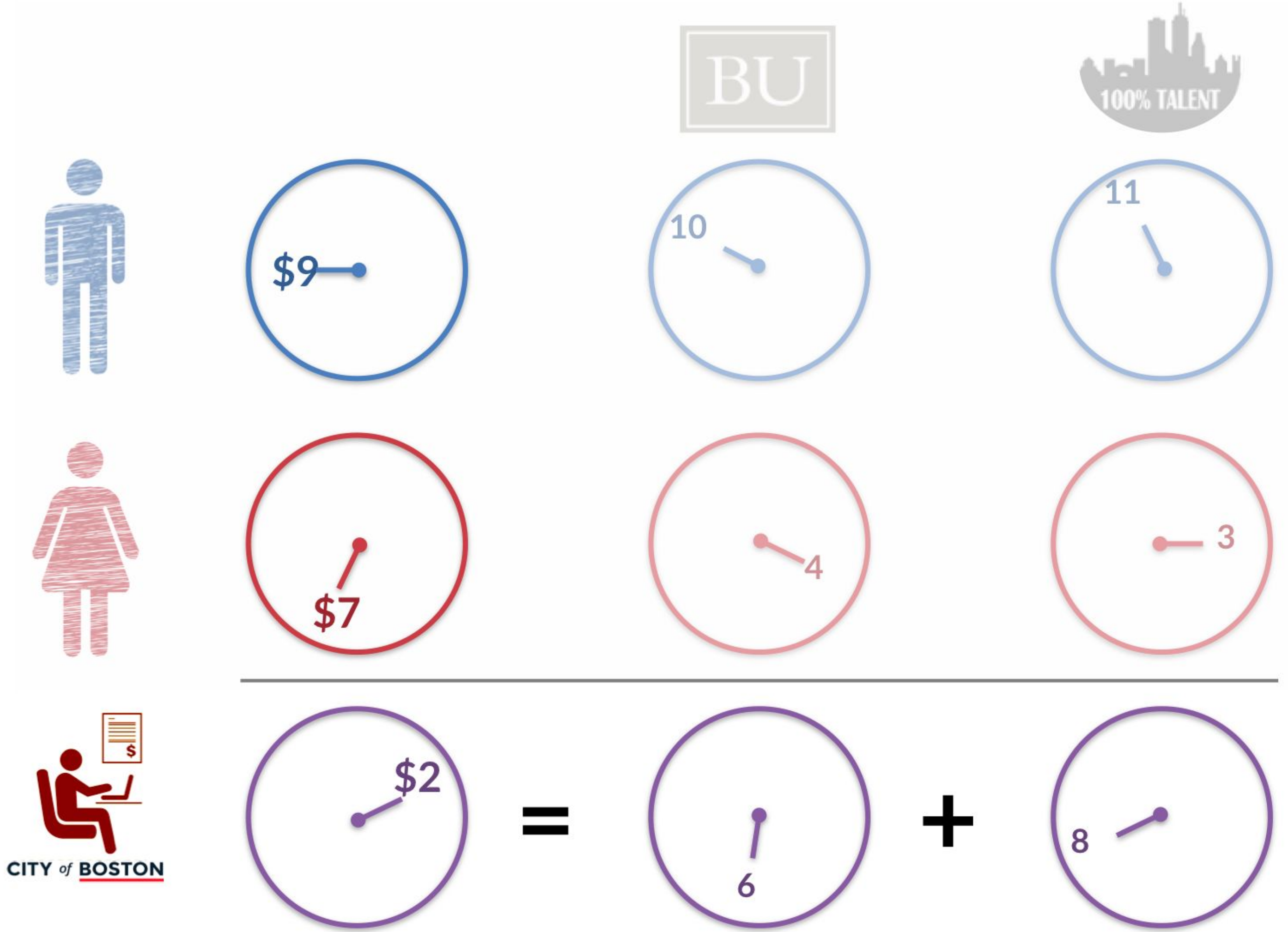

BU
100% TALENT
10
11
4
3

BU

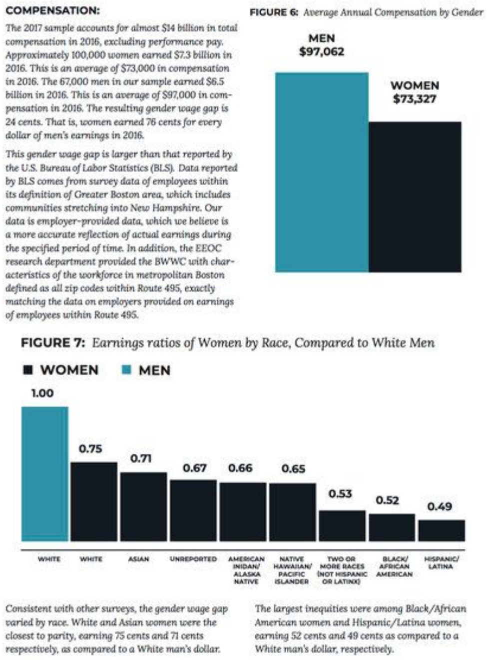
100% TALENT



CITY of BOSTON


BU
100% TALENT
$9
10
11
$7
4
3
CITY of BOSTON
$2
=
6
+
8

MPC at the scale of a city



COMPENSATION:

The 2017 sample accounts for almost $14 billion in total compensation in 2016, excluding performance pay. Approximately 100,000 women earned $7.3 billion in 2016. This is an average of $73,000 in compensation in 2016. The 67,000 men in our sample earned $6.5 billion in 2016. This is an average of $97,000 in compensation in 2016. The resulting gender wage gap is 24 cents. That is, women earned 76 cents for every dollar of men's earnings in 2016.

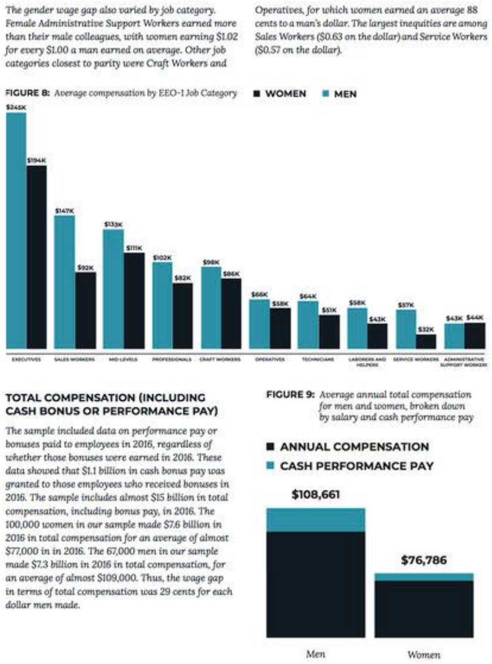
This gender wage gap is larger than that reported by the U.S. Bureau of Labor Statistics (BLS). Data reported by BLS comes from survey data of employees within its definition of Greater Boston area, which includes communities stretching into New Hampshire. Our data is employer-provided data, which we believe is a more accurate reflection of actual earnings during the specified period of time. In addition, the EEOC research department provided the BWWC with characteristics of the workforce in metropolitan Boston defined as all zip codes within Route 495, exactly matching the data on employers provided on earnings of employees within Route 495.

FIGURE 6: *Average Annual Compensation by Gender*

FIGURE 7: *Earnings ratios of Women by Race, Compared to White Men*

Consistent with other surveys, the gender wage gap varied by race. White and Asian women were the closest to parity, earning 75 cents and 71 cents respectively, as compared to a White man's dollar.

The largest inequities were among Black/African American women and Hispanic/Latina women, earning 52 cents and 49 cents as compared to a White man's dollar, respectively.



The gender wage gap also varied by job category. Female Administrative Support Workers earned more than their male colleagues, with women earning $1.02 for every $1.00 a man earned on average. Other job categories closest to parity were Craft Workers and Operatives, for which women earned an average 88 cents to a man's dollar. The largest inequities are among Sales Workers ($0.63 on the dollar) and Service Workers ($0.57 on the dollar).

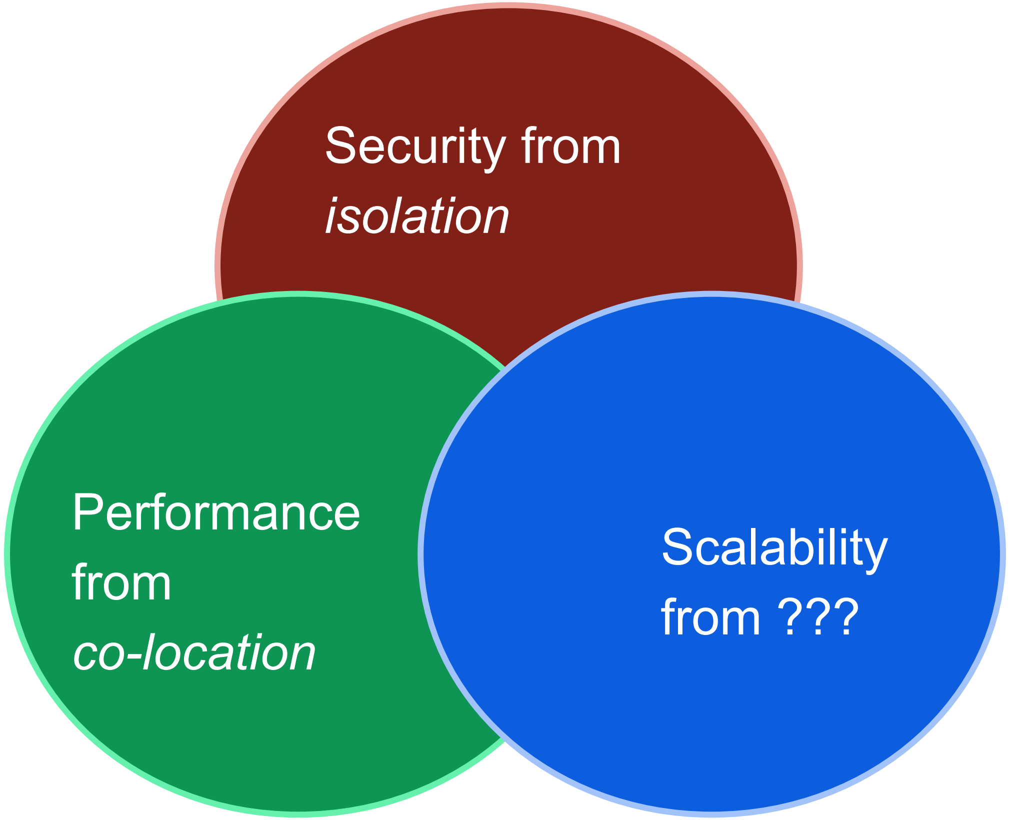
FIGURE 8: *Average compensation by EEO-1 Job Category*

TOTAL COMPENSATION (INCLUDING CASH BONUS OR PERFORMANCE PAY)

The sample included data on performance pay or bonuses paid to employees in 2016, regardless of whether those bonuses were earned in 2016. These data showed that $1.1 billion in cash bonus pay was granted to those employees who received bonuses in 2016. The sample includes almost $15 billion in total compensation, including bonus pay, in 2016. The 100,000 women in our sample made $7.6 billion in 2016 in total compensation for an average of almost $77,000 in in 2016. The 67,000 men in our sample made $7.3 billion in 2016 in total compensation, for an average of almost $109,000. Thus, the wage gap in terms of total compensation was 29 cents for each dollar men made.

FIGURE 9: *Average annual total compensation for men and women, broken down by salary and cash performance pay*

MPC gets...

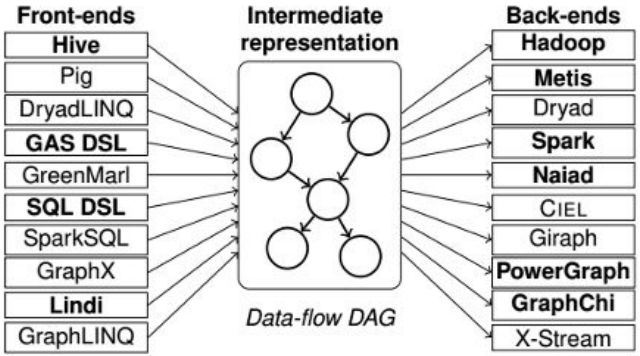

Security from *isolation*
Performance from *co-location*
Scalability from ???

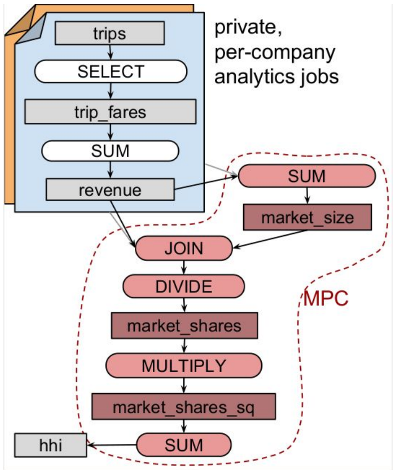
Conclave: Automating Secure Computation

SQL-like programming language
⇒ ***No MPC experience necessary***

Dispatcher executes jobs on available backends
⇒ ***No new infrastructure***

Compiler discerns boundaries of secure computing
⇒ ***No need for privacy experts***





Software available at github.com/multiparty/conclave
Paper available at arxiv.org/pdf/1902.06288.pdf

Conclave’s Query Specification

```
# state where the data lives
data = cc.defineTable(schema, at=[org-1, org-2, org-3])

# compute over the data as if it resided in one place
rev = data.project(["companyID", "price"])
          .sum("local_rev", group=[“companyID”], over="price")
          .project([0, "local_rev"])
market_size = rev.sum(“total_rev", over=“local_rev")
share = rev.join(market_size, left=[“companyID"],
                 right=[“companyID"])
           .divide("m_share", "local_rev", by="total_rev")
hhi = share.multiply(share, "ms_squared", "m_share")
           .sum(“hhi", on=“ms_squared”)
           .divide(“hhi", by=10k)
```

Conclave’s Static Analysis: Calculate Relations in the Clear

```
# state where the data lives
data = cc.defineTable(schema, at=[partyA, partyB, partyC])

# compute over the data as if it resided in one place
rev = data.project(["companyID", "price"])
          .sum("local_rev", group=[“companyID”], over="price")
          .project([0, "local_rev"])
market_size = rev.sum(“total_rev", over=“local_rev")
share = rev.join(market_size, left=[“companyID"],
                 right=[“companyID"])
           .divide("m_share", "local_rev", by="total_rev")
hhi = share.multiply(share, "ms_squared", "m_share")
           .sum(“hhi", on=“ms_squared”)
           .divide(“hhi", by=10k)
```

Integrating Conclave into the cloud

- Conclave runs in containers within each silo
- OpenShift/Kubernetes orchestrate the execution of Conclave jobs
- MPC jobs begin when data analysts make queries over aggregate data
- Benefit: improved performance of secure computing via co-location

How we arrived here

SAIL Software & Application Innovation Lab

- Both projects have common technology stack
- OpenShift
    - Isolation Techniques
    - Scaled job framework
    - Resource Management
        - CPU/Memory/Network/GPU

BU

Boston Children's Hospital

Red Hat

OPENSHIFT

openstack™

How MPC fit in Health Care ?

- When data is scarce
  - Boston Trauma Center
  - Understanding rare diseases
- Sharing patient data in the clear is restricted
  - Privacy laws
  - Hospital standard practices

Sharing Information w/o Sharing Data

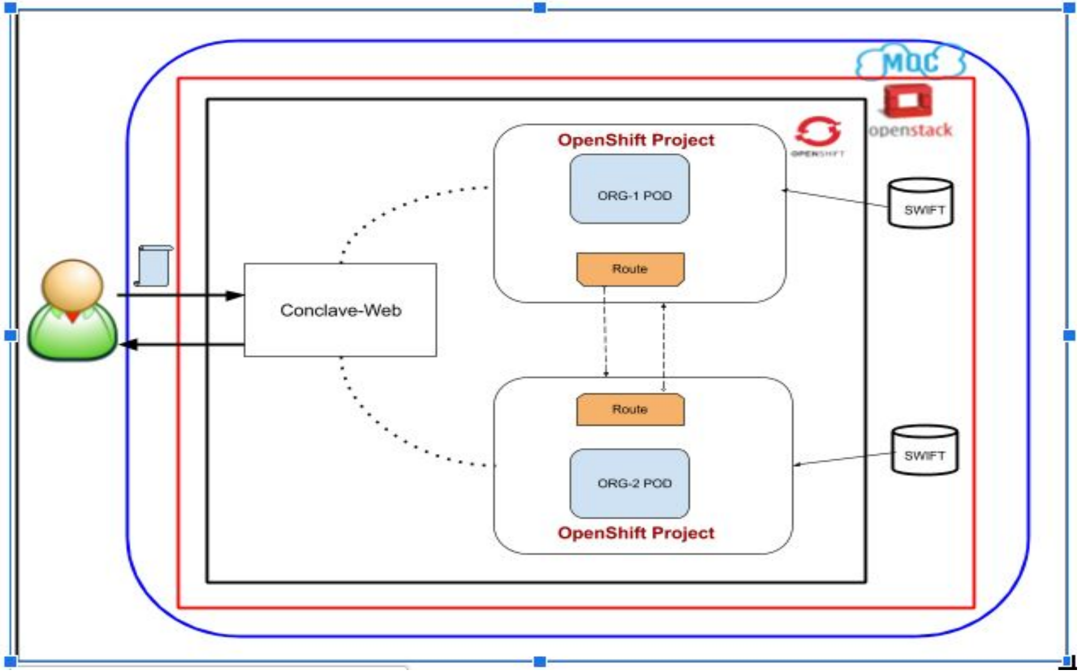
- Augment ChRIS with cryptographically secure Multi Party Computation
- OpenShift enabled **isolated computing environment**

Isolated Computing Environment

- What?
  - Run individual computations on pre configured secure nodes
  - Container segregation on a host level

Isolated Computing Environment in Cloud

- How?
  - Machine like virtualisation
    - Namespace
    - SeLinux
  - Project Isolation
  - Network Isolation



MPC Application in Health Care - Example

Analyse brain segment volume

Extrapolates the patient's brain volume against the population mean

Identify segments that have significant deviation from the population mean in terms of brain volume for the same age group

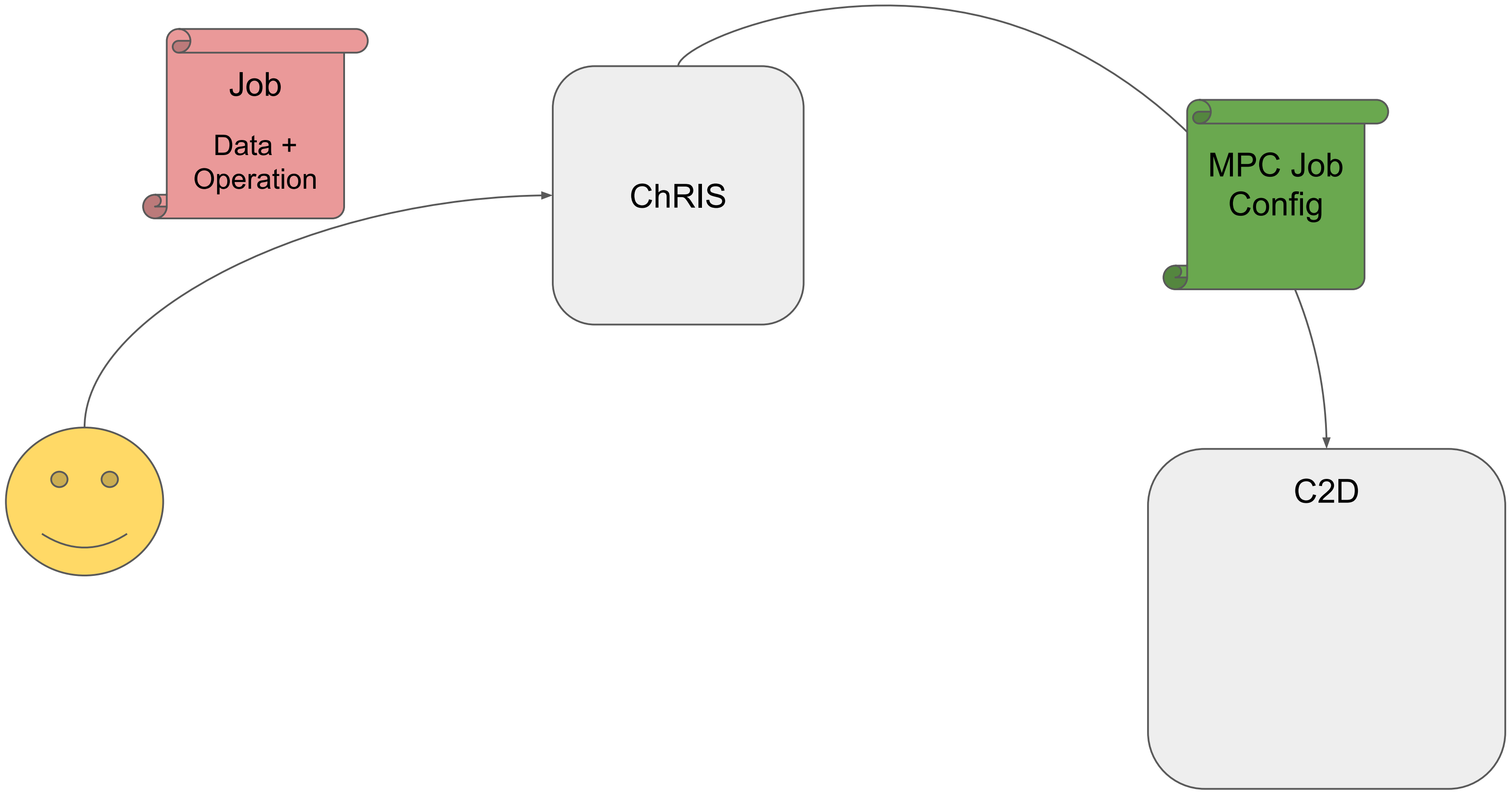
Medical Image Processor

MPC (C2D)

- Calculate population mean
- Calculate population standard deviation

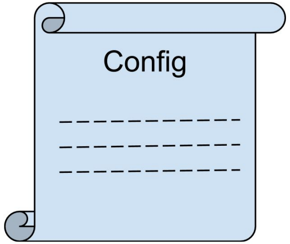
Non MPC (ChRIS)

- Project patient’s brain volume against the population mean
- Number of standard deviations from the mean the patient datapoint is


Job
Data +
Operation
ChRIS
MPC Job
Config
C2D

MPC job in C2D

- Calculate population mean
- Calculate population standard deviation



Base64 encoded linear algebra query

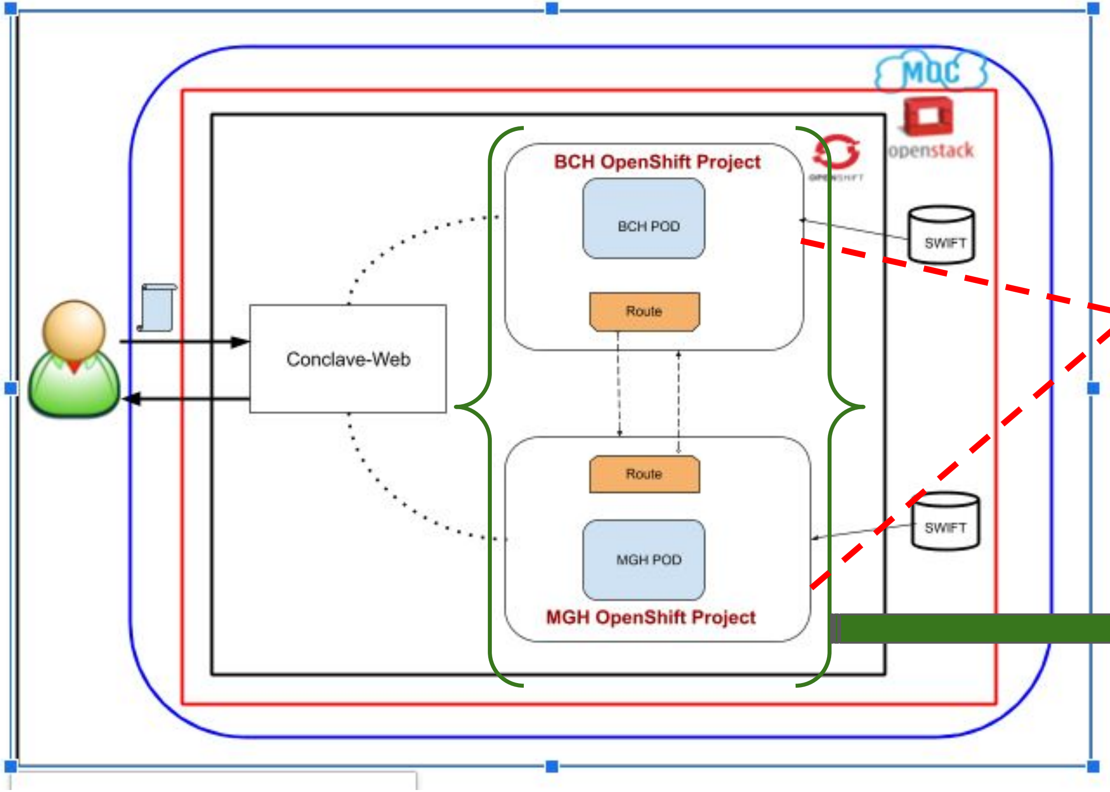
```
cols_in_a = [
    defCol('a', 'INTEGER', [1]),
    defCol('b', 'INTEGER', [1]),
    defCol('c', 'INTEGER', [1]),
]
cols_in_b = [
    defCol('a', 'INTEGER', [2]),
    defCol('b', 'INTEGER', [2]),
    defCol('c', 'INTEGER', [2]),
]

in1 = cc.create("in1", cols_in_a, {1})
in2 = cc.create("in2", cols_in_b, {2})

cc1 = cc.concat([in1, in2], 'cc1', ['a', 'b', 'c'])

agg1 = cc.aggregate(cc1, "agg1", ['a'], "b", "mean", "b")

cc.collect(agg1, 1)

return {in1, in2}
```

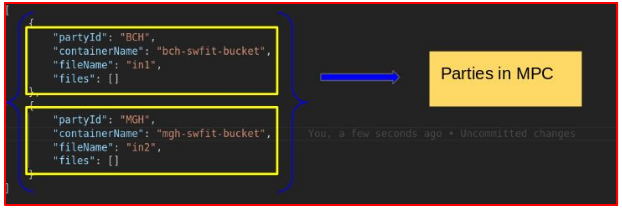

```
{
  "protocol": {
      "data": "aW1wb3J0IGNvbmNsYXZlLmxhbmcgYXMgY2MKZnJvbSBjb25jbGF2ZS51dGlscyBpbXBvcnQgZGVmQ29sCmZyb20gY29uY2xhdmUgaW1wb3J0IH
      "format": "b64"
  },
  "config":{
      "ID": "brain-volume-density",
      "backend": "swift"
  },
  "swift": {
      "endpoints":
      [
        {
          "partyId": "BCH",
          "containerName": "bch-swfit-bucket",
          "fileName": "in1",
          "files": []
        },
        {
          "partyId": "MGH",
          "containerName": "mgh-swfit-bucket",
          "fileName": "in2",
          "files": []
        }
      ]
  }
}
```


You, a few seconds ago • Uncommitted changes

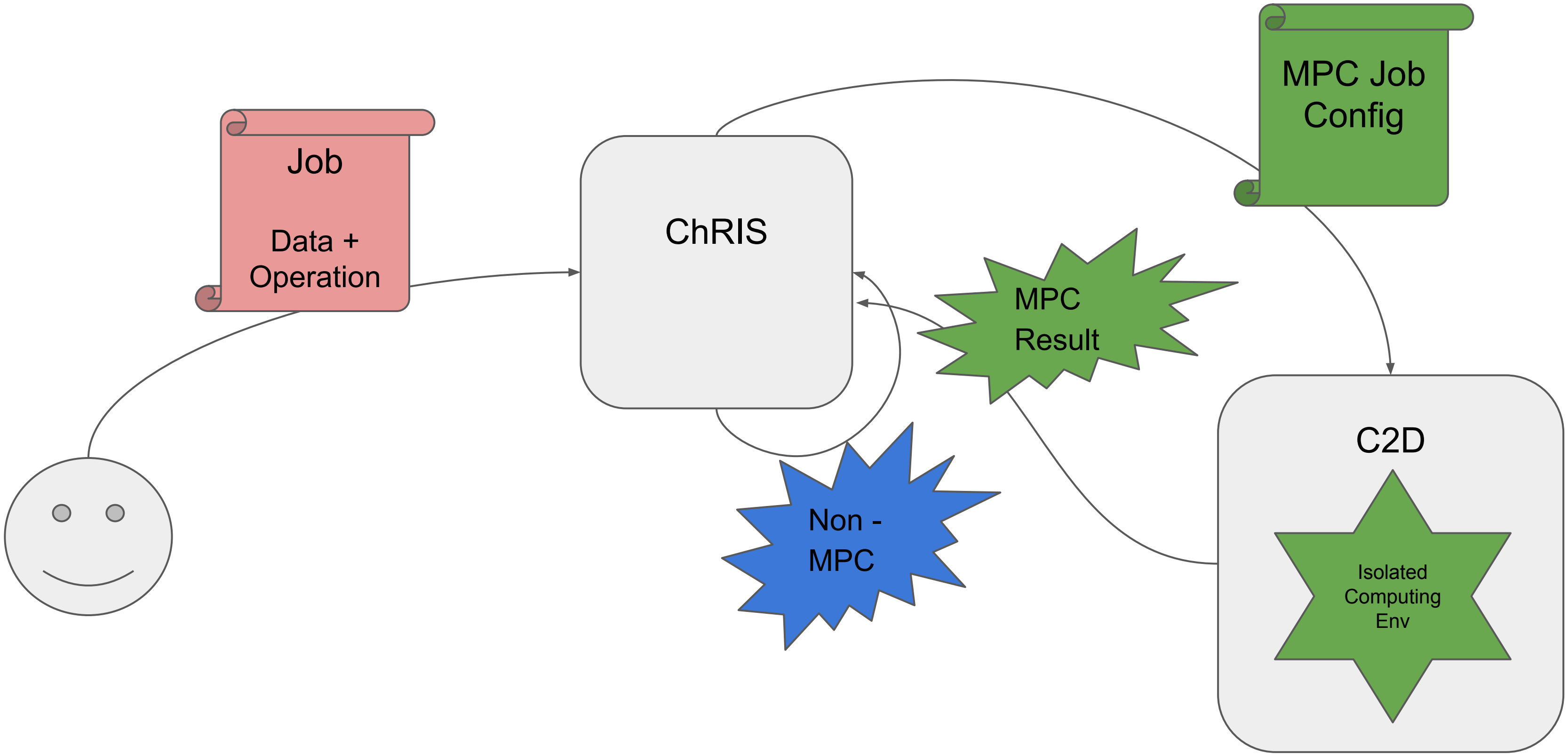
Parties in MPC

MPC job in C2D



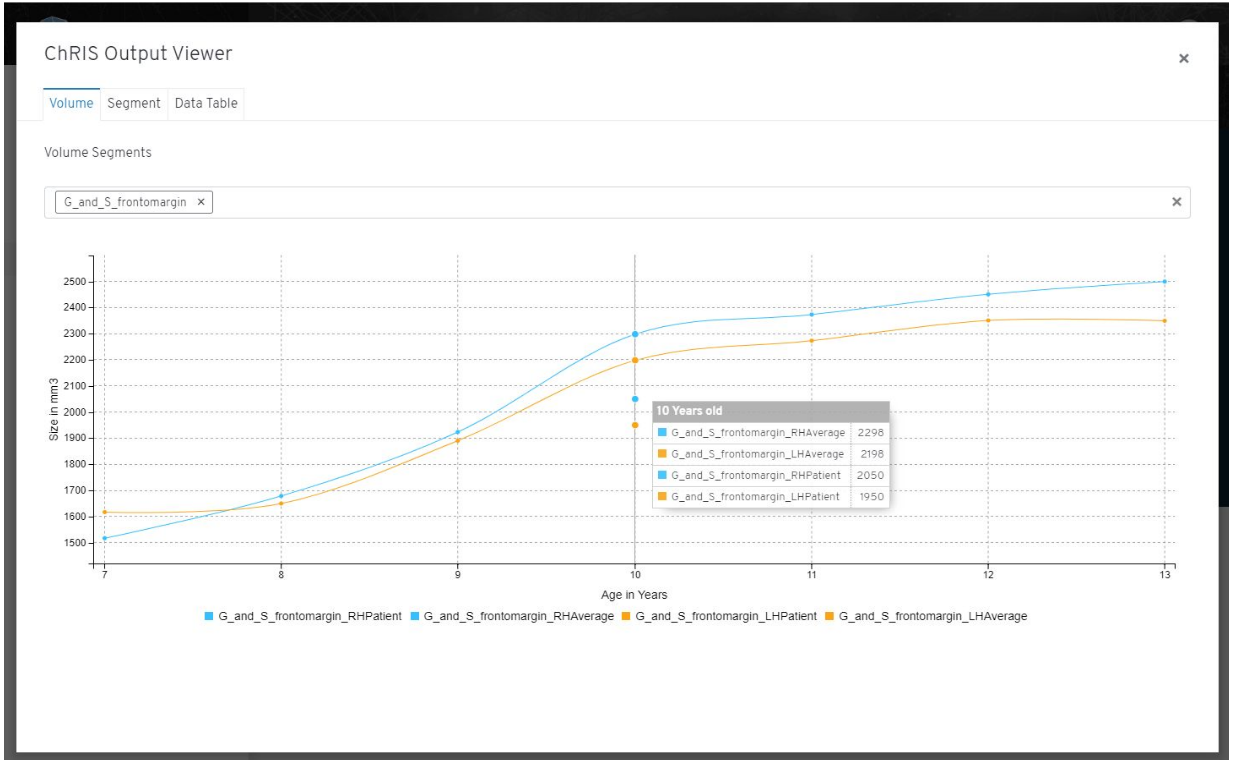


OpenShift enabled isolated computing environment.


Job
Data + Operation
ChRIS
MPC Job Config
MPC Result
Non - MPC
C2D
Isolated Computing Env

Non MPC job in ChRIS

- Extrapolates the patient's brain volume against the population mean of different age group

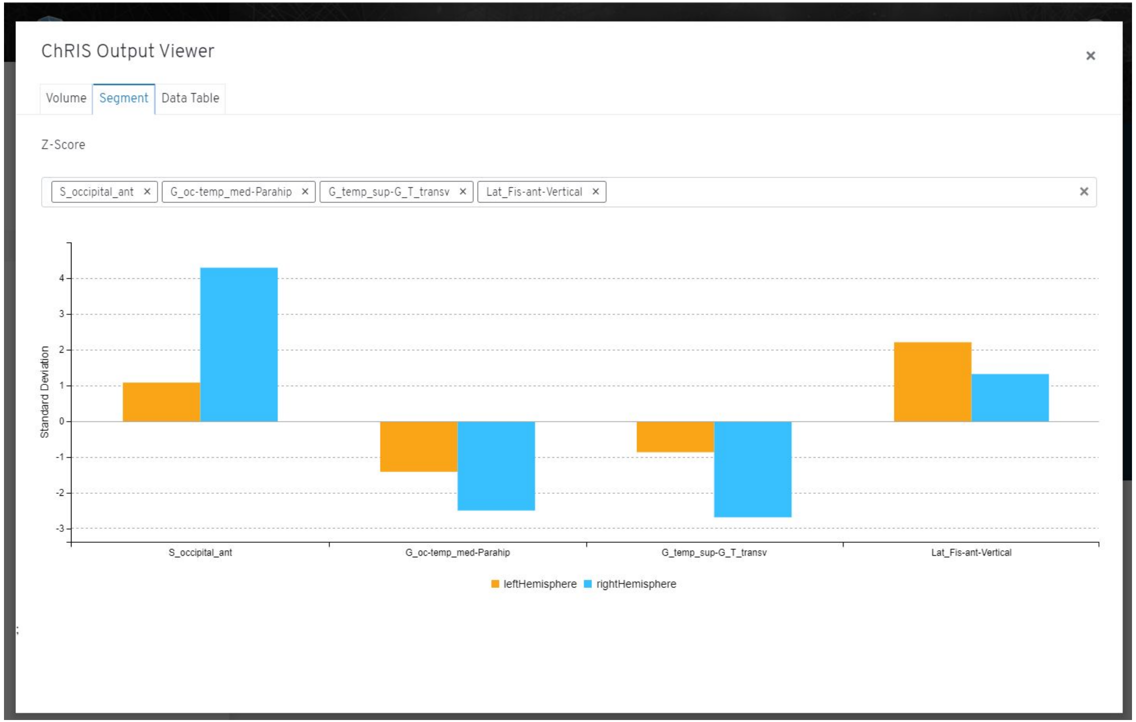


Non MPC job in ChRIS

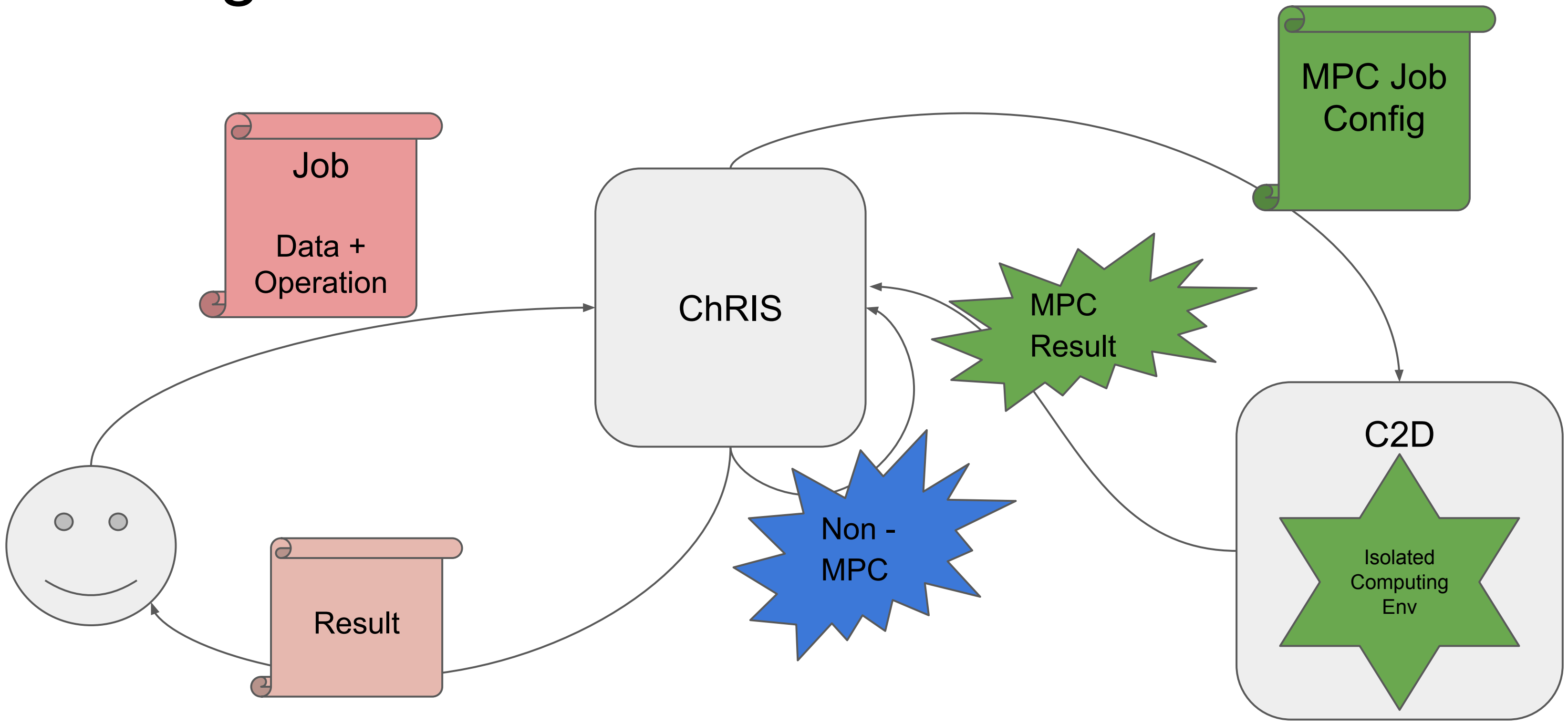
- Number of standard deviations from the mean a datapoint is

z score $z = (x - \mu) / \sigma$

μ : Population Mean
σ : Standard Deviation
x : Data value/ Score

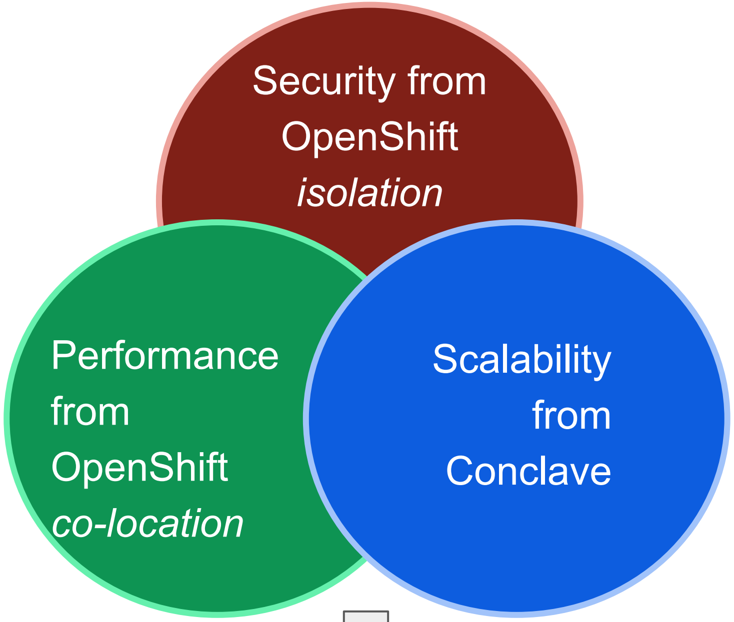


The Integration of ChRIS & C2D



The road ahead

- Collaboration when data is scarce
- MPC can make a huge impact in the medical landscape
- OpenShift provides the necessary features to build secure ecosystem for this collaboration
- MPC + OpenShift together can scale to big data



Questions?