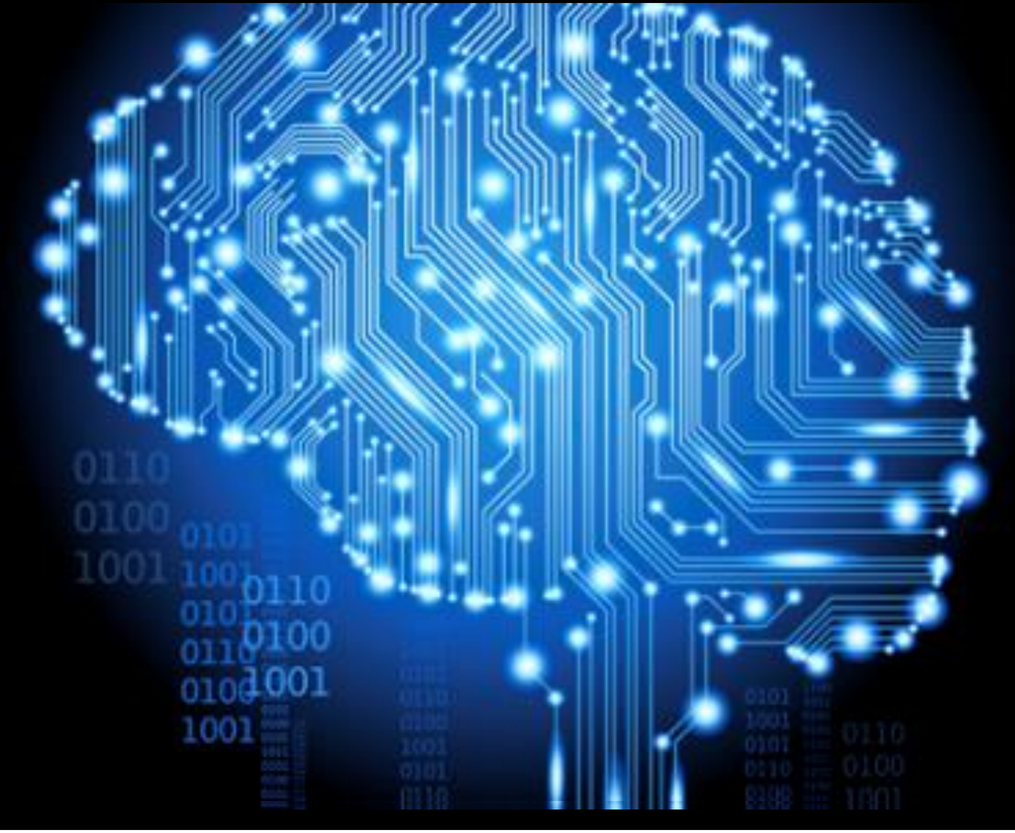
THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS



The Download: Community Tech Talks Episode 2

February 16, 2017

HPCC SYSTEMS®

RELX Group

Welcome!

- Please share: Let others know you are here with #HPCCTechTalks
- Ask questions! We will answer as many questions as we can following each speaker.
- Look for polls at the bottom of your screen. Exit full-screen mode or refresh your screen if you don't see them.
- We welcome your feedback - please rate us before you leave today and visit our blog for information after the event.
- Want to be one of our featured speakers? Let us know! techtalks@hpccsystems.com

Today’s Speakers

Dr. Flavio Villanustre

VP Technology, LexisNexis® Risk Solutions

Flavio.Villanustre@lexisnexis.com

Dr. Flavio Villanustre leads HPCC Systems®, and is also VP, Technology for LexisNexis Risk Solutions. In this position, he is responsible for Information and Physical Security, overall HPCC Systems® platform strategy and new product development.

Flavio has been involved with the open source community for over 15 years through multiple initiatives. Some of these include founding the first Linux User Group in Buenos Aires (BALUG) in 1994, releasing several pieces of software under different open source licenses, and evangelizing open source to different audiences through conferences, training and education. Prior to Flavio’s technology career, he was a neurosurgeon.

Fujio Turner

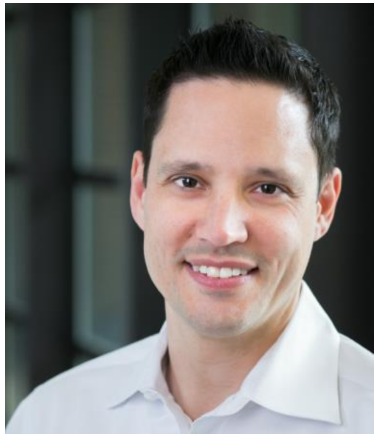
Solutions Architect, Couchbase

mail@fuj.io

Fujio Turner is a Couchbase Solutions Architect for Mobile and he specializes in high-speed data platforms. He began his IT career as a LAMP stack developer and soon became a MySQL developer and DBA. His attention turned to the high availability NoSQL systems of CouchDB/Couchbase in 2010.

With his personal philosophy, “In the future, there will be more data, not less,” HPCC Systems was a perfect fit for him. In his spare time, Fujio evangelizes HPCC Systems in the Silicon Valley area with the Meetup group, “Exabyte Big Data – HPCC Systems – Silicon Valley.” His list of current and future projects include 3DJSON and Virtual Reality and Big Data.

Today's Speakers



Jacob Pellock

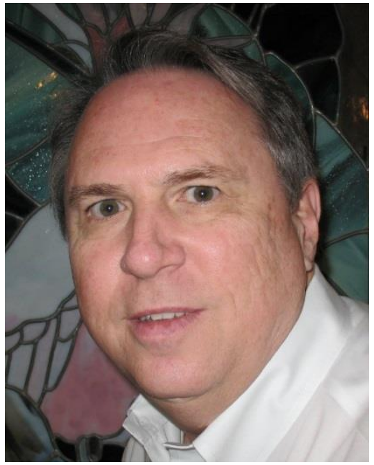
Sr Director Software Engineering,
LexisNexis® Risk Solutions
jacob.pellock@lexisnexisrisk.com

Jacob Pellock is a Sr. Director with LexisNexis Risk Solutions where he is responsible for supporting cross departmental Business Intelligence. He has been working at LexisNexis for 14 years building solutions to support analytics across multiple industries. He is particularly specialized in utilizing Big Data capabilities to support analysis and deployment of analytics capabilities into end user and system workflows.

Roger Dev

Sr Architect,
LexisNexis® Risk Solutions
roger.dev@lexisnexisrisk.com

Roger is a Senior Architect working with John Holt on the Machine Learning Team. He recently joined HPCC Systems from CA Technologies. Roger has been involved in the implementation and utilization of machine learning and AI techniques for many years, and he has over 20 patents in diverse areas of software technology.



Richard Taylor

Chief Trainer, HPCC Systems
LexisNexis® Risk Solutions
richard.taylor@lexisnexisrisk.com

Richard Taylor has worked with the HPCC Systems technology platform and the ECL programming language for over 15 years. He is the original author of the ECL documentation, developer and designer of the HPCC Systems Training Courses, and is the Chief Instructor for all classroom and remote based training.

THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS

HPCC SYSTEMS®

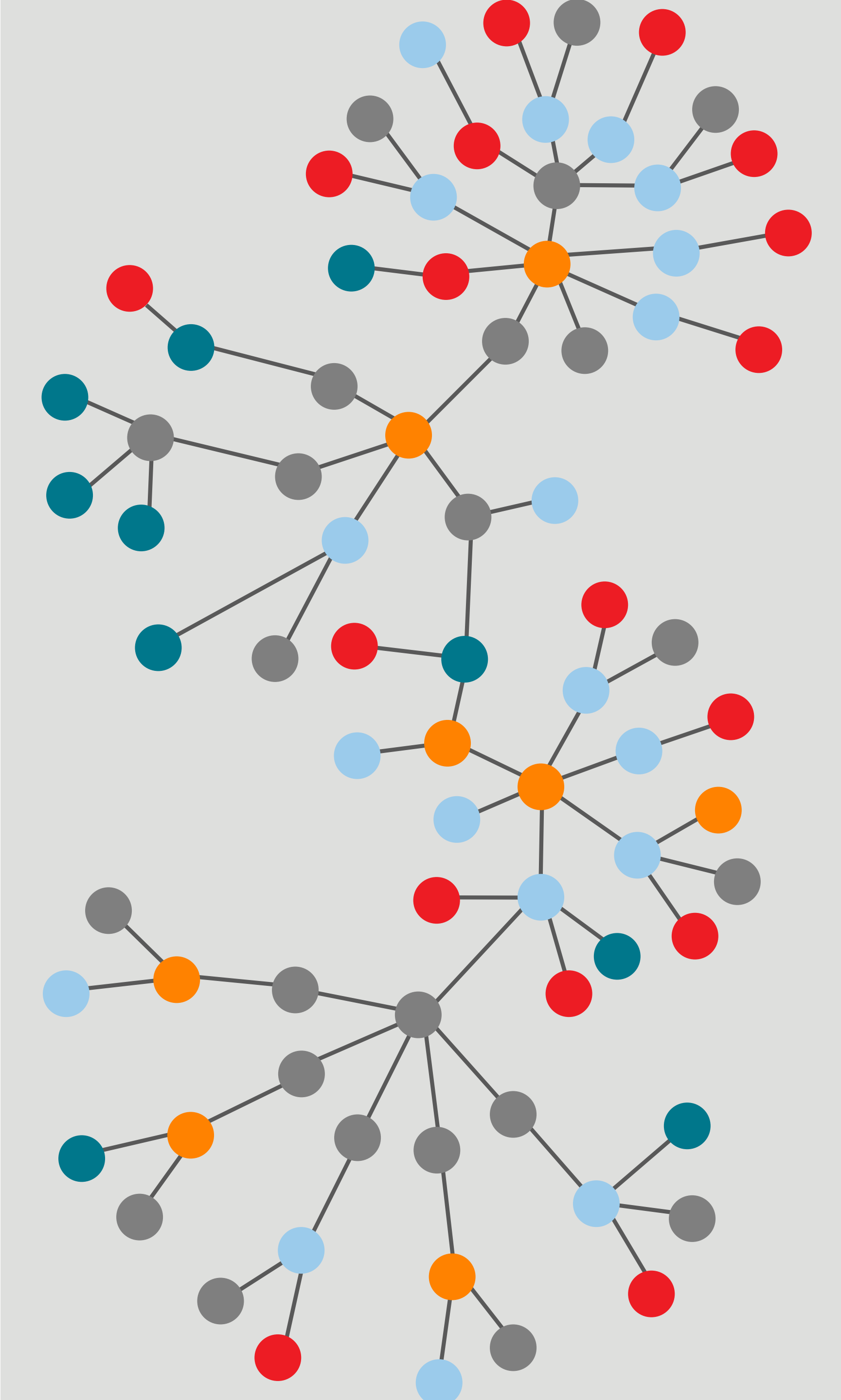
Mobile/IoT & HPCC Systems

Fujio Turner
Solutions Architect, Couchbase

Couchbase

Quick poll: Do you see mobile data being a bigger part of your business?

See poll on bottom of presentation screen



HPCC SYSTEMS®

Why Should We Care

MOBILE

- User Primary Information Hub
- User Primary Social Hub
- User Secondary Purchasing Hub

IOT

Better Predict Systems Behaviors

Micro / Macro

- Safety
- Automation
- Customer View

Problems

- Syncing Data
- Speed Data
- In-Accurate/Stale Data

- Collection
- Transport
- Storage
- Analyze

Why Should We Care

MOBILE

- User Primary Information Hub
- User Primary Social Hub
- User Secondary Purchasing Hub

IOT

Better Predict Systems Behaviors

Micro / Macro

- Safety
- Automation
- Customer View

Problems

- Syncing Data
- Speed Data
- In-Accurate/Stale Data

- Collection
- Transport
- Storage
- Analyze

What People are Saying and Asking

“My data is scattered.”

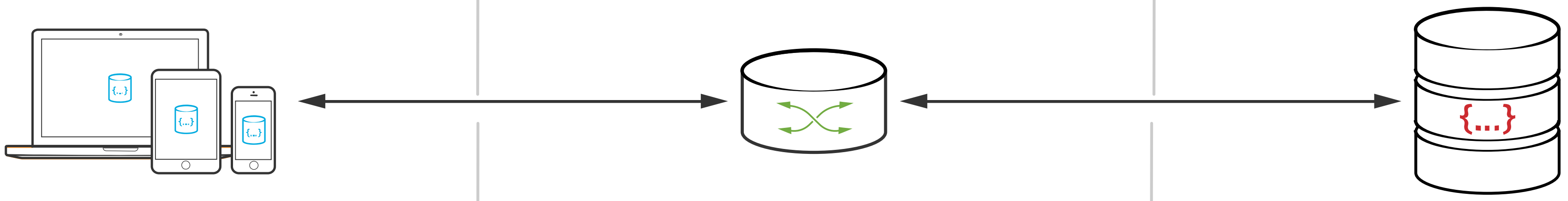
“Personalization / Customization.”

“My data needs to do more for me.”

“Move faster , More Agile with Data”

- Consolidate Data
- HPCC Systems Machine Learning
- Easy Analytics
- Couchbase / HPCC Systems

Couchbase Mobile: The Complete Mobile Database Solution



Couchbase Lite

Sync Gateway

Couchbase Server

EMBEDDED DATABASE

SYNCHRONIZATION

DATABASE SERVER

Lightweight Local NoSQL Database

- *CRUD*
- *Query Functionality*

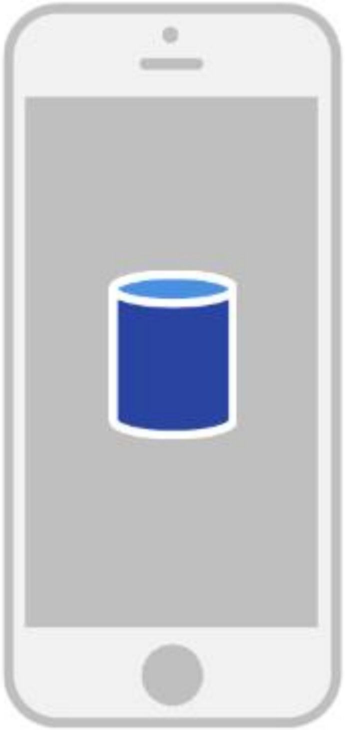
Secure Web Gateway

- *REST*
- *Sync*
- *Stream Batch*
- *Event APIs*

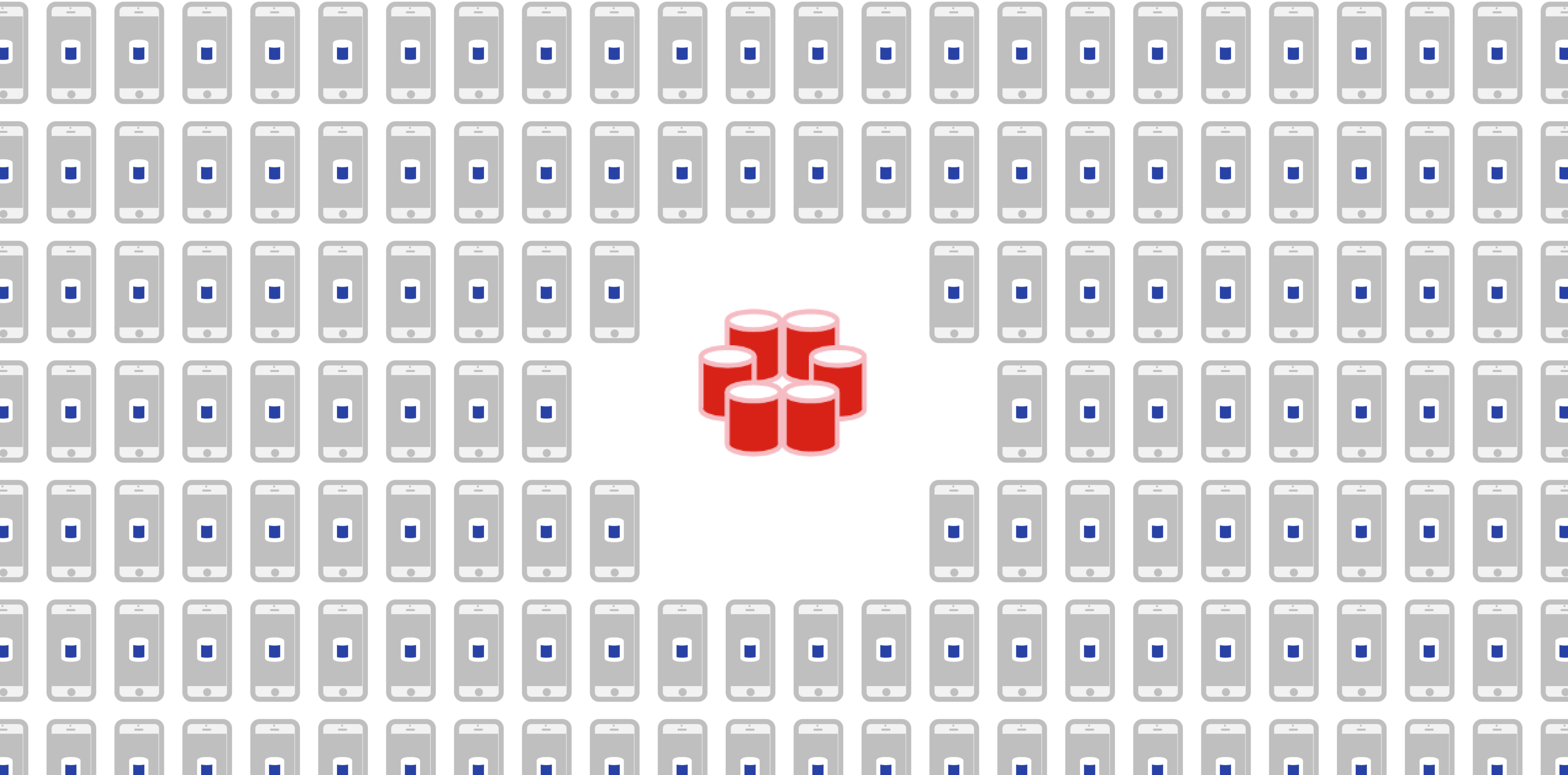
NoSQL Database

- *Highly Scalable*
- *Highly Available*
- *High Performance*
- *Key/Value & SQL++*

Syncing a Single User's Database is *EASY*

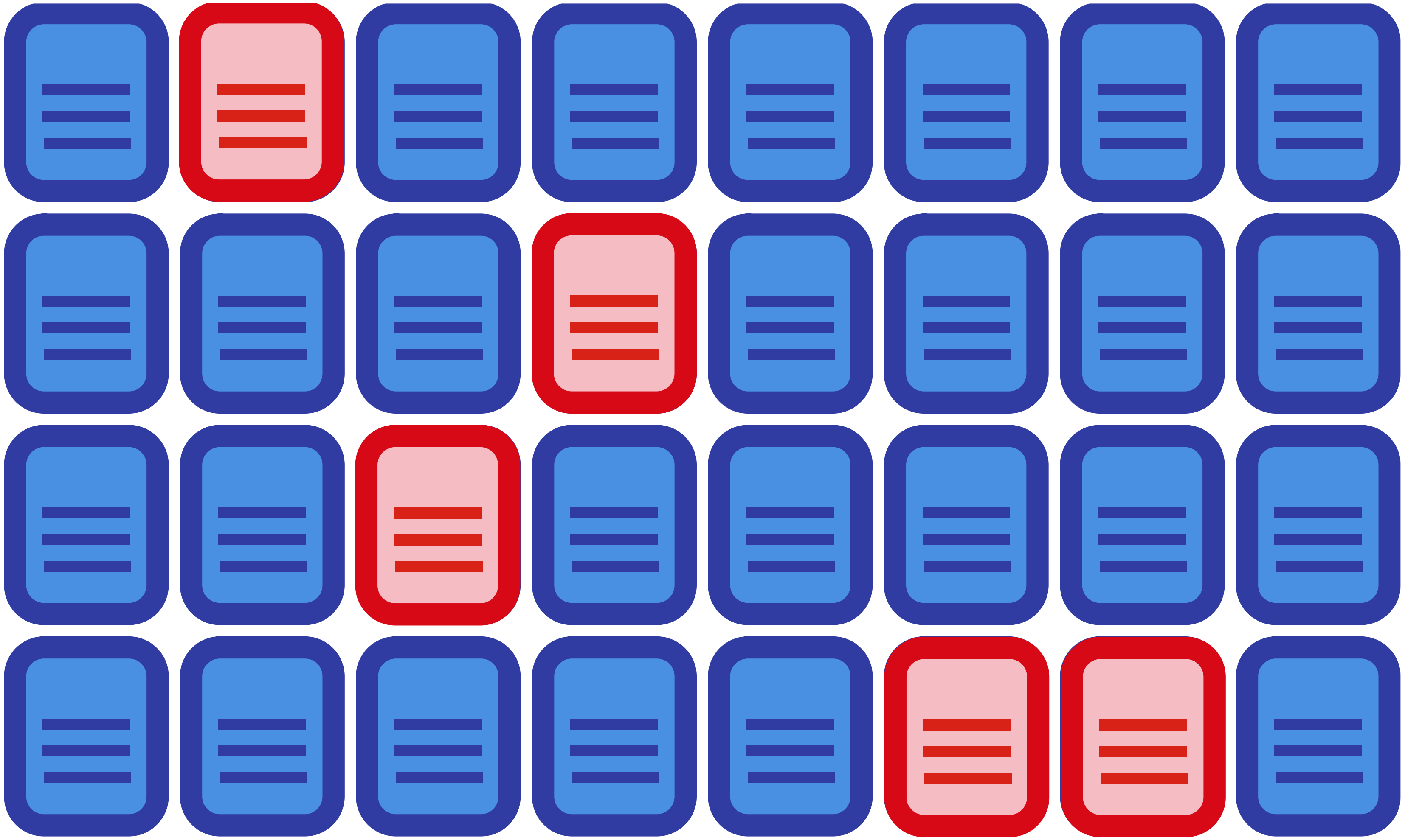


What About Syncing x,000,000?



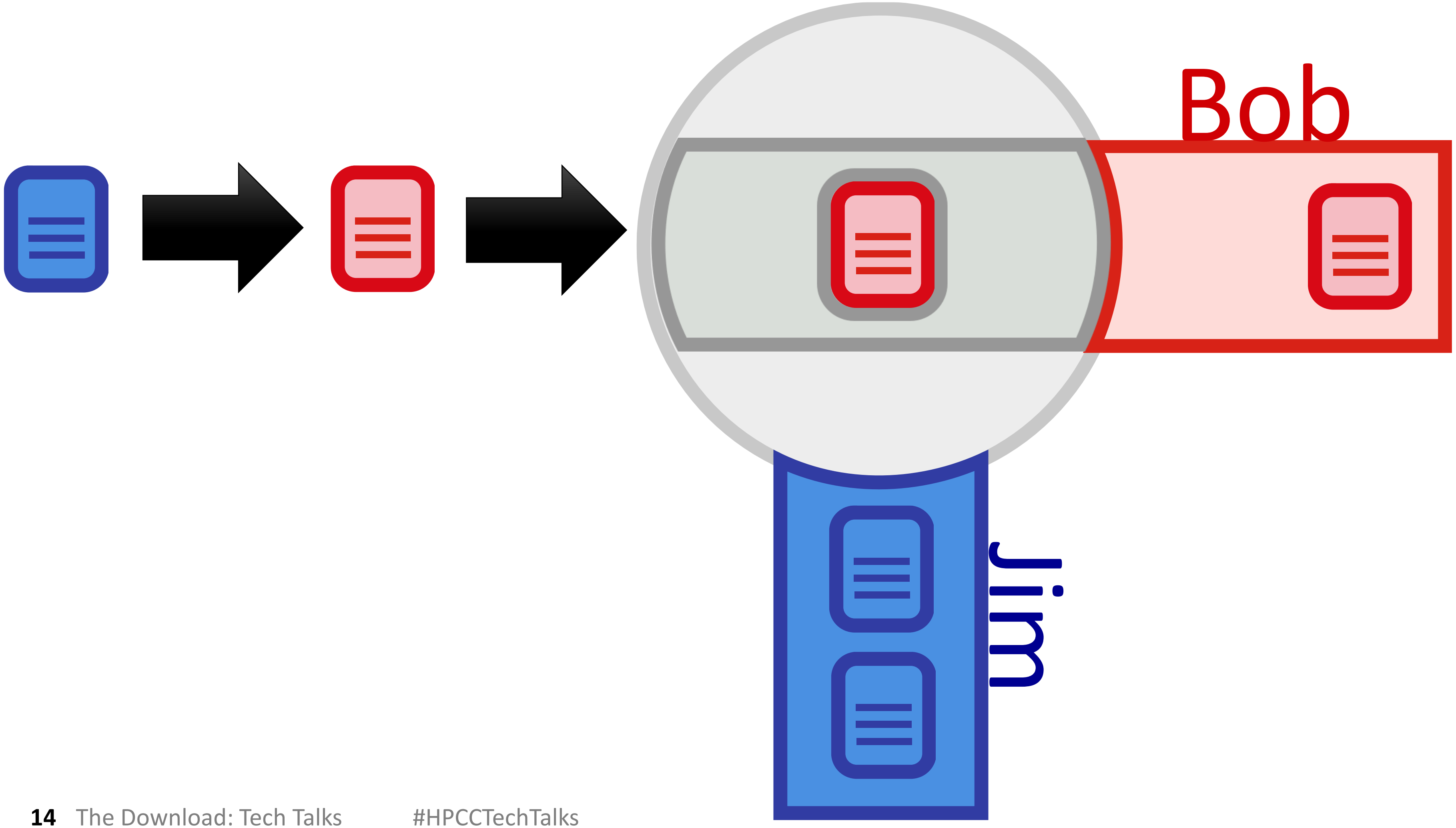
Read-side Filtering

```
doc.owner == 'bob'
```

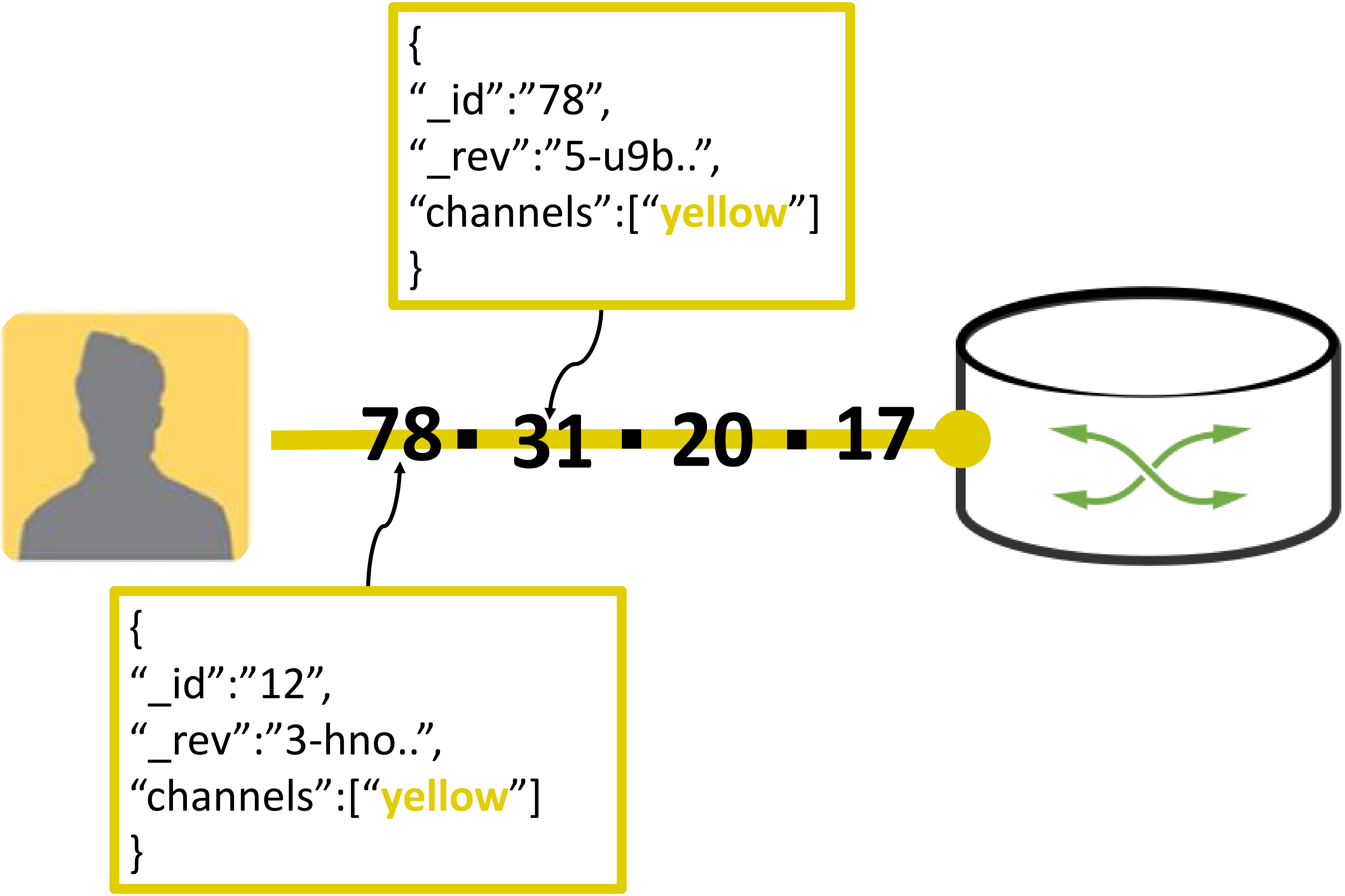


Partition Data During Writes Instead

`channel(doc.owner)`



Sync Gateway & Channels – Pull Feed

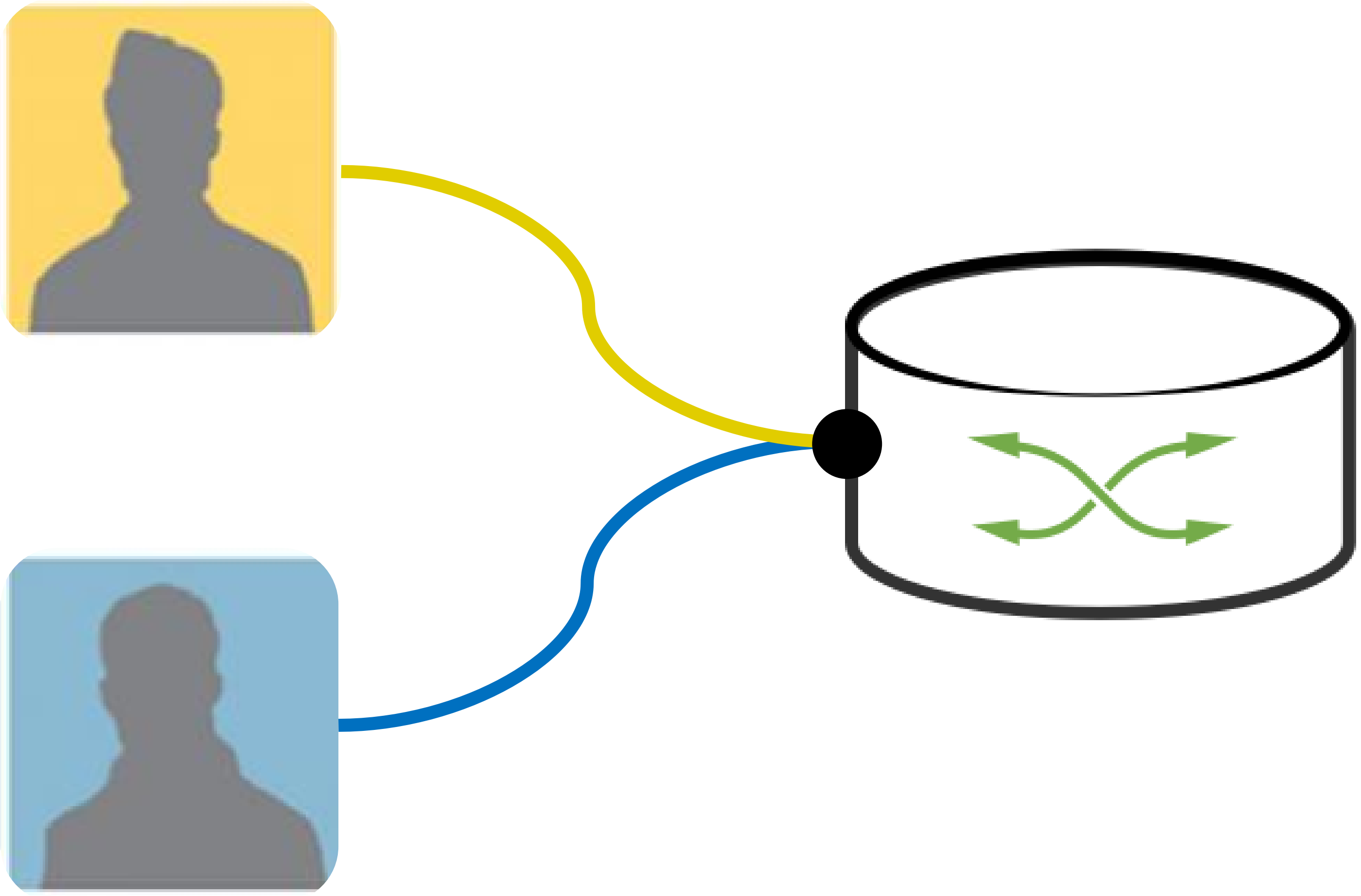


```
{
“_id”:”78”,
“_rev”:”5-u9b..”,
“channels”:[“yellow”]
}
```

```
{
“_id”:”12”,
“_rev”:”3-hno..”,
“channels”:[“yellow”]
}
```

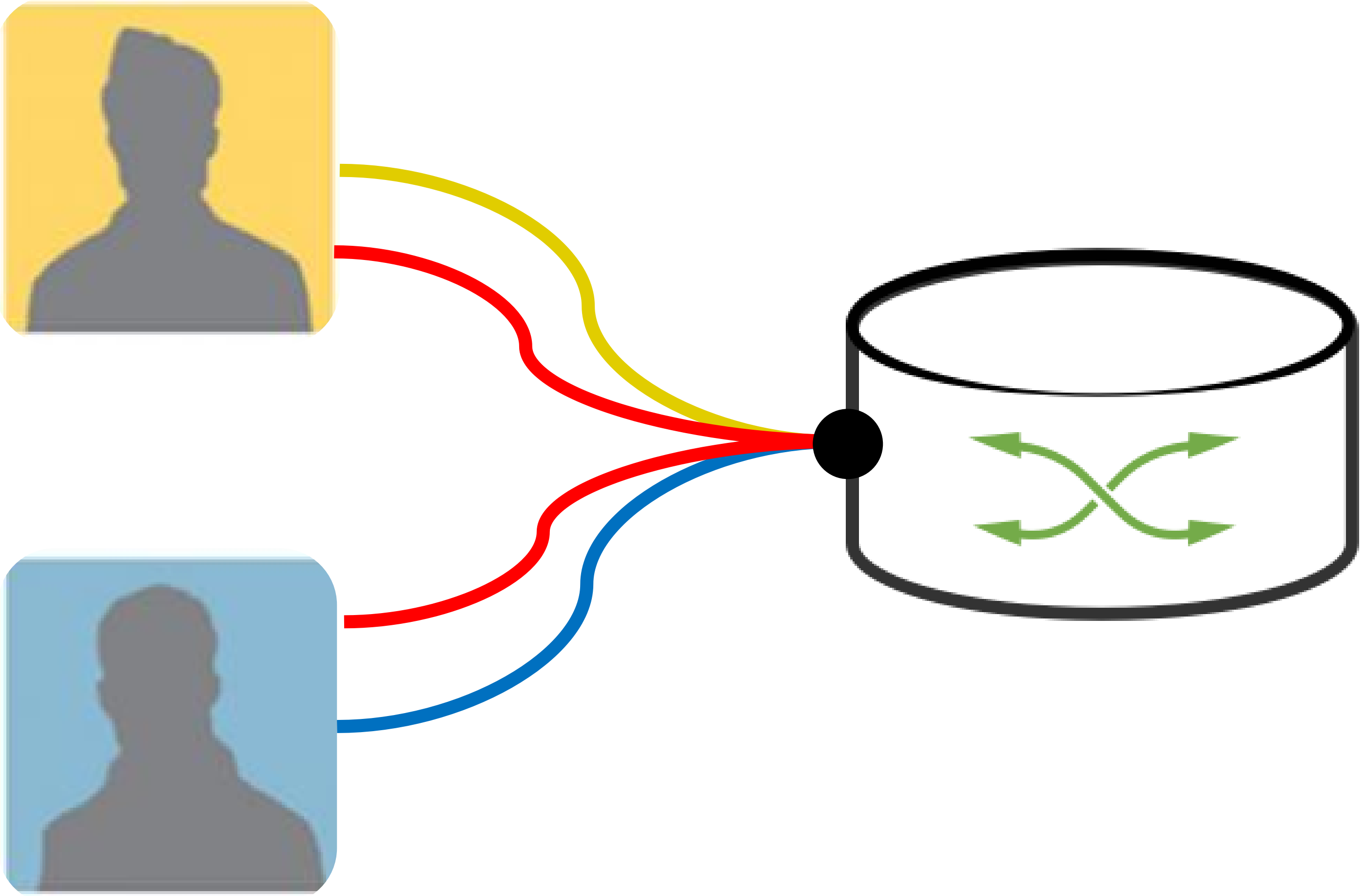
Sync Gateway & Channels

- Private



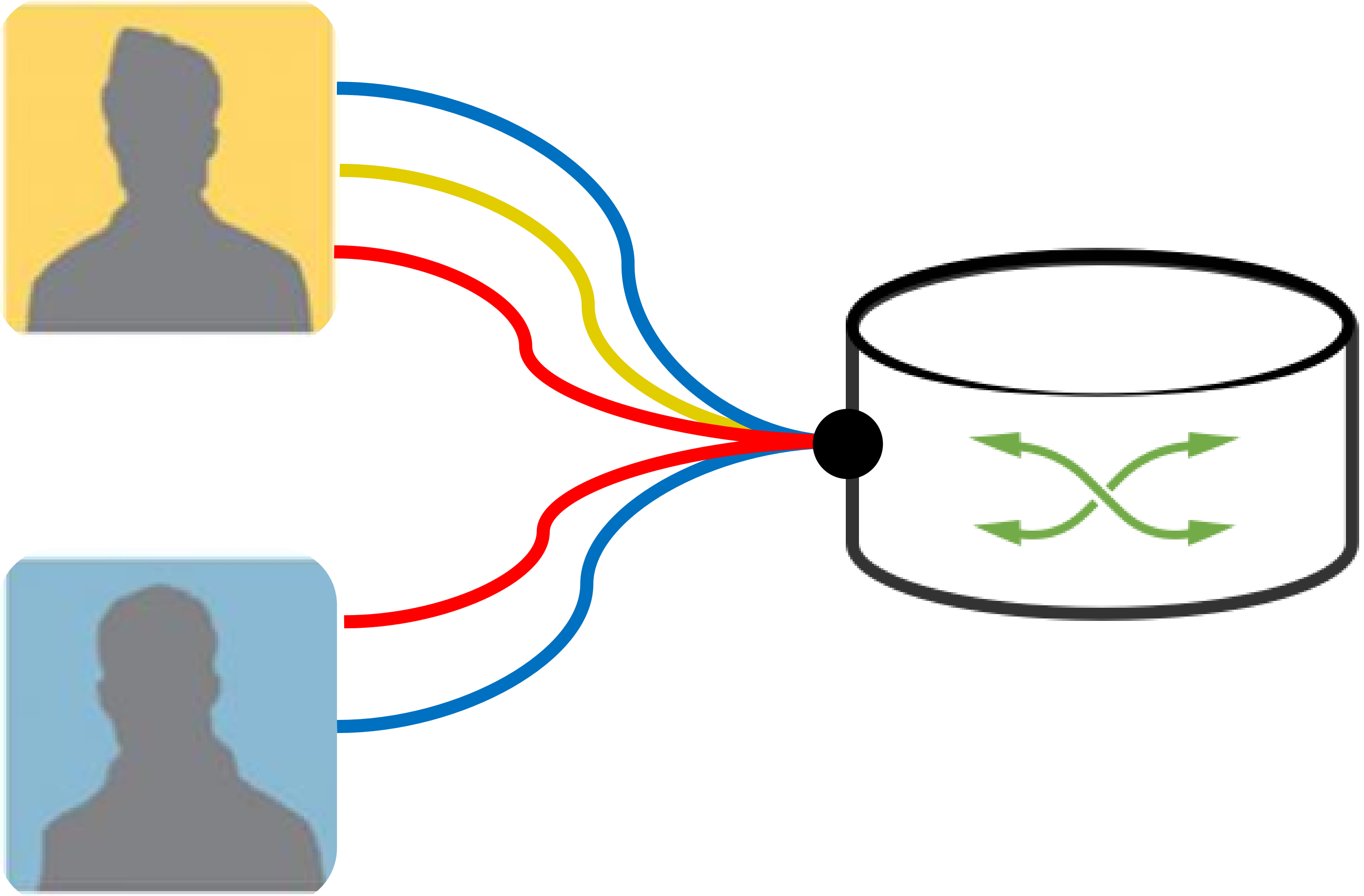
Sync Gateway & Channels

- Private
- Public / Group

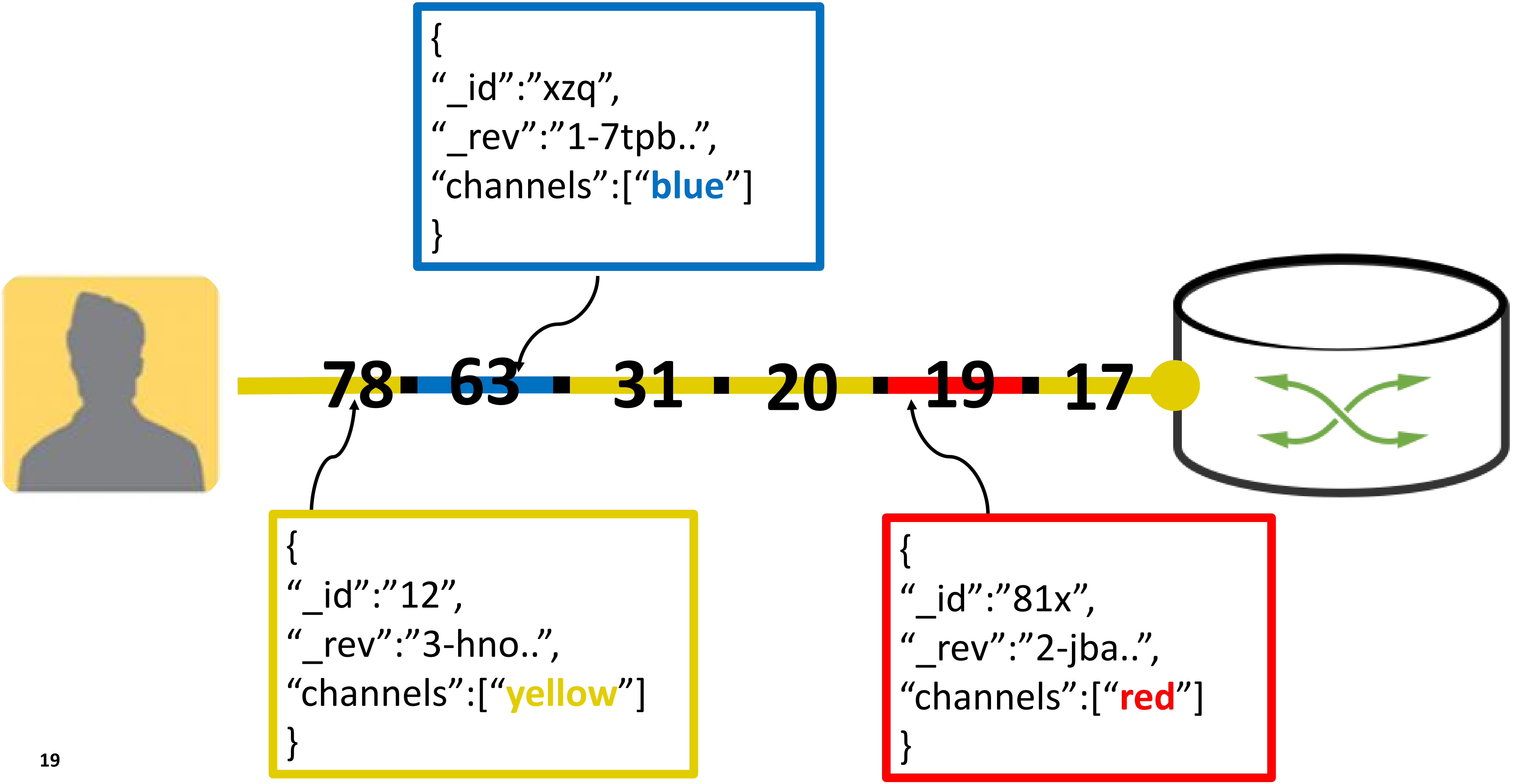


Sync Gateway & Channels

- Private
- Public / Group
- Share Private



Sync Gateway & Channels – Multi-channel Feed

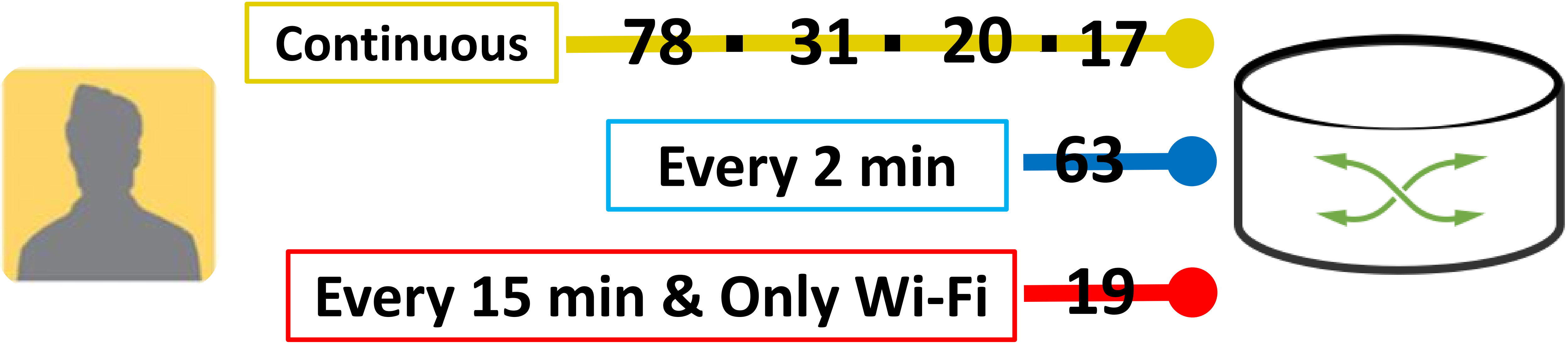


```
{
“_id”:”xzq”,
“_rev”:”1-7tpb..”,
“channels”:[“blue”]
}
```

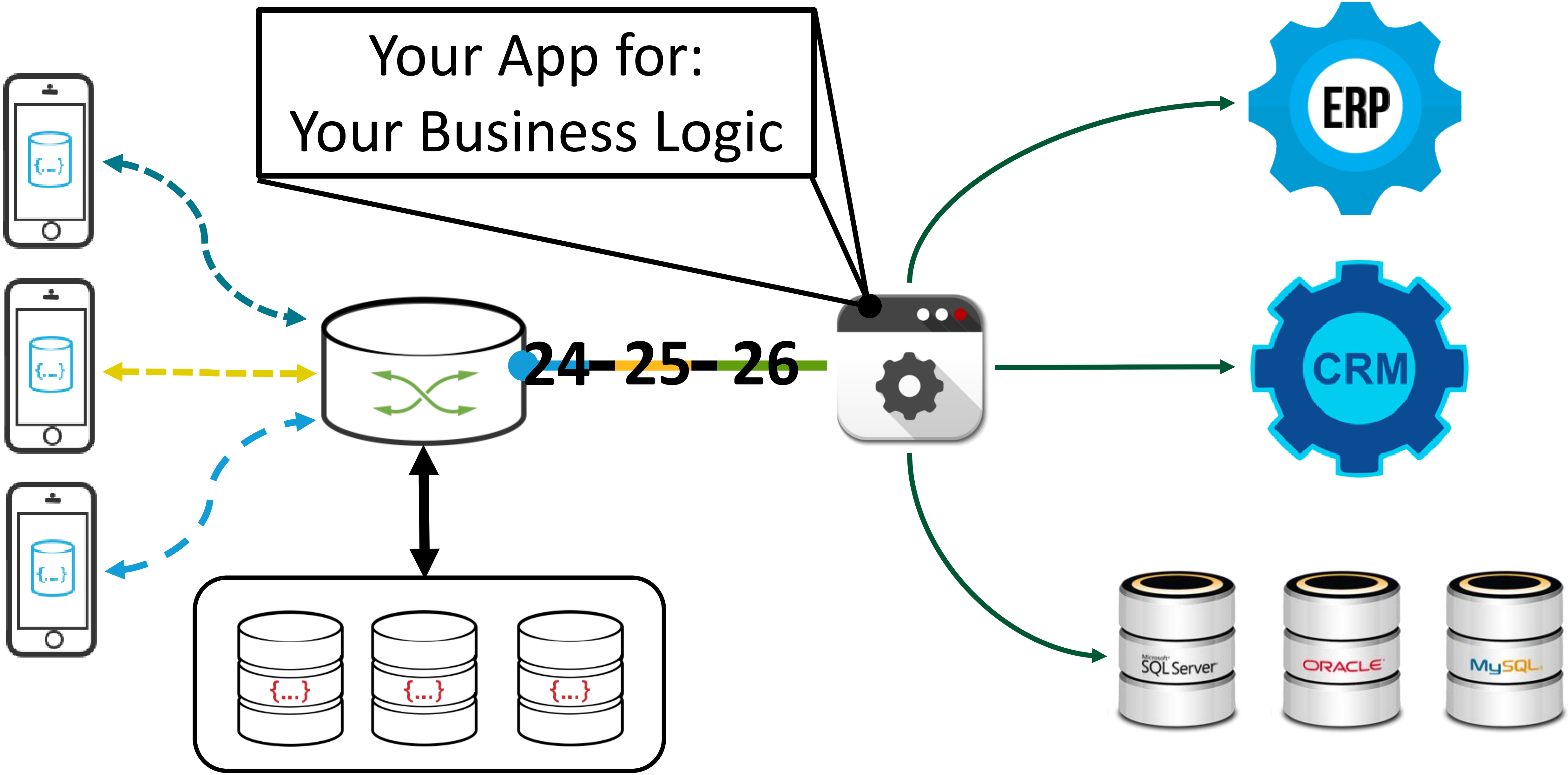
```
{
“_id”:”12”,
“_rev”:”3-hno..”,
“channels”:[“yellow”]
}
```

```
{
“_id”:”81x”,
“_rev”:”2-jba..”,
“channels”:[“red”]
}
```

Sync Gateway & Channels – *Filtered Feed by Channel(s)*



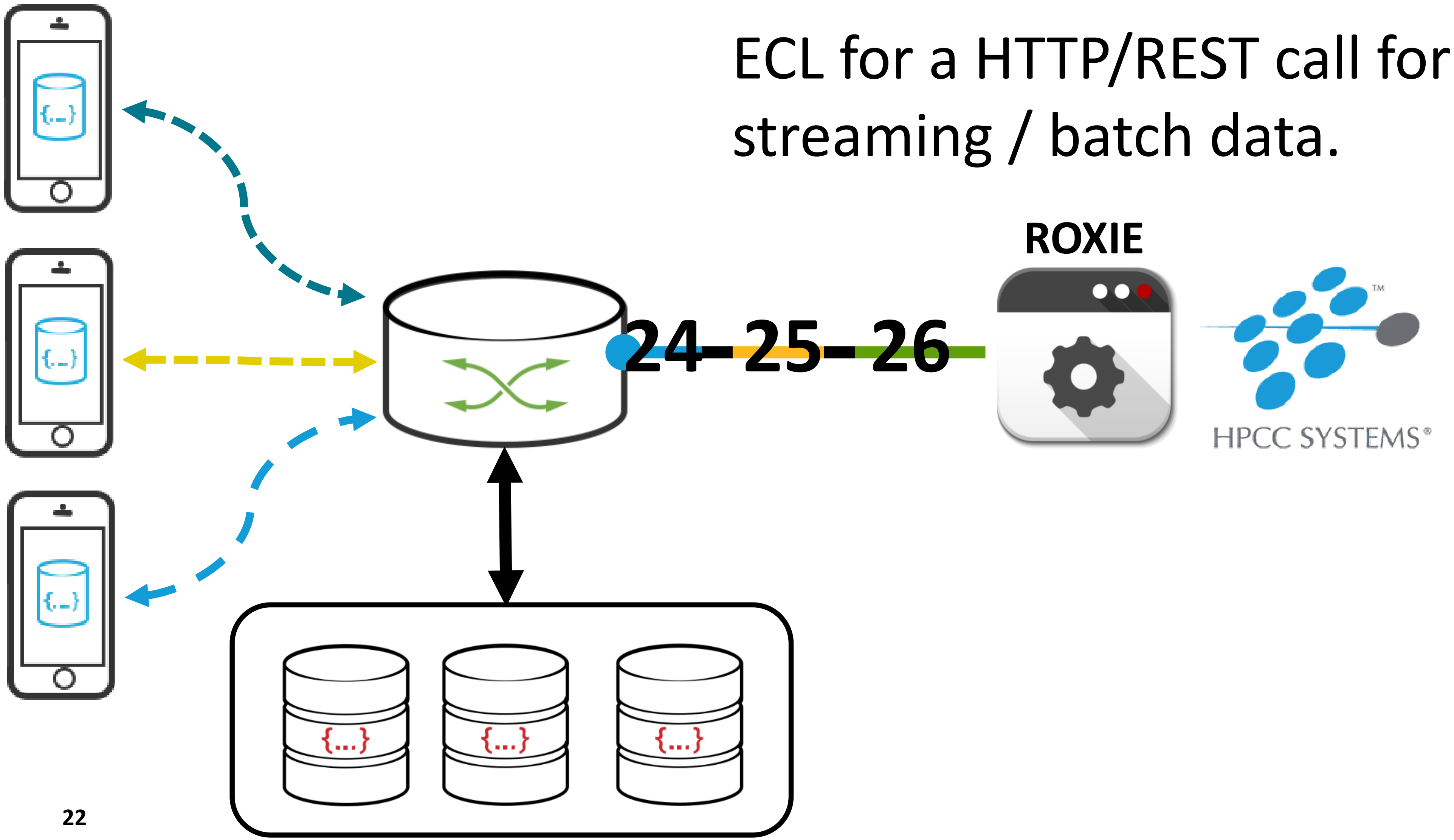
Couchbase Mobile to My Current Systems



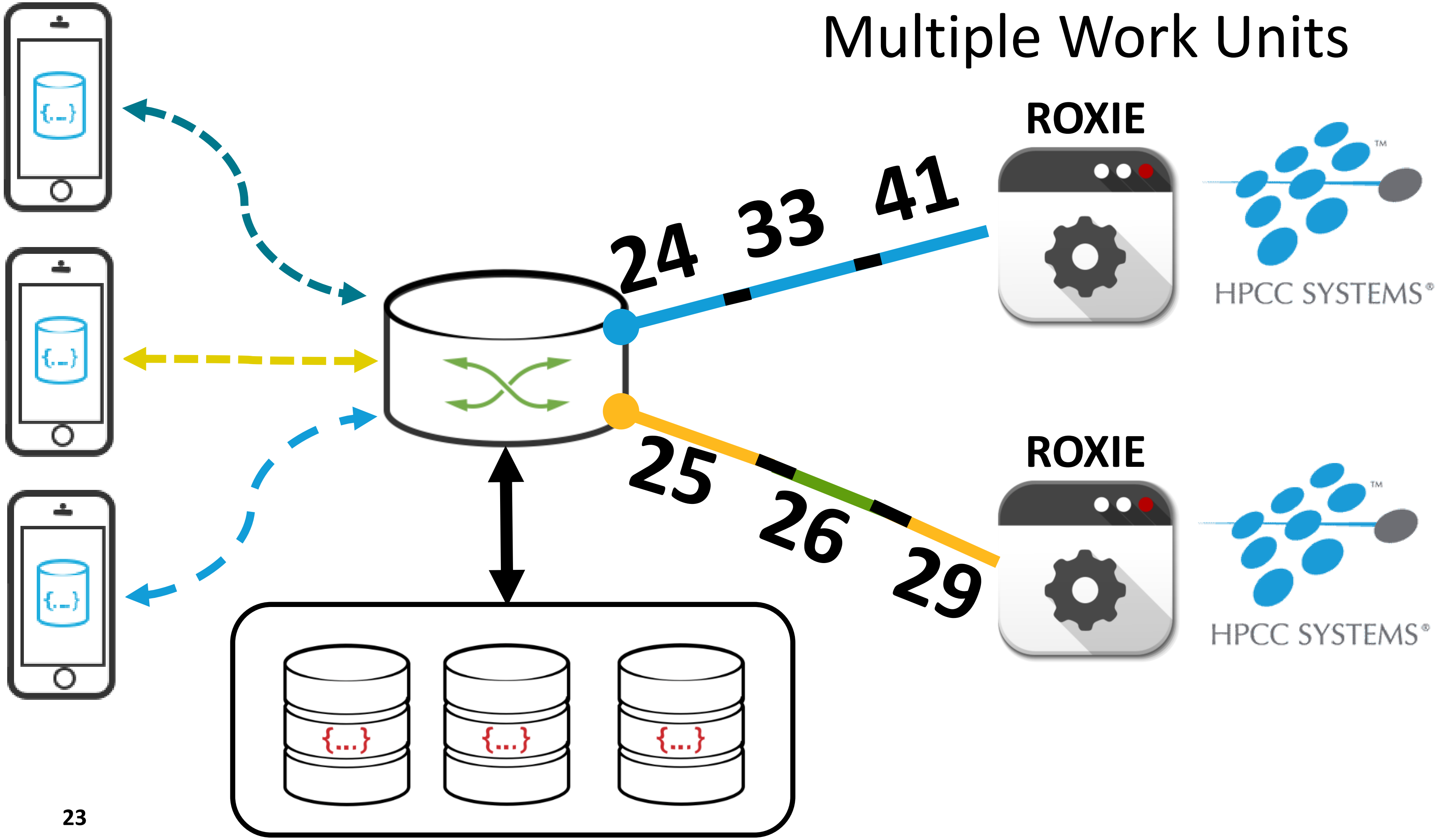
Your App for:
Your Business Logic

Couchbase Mobile to HPCC Systems

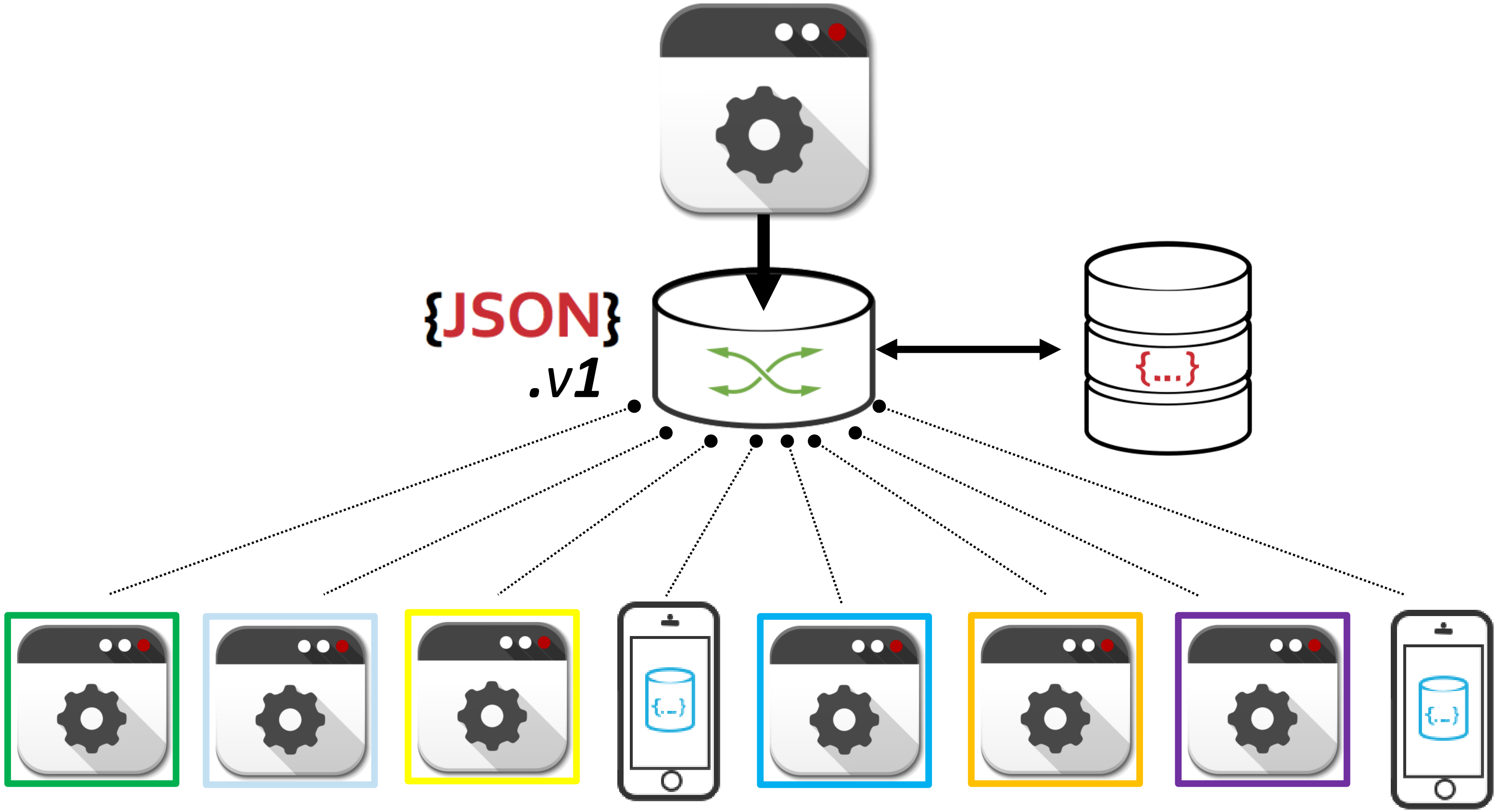
ECL for a HTTP/REST call for streaming / batch data.



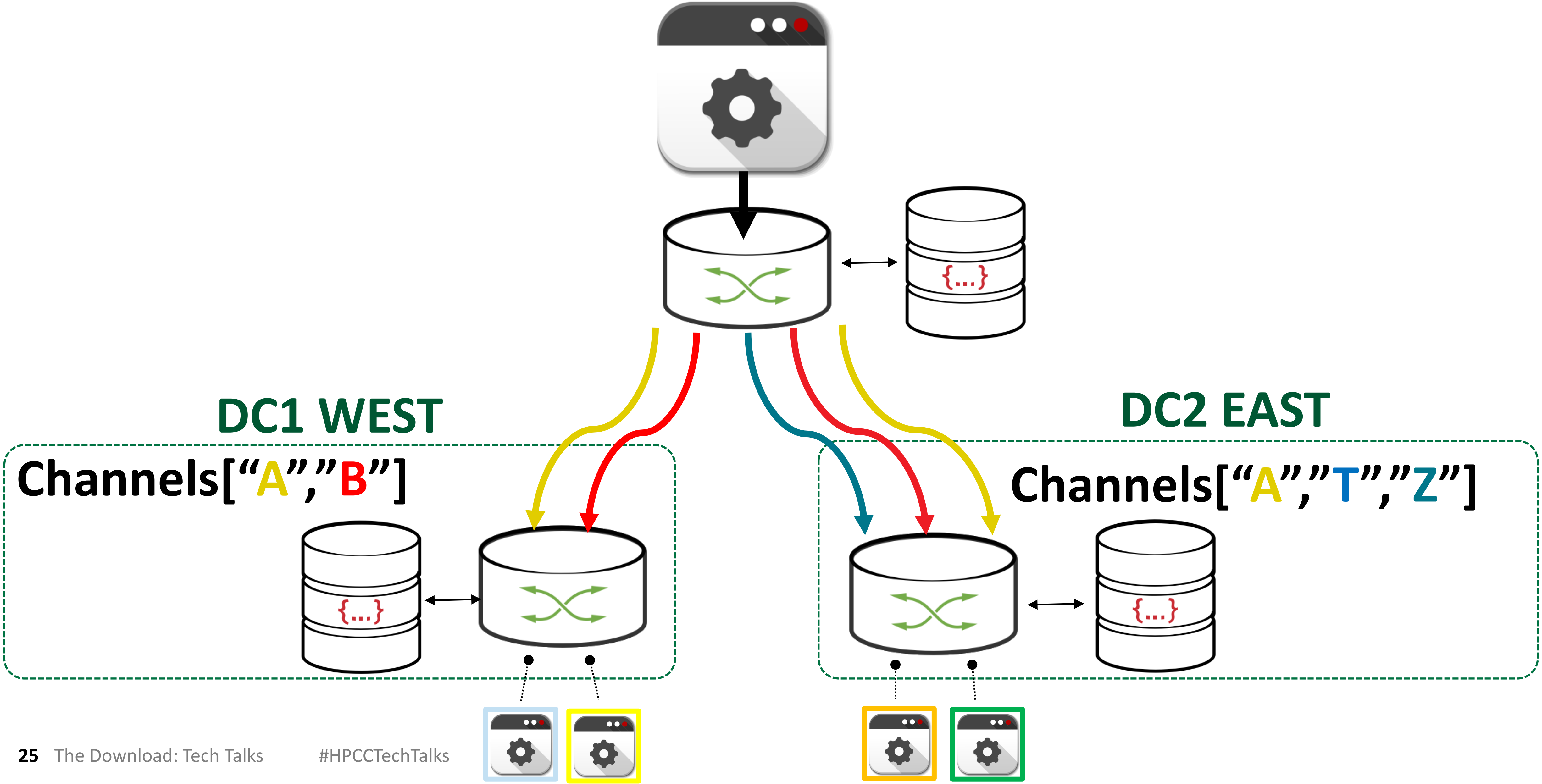
Couchbase Mobile to HPCC Systems



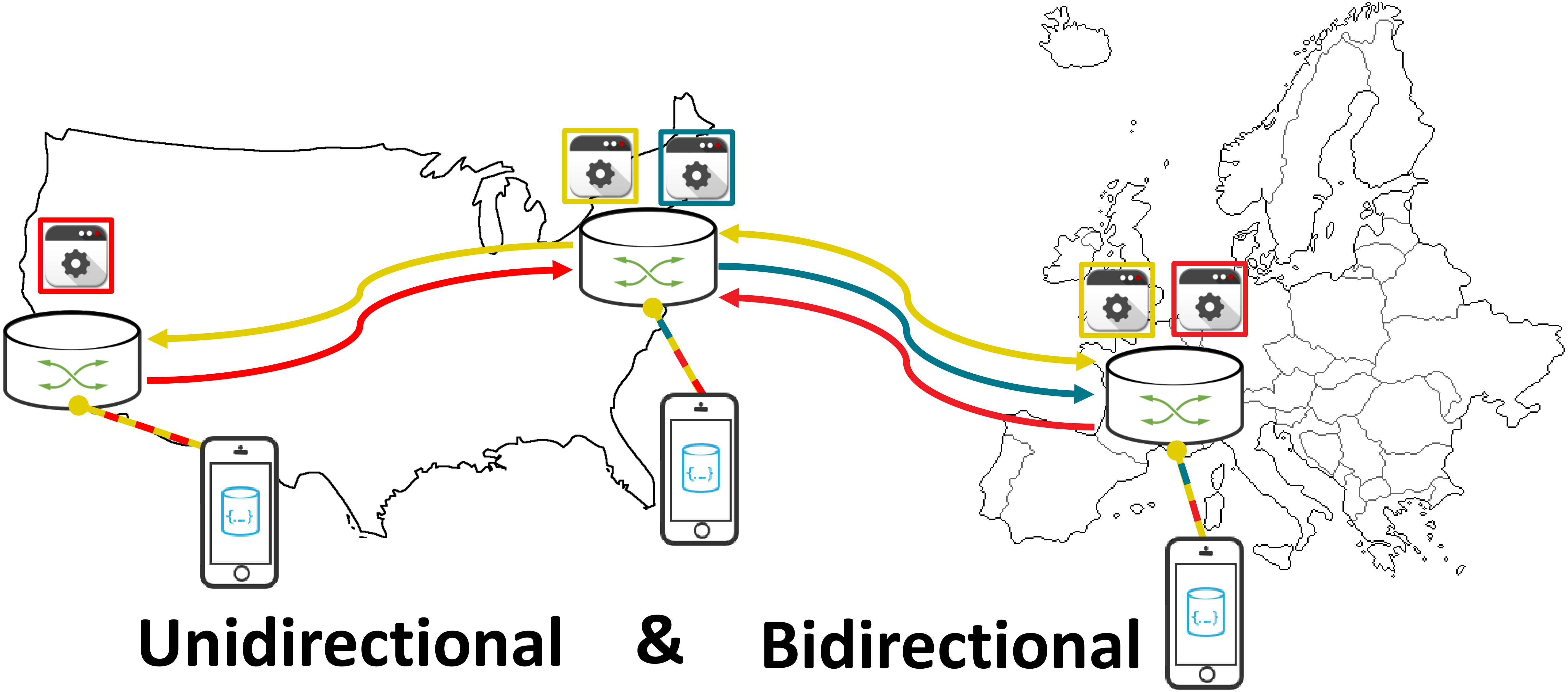
Sync Gateway – the “Truth”



“Chaining” Sync Gateways – Data Locality & Filtering

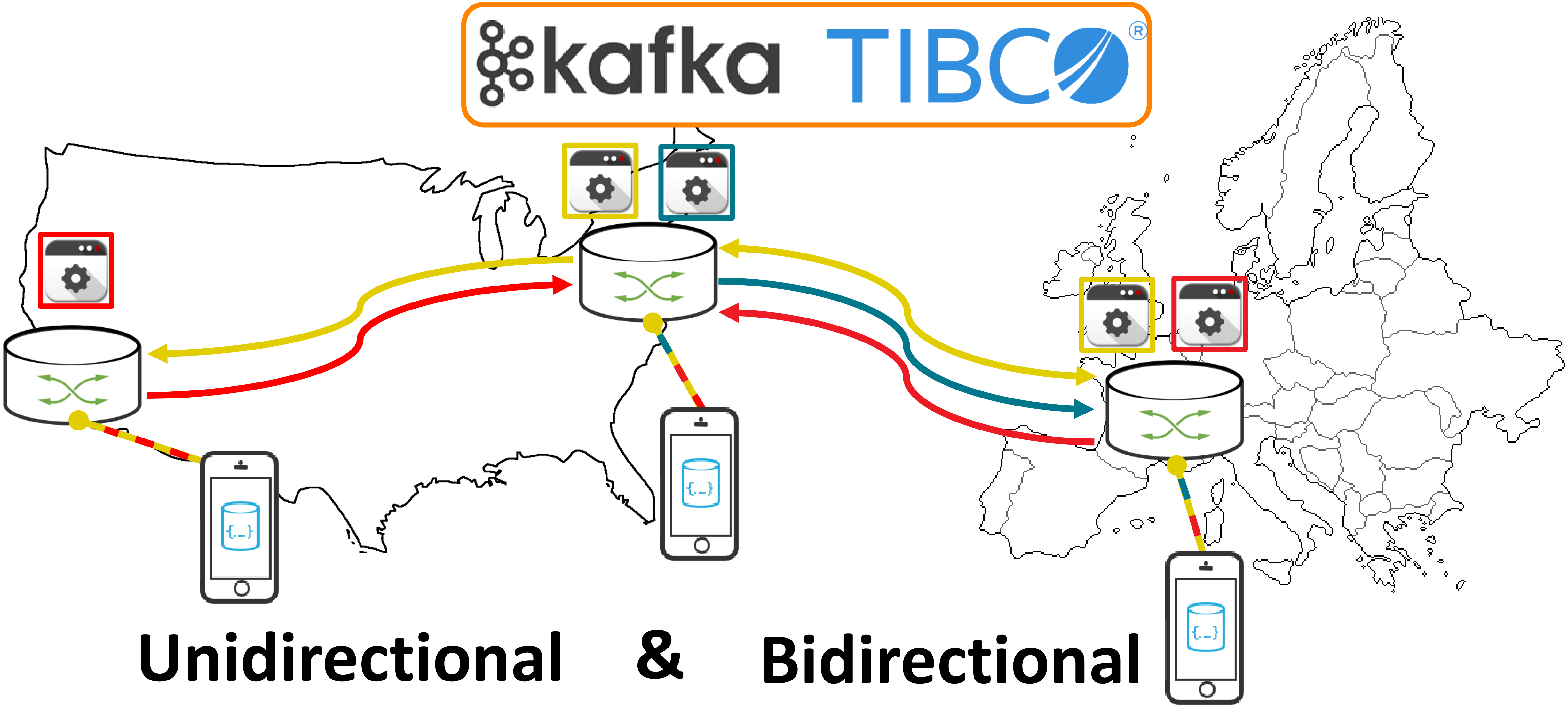


Data Pipeline w/ Sync Gateways - *Options & Future*



Unidirectional & Bidirectional

Data Pipeline w/ Sync Gateways - *Options & Future*

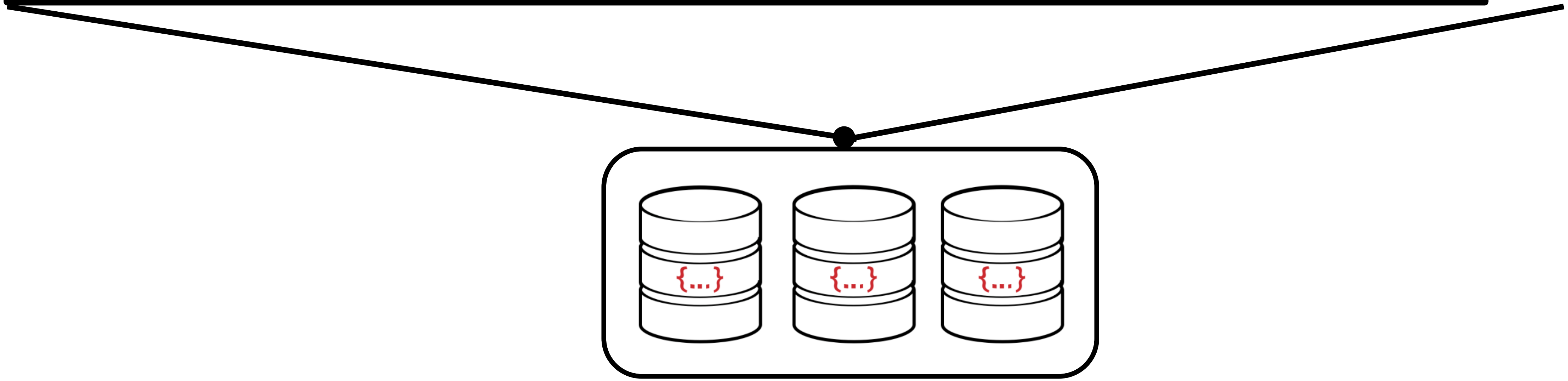


Unidirectional & Bidirectional

N1QL – SQL for JSON - SQL-92

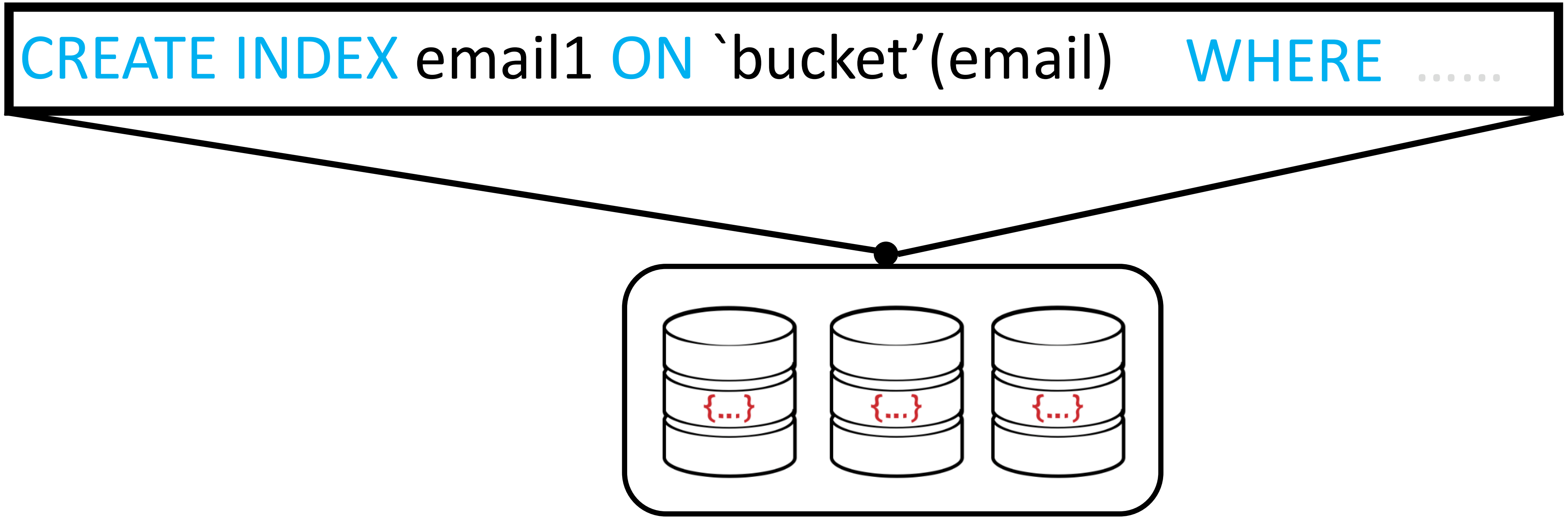
```
{
“type”:”profile”,
“email”:”ted@gmail.com”,
“friends”:[{“name”:”Bob”}
           ,{“name”:”Kevin”}]
}
```

```
SELECT * FROM `bucket’ WHERE email LIKE “%gmail.com”;
```



N1QL – Creating Indexes

```
{
“type”:”profile”,
“email”:”ted@gmail.com”,
“friends”:[{“name”:”Bob”}
              ,{“name”:”Kevin”}]
}
```

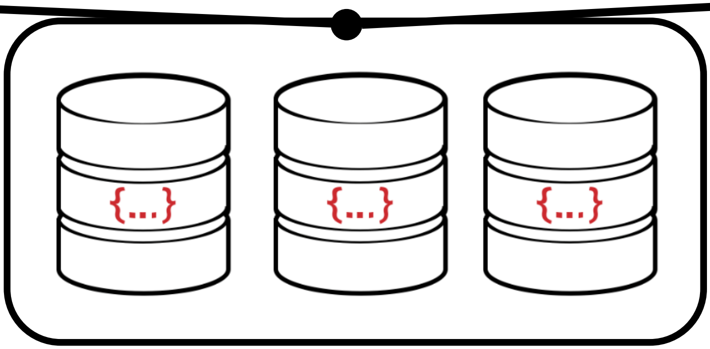


CREATE INDEX email1 ON `bucket’(email)   WHERE ……

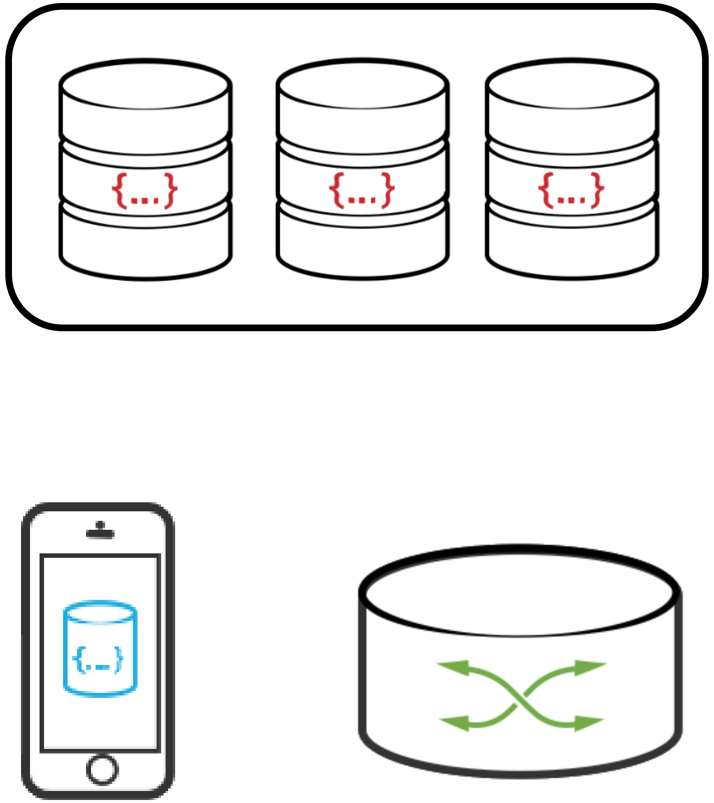
Full Functional SQL for JSON

```
{
“type”:”profile”,
“email”:”ted@gmail.com”,
“friends”:[{“name”:”Bob”},
            {“name”:”Kevin”}]
}
```

```
SELECT * FROM `bucket’ WHERE
                    ANY x IN friends SATIFIES x.name = “Bob”
END;
```



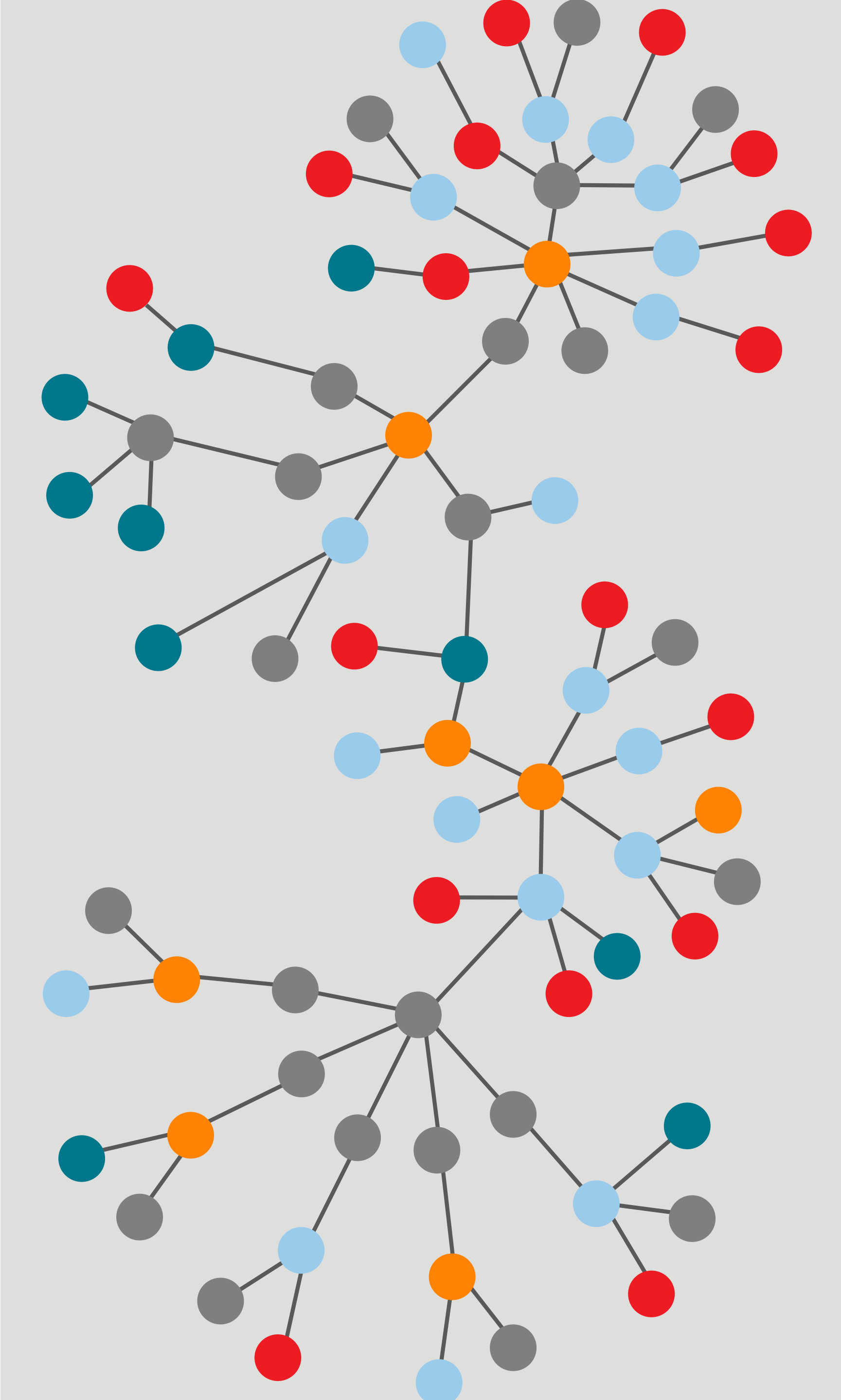
Consolidation



Key/Value & Document	mongoDB cassandra
Query & Index	MySQL ORACLE Microsoft SQL Server
Mobile / Data Flow	SQLite kafka TIBCO®
Search (DP2)	elastic Apache Solr
Analytics (DP1)	APACHE Spark™

Quick poll: Do you see more data coming from Mobile or IoT?

See poll on bottom of presentation screen



HPCC SYSTEMS®

Questions?

Couchbase

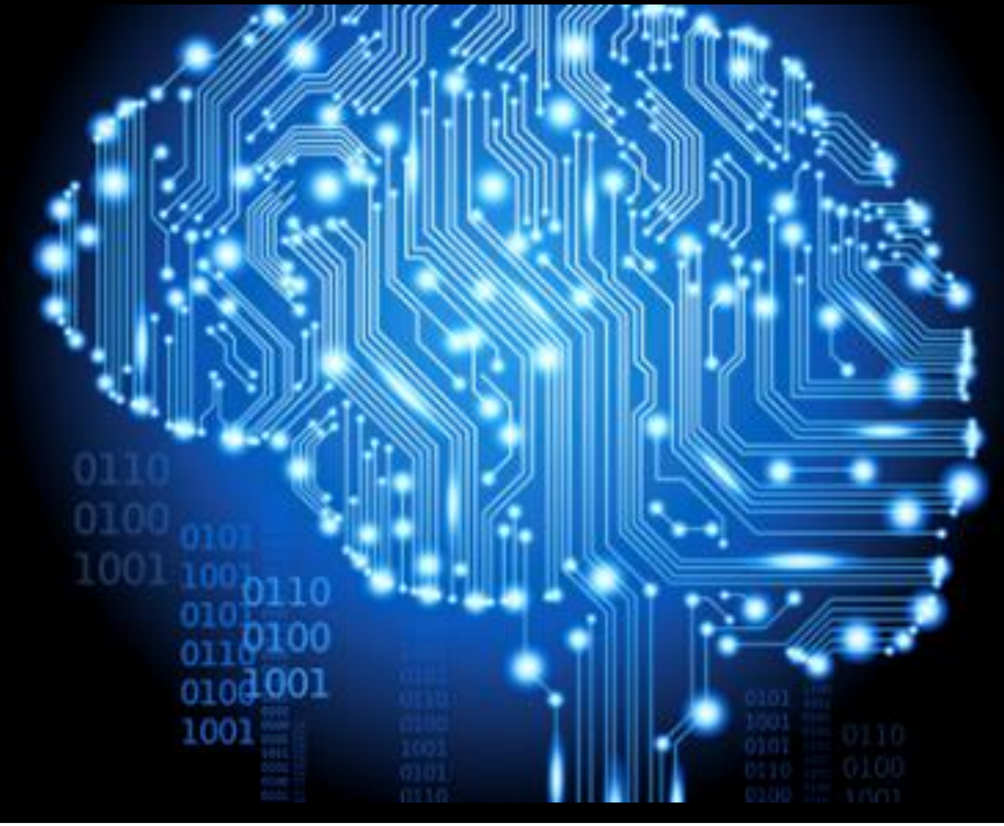
Fujio Turner

Solutions Architect, Couchbase

mail@fuj.io

THE DOWNLOAD

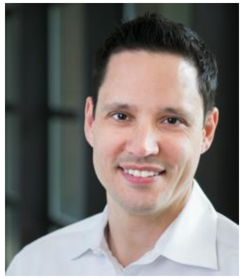
TECH TALKS BY HPCC SYSTEMS



HPCC SYSTEMS®

Operationalizing Your HPCC Systems Environment, Part 1

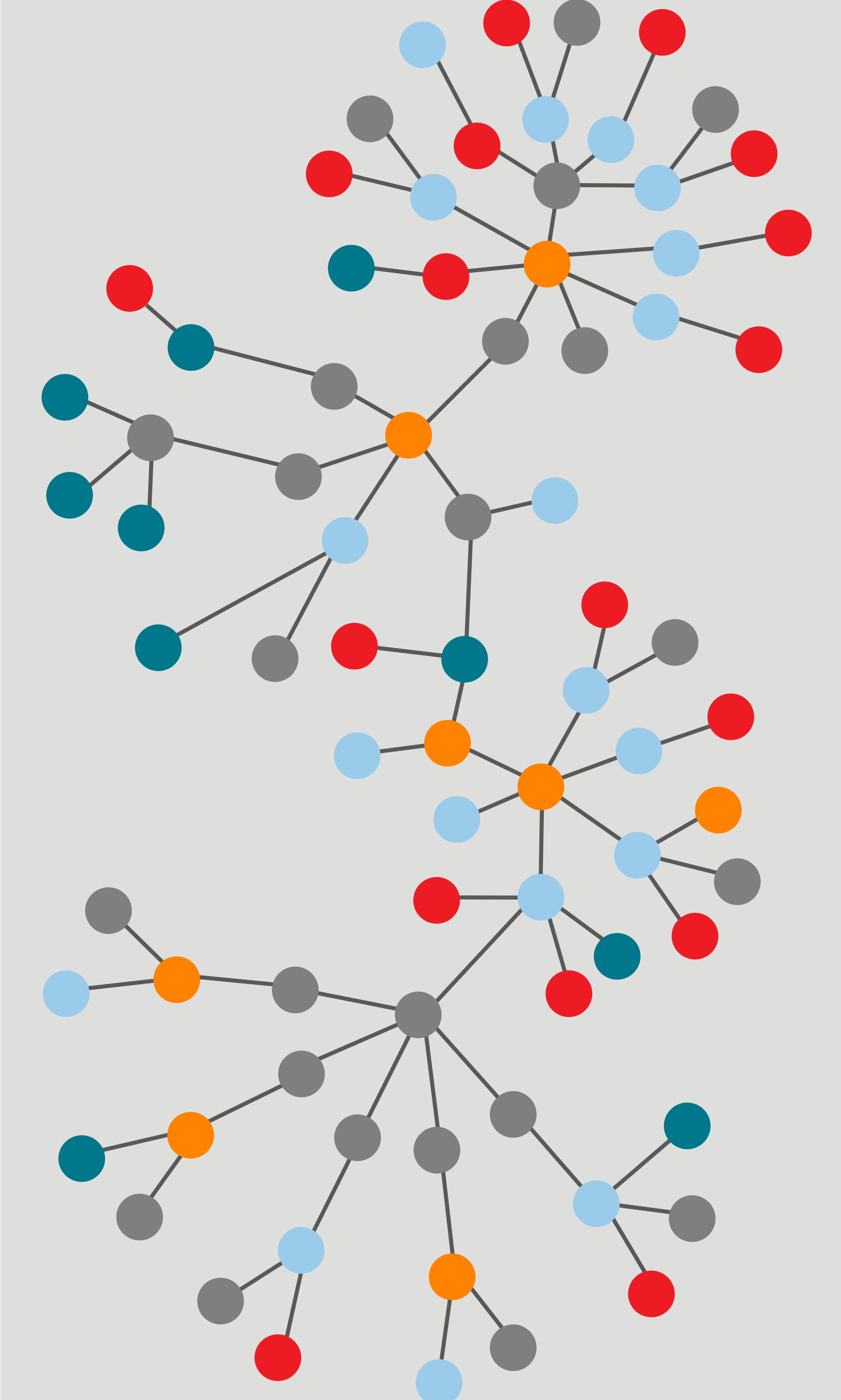
Jacob Pellock
Sr Director Software Engineering



LexisNexis® RISK SOLUTIONS

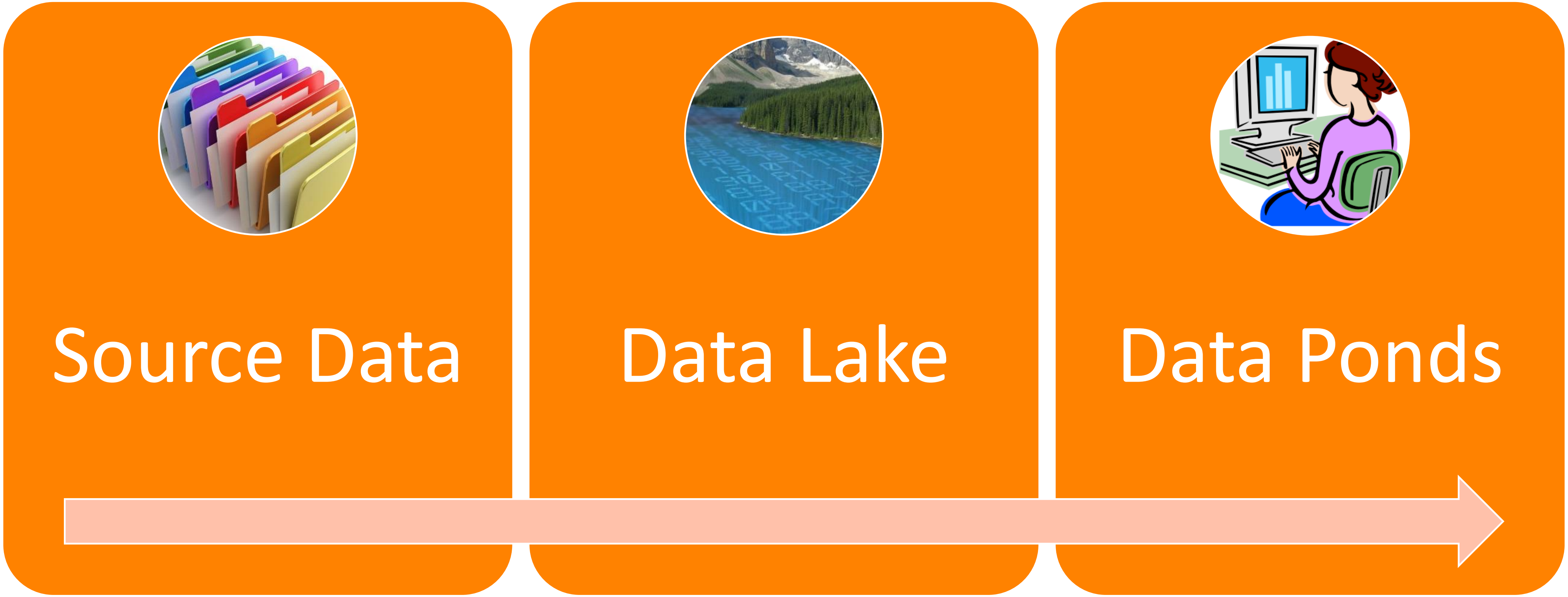
Quick poll: What stage are you in with your HPCC Systems deployment?

See poll on bottom of presentation screen



HPCC SYSTEMS®

Background on our Team



Source Data

Data Lake

Data Ponds

Technologies Used

- HPCC/ECL – warehouse data integration/transformation/distribution
- Git – source code repository
- Python – glue
- HPCC Client Tools – remote job submission

HPCC Client Tools (https://hpccsystems.com/download/developer-tools/client-tools)

- eclcc – ECL compiler
- ecl – Command line interface for job submission

ECLCC Usage

```
$ eclcc
Usage:
   eclcc <options> queryfile.ecl
General options:
   -I <path>     Add path to locations to search for ecl imports
   -L <path>     Add path to locations to search for system libraries
   -o <file>     Specify name of output file (default a.out if linking to executable, or stdout)
   -manifest     Specify path to manifest file listing resources to add
   -foption[=value] Set an ecl option (#option)
   -main <ref>   Compile definition <ref> from the source collection
   -syntax       Perform a syntax check of the ECL
   -target=hthor Generate code for hthor executable (default)
   -target=roxie Generate code for roxie cluster
   -target=thor  Generate code for thor cluster
```

ECLCC Usage (cont.)

Output control options

-E	**Output preprocessed ECL in xml archive form**
-q	Save ECL query text as part of workunit
-wu	Only generate workunit information as xml file

ECLCC Example

```
eclcc -I ./my_code -E -o ./my_archive.xml ./my_code/my_job.ecl
```

ECL Run Usage

```
$ ecl
Usage:
    ecl [--version] <command> [<args>]
Commonly used commands:
   deploy      create a workunit from an ecl file, archive, or dll
   publish     add a workunit to a query set
   unpublish   remove a query from a query set
   run         run the given ecl file, archive, dll, wuid, or query
   activate    activate a published query
   deactivate  deactivate the given query alias name
   queries     show or manipulate queries and querysets
Run 'ecl help <command>' for more information on a specific command
```

ECL Run Usage (cont.)

```
$ ecl help run
Usage:

The 'run' command exectues an ECL workunit, text, file, archive, shared
object, or dll on the specified HPCC target cluster.

Query input can be provided in xml form via the --input parameter.  Input
xml can be provided directly or by refrencing a file

ecl run [--cluster=<val>][--input=<file|xml>][--wait=<ms>] <wuid>
ecl run [--cluster=<c>][--input=<file|xml>][--wait=<ms>] <queryset> <query>
ecl run [--cluster=<c>][--name=<nm>][--input=<file|xml>][--wait=<i>] <dll|->
ecl run --cluster=<c> --name=<nm> [--input=<file|xml>][--wait=<i>] <archive|->
ecl run --cluster=<c> --name=<nm> [--input=<file|xml>][--wait=<i>] <eclfile|->
```

ECL Run Usage (cont.)

-	specifies object should be read from stdin
<wuid>	workunit to publish
<archive|->	**archive to publish**
<ecl_file|->	ECL text file to publish
<so|dll|->	workunit dll or shared object to publish

ECL Run Usage (cont.)

Options:

-cl, --cluster=<val>	**cluster to run job on (defaults to cluster defined inside workunit)**
-n, --name=<val>	job name
-in,--input=<file|xml>	file or xml content to use as query input
--wait=<ms>	**time to wait for completion**
-v, --verbose	output additional tracing information
-s, --server=<ip>	**ip of server running ecl services (eclwatch)**
--port=<port>	ecl services port
-u, --username=<name>	**username for accessing ecl services**
-pw, --password=<pw>	**password for accessing ecl services**
--main=<definition>	definition to use from legacy ECL repository
--ecl-only	send ecl text to hpcc without generating archive
--limit=<limit>	sets the result limit for the query, defaults to 100

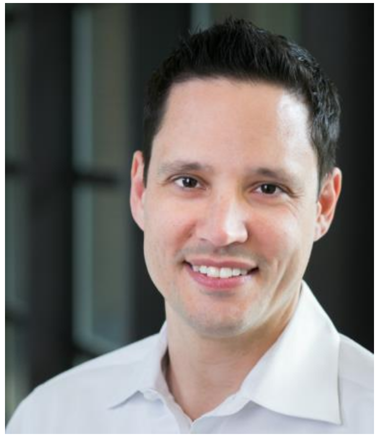
ECL Run Example

```
ecl run --cluster=thor --name=my_thor_job --username=my_username --password=my_password --server=127.0.0.1 --wait=10000 my_archive.xml
```

Other ECL Client Tools

- eclplus – legacy command line tool for executing ECL commands
- dfuplus – command line tool for filesystem operations

Questions?



Jacob Pellock

Sr Director Software Engineering,
LexisNexis® Risk Solutions
jacob.pellock@lexisnexisrisk.com

HPCC SYSTEMS®

THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS

HPCC SYSTEMS®

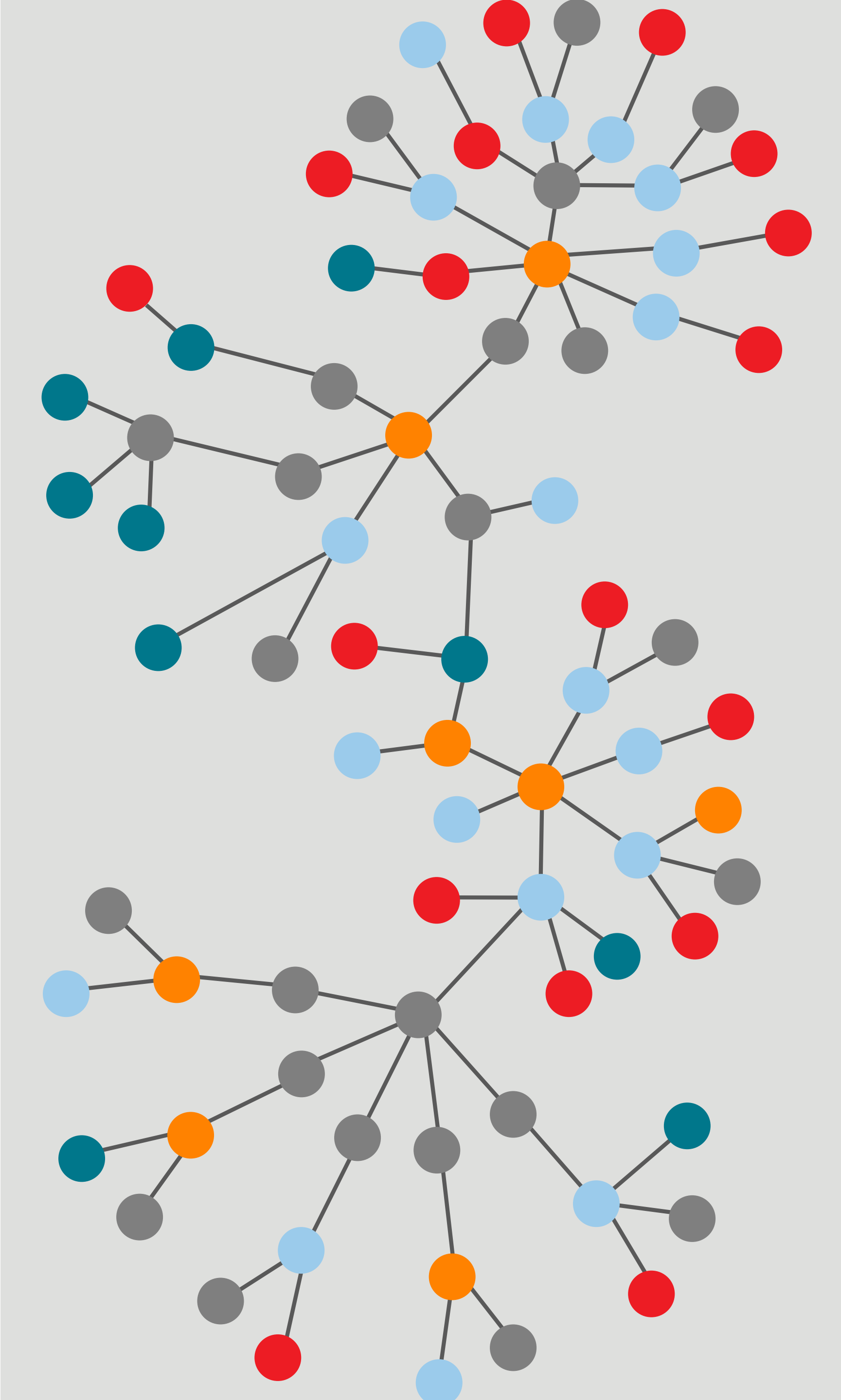
Basic Linear Algebra Subsystem (BLAS) and Parallel Block BLAS (PBBlas) Libraries for HPCC Systems

Roger Dev
Sr Architect, LexisNexis® Risk Solutions

LexisNexis®
RISK SOLUTIONS

Quick poll: Have you had occasion to use Linear Algebra in your work?

See poll on bottom of presentation screen



HPCC SYSTEMS®

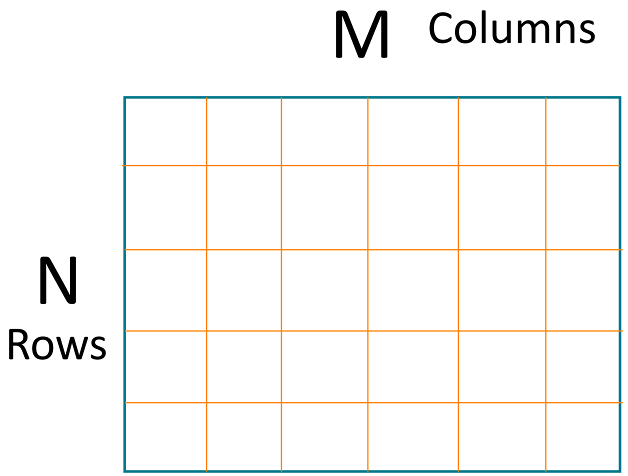
BLAS and PBBlas

- BLAS (Basic Linear Algebra Subprograms) is an industry de-facto standard interface for Linear Algebra operations
  - Very mature – many implementations
  - Highly optimized for different hardware architectures
  - As of HPCC 6.2.0, BLAS is a part of the Std Library
    - IMPORT Std.BLAS
- PBblas – Parallel Block BLAS, unique to HPCC, provides a BLAS-like interface that can:
  - Scale to HUGE matrixes
  - Balance workload across the nodes in an HPCC cluster
  - Simultaneously perform independent operations on many matrixes in parallel
  - PBblas is an installable bundle at the top-level of HPCC-Systems organization (Github)

Quick Review

How can we parallelize?

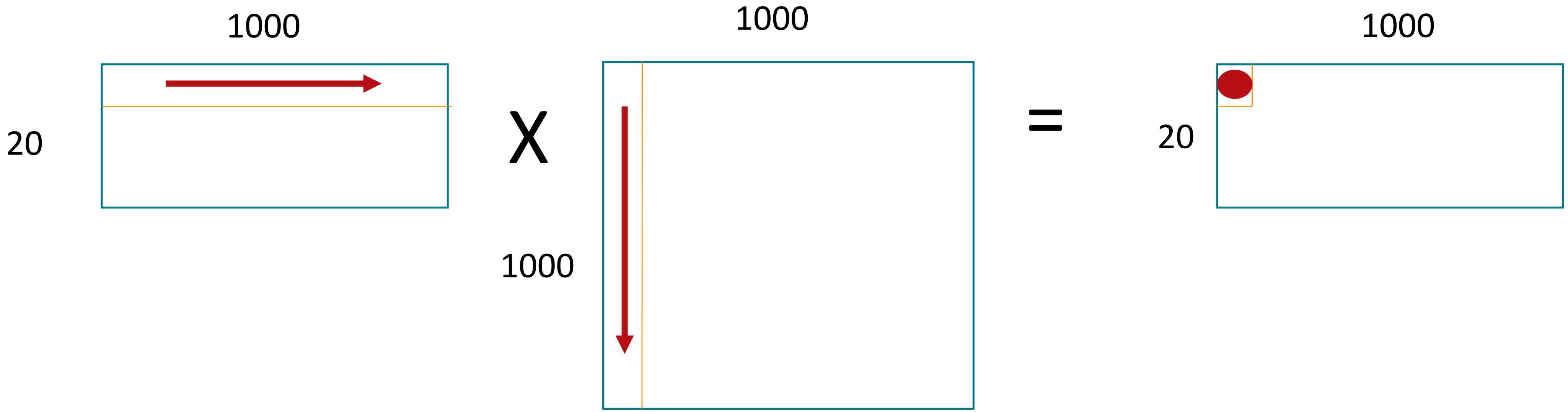
Matrixes



N x M Matrix

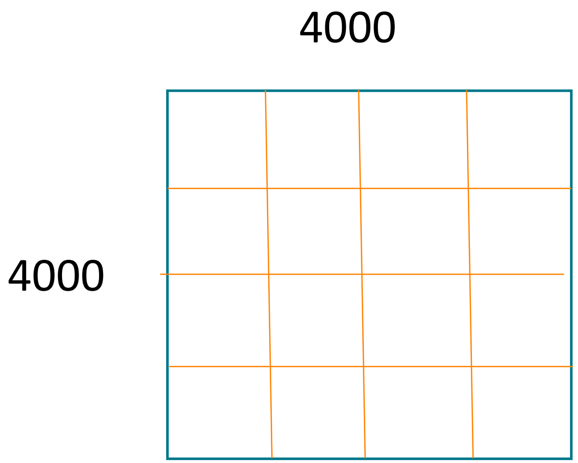
5 x 6 Matrix

Matrix Multiplication



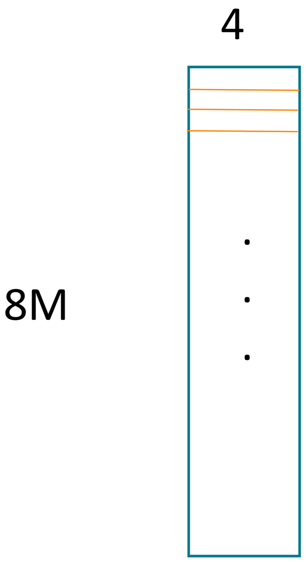
Block Partitioning

“Auto Partitioning”



16 Million Cells

- 4 X 4 Block Partitions
- 1M cells per block
- Each block = 1000 x 1000
- “Square Partitioning”

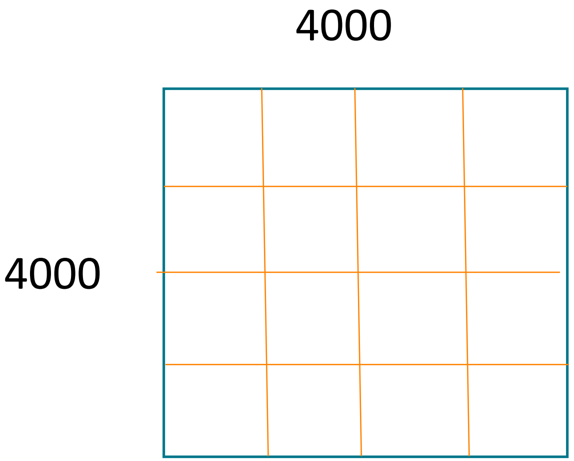


32 Million Cells

- 250K x 1 Block Partitions
- 1M cells per block
- Each block = 250K X 4
- “Row / Column Partitioning”

Cluster Optimization

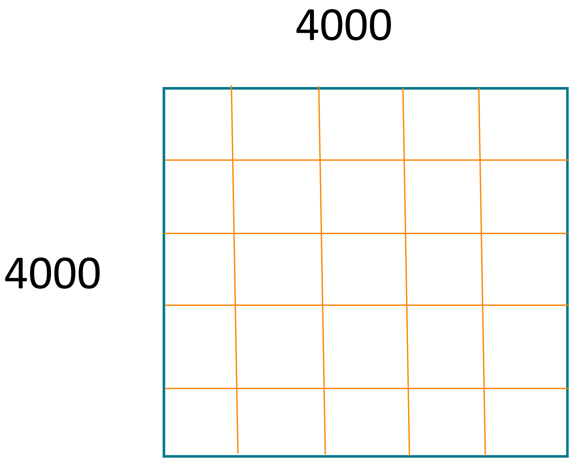
“Workload Balancing”



16 Million Cells

- 4 X 4 Block Partitions
- 1M cells per block
- Each block = 1000 x 1000

What if I’m running on a 25 node cluster?



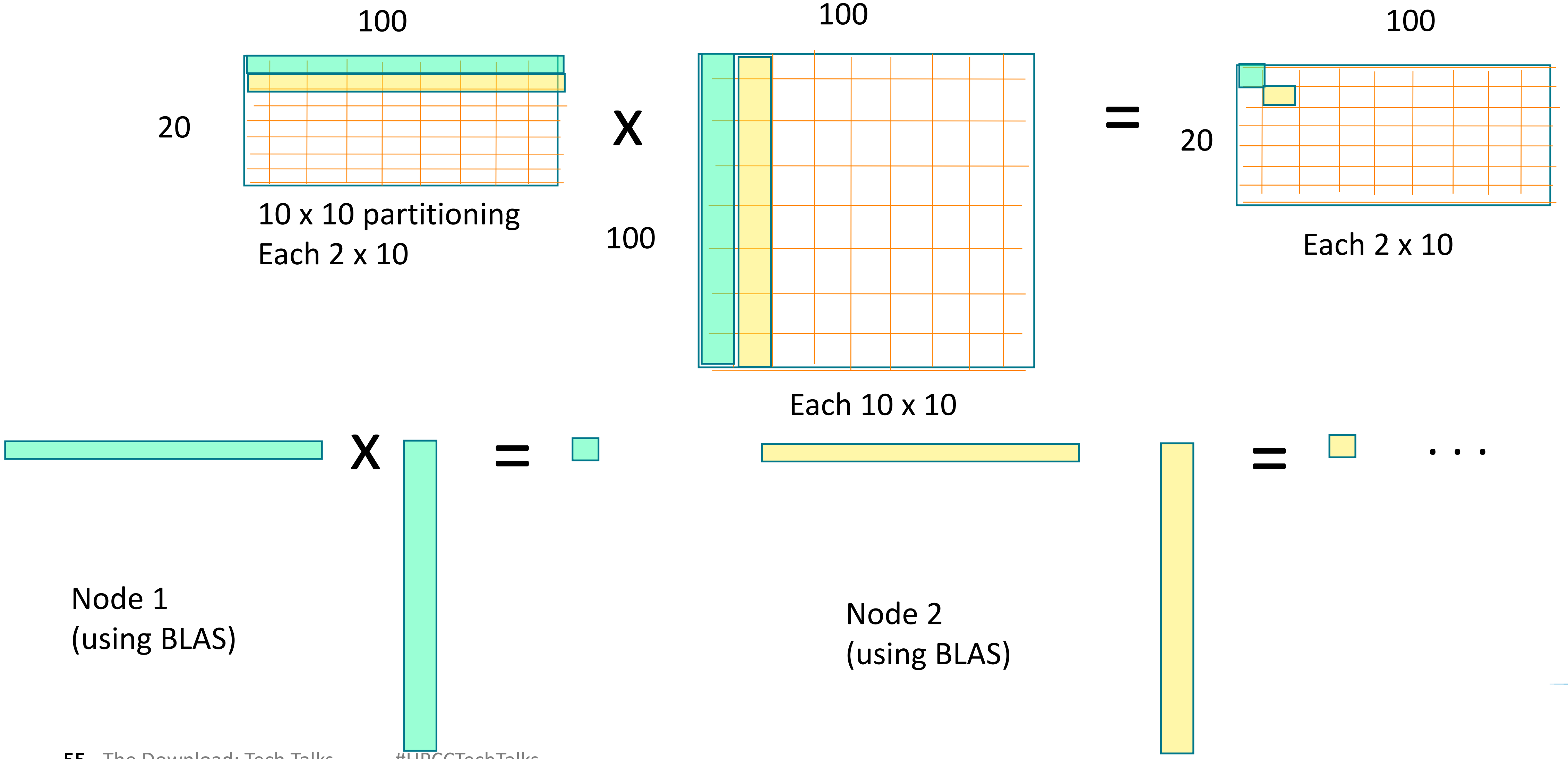
16 Million Cells

- 5 X 5 Block Partitions
- 640K cells per block
- Each block = 800 x 800

Distributed Operations

"Operation Localization"

Matrix Multiplication



10 x 10 partitioning
Each 2 x 10

Each 10 x 10

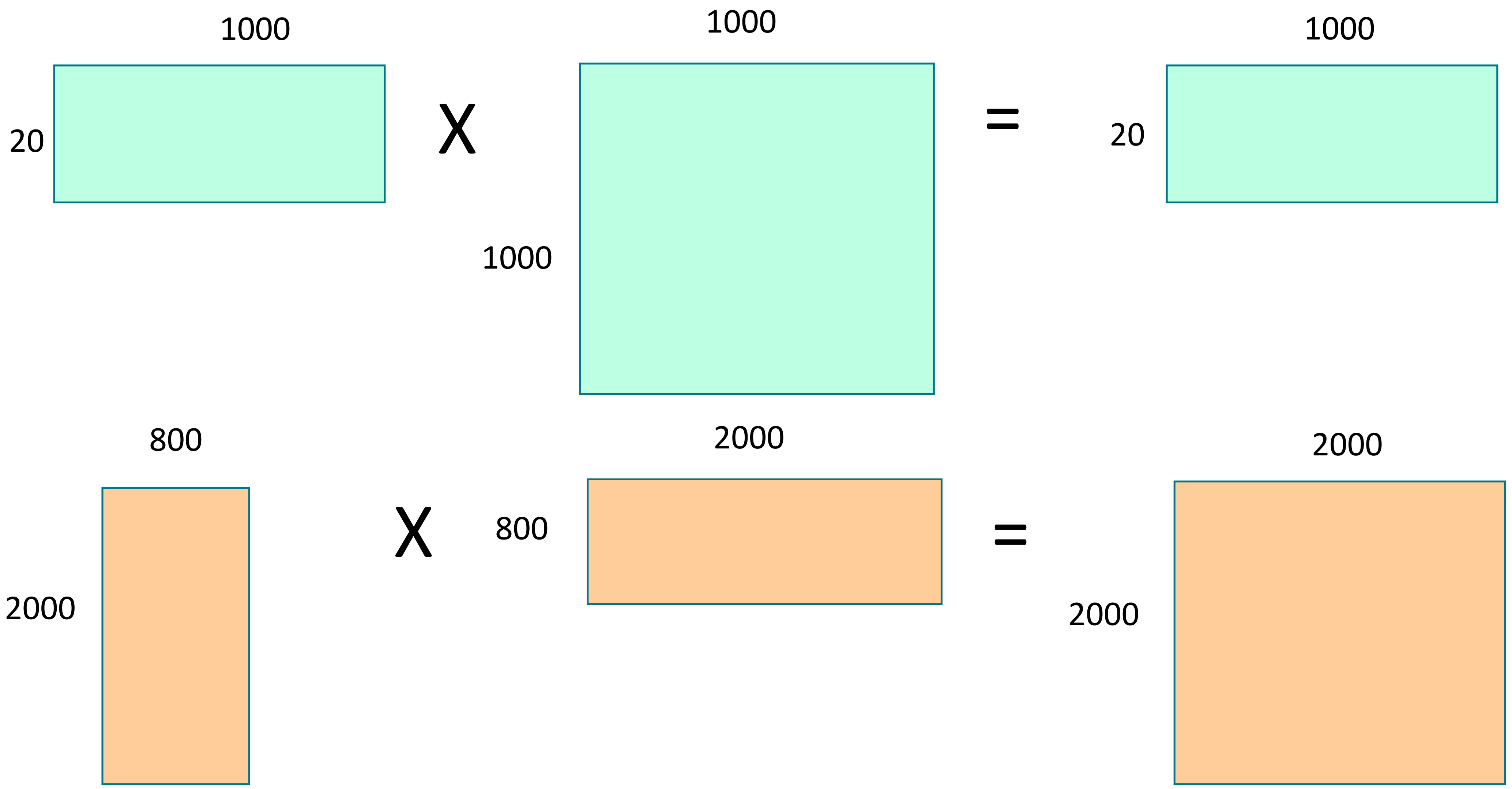
Each 2 x 10

Node 1
(using BLAS)

Node 2
(using BLAS)

Multiple Smaller Operations

“Myriad Operations”



BLAS and PBBlas – Composite Operations

- BLAS insight:
  - Many operations a nearly free when done in tandem with other operations
- Example:
  - gemm: Alpha * TRANSPOSE(A) * TRANSPOSE(B) + Beta * C

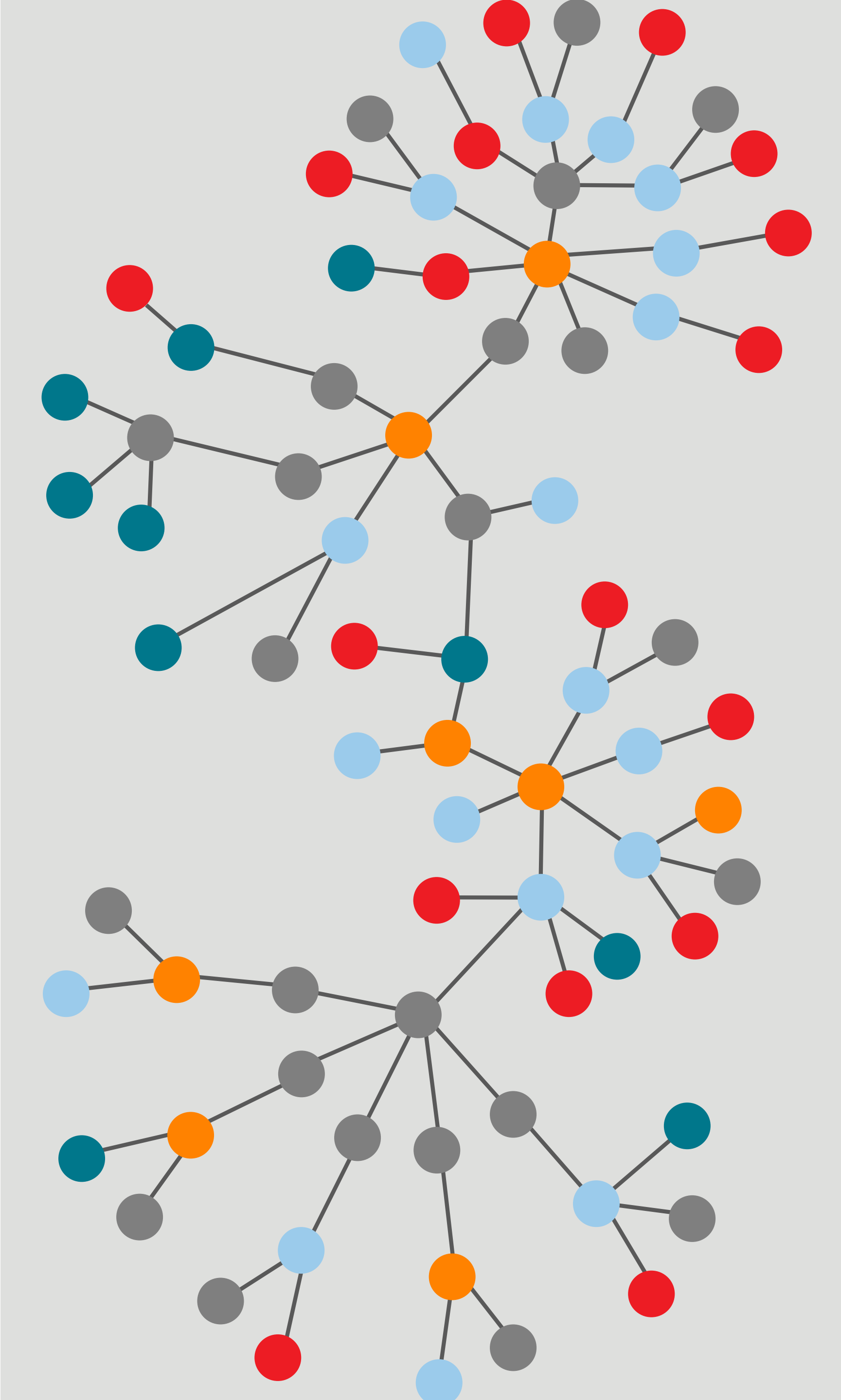
In Summary

- BLAS for optimized local operations on moderate sized matrices
- PBblas for:
  - Operations on large matrixes
  - Efficient utilization of cluster resources
  - Multiple operations in parallel

Unless you have an overriding reason to use BLAS directly, use PBBlas on HPCC clusters.

Quick poll: Do you think you may have a use for BLAS or PBblas in the future?

See poll on bottom of presentation screen



Questions?

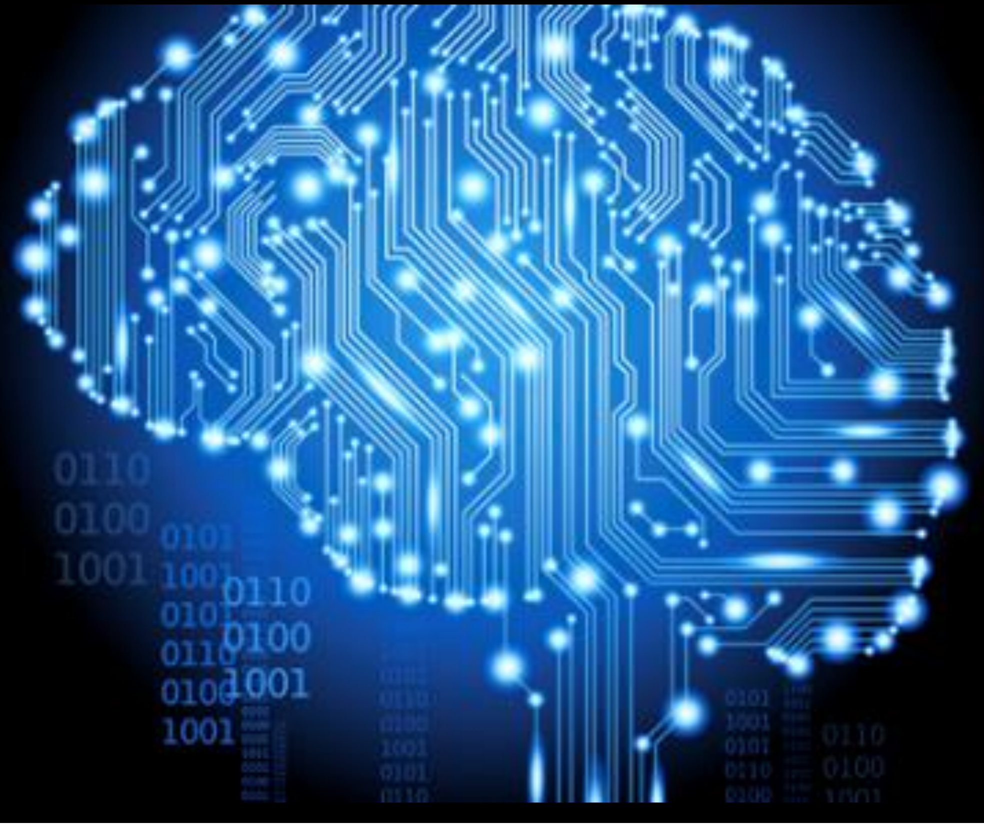
Roger Dev

Sr Architect, LexisNexis® Risk Solutions

roger.dev@lexisnexisrisk.com

THE DOWNLOAD

TECH TALKS BY HPCC SYSTEMS



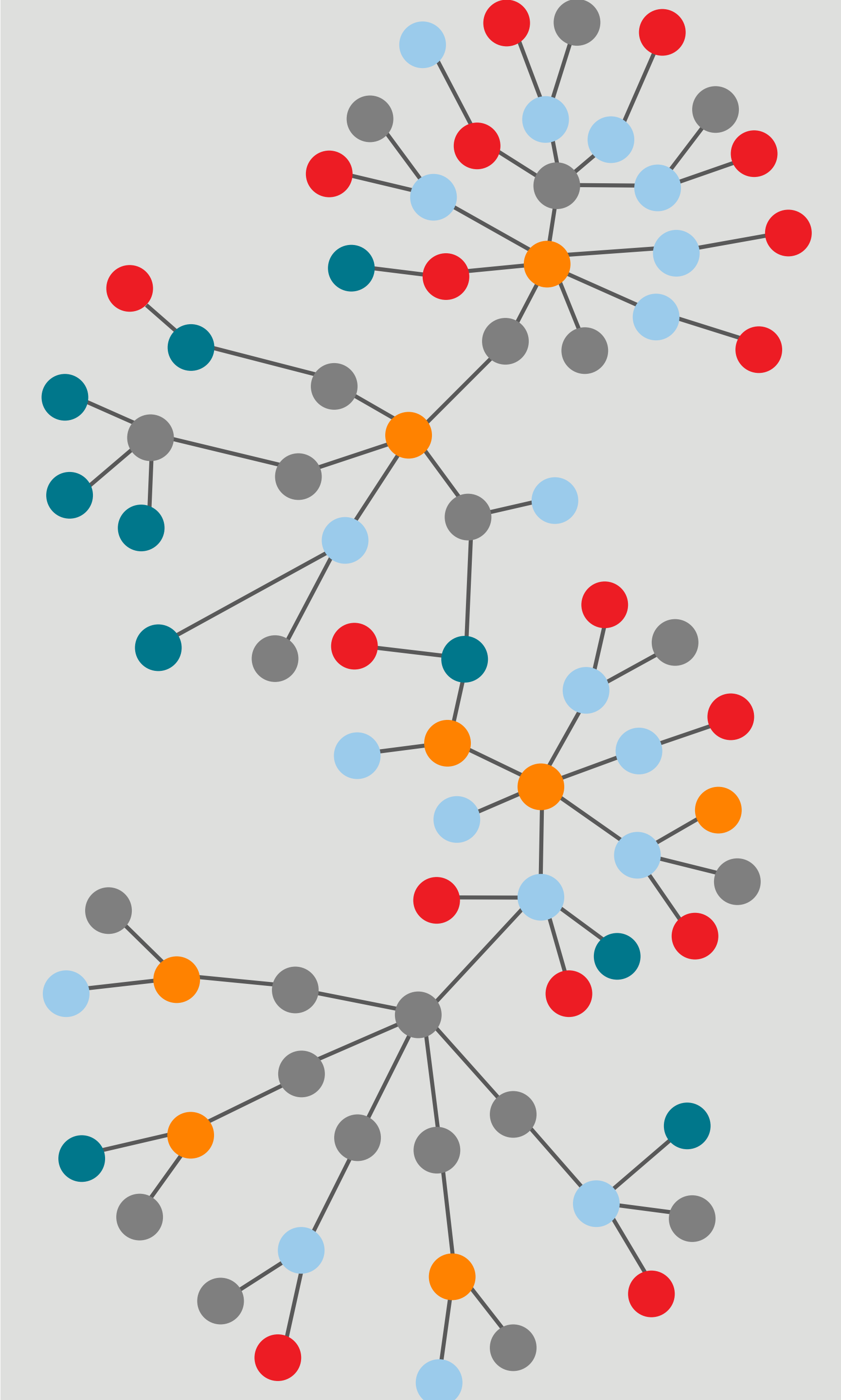
HPCC SYSTEMS®

HPCC Systems Training: Updates and Deep Dives on Cool Code

Richard Taylor
Chief Trainer, HPCC Systems

LexisNexis®
RISK SOLUTIONS

Quick poll: How many different ways do we deliver our ECL courses?



HPCC SYSTEMS®

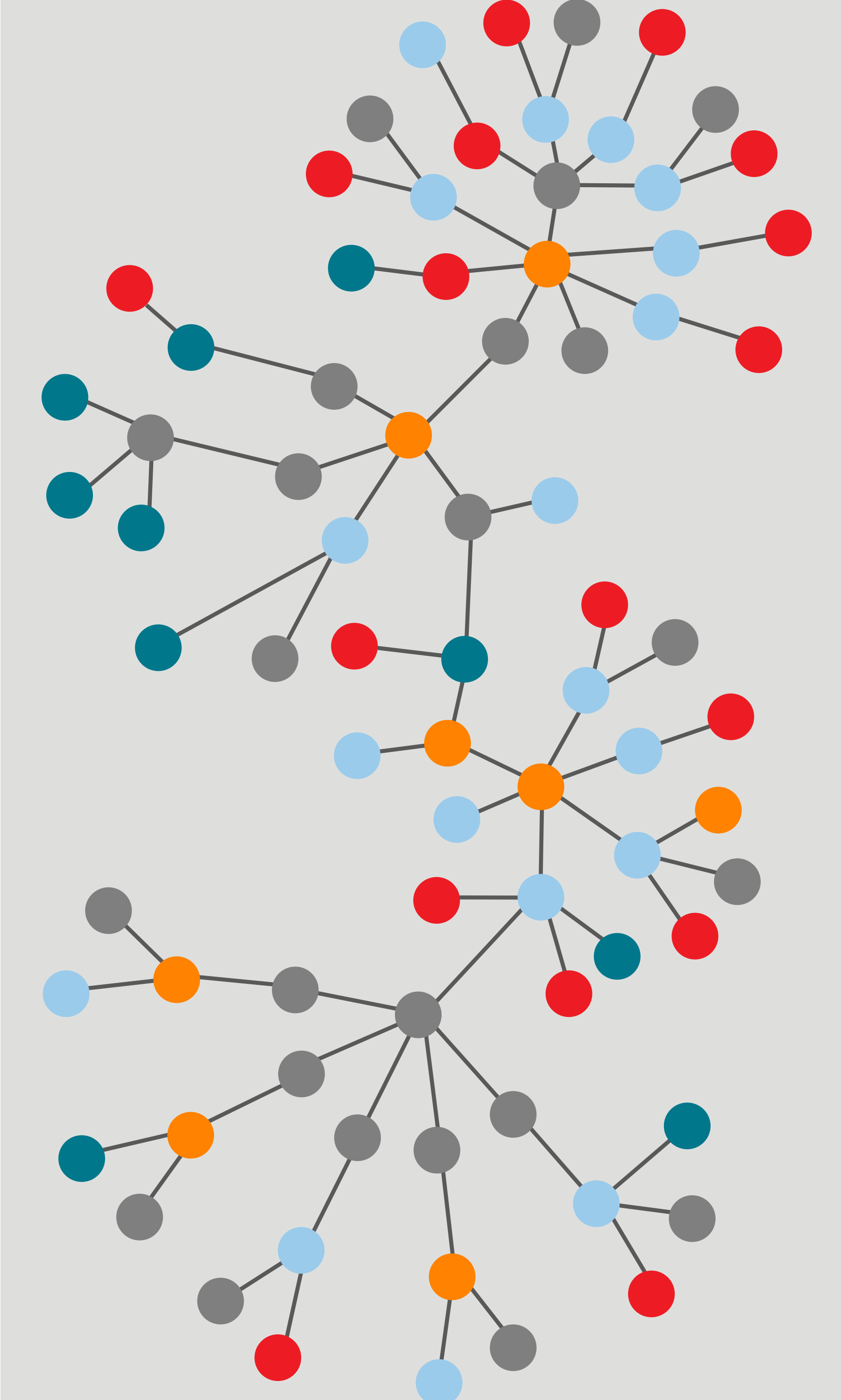
HPCC Systems Training

- **ECL Courses Renamed:**
  - Introduction to ECL (Part 1)
  - Introduction to ECL (Part 2)

  - Advanced ECL (Part 1)
  - Advanced ECL (Part 2)

  - Roxie ECL (Part 1)
  - Roxie ECL (Part 2)

Quick poll: Have you taken all the ECL courses you'd like to?



HPCC SYSTEMS®

HPCC Systems Training

- **FOUR ECL Course Delivery Methods:**
  - **On-site**, live instructor-led
    - Sign-up requires discount code
  - **Remote** (Lync), live instructor-led
    - Sign-up requires discount code
  - **Online**, pre-recorded, self-paced
    - Sign-up for Advanced, Roxie, and SALT courses requires discount code
  - **Mobile app**, pre-recorded, self-paced
    - Sign-up for Advanced and Roxie courses requires discount code

HPCC Systems Training

- **On-site course Schedule:**
  - First month of each quarter in **Alpharetta**
    - Two weeks, 8 class days
      - Introduction to ECL (parts 1 & 2)
      - Advanced ECL (parts 1 & 2)
  - Second month of each quarter in **Sutton**
    - Two weeks, 10 class days
      - Introduction to ECL (parts 1 & 2)
      - Advanced ECL (parts 1 & 2)
      - Roxie ECL (parts 1 & 2)

- Sign up here: https://hpccsystems.com/getting-started/training-classes

HPCC Systems Training

- **On-site courses can be scheduled:**
  - Anywhere in the world
  - 6 student minimum
  - Expenses that go to your cost center:
    - Instructor travel
    - Printing
    - No other costs for RELX Group
  - Negotiable:
    - Location
    - Courses taught
    - Timeframe

HPCC Systems Training

- **Remote course Schedule:**
  - Third month of each quarter
    - Three weeks, 12 class days
      - Introduction to ECL (parts 1 & 2)
      - Advanced ECL (parts 1 & 2)
      - Roxie ECL (parts 1 & 2)
- Sign up here: https://hpccsystems.com/getting-started/training-classes

HPCC Systems Training

- **Remote courses can also be scheduled:**
  - Anywhere in the world
  - 4 student minimum
  - No cost for RELX Group
  - Courses taught and timeframe are negotiable

HPCC Systems Training

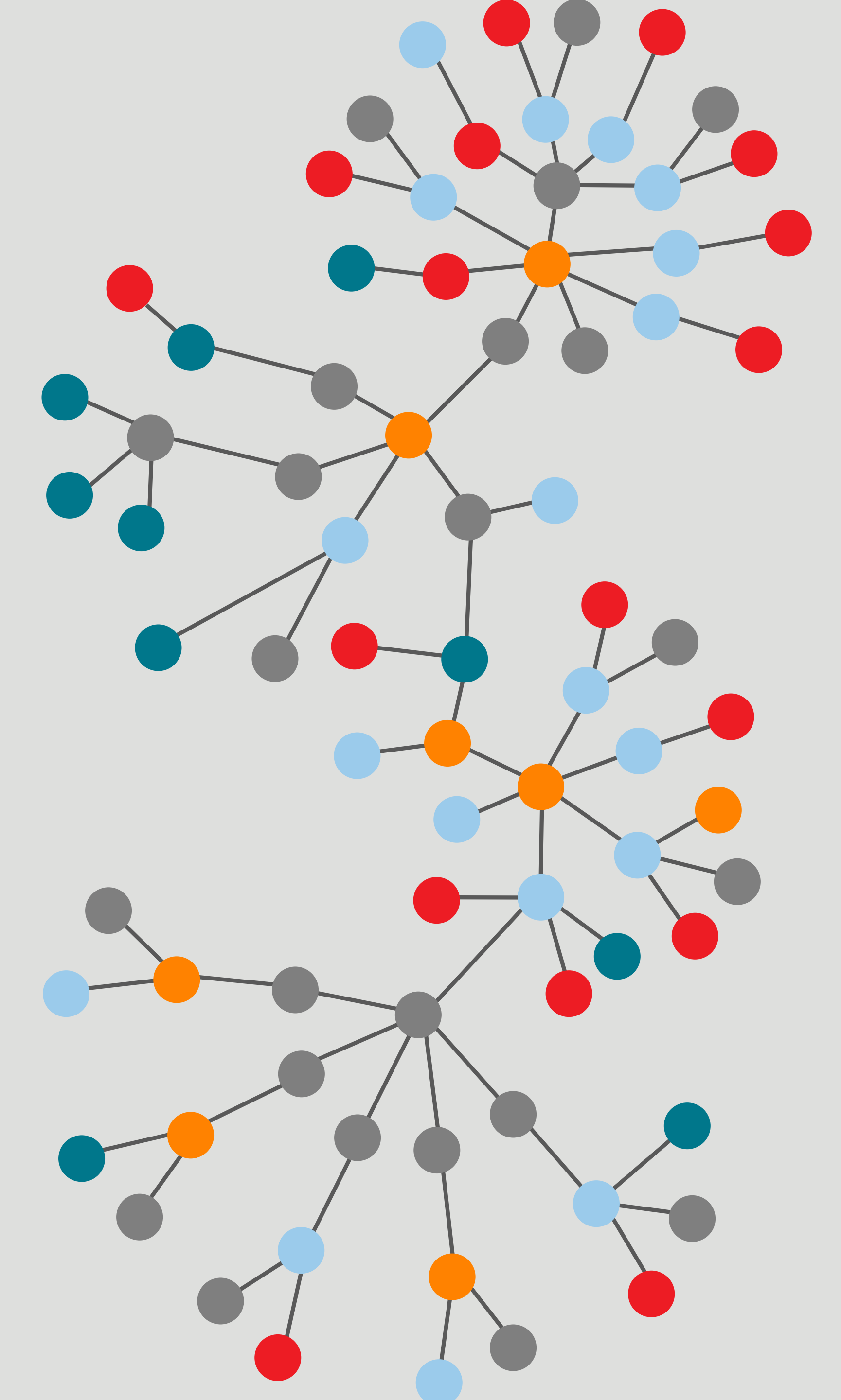
- **Online courses:**
  - Always available and Self-paced
  - Courses available:
    - HPCC for Managers – Free to all
    - HPCC Systems Administration – Free to all
    - Introduction to ECL (parts 1 & 2) – Free to all
    - Advanced ECL (parts 1 & 2) – Free with Discount code
    - Roxie ECL (parts 1 & 2) – Free with Discount code
    - Applied ECL: Code Generation – Free with Discount code
    - Introduction to SALT – Free with Discount code
    - Advanced SALT – Free with Discount code
- Sign up here: https://learn.lexisnexis.com/hpcc

HPCC Systems Training

- **Mobile App courses:**
  - Always available and Self-paced
  - Courses available:
    - Introduction to ECL (parts 1 & 2) – Free to all

- To install the mobile app:
  - For Apple: https://itunes.apple.com/gb/app/hpcc-systems/id1114796489?mt=8
  - For Android: https://play.google.com/store/apps/details?id=uk.co.cple_learning.lexisnexusmobileconsole&hl=en

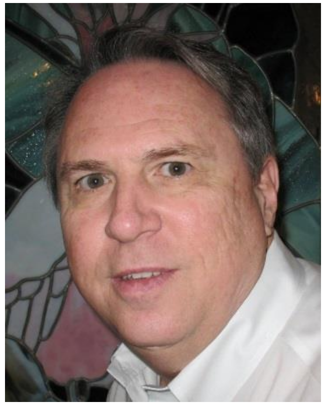
Quick poll: Have you used the ECL LOOP function in your code?



Demo

Let’s explore the LOOP function.

Questions?



Richard Taylor

Chief Trainer, HPCC Systems

richard.taylor@lexisnexisrisk.com

Submit a Talk for an Upcoming Episode!

- Have a new success story to share?
- Want to pitch a new use case?
- Have a new HPCC Systems application you want to demo?
- Want to share some helpful ECL tips and sample code?
- Have a new suggestion for the roadmap?
- Be a featured speaker for an upcoming episode! Email your idea to Techtalks@hpccsystems.com

Visit The Download Tech Talks wiki for more information:
https://wiki.hpccsystems.com/display/hpcc/HPCC+Systems+Tech+Talks

HPCC SYSTEMS®

Thank You!

HPCC SYSTEMS®

RELX Group

A copy of this presentation will be made available soon on our blog: hpccsystems.com/blog