

This page intentionally left blank.



Buy ▾ Sell ▾ Send ▾ Business

[Log In](#)

Settle up as you settle in.

[Request Money Now](#)

High Performance Data Analytics At PayPal

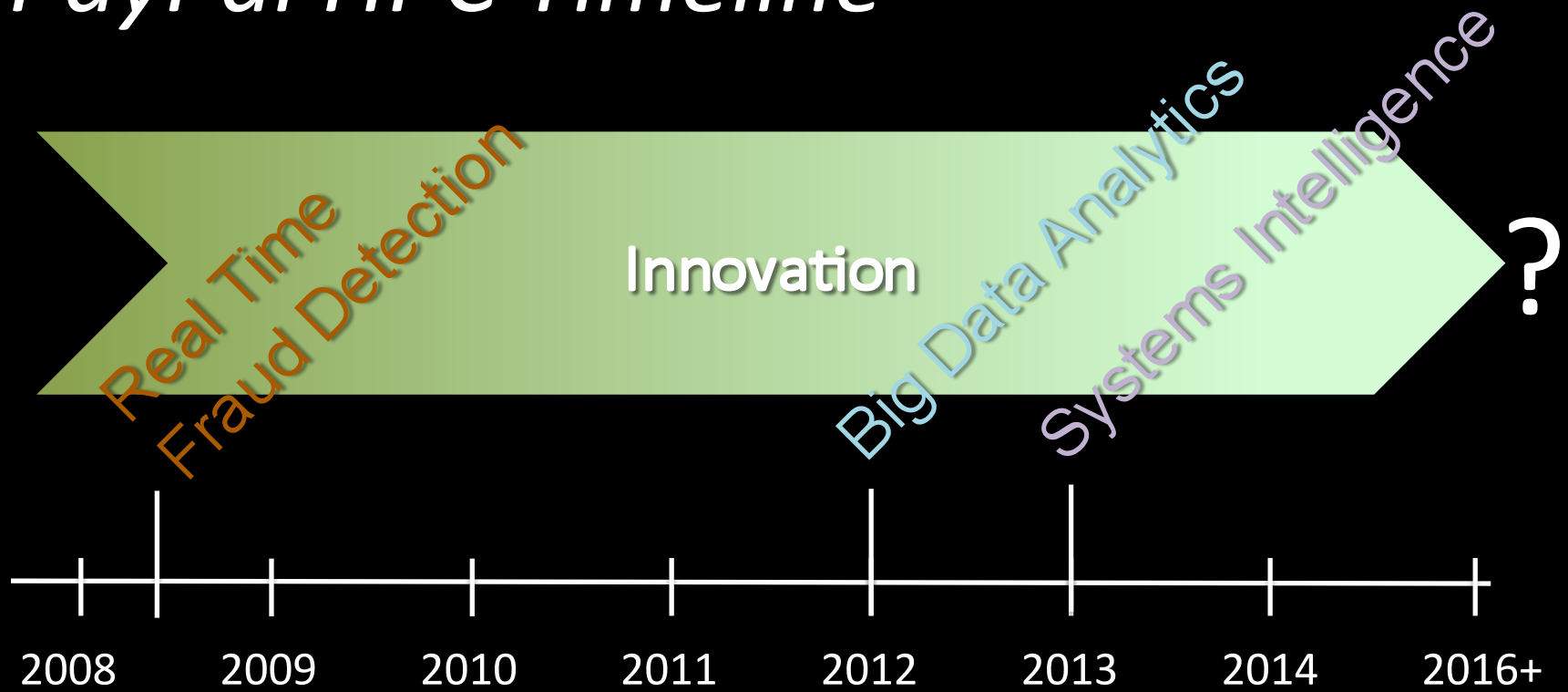
Arno Kolster
Advanced Technology Group

September 29, 2016

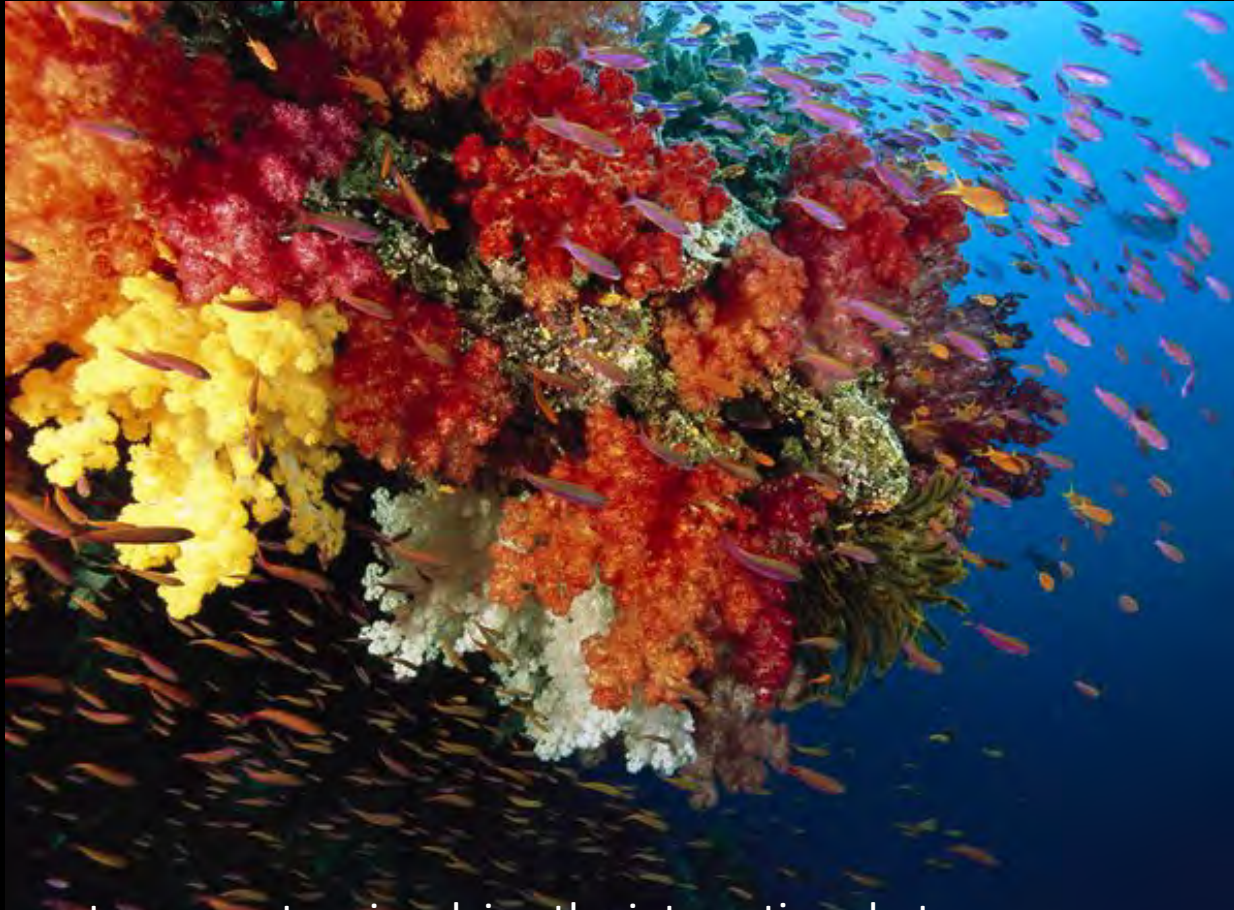




PayPal HPC Timeline



A FINELY TUNED ECOSYSTEM



Ecosystem: a system involving the interactions between a community of living organisms in a particular area and its nonliving environment



What are we trying to accomplish with operational analytics?

Gather an holistic view of PayPal's ecosystem. (i.e. interactions between physical sites, infrastructure, applications and customers). Think "Internet of Things" inside the data center.

Create a self-healing environment through the use of predictive analytics, event correlation and behavior and remediation rule sets.

Model the entire ecosystem's capacity and capabilities for growth, performance and efficiency. Very "HPC like."

Leverage real time streaming analytics with dynamic models built offline to recognize patterns and take appropriate actions.

Reduce costs by minimizing power consumption, operational deficiencies and intelligent deployment.

What are we up against?

Operations analytics in real time...

Real time anomaly detection in correlated event streams using predictive analytical models based on historical data sets.

Streams include application logs, server machine data, data center metrics and social media.

What are we up against with real time?

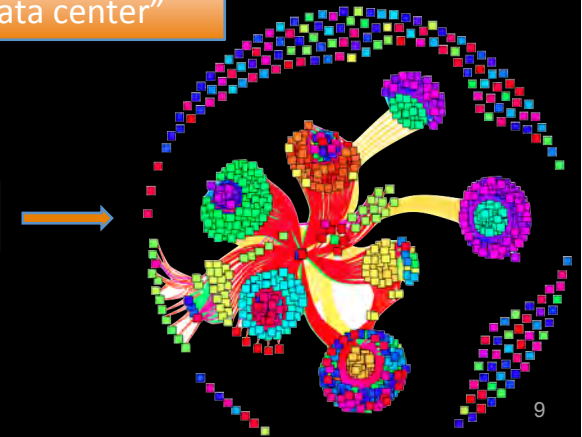
3 Million events / sec from 1000s of sources in our data centers. "IoT in the data center"

25Tb of data ingested every hour

20Mb / sec machine data

50K metadata relationships

Increasing social media trends / customer interaction per day



Our Solution – Systems Intelligence

A Blend of HPC and Enterprise Architectures

Common ontology for concepts and relationships.

Purpose built systems driven by underlying technology.

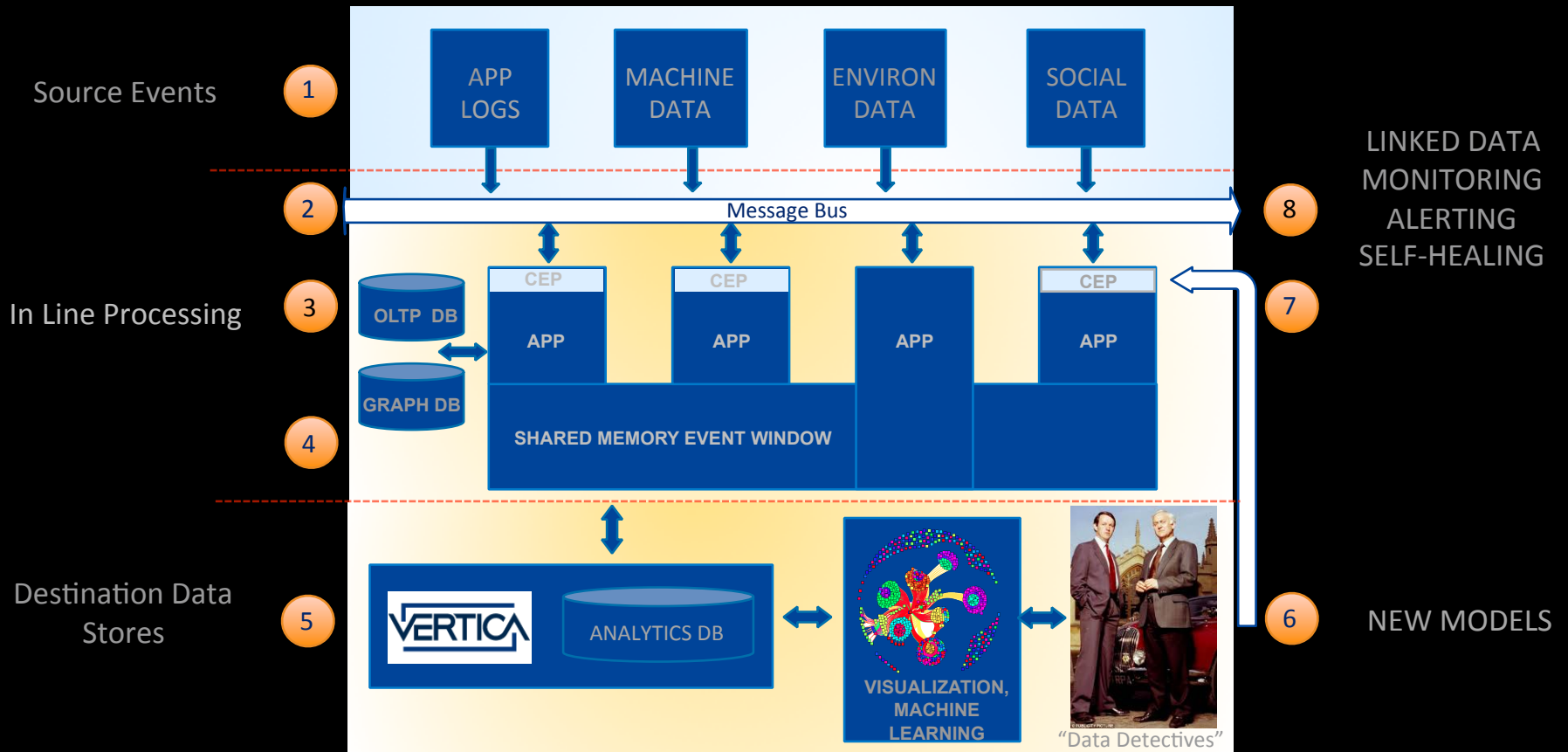
RDMA, clustered file systems for reduced copy times.

Inline stream analytics using CEP, ESP and patented technologies.

Downstream analytics environments for model building and further processing.



Systems Intelligence Flow



“BABAR” BLENDED SYSTEM

HPC infrastructure:

- 3712 cores
- 22Tb memory
- 1.2 P storage
- Infiniband (FDR/QDR)

Enterprise technologies:

- messaging
- OLTP/OLAP
- command & control
- security
- hadoop environment



Those were the big machines....

What about the little ones?

Innovative idea evolved with the m800 cartridge..

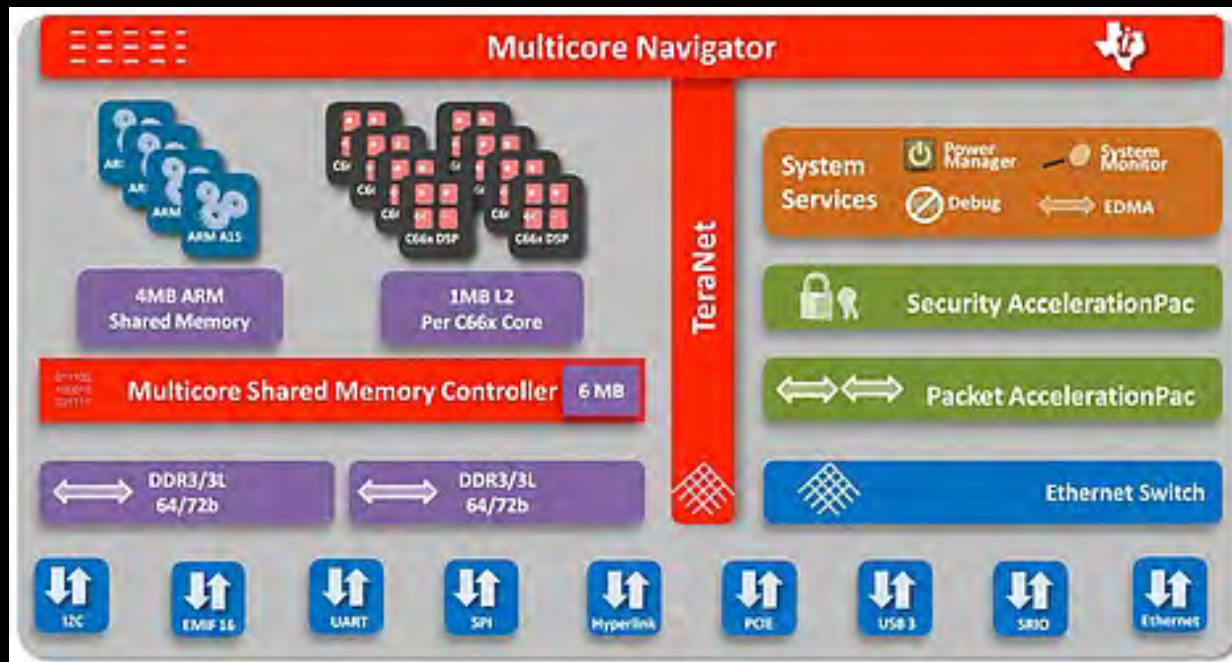


**Hewlett Packard
Enterprise**



©2016 PayPal Inc. All rights reserved. Confidential and proprietary.

... HPC in a SoC ...



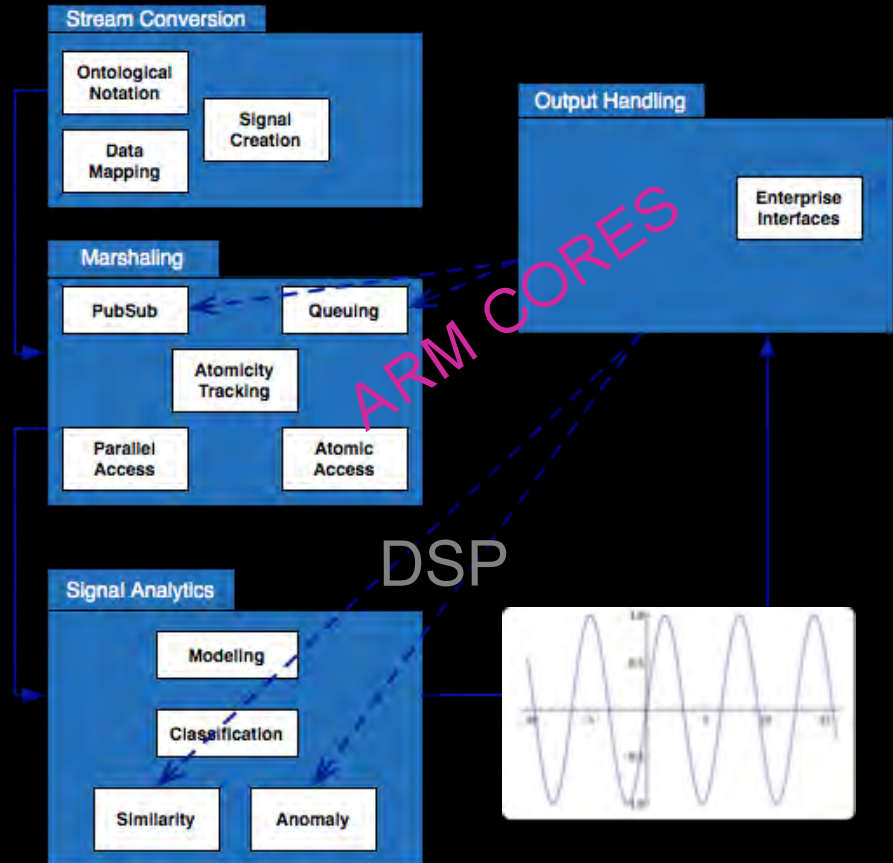
Complex Event Processing as Digital Signals

Familiar Systems Integration (ARM)

- Linux for general purpose work
 - integrating with enterprise systems (databases, marshaling, command & control)
 - short development learning curve (python, java, openCL, openMPI)

Efficient, Real-Time Parallel Processing

- Implement signal analysis in hardware
 - solve encoding, marshaling, atomicity
 - apply both global shared memory and scale-out process best practices
 - leverage cross-platform development to decrease ramp-up and testing time (openCL)



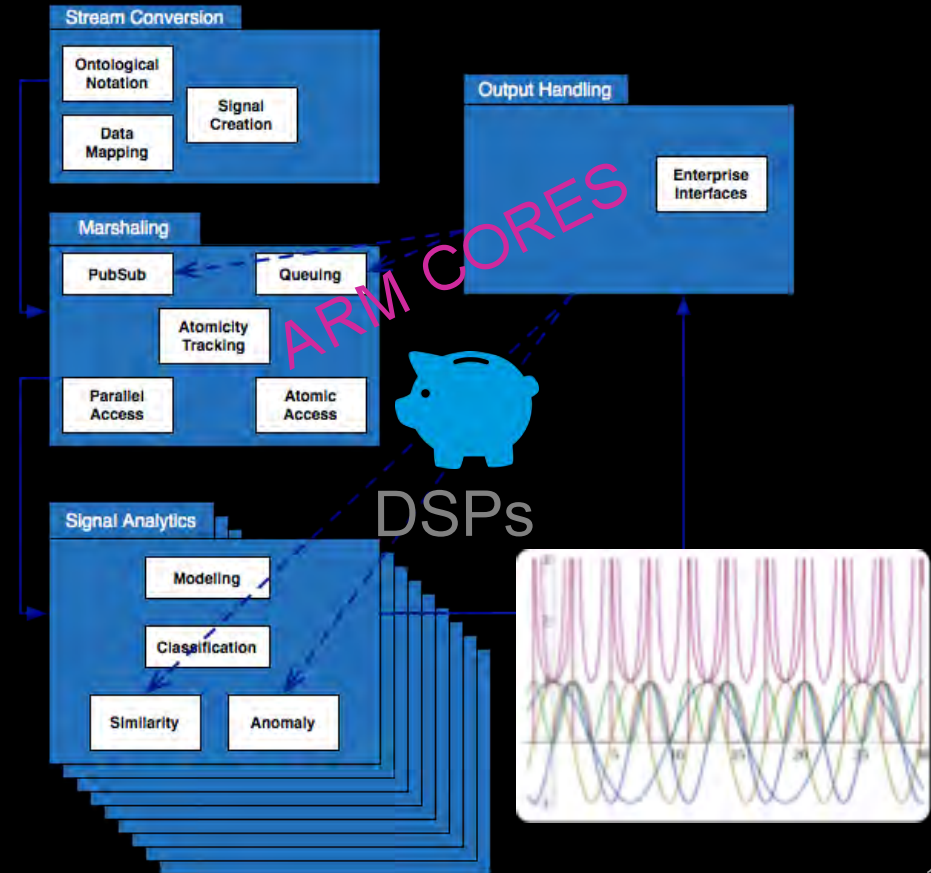
Complex Event Processing as Digital Signals

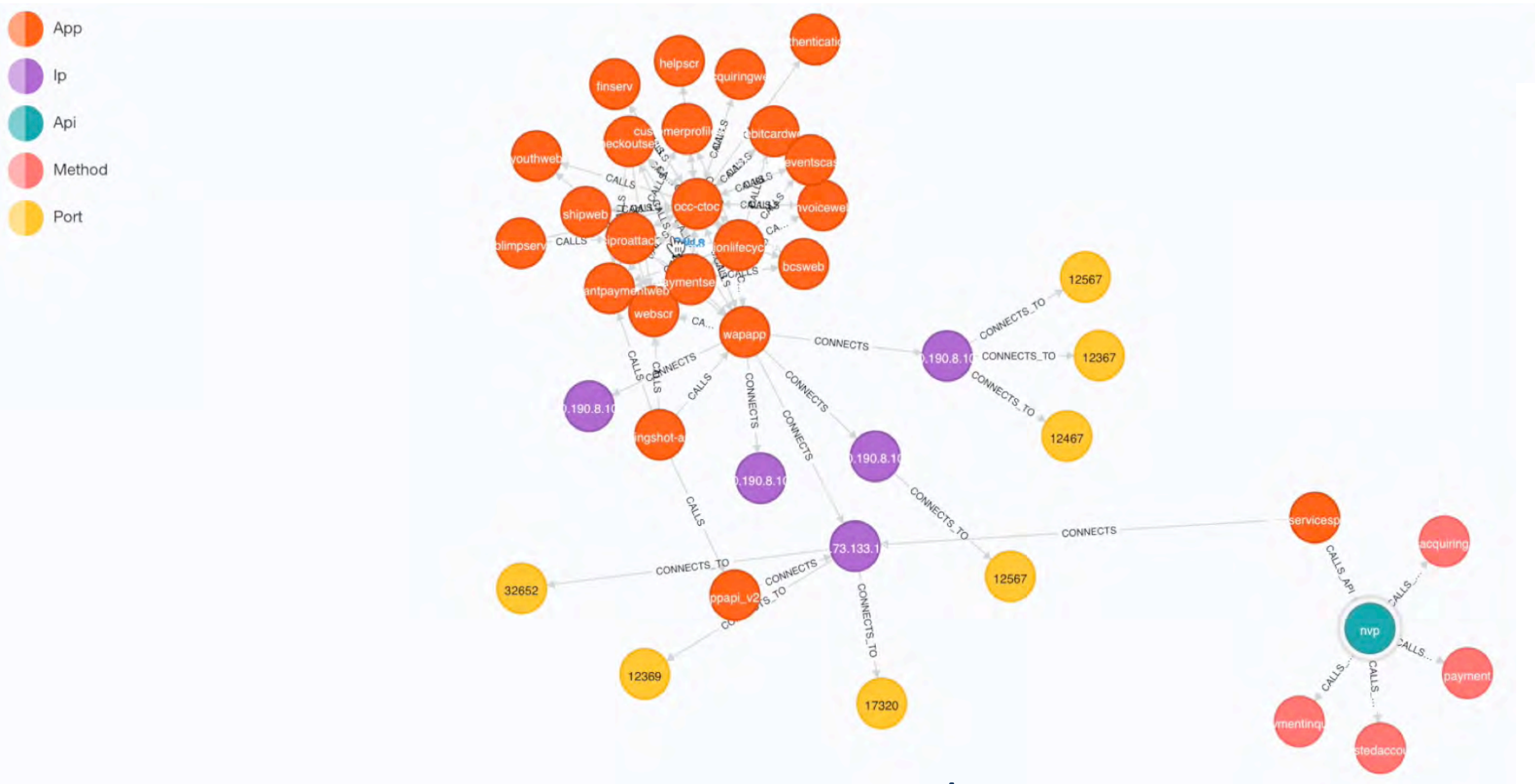
Parallel, True Real-Time Analytics

- Multiple filters/atomic event stream
- Multiple streams/filter
- Multiple filters/multiple streams
- Pattern recognition (outliers, clusters, frequency matrices, etc)
- Rich library of functions (notch/high pass/low pass filters, DFT/FFT, z-,bilinear- transform, etc.)

HPC and Enterprise Best Practices

- Multicore implementation
- Tiered shared memory and queuing
- High-speed, low-latency transports inter/intra SoC
- Support for common development libraries and standards (openCL, openMP/MPI)
- Efficient, low-power solutions
- Extreme performance (**11.2 GF/watt**)





Use Case: Flow Analysis/Graphing

Benefits derived from HPC & Systems Intelligence

Visibility into the health of the complete ecosystem all the time.

Self healing of potential issues through predictive models and remediation rule sets.

Modeling the entire PayPal system into a Linked Data paradigm for future use.

Less reliance on humans. Computers don't need sleep.

Ecosystem has become too complex for humans to comprehend.

Cost benefit to business – running a more efficient system, up to its capabilities, not its capacity.



HPC and enterprise technologies work together,
but it hasn't been easy.....

“Most life sciences researchers are not HPC savvy.”

Dr. Jack Collins, National Cancer Institute – Oct 2015

Is it really worth it?

Challenges

Technical expertise
Reliability first, performance second
Social aspects of alternate tech

“Point of Spear” Syndrome

Victories

Revolutionary technology solutions
Resets scaling base, allows new problems
Leverage new resources across the organization

Challenge: Social Aspects

Misunderstanding of what HPC is.

“I just bought a cloud... why do I to buy this HPC thing now?”

People are comfortable with what they know and very resistant to change.

No ability to apply new technology to existing problems in a different way.

In terms of analytics – shortage of analysts with technical skills.

A world map with numerous blue circular data points of varying sizes scattered across the continents. The points are labeled with numbers, representing data density or volume in different geographical regions. The map is overlaid with text and a logo.

What are we up to next in analytics?

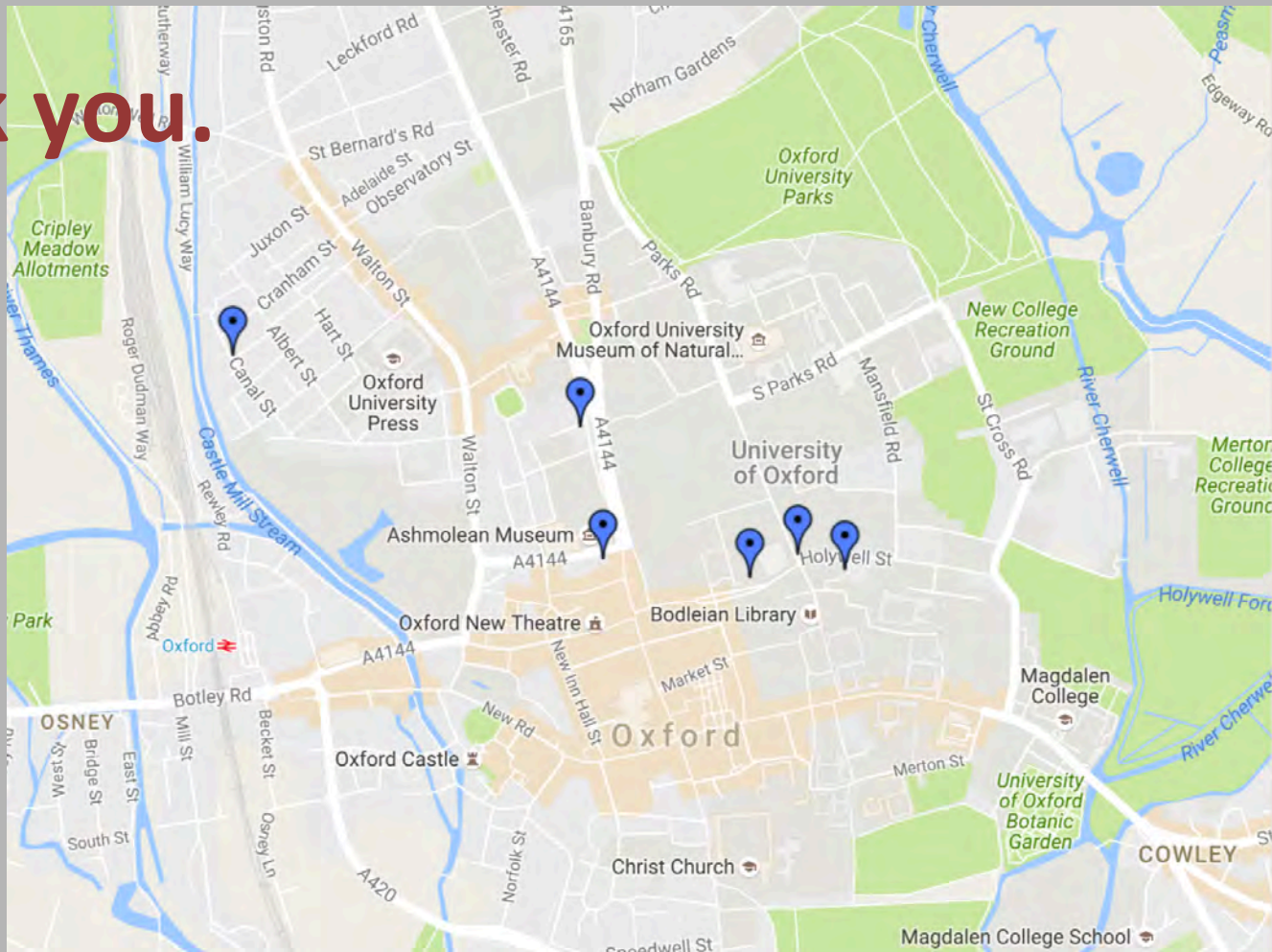
Paradigm shift in graph technologies for real time analytics.

The wire as a data space. “Analytics on data in motion.”

Preality. “Pre-reality” – encompassing predictive analytics with real time event processing.

Will continue to require hybrid enterprise and HPC architectures.

Thank you.



akolster@paypal.com