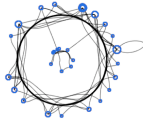
ata-driven pipelines high performance computing machine and deep learning relationship an
raph analytics streaming data analytics iot technology consulting high performance analytic
rchitecture learned and intelligent systems technology vision, strategy, and practice blockchai
nd smart contracts data ontology and design data-driven pipelines high performance computin
achine and deep learning relationship and graph analytics streaming data analytics ic
echnology consulting high performance analytics architecture infrastructure intelligent system
eal-time event-driven systems blockchain and smart contracts data ontology and design data
riven pipelines high performance computing machine and deep learning relationship and grap
nalytics streaming data analytics iot technology consulting high performance analytic
rchitecture infrastructure intelligent systems event-driven real-time analytics blockchain an
mart contracts data ontology and design data-driven pipelines high performance computin
achine messaging topologies relationship and graph analytics streaming data analytics ic
echnology consulting high performance analytics architecture infrastructure intelligent system
eal-time event-driven systems blockchain and smart contracts data ontology and design data
riven pipelines high performance computing machine and deep learning relationship and grap
nalytics streaming data analytics iot technology consulting high performance analytic
rchitecture infrastructure intelligent systems real-time event-driven systems blockchain an

Providentia Worldwide



Perspectives On HPC And Enterprise High Performance Data Analytics

Arno Kolster

CRAY
THE SUPERCOMPUTER COMPANY

Providentia Worldwide

May 10, 2017

A Little Bit About Me…

30 years in database architecture and infrastructure operations

Until recently, at PayPal for 13 years with a focus on database, analytics and operations architecture

Career interest turned to HPC to solve real time, web scale analytics problems

Co-recipient of IDC HPC Innovation Award at SC 12 and SC 14 with colleague Ryan Quick

HPC User Forum and Leverage Big Data steering committee member

TECH.CO

Tech Startup News, Events & Resources

NEWS CHANNELS CITIES EVENTS ECOSYSTEM RESOURCES JOBS

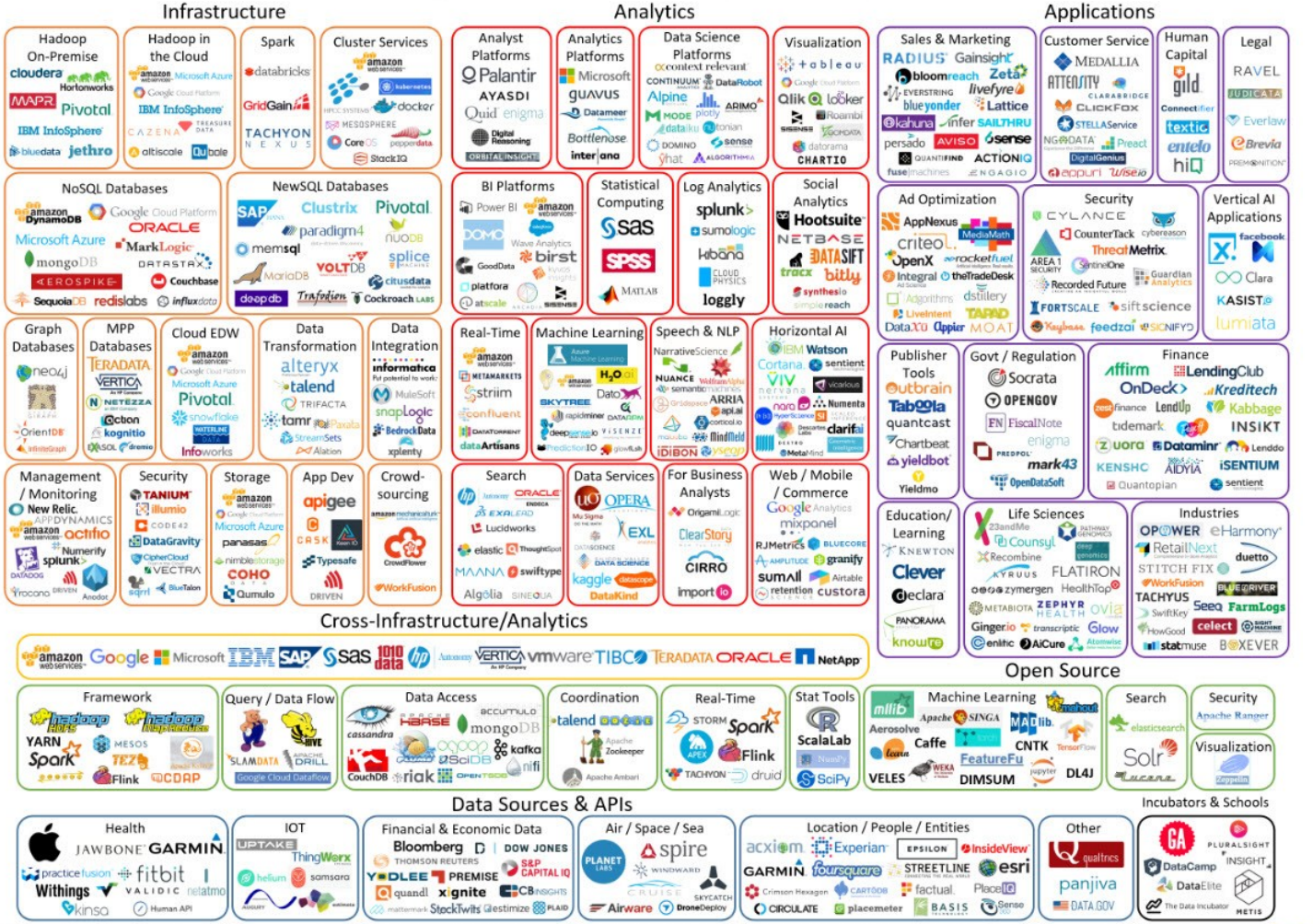
Search

Facebook Admits AI Can't Do Everything Yet and Hires 3,000 Editors

Facebook


LIFE

Big Data Landscape 2016 (Version 3.0)



Last Updated 3/23/2016

FIRSTMARK

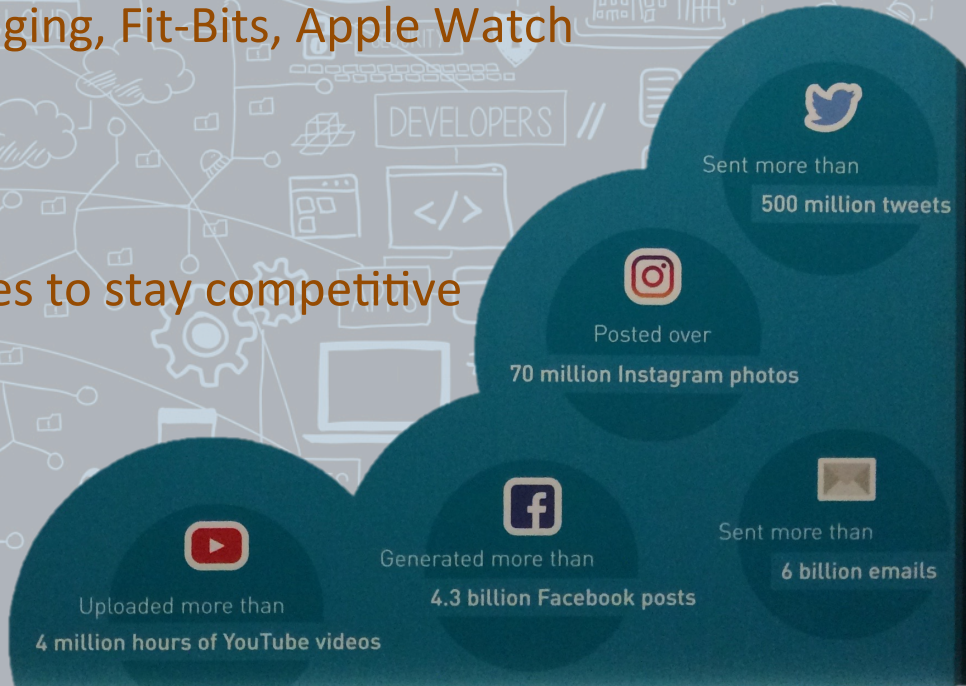
So What's Happened?

Huge amount of data being generated by humans and machines

- Rich social media and data streams
- Sensor data from IoT, medical imaging, Fit-Bits, Apple Watch
- Autonomous vehicle data
- Smart cities

Need to analyze data faster for businesses to stay competitive

- Customer insights
- Real time marketing
- Fraud detection



A FINELY TUNED ECOSYSTEM

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

Why The Convergence Of Enterprise And HPC?

Enterprise is analyzing increasing larger data sets rivaling HPC centers

Hardware and software within budgets to tackle some of these problems

Open source and commercial product availability

Lower latency, real time analytics driven by the need to make quicker decisions

Trend toward Machine Learning, Deep Learning and AI fueling innovation and First-To-Market

Enterprise Needs HPC

Ability to use shared memory for large data sets

Apply modeling and simulation to traditional reactive analytics

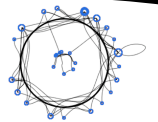
Global, parallel file systems can be accessed from disparate systems

RDMA and distributed memory access

Leverage high speed interconnects to move data faster

Capability of running atomic jobs in parallel

Rich collaboration history



HPC Needs Enterprise

Enterprise has resiliency and availability ‘down’

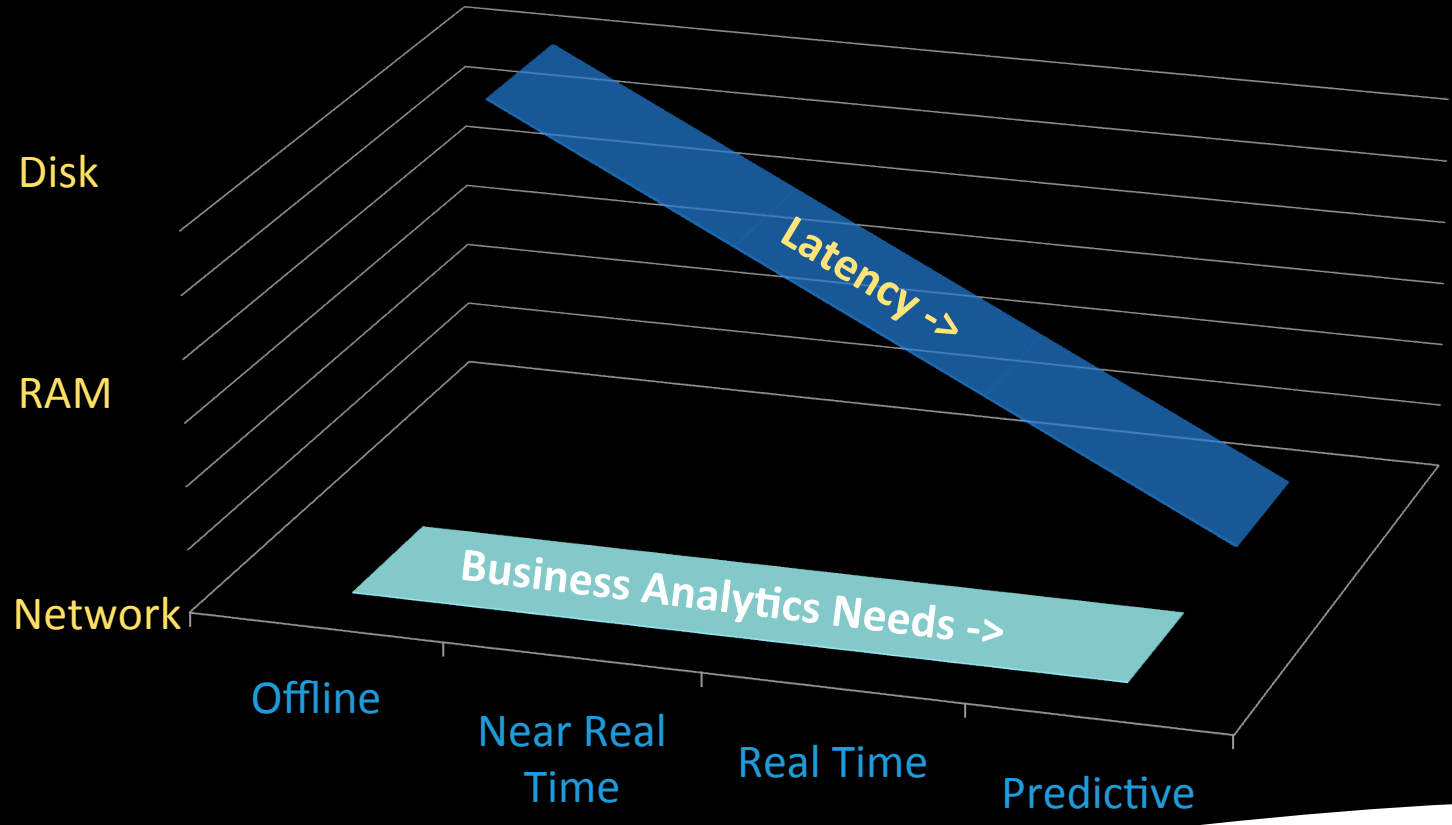
Checkpointing and restarting are antiquated standards and better alternatives exist

Open source and shift in development paradigms provide much faster time-to-market

Distributed computing i.e. cloud is now second nature

Enterprises by their nature create new markets which provide opportunity to HPC vendors

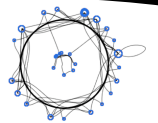
Business Is Driving The Speed Of Analytics



Providentia Worldwide

Let’s Talk About Real Time

“Complex Event Processing as Digital Signals”



Real-Time Analogy

Everyone likes to go to concerts…

The Concert Experience

You’re at the concert listening to your favorite piece of music or song.

You’re really enjoying yourself, you’ve had a glass of wine, you’re feeling the groove with the musicians…

Suddenly you hear a bad note.

But it’s a concert, the show goes on, you ignore it and you have another sip of wine.

The Concert Experience

But…what just happened?

You analyzed data in real time.

Not a second later, not a minute later, not a day later…

But at the instant the event occurred.

You used predictive models to look for anomalies in the event stream, **in REAL TIME .**

So… how do we do that?

Faster 

Cheaper 

Greener 

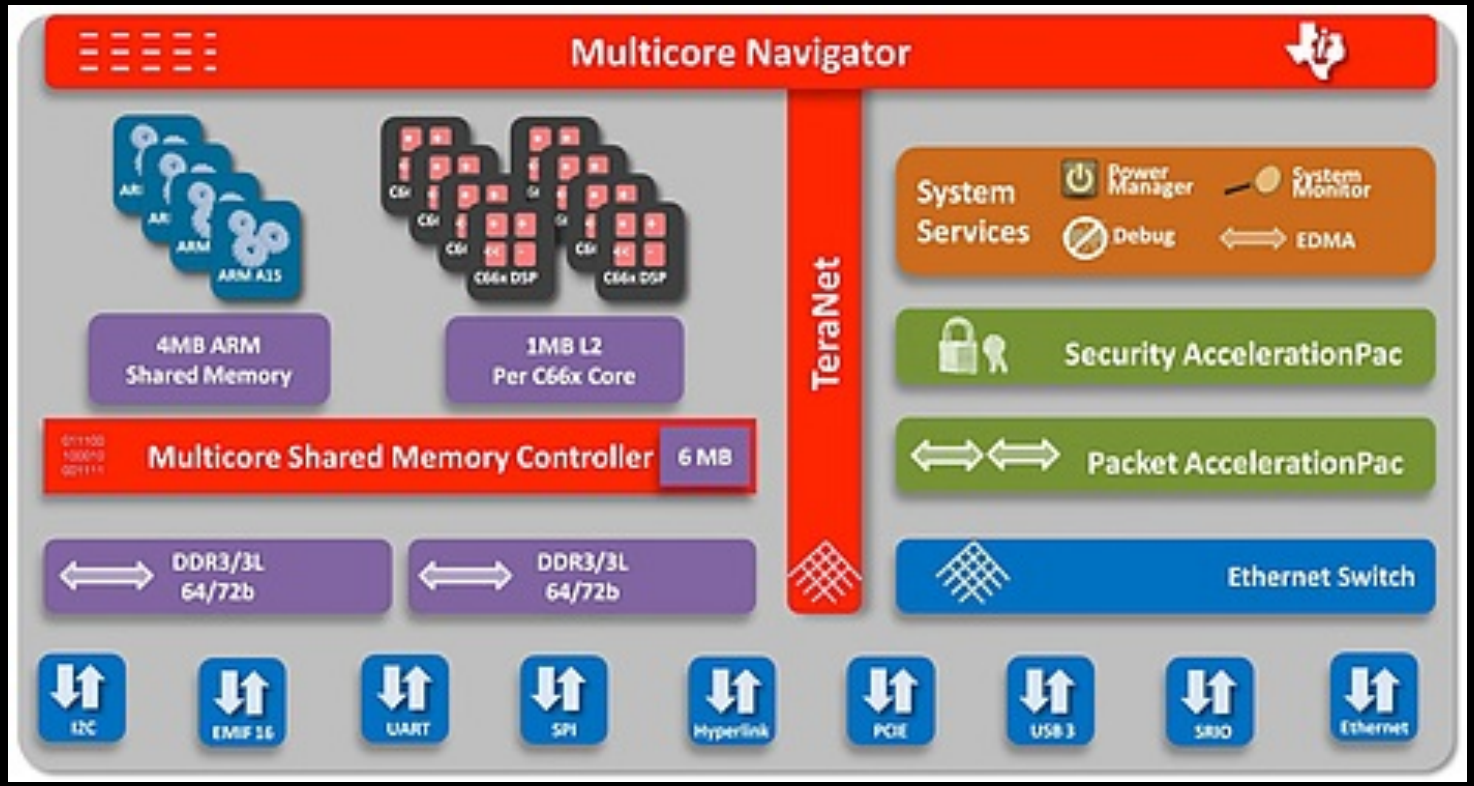
You just add little machines to the BIG ones…

Innovative idea evolved with HPE's m800 cartridge..

66AK2Ex
KeyStone
Multicore DSP+ARM®

Texas Instruments

… HPC in a SoC …


Multicore Navigator
ARM A15
C66x DSP
4MB ARM
Shared Memory
1MB L2
Per C66x Core
TeraNet
System
Services
Power
Manager
System
Monitor
Debug
EDMA
Security AccelerationPac
Packet AccelerationPac
Multicore Shared Memory Controller
6 MB
DDR3/3L
64/72b
DDR3/3L
64/72b
Ethernet Switch
I2C
EMIF 16
UART
SPI
Hyperlink
PCIE
USB 3
SRIO
Ethernet

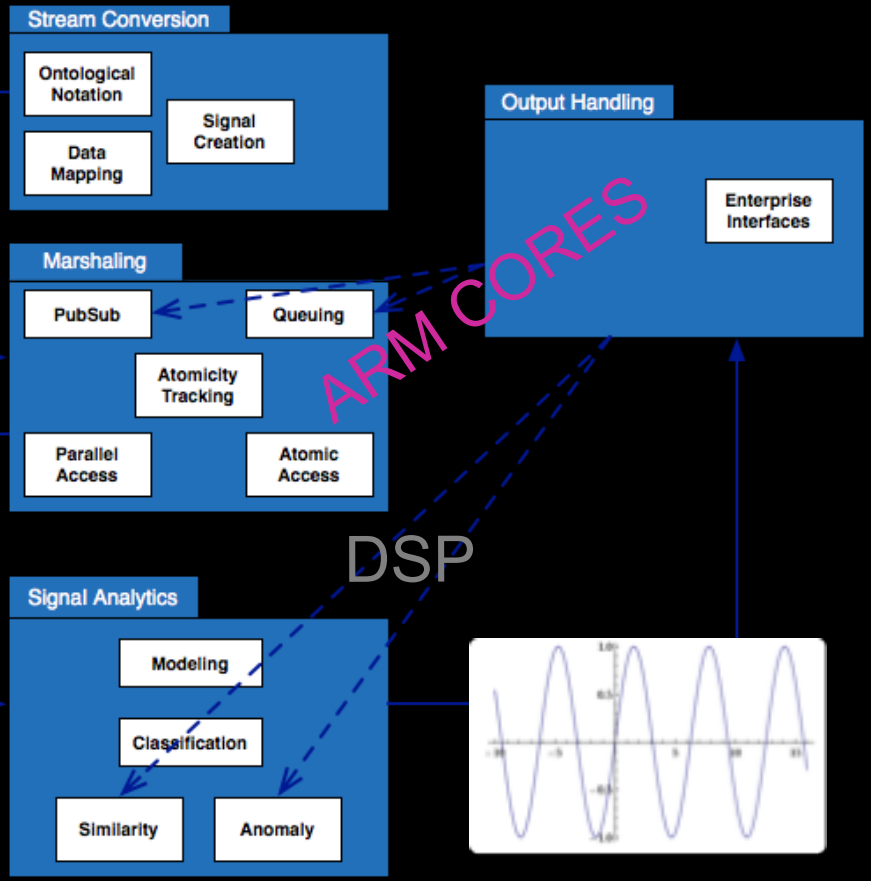
Complex Event Processing as Digital Signals

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)



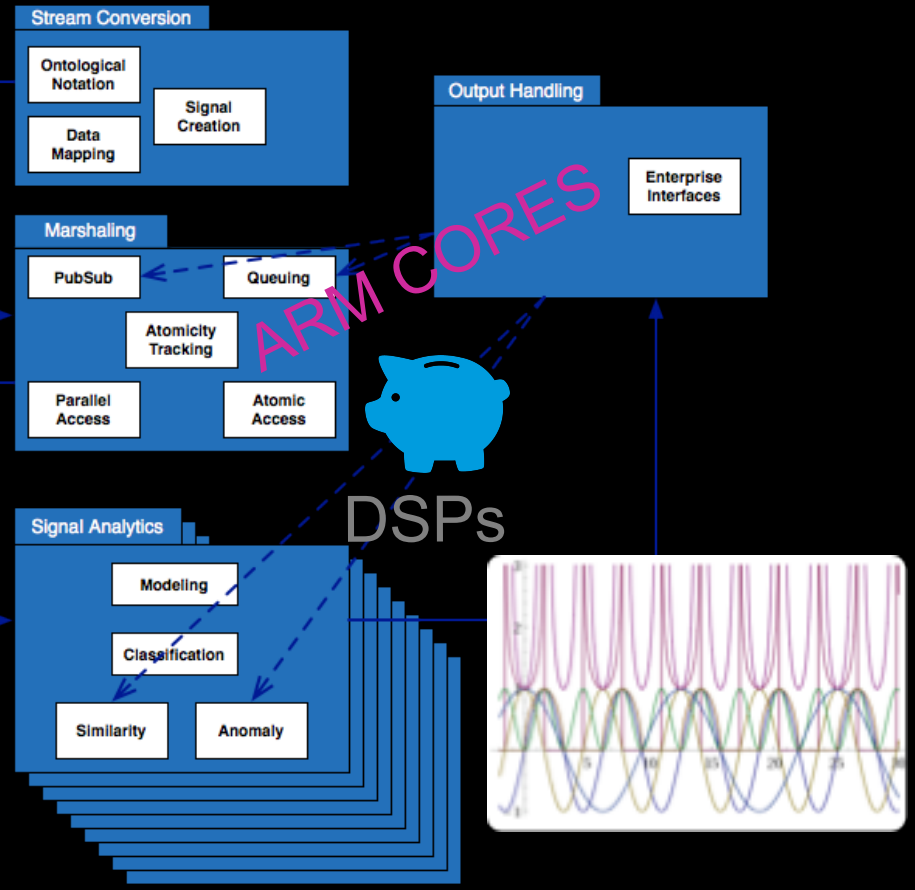
Complex Event Processing as Digital Signals

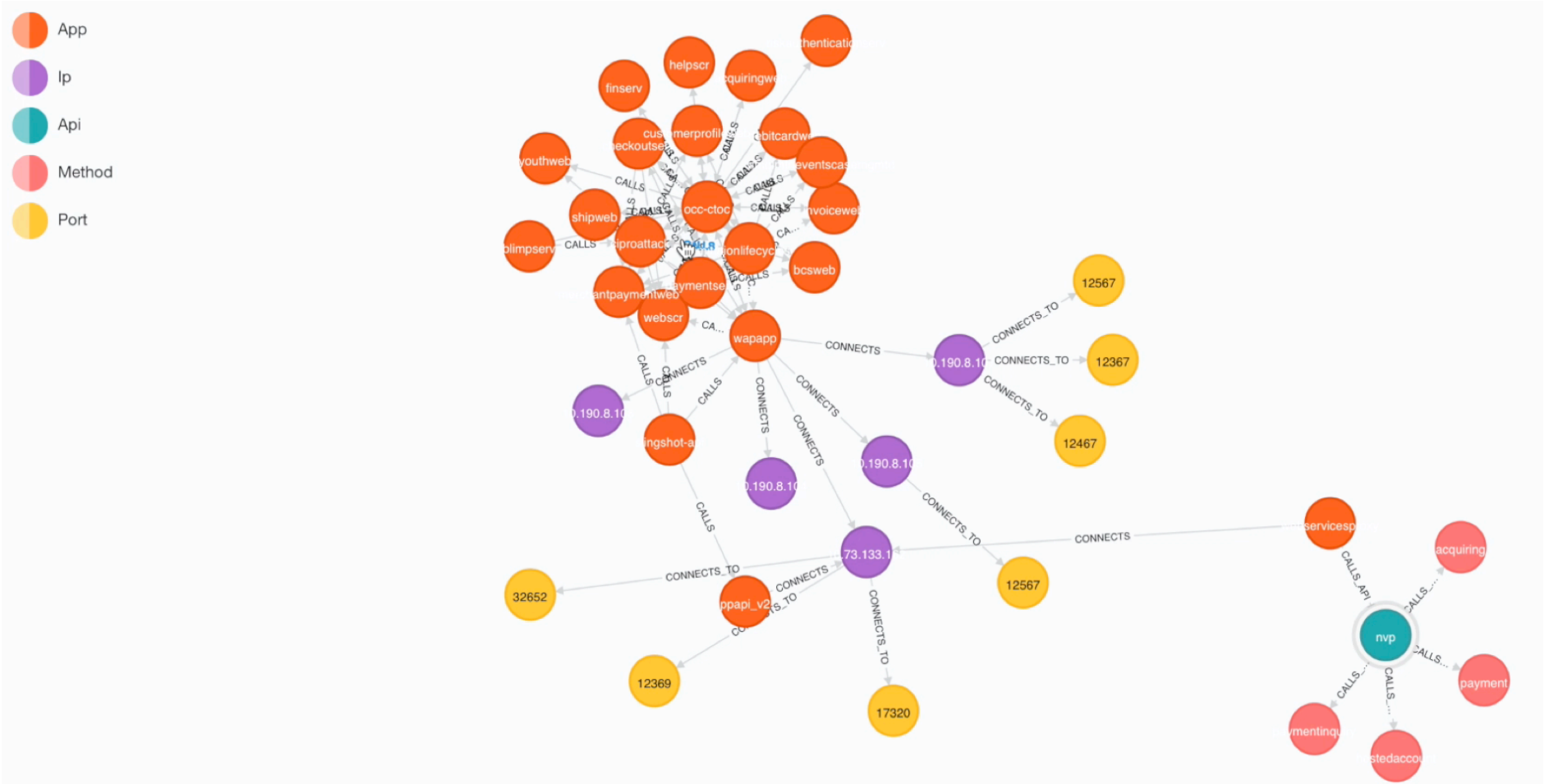
Parallel, True Real-Time Analytics

- 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 (**~55W/cartridge (4 SoCs / cartridge**)
- Extreme performance (**11.2 GF/watt**)





Use Case: Flow Analysis/Graphing

Cray Urika-GX

Cray Graph Engine (CGE) - semantic graph, RDF

Supercomputer technology - large shared memory, Aries fabric

Multi-node clusters up to 48 nodes, 1728 cores

Built-in open source frameworks for Big Data – Hadoop, Spark, Mesos, OpenStack

Bridging the gap between HPC and enterprise workloads

Graph Analytics

Why are graphs so cool?

They represent relationships between entities

Easily traversed with a variety of different algorithms

Visualization of a graph shows areas of interest quickly

Allow proper modeling and simulation of populations, behaviors, paths flows, etc.

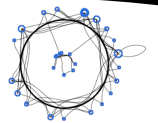
Graph Analytics Challenges

Ingestion rates still not suitable for some real time applications

Temporal analysis of networks still extremely difficult at scale

Disparate offline data sets need to be ‘massaged’ into the existing graph

Augmentation of retrieved data difficult, requires subsequent processing



The Connected World

Or

We Thought We Had Data Challenges Today….

Just Wait.

Internet of More Stuff

Estimates range from 50 to 100 billion data generating devices by 2021

Distributed networks of micro services and imbedded devices

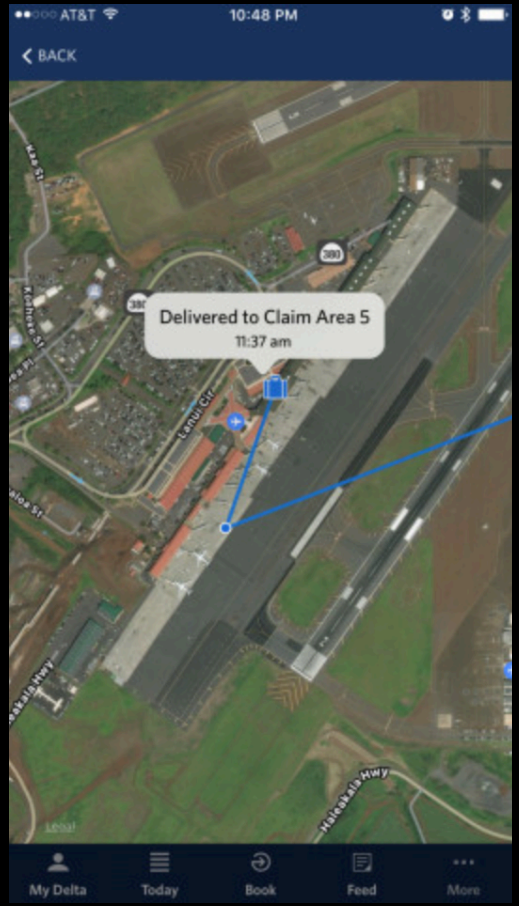
Connected devices - phones, homes, humans and toaster ovens

Security vulnerabilities and privacy concerns abound

Standards for wireless protocols and data format specs are still in their infancy

Will we aggregate and analyze at the edge? Most definitely.


CANADA
BRITISH COLUMBIA
SASKATCHEWAN
MANITOBA
ONTARIO
QUEBEC
NEW BRUNSWICK
Lake Winnipeg
Lake Superior
Lake Michigan
Lake Huron
Lake Ontario
Lake Erie
Vancouver
Victoria
Calgary
Regina
Swift Current
Winnipeg
Moosonee
Timmins
Amos
Quebec
Fredericton
Montreal
Ottawa
Kingston
Toronto
WASHINGTON
Seattle
Olympia
Spokane
Yakima
Vancouver
Portland
Salem
OREGON
Medford
Klamath Falls
Eureka
Redding
Chico
IDAHO
Lewiston
Boise
Idaho Falls
Twin Falls
MONTANA
Missoula
Great Falls
Helena
Butte
Billings
Miles City
Missouri
NORTH DAKOTA
Minot
Bismarck
Fargo
Grand Forks
SOUTH DAKOTA
Rapid City
Pierre
Sioux Falls
WYOMING
Sheridan
Casper
Rock Springs
Cheyenne
Buffalo
NEVADA
Reno
Carson City
Elko
Ely
Las Vegas
UTAH
Salt Lake City
Cedar City
Monticello
Marble Canyon
COLORADO
Fort Collins
Boulder
Denver
Grand Junction
Colorado Springs
Pueblo
Durango
Colorado
Arkansas
CALIFORNIA
San Francisco
San Jose
Sacramento
Oakland
Modesto
Monterey
Fresno
Visalia
Bakersfield
Santa Maria
Los Angeles
San Bernardino
San Diego
ARIZONA
Flagstaff
Phoenix
Tucson
NEW MEXICO
Santa Fe
Albuquerque
Roswell
Las Cruces
Carlsbad
El Paso
NEBRASKA
Scottsbluff
Omaha
Lincoln
KANSAS
U.S.A.
Topeka
Dodge City
Wichita
Kansas City
OKLAHOMA
Guymon
Tulsa
Oklahoma City
Canadian
Red
TEXAS
Amarillo
Wichita Falls
Dallas
Big Spring
Odessa
Waco
Temple
Austin
San Antonio
Houston
Beaumont
Galveston
Del Rio
Laredo
Corpus Christi
Texarkana
Pecos
Rio Grande
MINNESOTA
Duluth
St. Paul
Minneapolis
Mankato
IOWA
Sioux City
Mason City
Ames
Des Moines
WISCONSIN
Ironwood
Marinette
Green Bay
Madison
Milwaukee
Mississippi
MICHIGAN
Marquette
Lansing
Detroit
Ann Arbor
ILLINOIS
Chicago
Bloomington
Springfield
MISSOURI
Jefferson City
St. Louis
Poplar Bluff
ARKANSAS
Fayetteville
Little Rock
Hot Springs
Pine Bluff
LOUISIANA
Shreveport
Baton Rouge
New Orleans
MISSISSIPPI
Greenville
Jackson
INDIANA
Fort Wayne
Indianapolis
OHIO
Cleveland
Columbus
Cincinnati
Ohio
KENTUCKY
Louisville
Frankfort
Owensboro
TENNESSEE
Nashville
Knoxville
Chattanooga
Memphis
ALABAMA
Huntsville
Birmingham
Montgomery
Mobile
GEORGIA
Atlanta
Augusta
Macon
Savannah
Albany
FLORIDA
Tallahassee
Pensacola
Jacksonville
Gainesville
Daytona Beach
Orlando
Tampa
Sarasota
West Palm Beach
Miami
Key West
SOUTH CAROLINA
Greenville
Columbia
Charleston
NORTH CAROLINA
Asheville
Charlotte
Greensboro
Raleigh
VIRGINIA
Roanoke
Richmond
Norfolk
WEST VIRGINIA
Charleston
Washington, D.C.
MARYLAND
Baltimore
Annapolis
DELAWARE
Dover
PENNSYLVANIA
Harrisburg
Pittsburgh
Philadelphia
Wilmington
Erie
NEW JERSEY
Trenton
NEW YORK
New York City
Albany
Rochester
Buffalo
VERMONT
Montpelier
NEW HAMPSHIRE
Concord
MAINE
Bangor
Augusta
Portland
MASSACHUSETTS
Boston
RHODE ISLAND
Providence
CONNECTICUT
Hartford
Nassau
MEXICO
Mexicali
Hermosillo
Chihuahua
Saltillo
Monterrey
Pacific Ocean
Atlantic Ocean
Gulf of Mexico
RUSSIA
ALASKA
CANADA
Anchorage
Juneau
0 500 KM
0 500 Miles
HAWAII
Honolulu
0 100 KM
United States of America
State Boundary
Road
River
National Capital
State Capital
City or Town
0 300 KM
0 300 Miles
© 2007 Geology.com



Delta's Checked Bag Tracker

RFID baggage tags

180 Million* passengers annually

50% of passengers check 1 bag – 90M

50% use the app – 45M

check-in, on plane, off place, bag claim - 140M recs

~4.5 recs/sec

*http://news.delta.com/corporate-stats-and-facts

What Does This Allow Delta To Do?

Minimize bag loss

Customer now able to feel more confident (or not) about bag location

Data Analytics With Enterprise:

- Number of bags per passenger / per flight
- Near real time reporting on bag delivery
- Predictive ground crew resource management
- Long term trend analysis

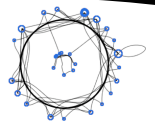
But Add In HPC…

Modeling and simulation of bag handling and impact on carousels, cargo holds

Provide bag manufacturers with information on wear and tear areas based upon analysis of bag images

Cargo loading can be based upon image data for size, shape and cargo hold volume

Discover a potential correlated attack on multiple flights based on similarity of bags



Governance of Data

This will become increasingly important

Curation and provenance of data

Where and how is it being stored?

Where did it come from?

Data integrity, privacy and security

Audit trail logging

Meta-data requirements

Barriers To Acceptance/Social Challenges

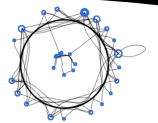
Misunderstanding of what HPC is.

"I just bought a cloud… why do I need 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.
Limited technical vision.

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



What's Next In Analytics?

Paradigm shift in graph technologies to enable 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. ML is driving AI.

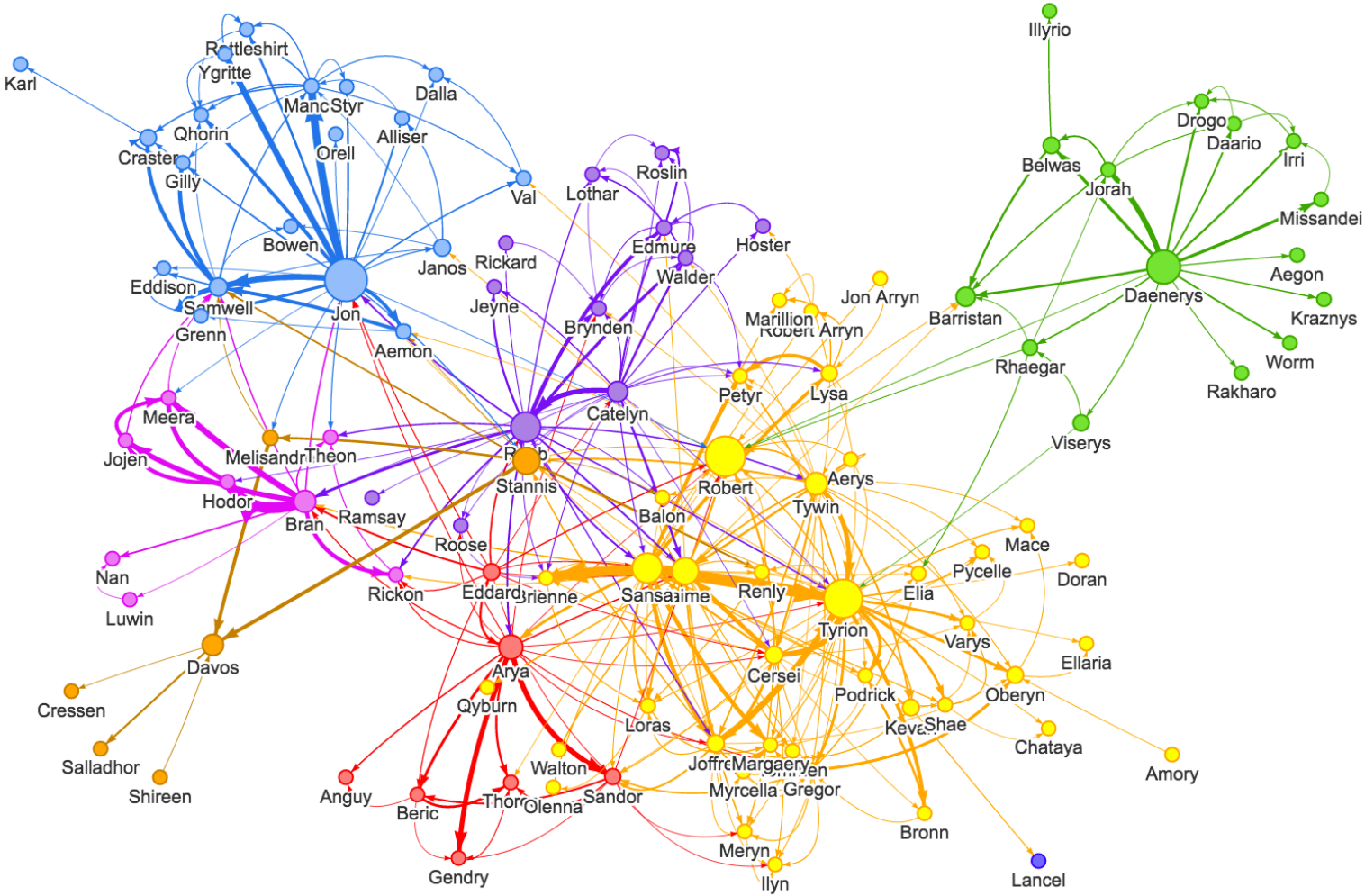
More acceptance of HPC in the enterprise with hybrid architectures.

IoT will further drive analytics on the edge.

More collaboration between the HPC centers and industry. (NSCI initiative)

http://onemilliontweetmap.com


Real time streaming
text analytics
Machine learning
Pattern matching
Graph?
GEORGE R.R. MARTIN
A GAME OF THRONES
GEORGE R.R. MARTIN
A CLASH OF KINGS
GEORGE R.R. MARTIN
A STORM OF SWORDS
GEORGE R.R. MARTIN
A FEAST FOR CROWS
GEORGE R.R. MARTIN
A DANCE WITH DRAGONS


Illyrio
Karl
Rattleshirt
Ygritte
Manc
Styr
Dalla
Qhorin
Orell
Alliser
Craster
Gilly
Val
Lothar
Roslin
Drogo
Daario
Irri
Belwas
Jorah
Missandei
Bowen
Janos
Rickard
Edmure
Hoster
Walder
Eddison
Samwell
Jon
Jeyne
Brynden
Aegon
Daenerys
Grenn
Aemon
Marillion
Jon Arryn
Barristan
Kraznys
Worm
Rhaegar
Rakharo
Petyr
Lysa
Catelyn
Viserys
Meera
Jojen
Melisandr
Theon
Stannis
Robert
Aerys
Tywin
Hodor
Bran
Ramsay
Balon
Mace
Pycelle
Doran
Roose
Nan
Rickon
Eddard
Sansa
Renly
Elia
Luwin
Tyrion
Varys
Ellaria
Davos
Arya
Cersei
Cressen
Qyburn
Podrick
Oberyn
Loras
Chataya
Salladhor
Walton
Myrcella
Gregor
Amory
Shireen
Anguy
Beric
Sandor
Olenna
Bronn
Meryn
Gendry
Ilyn
Lancel

Thank you.

arno.kolster@providentiaworldwide.com

@providentia_ww

www.providentiaworldwide.com

Special Thanks To The CUG Program Committee and

David Hancock
Yun (Helen) He
Jim Rogers