Accelerating Data Analysis with Machine Learning

BCBSM AI Case Study

Todd Alcock
alcock@lighthousecs.com

Rizwan Mallal
rmallal@forumsys.com

Mamoon Yunus
myunus@forumsys.com

Session Outline

- Company Intro and History with BCBSM
- Cisco Manual Splunk Analysis (October 2018)
- Machine Learning Techniques
- F-5 ML-assisted Analysis (June 2019)
- Quantum Similarity Platform
- Platform Demo
- Conclusions & Recommendations

Company Overview

Forum Systems 18 Years of Expertise (May 2001)

Our DNA: API, Security, Cloud, AI

Four Product lines

- Forum Sentry – API Security Gateway
- SOAPSonar – API Testing
- CloudPort – API Simulation
- *Quantum Similarity*

Locations

- HQ: Needham, MA
- Regional Office: Marina del Ray, CA
- International: 4 (DE, NL, APAC, CA),

Patents & Certifications

- API Network Appliance
- API Pen Testing
- FIPS, DoD PKI, NDPP

Sample Forum Sentry Customers

IRS
All US Electronic Tax Filings (Corporate and Individual) are processed through Forum Sentry Gateways (over $15 trillion processed without compromise since 2005).

TSYS
World's biggest credit card processing company. Forum Sentry provides data integrity, information assurance, and secure protocol translation for back-end credit card processing transactions in US and central Europe.

Kansas City Power & Light
Energy company servicing the Midwest region of the United States. Forum Sentry provides security for real-time SmartGrid communication including PKI encryption, data validation, and identity authorization.

Cincinnati Insurance Companies
Forum Sentry deployed to secure all inbound and outbound APIs for auto, home, and life insurance services.

Synovus Bank
Forum Sentry is deployed for over 100 web services and the mobile online banking infrastructure. This infrastructure services ~ 200,000 users per hour at peak times and over 11 billion transactions per year.

CDL Insurance UK
Insurance software company in UK who utilize Forum Sentry to secure all SOAP web services and mediate between HTTPS and SFTP protocols.

American Century Investments
Forum Sentry deployed to secure all inbound and outbound APIs for online trading platform.

Sample Forum Sentry Customers

Vodafone Germany
Forum Sentry enables SAML SSO and 2-factor authentication for their internal CRM system SalesForce to access internal employees and external partners.

Deutsche Telekom / T-Mobile
Forum Sentry is used to secure inbound and outbound SOAP and REST APIs utilized across various business units (mobile phone activations, account management, billing systems, etc.).

Synchrony Financial
Largest issuer of private label credit cards in the U.S. Forum Sentry provides mobile security, and identity federation integration – i.e. Lowes for Pros.

British Home Office
Forum Sentry is used for Identity and Access Management and securing various API's. UK Immigration, HMPO, UK Border Security, British Police, UK Office of Biometrics, UK EU Terrorist Monitoring.

Network Railway UK
Mobile Computing REST APIs for rapid infrastructure repair. These apps check for flaws in railway segments for rapid repair and to prevent failure.

WellCare
Medicare and Medicaid managed health care for over 4 million members. Forum Sentry provides REST API security for mobility initiatives and SSO Federation for pharmacies and healthcare exchanges.

Partners

- Strategic Partners
  - HPE (DXC) – Instrumental in generating new UK business opportunities
  - Accenture – Responsible for our largest deployment @ the IRS
  - IBM Global Services – UK
  - Harris – FAA
  - Lighthouse

- Technology Partners
  - Thales – HSM Card
  - MBX – Appliance

Security Certifications

FIPS 140-2 Level II Certified
The only certified gateway

NDPP (Network Device Protection Profile)
The only certified gateway

EAL 4+ Integrated Hardened Security
FIPS 140-2 Level III HSM

U.S. DoD PKI Component

Products

Forum Sentry
API Security Gateway

NEW!

Quantum Similarity

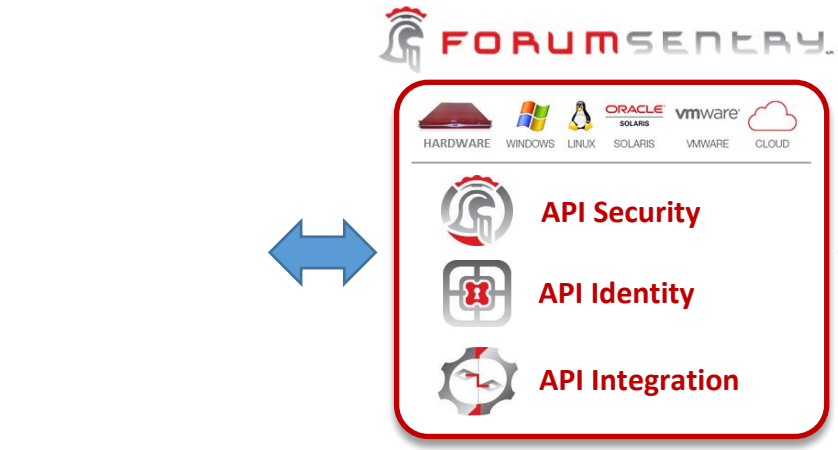
SOAPSonar
API Testing

CloudPort
Cloud Modeling

Forum Sentry – API Security Gateway

API Consumers	API Security Management	API Producers

People
Devices
Systems



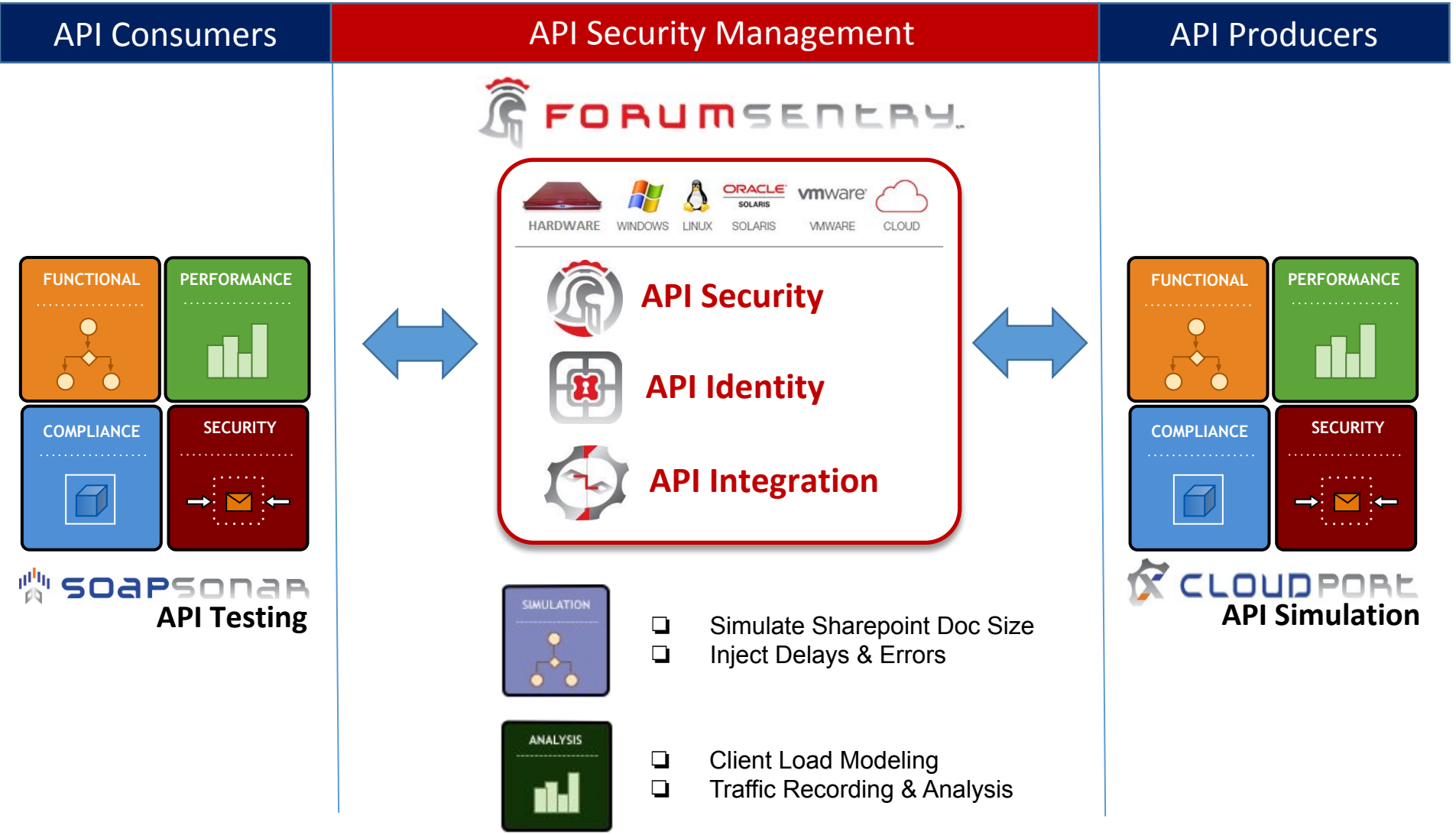
API Security:
- Threat Mitigation → AV Scanning, Rate & Size Throttling, URI Cloaking, DLP
- Privacy & Integrity → Granular message encryption and signatures

API Identity:
- Consumer & Produce Identity Tokens → OAuth, SAML, X.509
- Integration with Identity Stores → LDAP, SiteMinder, AD, Okta, Custom DB
- Map Identity Tokens → Strip/Enrich, Content-based Access

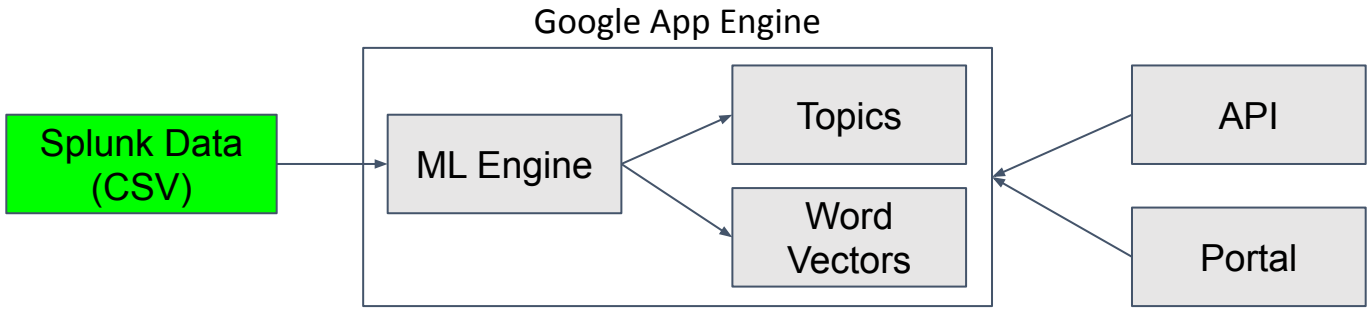
API Integration:
- Message Formats → XML, JSON, HTML, SOAP
- Transport Protocols → HTTP(S), JMS, IBM MQ, FTP(S), SMTP
- Mapping & Transformation → Bi-directional Message + Protocol

Applications
Services
Data

SOAPSonar + CloudPort – API Tools



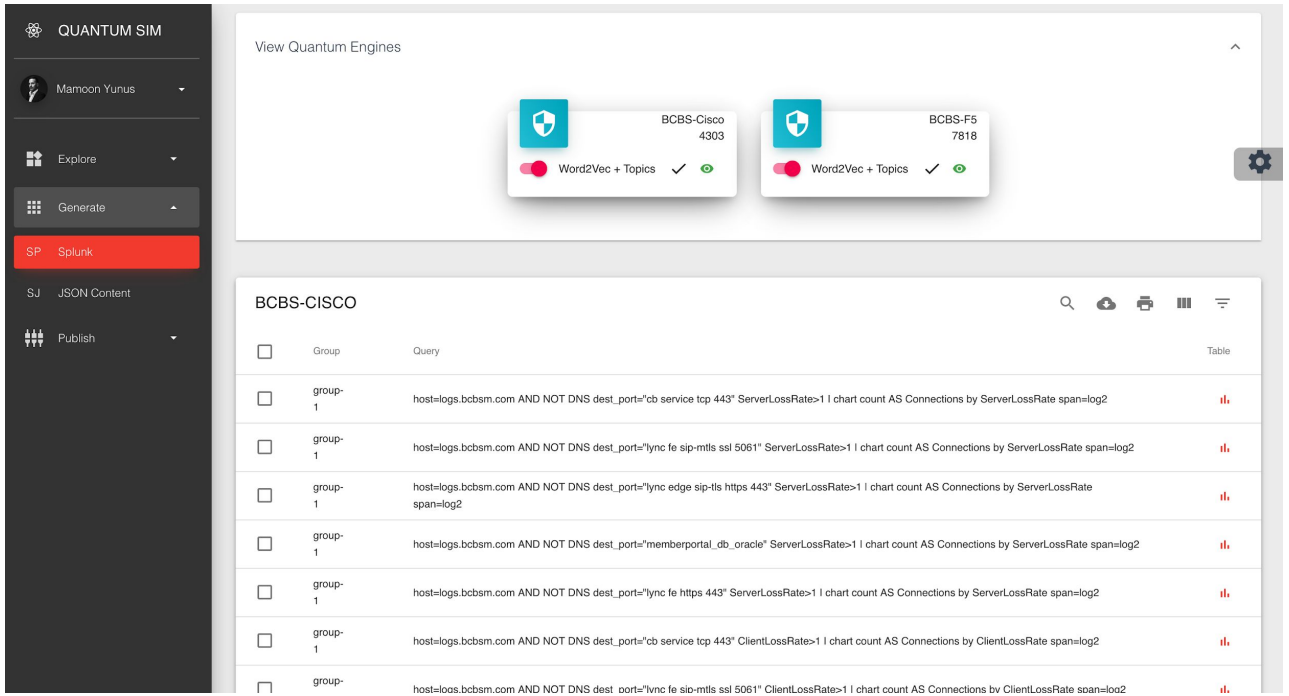
Quantum Similarity – Deep Learning Platform



- ❑ Splunk, API, DB
- ❑ CSV, JSON

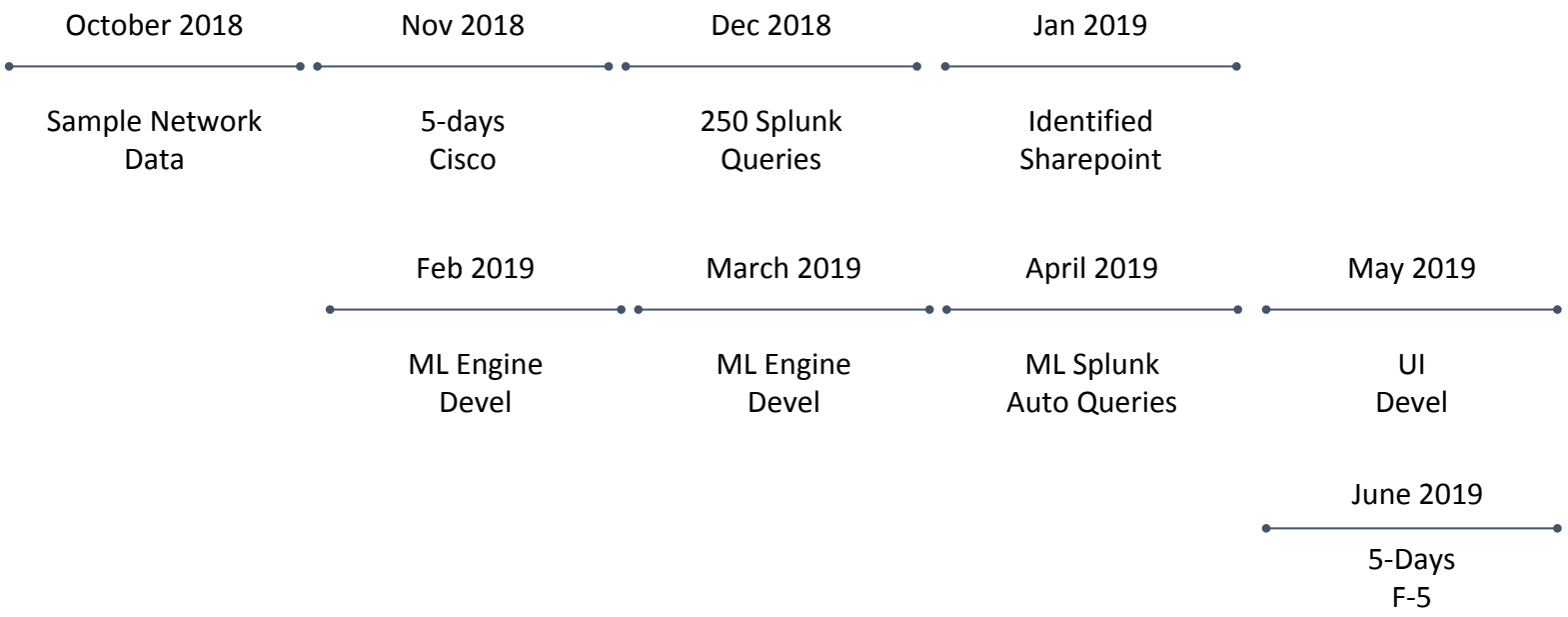
- ❑ Multi-terabyte training
- ❑ Cloud Scalable

- ❑ Portal for Reporting
- ❑ API Integration



Timeline & Milestones

Bill Fandrich engages Forum in Summer 2018 for 3rd party Network Performance Analysis



- Validated ML-generated vs Manual Splunk Queries
- Enhanced Insight from ML-generated Topics and Queries
- Completed V1 of ML Platform

Cisco Manual Analysis (October 2018)

- Manual Analysis of Splunk Data
  - Built 250+ queries across 50+ application
  - Over 200+ person-hours of analysis and query building

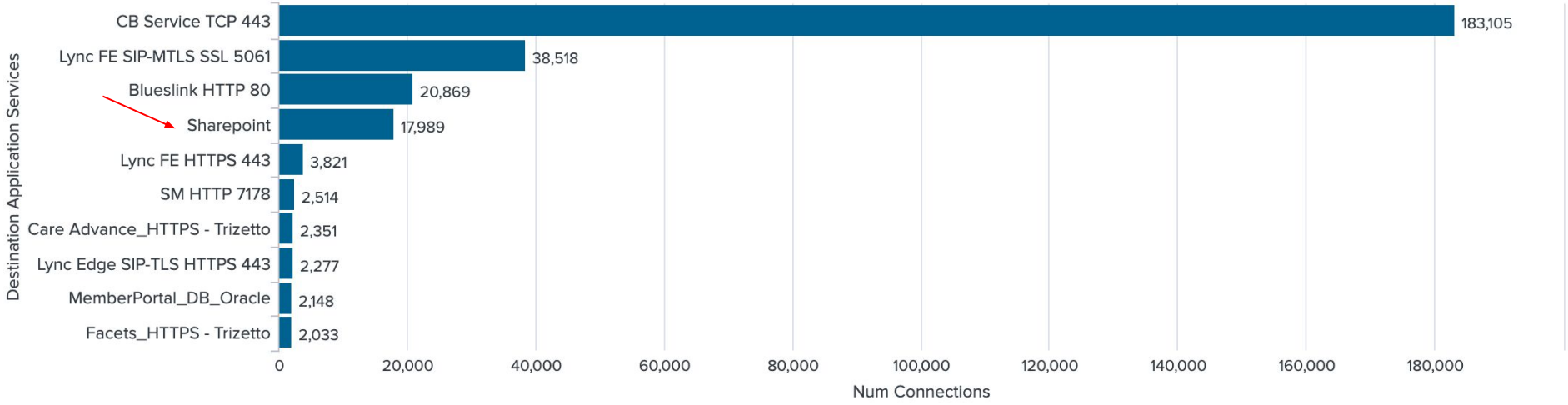
- Identified top 10 applications
  - Bytes Transferred
  - Connections

- Identified Sharepoint as primary optimization target

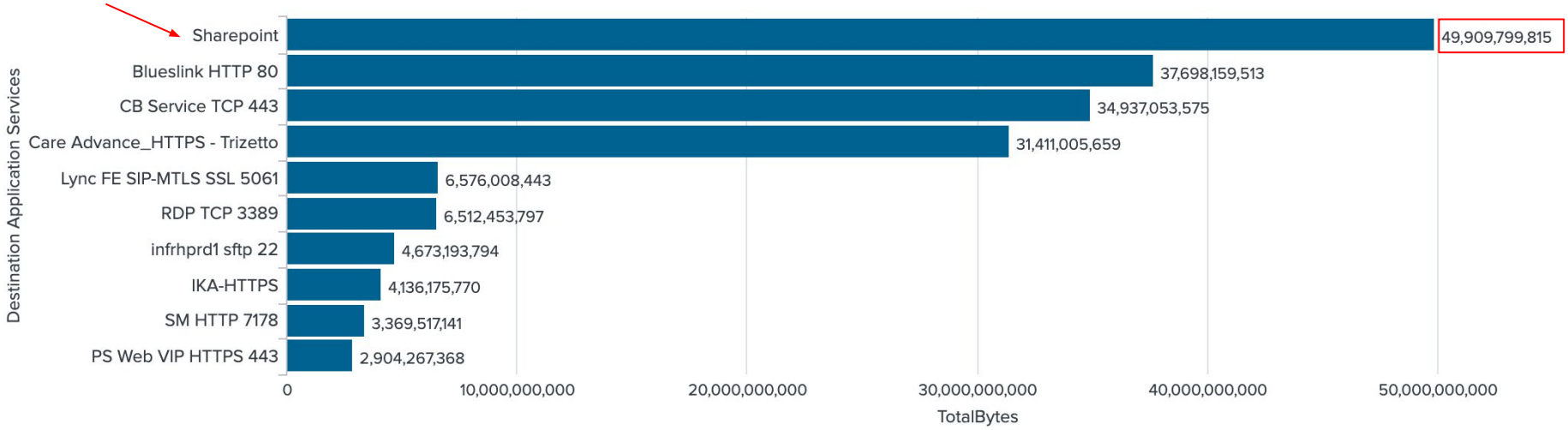
Top Applications (October 2018)

What applications are most active?

Connections

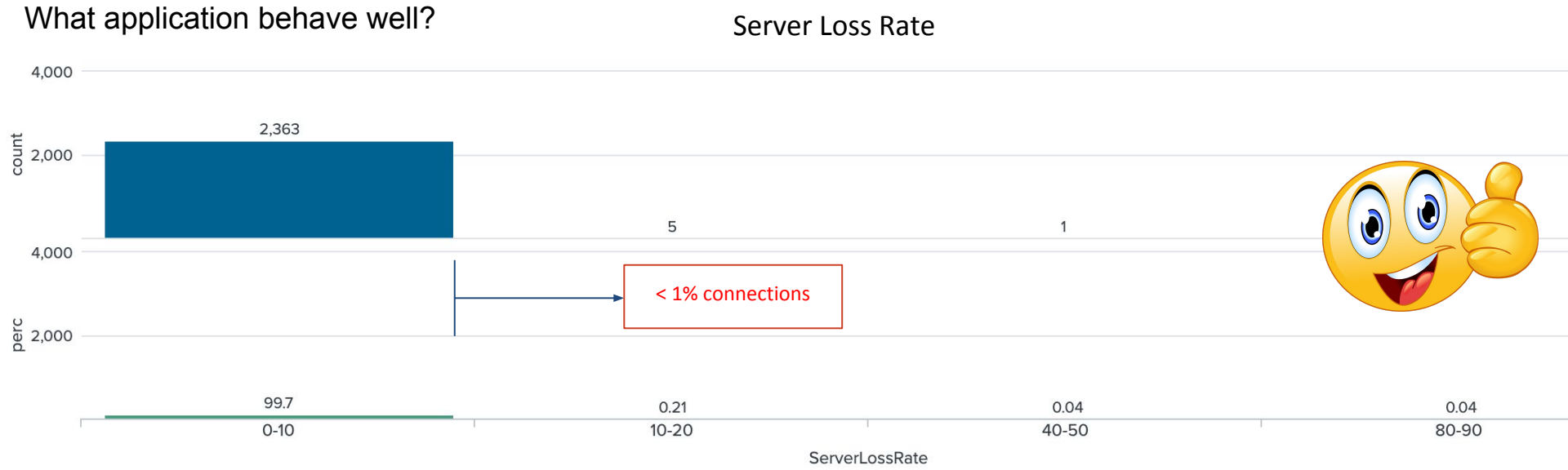


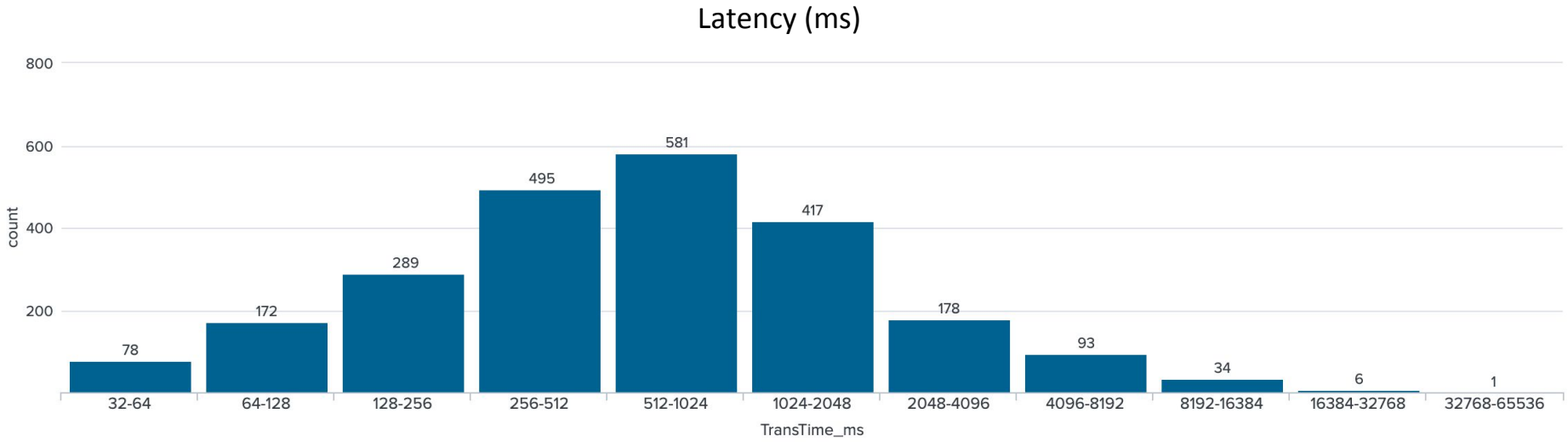
Bytes



Care Advance Trizetto (October 2018)

What application behave well?

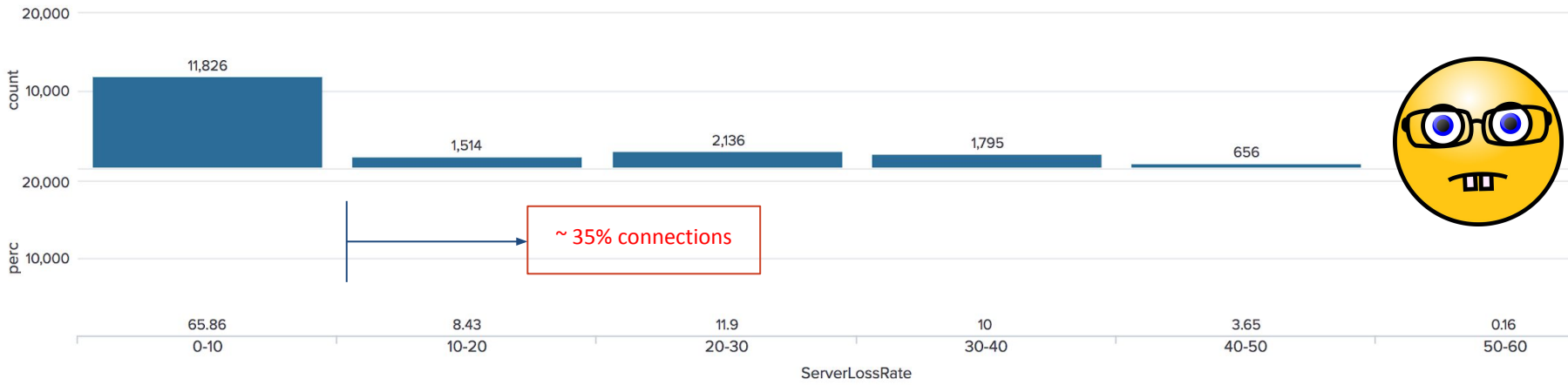




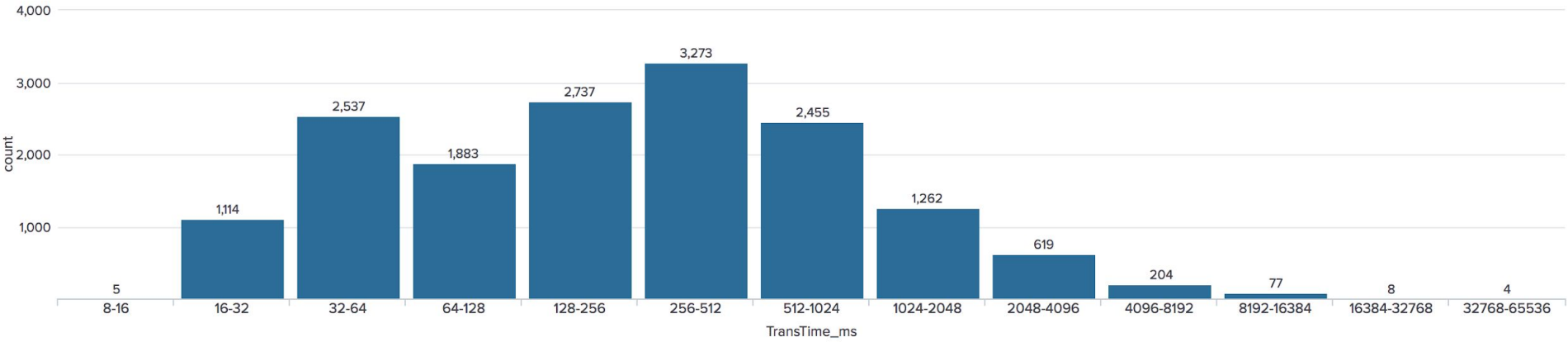
Sharepoint with Cisco (October 2018)

How does Sharepoint behave?

Server Loss Rate

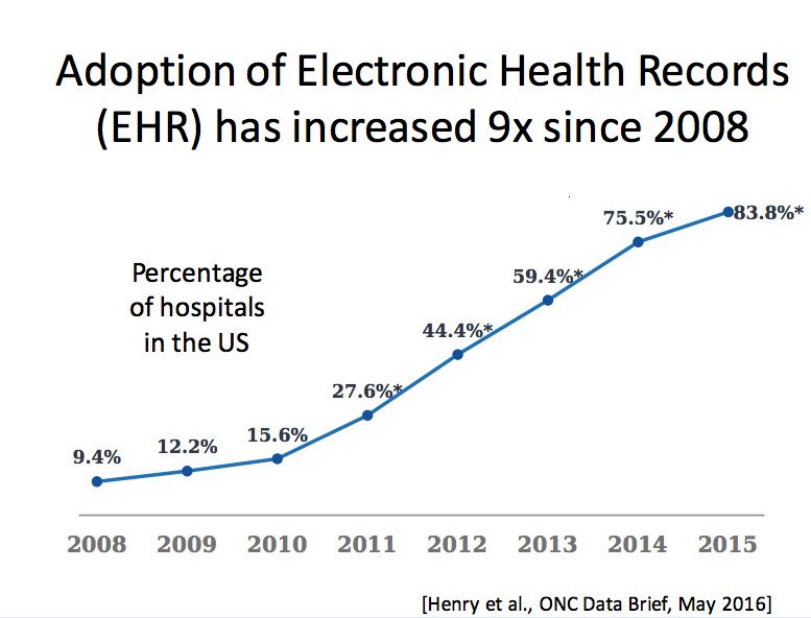


Latency (ms)



Motivation for Machine Learning in Healthcare

I. Data growth



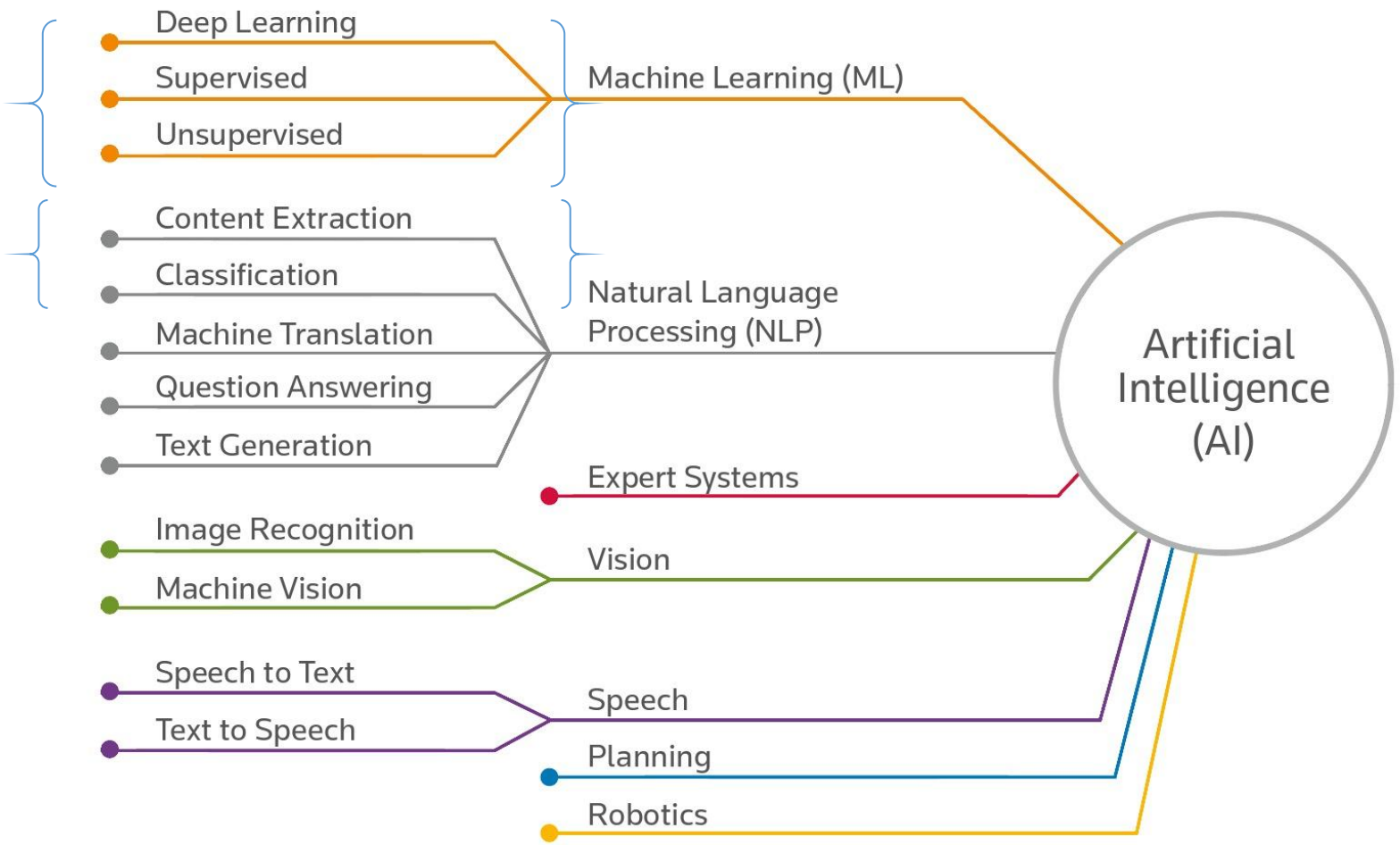
II. Affordable Computation

- ➔ IaaS multi-core CPU
- ➔ GPUs
- ➔ Quantum Computing

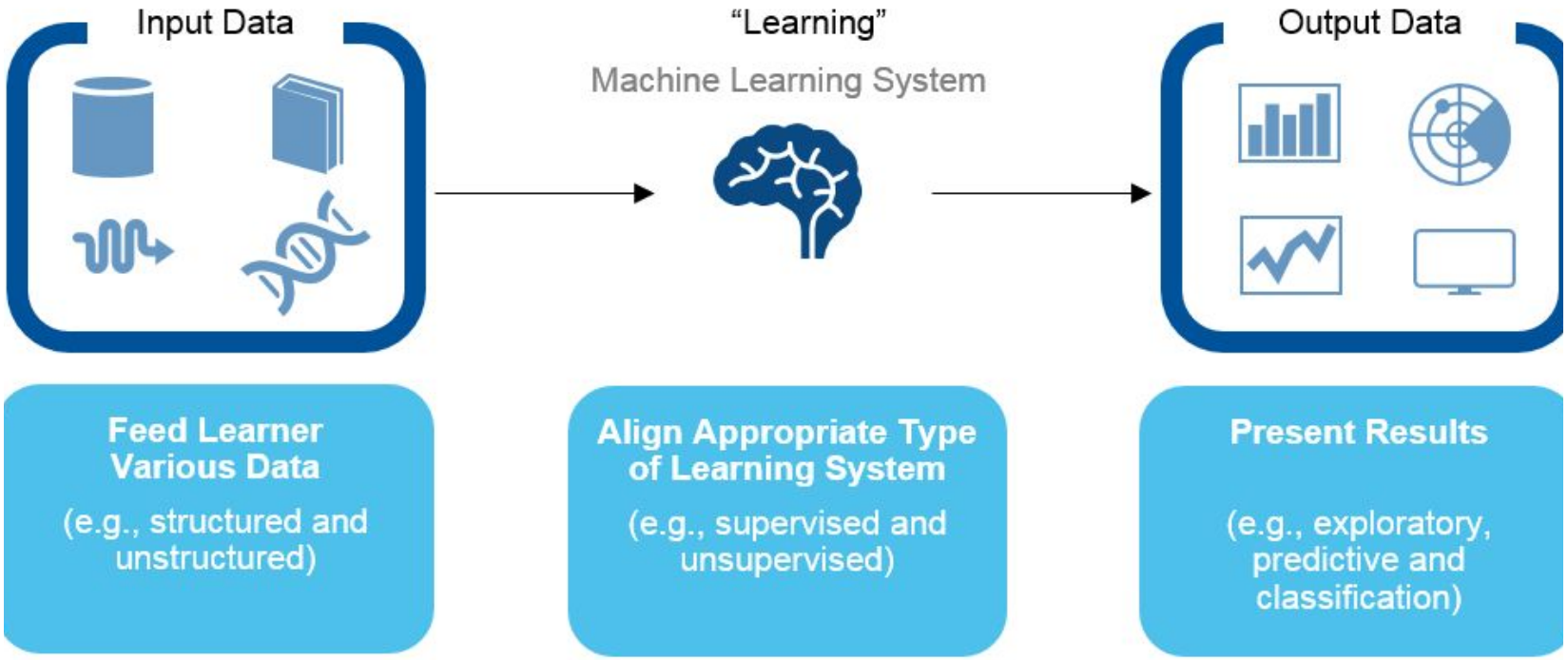
III. Democratization of ML

- ➔ high quality open-source software
- ➔ Unsupervised Learning
- ➔ Deep Learning Techniques → Transfer learning

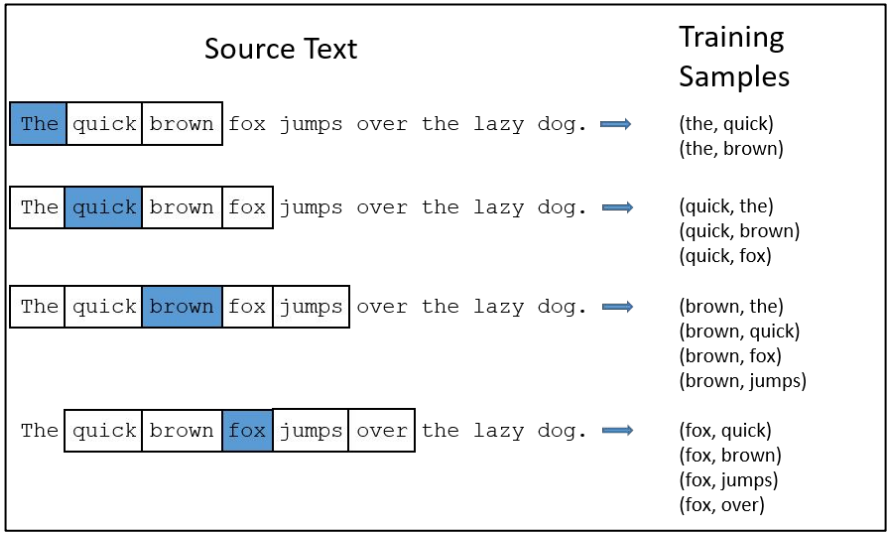
From AI to Deep learning

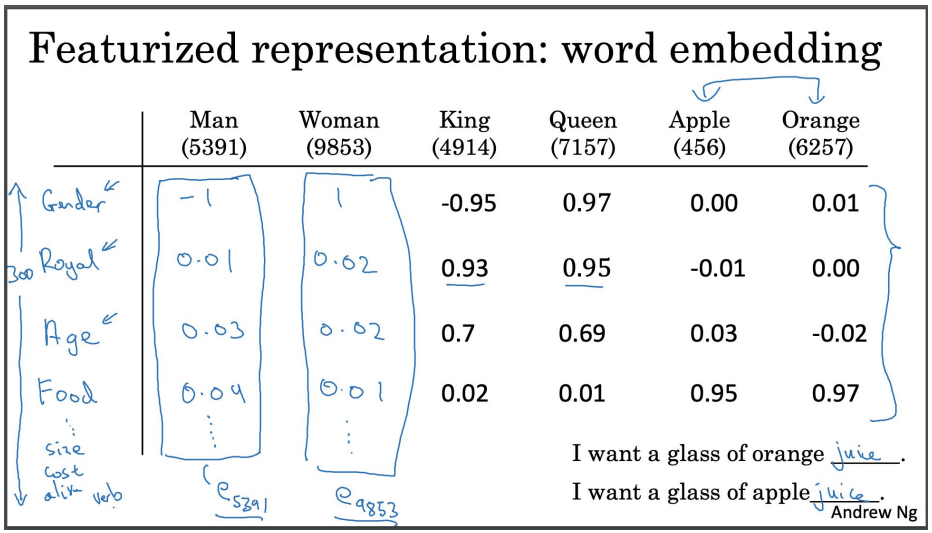


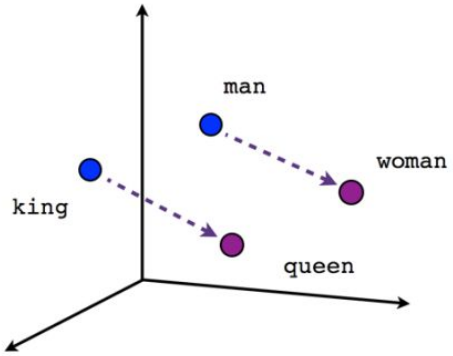
Basics of Machine Learning



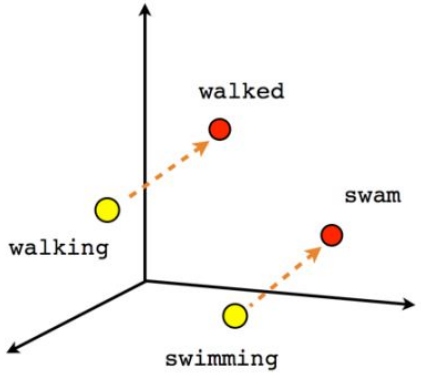
ML Technique - Word Embedding



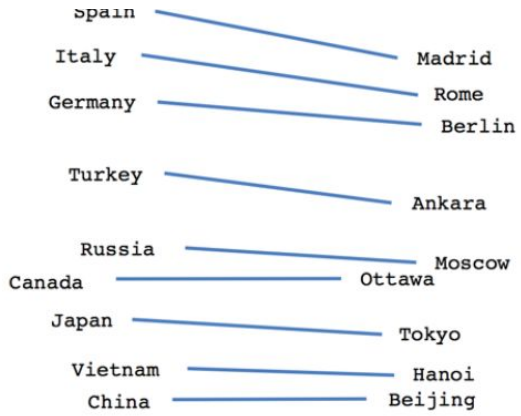




Male-Female

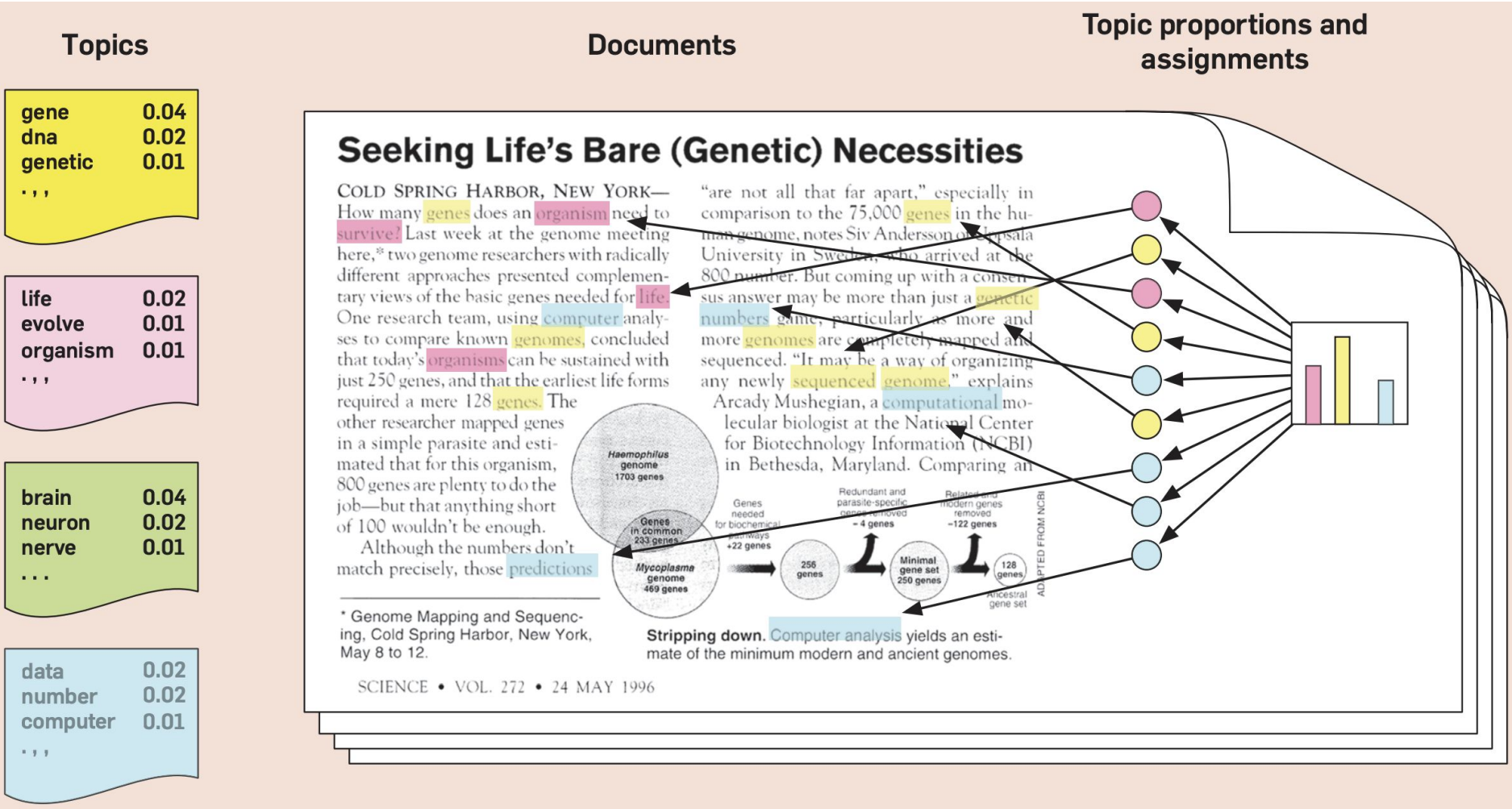


Verb tense



Country-Capital

ML Technique - Topics



Product Development: ML Platform Advantage

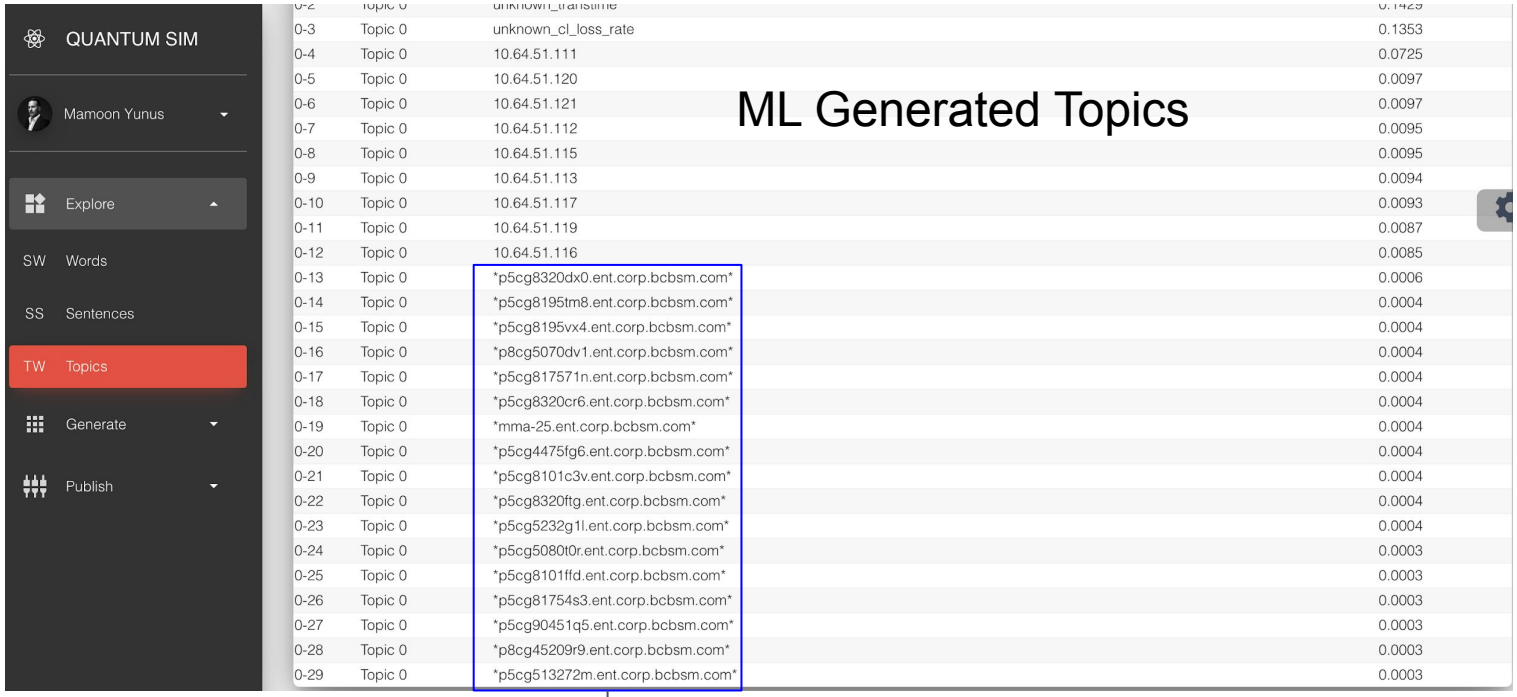
Problem: Enterprises generate multi-terabyte data per day with only a small fraction analyzed

- Writing Splunk queries is a manual process → Trails data generation velocity
- Query writers told where & what to search → Biased
- Data changes, queries don't → Inflexible non-adaptive queries

Solution: ML generated Splunk queries that are derived from enterprise dataset

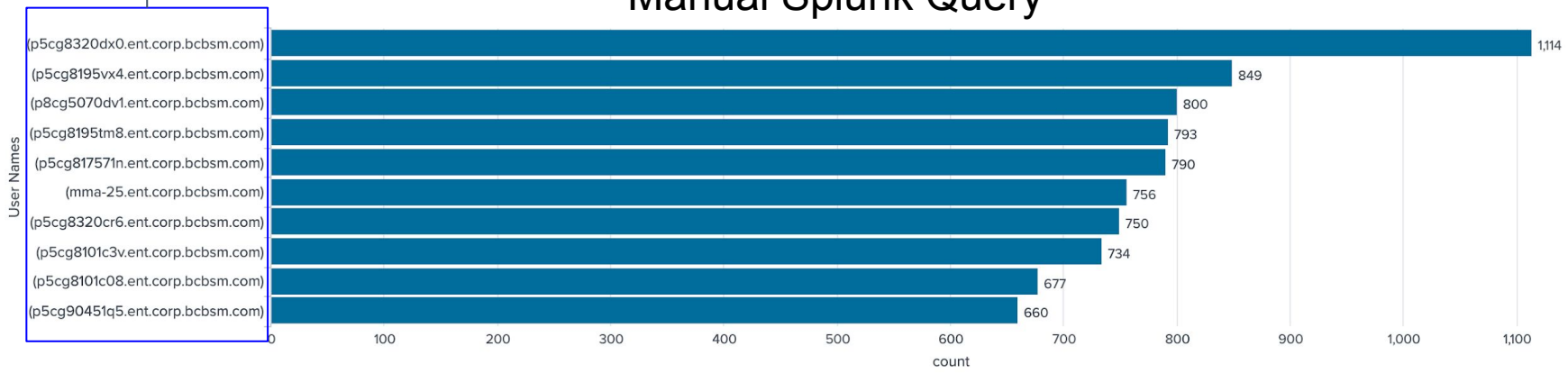
- Accelerated → Automated query generation through rapid ML training
- Agnostic → Unbiased queries generated purely from data
- Adaptive → Queries evolve with data, offer deeper insight
- Intelligent → Expert augment queries by selecting priority attributes

F-5 ML-Assisted Splunk Analysis



0-3	Topic 0	unknown_cl_loss_rate	0.1353
0-4	Topic 0	10.64.51.111	0.0725
0-5	Topic 0	10.64.51.120	0.0097
0-6	Topic 0	10.64.51.121	0.0097
0-7	Topic 0	10.64.51.112	0.0095
0-8	Topic 0	10.64.51.115	0.0095
0-9	Topic 0	10.64.51.113	0.0094
0-10	Topic 0	10.64.51.117	0.0093
0-11	Topic 0	10.64.51.119	0.0087
0-12	Topic 0	10.64.51.116	0.0085
0-13	Topic 0	*p5cg8320dx0.ent.corp.bcbsm.com*	0.0006
0-14	Topic 0	*p5cg8195tm8.ent.corp.bcbsm.com*	0.0004
0-15	Topic 0	*p5cg8195vx4.ent.corp.bcbsm.com*	0.0004
0-16	Topic 0	*p8cg5070dv1.ent.corp.bcbsm.com*	0.0004
0-17	Topic 0	*p5cg817571n.ent.corp.bcbsm.com*	0.0004
0-18	Topic 0	*p5cg8320cr6.ent.corp.bcbsm.com*	0.0004
0-19	Topic 0	*mma-25.ent.corp.bcbsm.com*	0.0004
0-20	Topic 0	*p5cg4475fg6.ent.corp.bcbsm.com*	0.0004
0-21	Topic 0	*p5cg8101c3v.ent.corp.bcbsm.com*	0.0004
0-22	Topic 0	*p5cg8320ftg.ent.corp.bcbsm.com*	0.0004
0-23	Topic 0	*p5cg5232g1l.ent.corp.bcbsm.com*	0.0004
0-24	Topic 0	*p5cg5080t0r.ent.corp.bcbsm.com*	0.0003
0-25	Topic 0	*p5cg8101ffd.ent.corp.bcbsm.com*	0.0003
0-26	Topic 0	*p5cg81754s3.ent.corp.bcbsm.com*	0.0003
0-27	Topic 0	*p5cg90451q5.ent.corp.bcbsm.com*	0.0003
0-28	Topic 0	*p8cg45209r9.ent.corp.bcbsm.com*	0.0003
0-29	Topic 0	*p5cg513272m.ent.corp.bcbsm.com*	0.0003

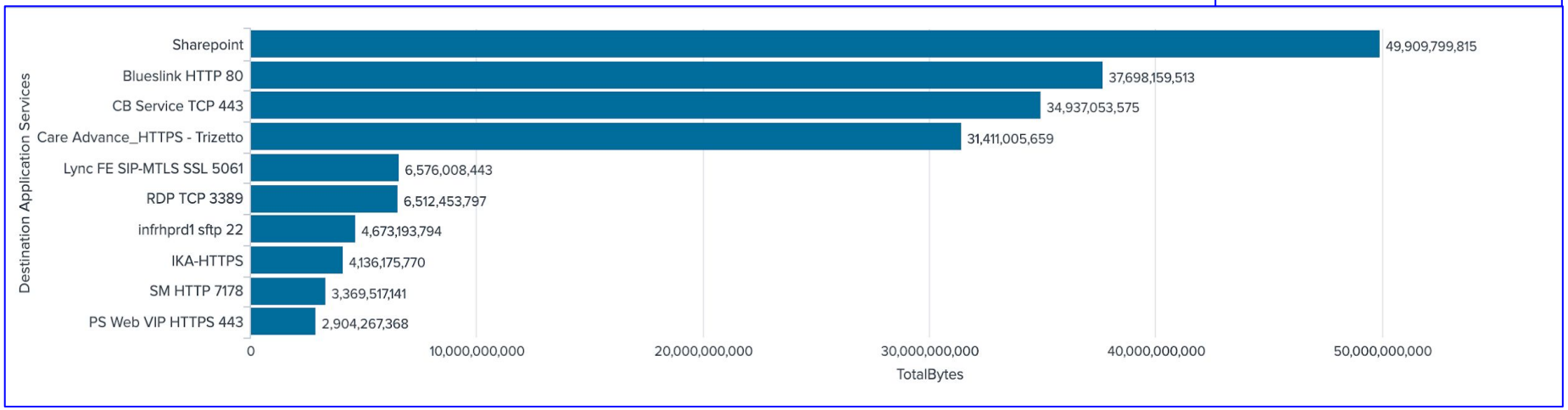
Manual Splunk Query



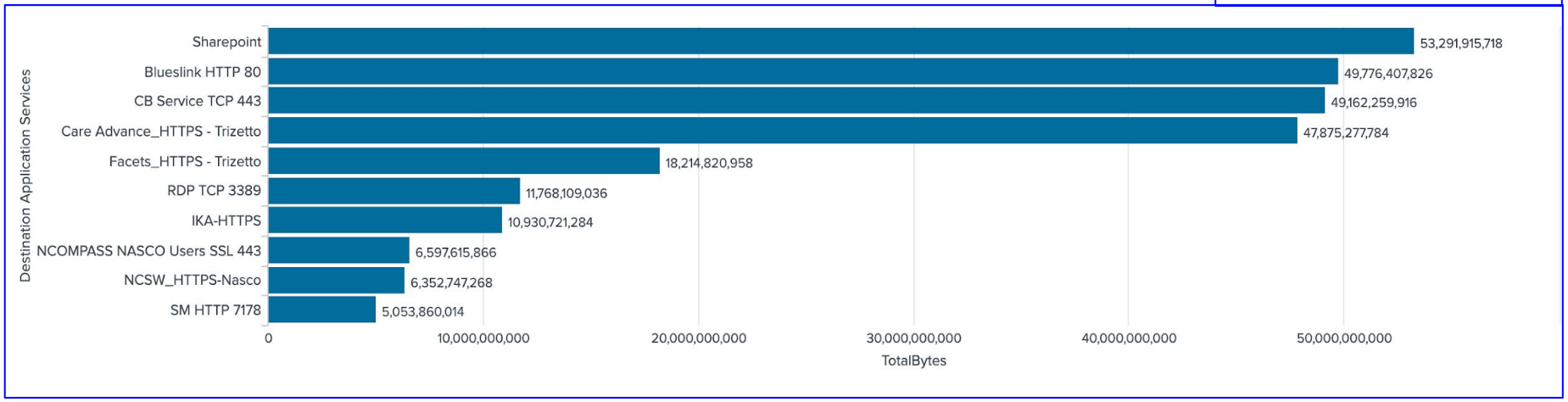
Top 10 Applications → Cisco vs F5

Did the Top 10 Applications Change?

Cisco (October 2018)



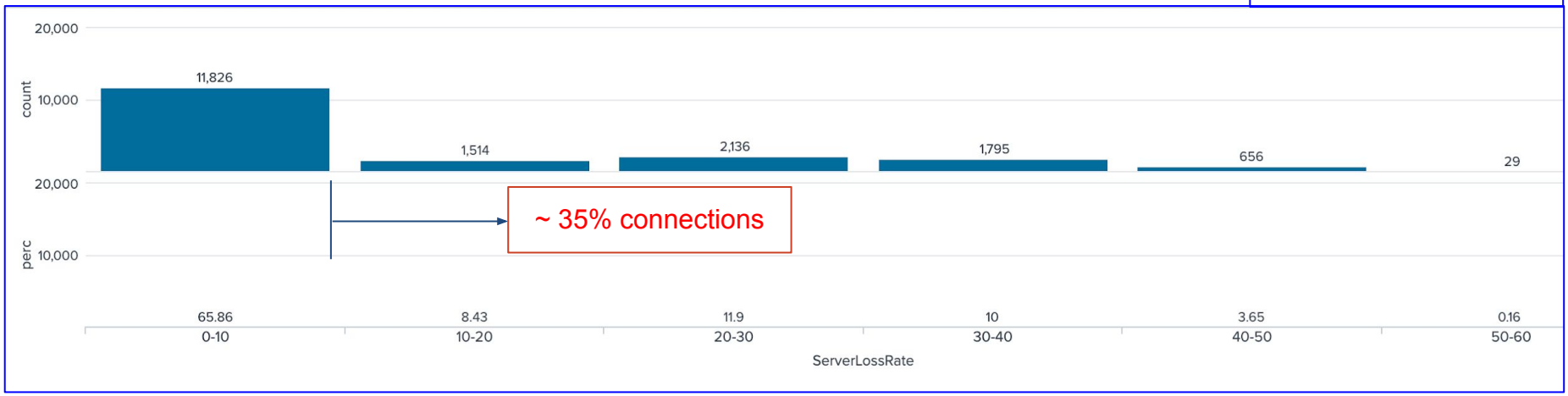
F5 (June 2019)



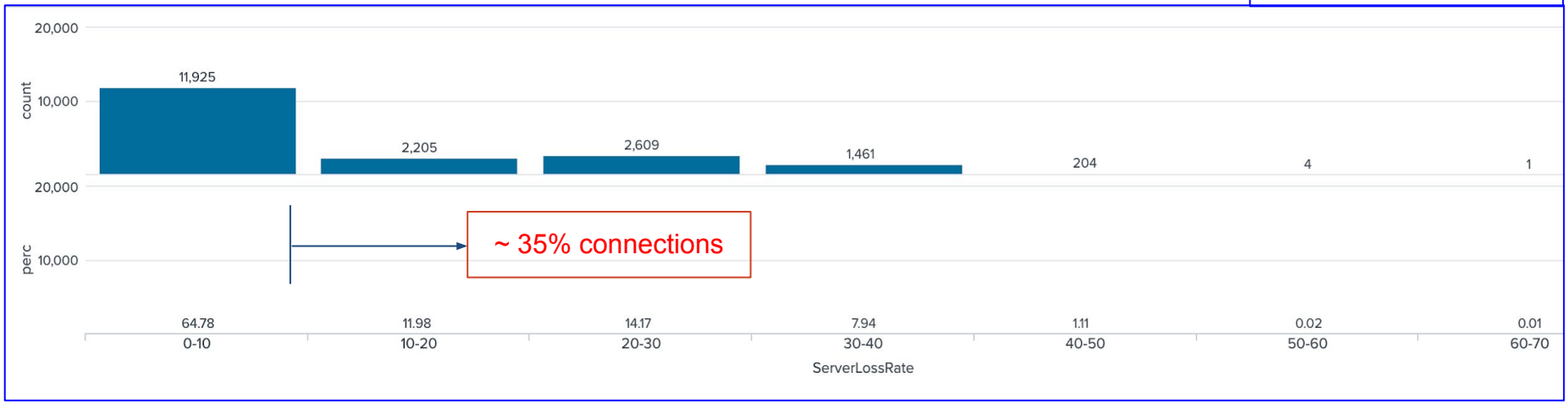
Sharepoint Server Loss Comparison

Did the Sharepoint Server Loss Change?

Cisco (October 2018)



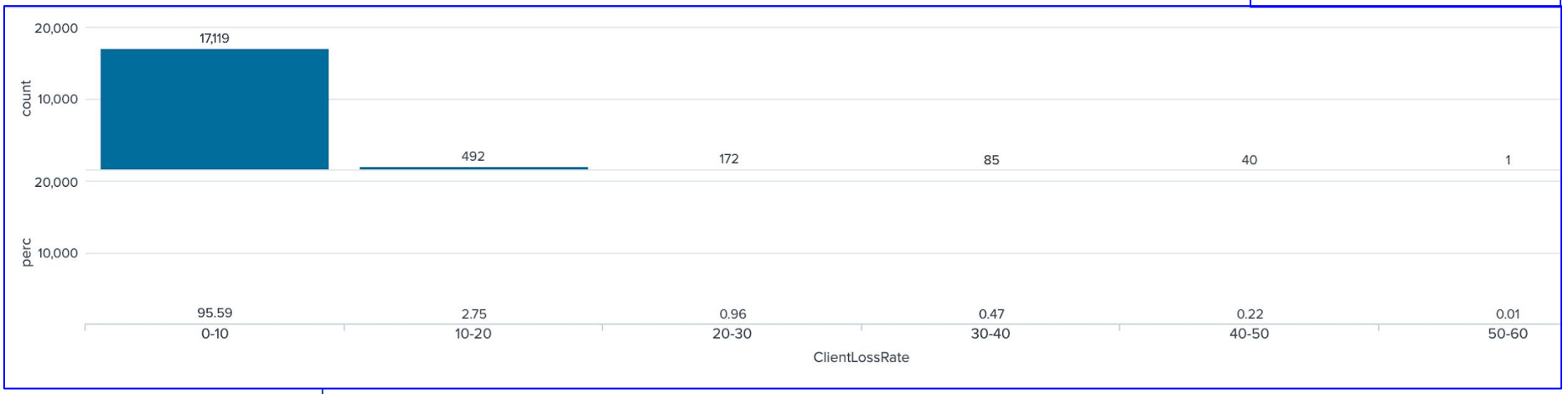
F5 (June 2019)



Sharepoint Client Loss Comparison

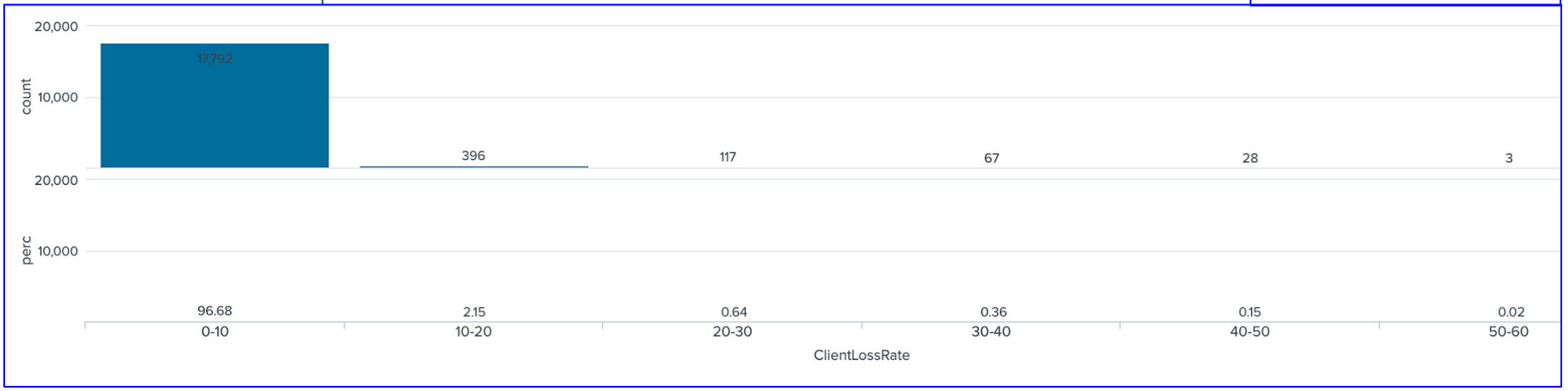
Did the Sharepoint Client Loss Change?

Cisco (October 2018)

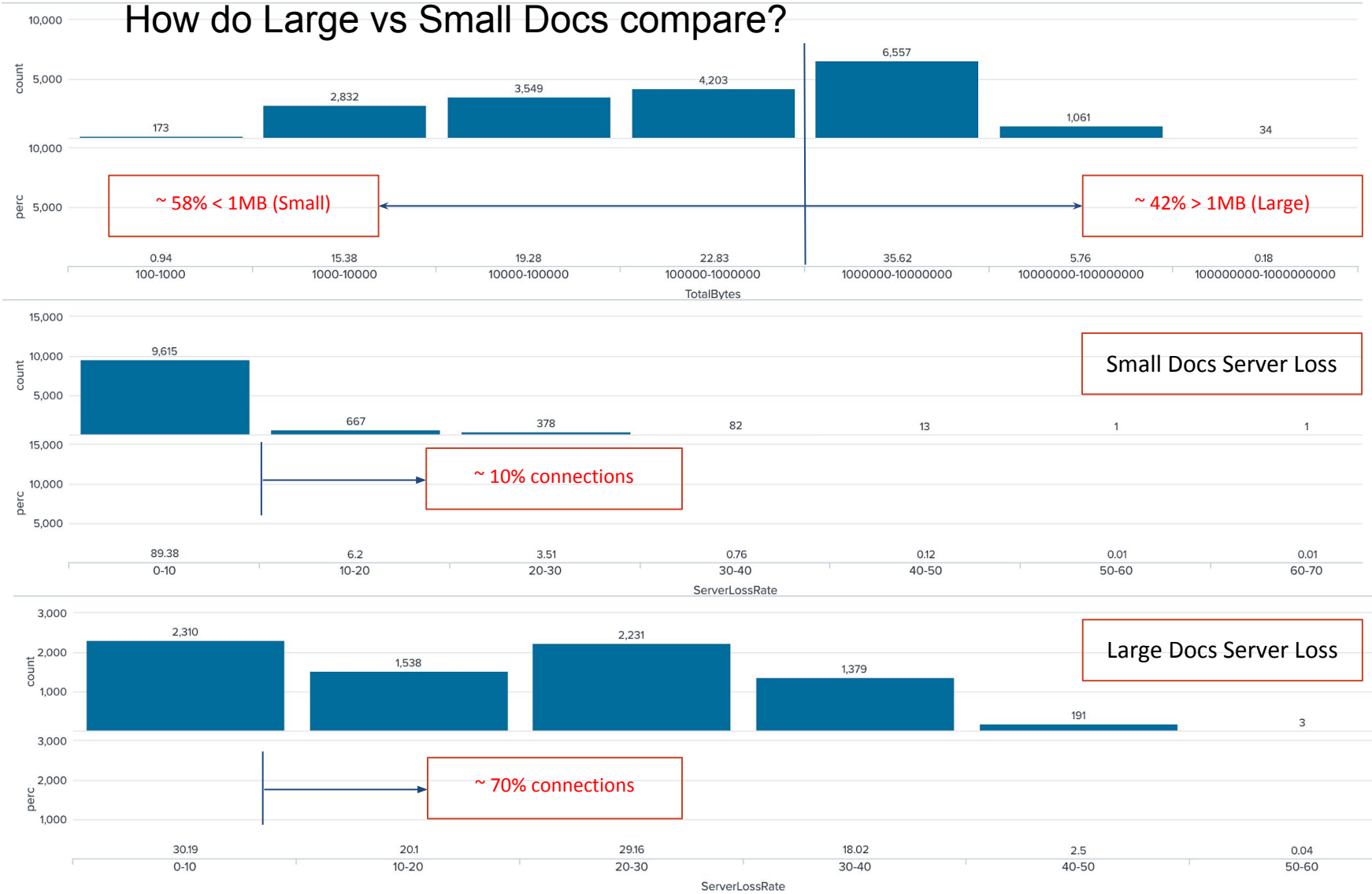


~ 5% connections

F5 (June 2019)



Sharepoint Data Distribution



Sharepoint Client Segmentation

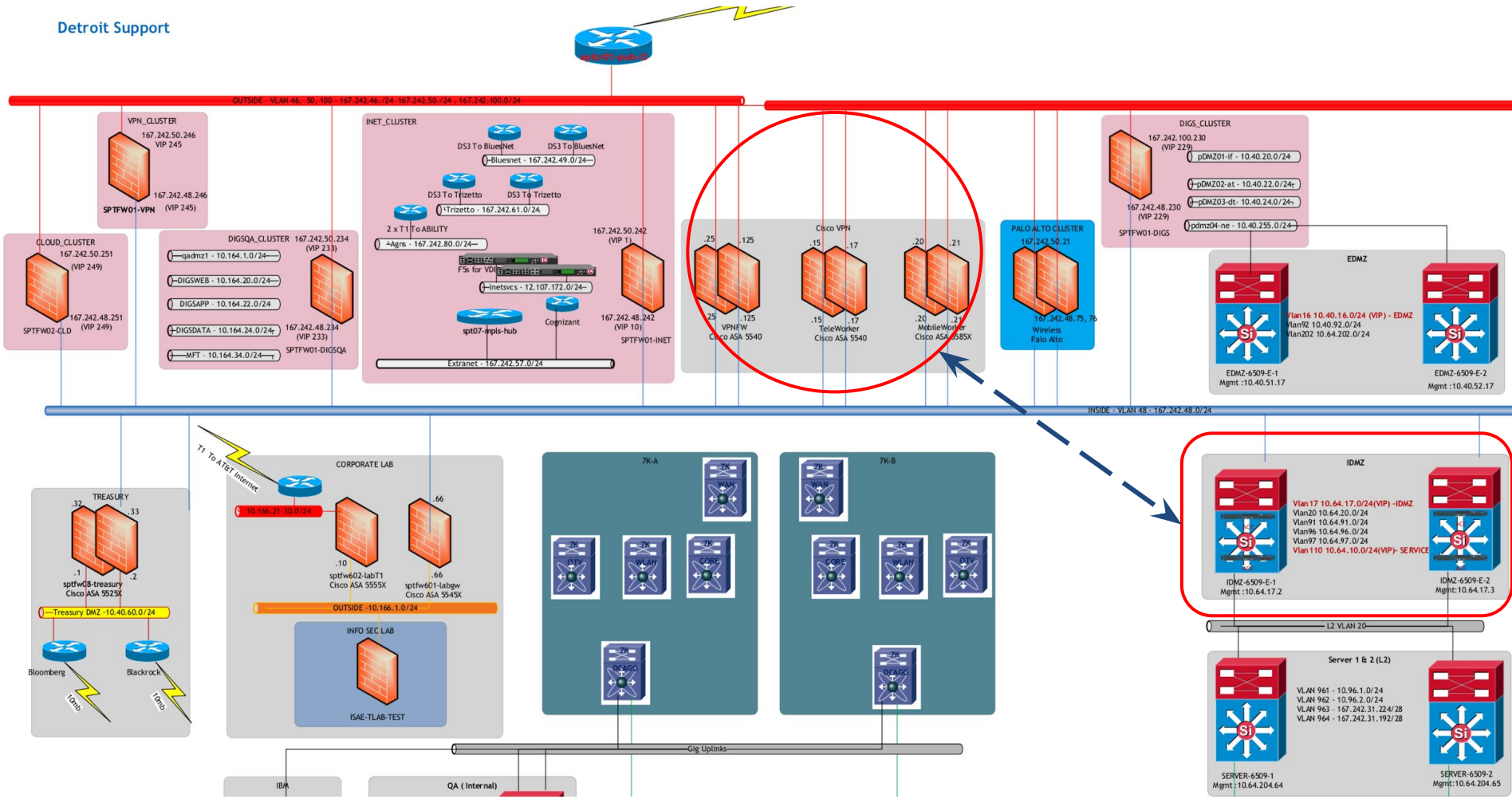
Who are the top active users with lowest server loss?

	begT	user_name	TotalBytes	ServerLossRate
1	6/3/19 16:00	10.17.0.68 (p5cg8195vjj.ent.corp.bcbsm.com)	14,476,824	0.0
2	6/3/19 14:00	10.17.2.47 (p5cg5090jmx.ent.corp.bcbsm.com)	35,345,912	0.0
3	6/3/19 14:00	10.17.0.94 (p5cg50647rn.ent.corp.bcbsm.com)	18,671,288	0.0
4	6/3/19 13:00	10.17.0.118 (p5cg5080t5m.ent.corp.bcbsm.com)	14,680,562	0.0
5	6/3/19 13:00	10.17.0.69 (p5cg44016rb.ent.corp.bcbsm.com)	12,746,243	0.0
6	6/3/19 12:00	10.20.11.128	44,158,532	0.0
7	6/3/19 11:00	10.20.9.216 (p5cg4520969.ent.corp.bcbsm.com)	29,057,884	0.0
8	6/3/19 10:00	10.17.0.69 (p5cg44016rb.ent.corp.bcbsm.com)	14,014,132	0.0
9	6/3/19 09:00	10.17.0.119 (p5cg8320bkv.ent.corp.bcbsm.com)	21,180,880	0.0
10	5/31/19 15:00	10.17.1.53 (p5cg4400s7p.ent.corp.bcbsm.com)	22,360,888	0.0

Who are the top active users with highest server loss?

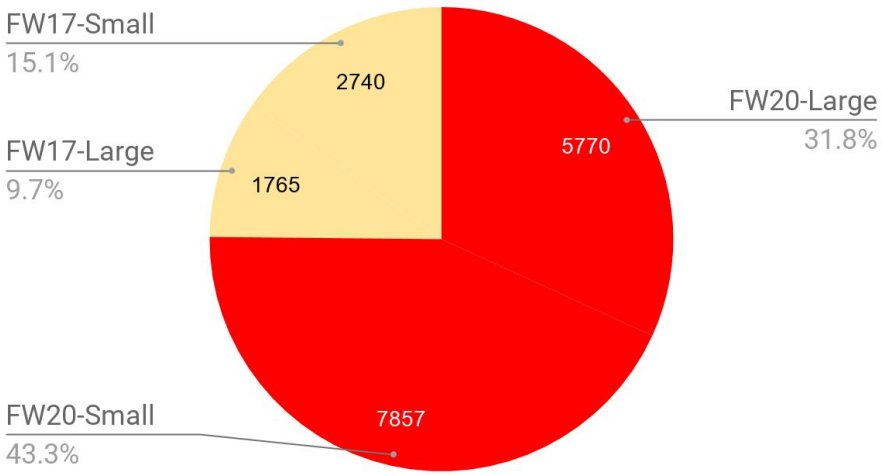
1	5/31/19 11:00	10.20.0.66 (p5cg4514tyh.ent.corp.bcbsm.com)	18,594,476	52 %
2	5/31/19 10:00	10.20.8.193 (p5cg4520774.ent.corp.bcbsm.com)	16,927,472	48 %
3	5/31/19 11:00	10.20.7.1 (p5cg81753pn.ent.corp.bcbsm.com)	11,130,323	48 %
4	5/30/19 13:00	10.20.4.163 (p5cd8400mc1.ent.corp.bcbsm.com)	129,762,688	47 %
5	5/30/19 14:00	10.20.1.190 (p5cg5080sw0.ent.corp.bcbsm.com)	27,223,860	47 %
6	5/28/19 17:00	10.20.1.68 (p5cg513272m.ent.corp.bcbsm.com)	12,117,770	46 %
7	6/3/19 14:00	10.20.11.66 (p5cg8195pgd.ent.corp.bcbsm.com)	30,597,380	46 %
8	5/31/19 09:00	10.20.4.226 (p5cg4514v7m.ent.corp.bcbsm.com)	29,709,430	46 %
9	5/31/19 09:00	10.20.5.188 (p5cg51251kz.ent.corp.bcbsm.com)	21,622,176	45 %
10	5/28/19 15:00	10.20.4.165 (p5cd819494f.ent.corp.bcbsm.com)	28,470,988	44 %

Network Topology



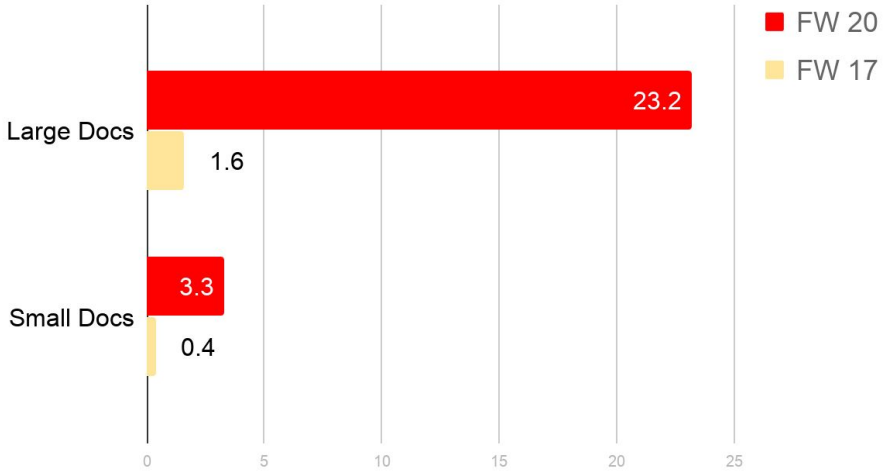
Firewall Profiles for Sharepoint Traffic

Connections



25% Tele Workers vs. 75% Mobile Workers

Average Server Loss Rate (%)

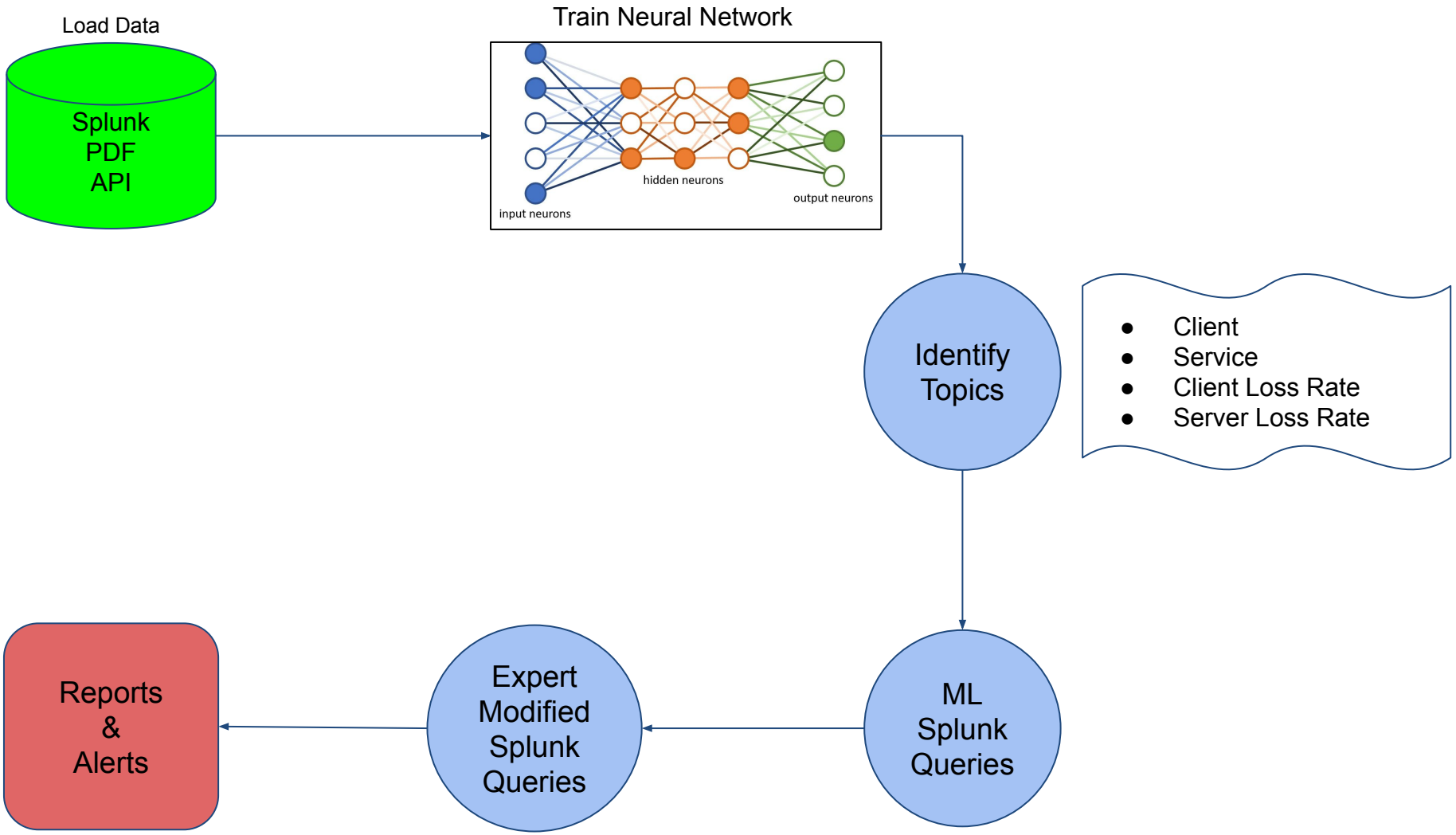


Large vs. Small Doc Experience

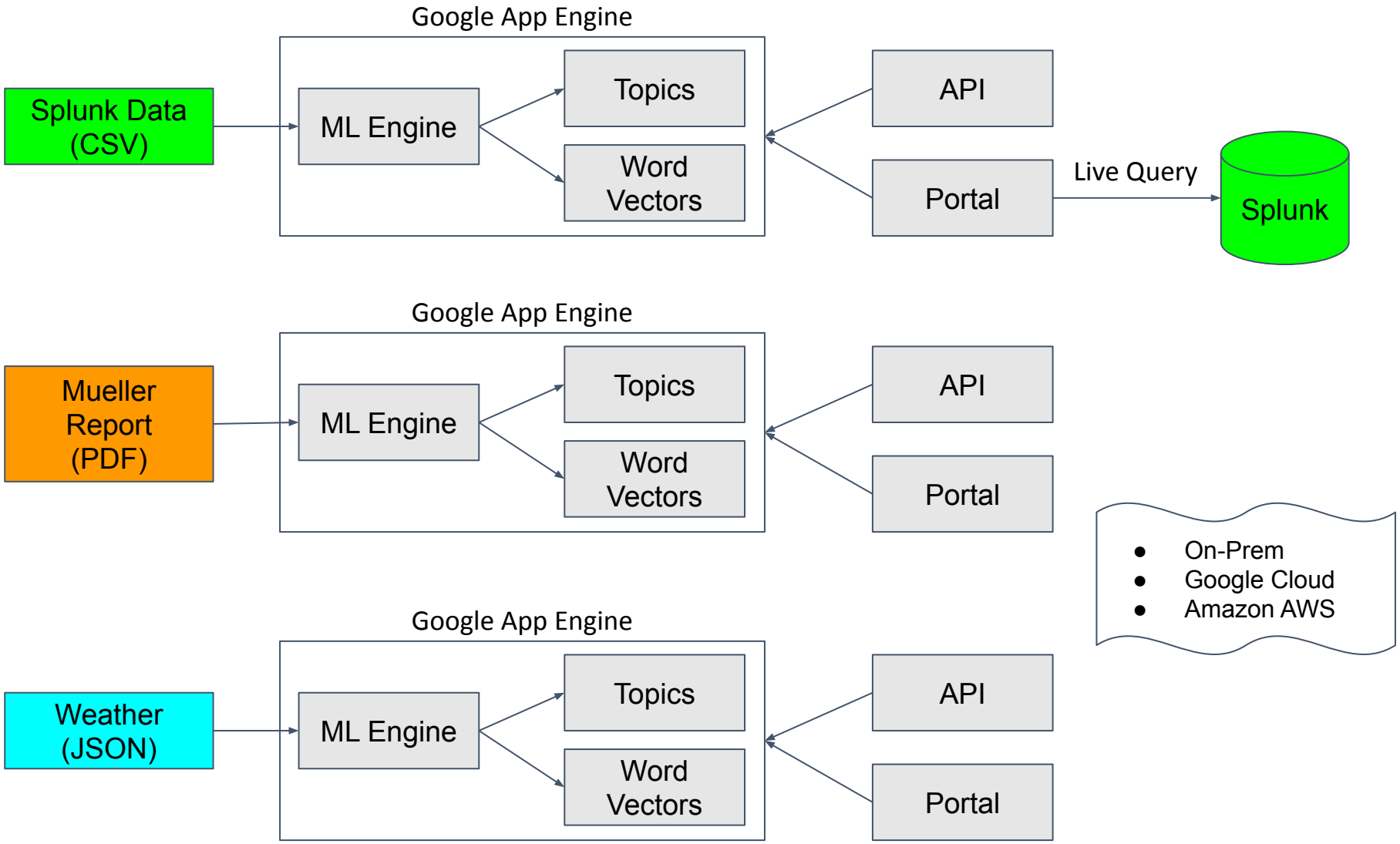
Nomadic Workers

	Intersected Sharepoint Users	10_20_FW_Avg_SLoss	10_17_FW_Avg_SLoss
1	(p5cg5043lnt.ent.corp.bcbsm.com)	33.3	0.0
2	(p5cg8195tlq.ent.corp.bcbsm.com)	30.6	0.0
3	(p5cg8320d8x.ent.corp.bcbsm.com)	28.0	0.0
4	(p5cd8400mc1.ent.corp.bcbsm.com)	26.4	0.6
5	(p5cg51251kn.ent.corp.bcbsm.com)	20.3	0.0
6	(p5cg81757mk.ent.corp.bcbsm.com)	20.2	1.3
7	(p5cg4475cjf.ent.corp.bcbsm.com)	19.8	2.0
8	(p5cg5080thq.ent.corp.bcbsm.com)	19.6	1.8
9	(p5cg51718mv.ent.corp.bcbsm.com)	18.6	2.3
10	(p5cd8100p8m.ent.corp.bcbsm.com)	18.3	0.1

Machine Learning Lifecycle



Quantum Similarity Architecture



Demo

Conclusions & Recommendations

- Optimize Sharepoint
  - end-to-end performance analysis
  - trace docs from various clients
  - dashboard for Sharepoint (weekly, monthly, quarterly profiles)

- Focus on FW20

- Extend Analysis to other strategic BCBSM Applications

- Incorporate AI in all aspects of your business → Force Multiplier
  - Forum Outsourced Services → Rapid ML models for Enterprise Data
  - Ongoing reporting, monitoring and remediation

- BCBSM Network of the Future
  - Identify strategic assets based on AI-assisted analysis
  - Simulate traffic patterns for cloud migration