Bigstream Hyperacceleration

The Analytics Performance Advantage
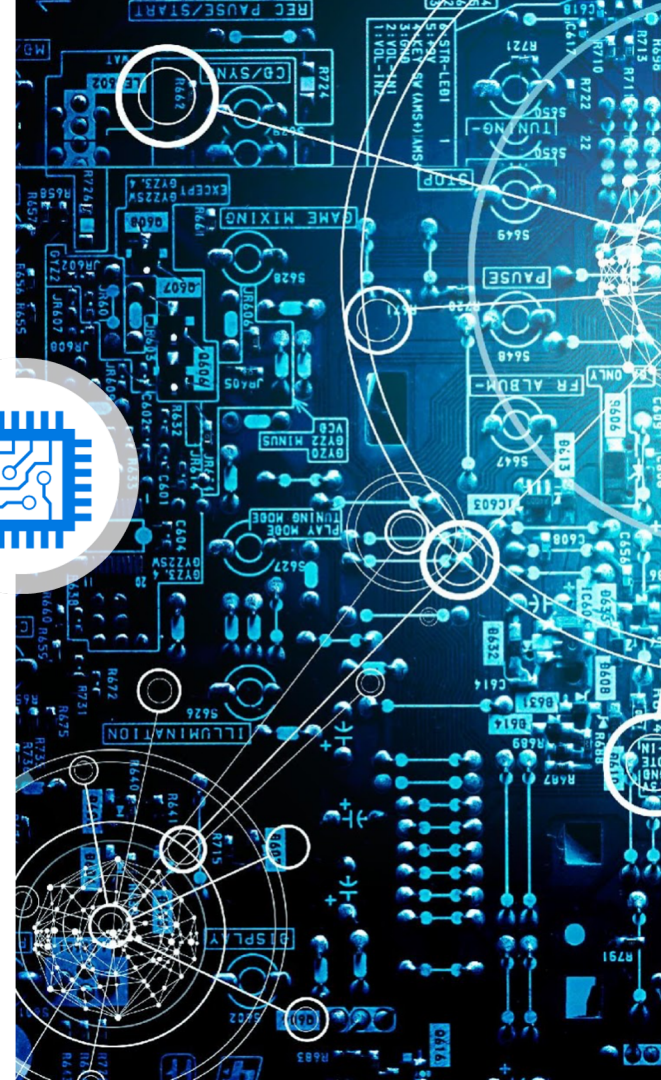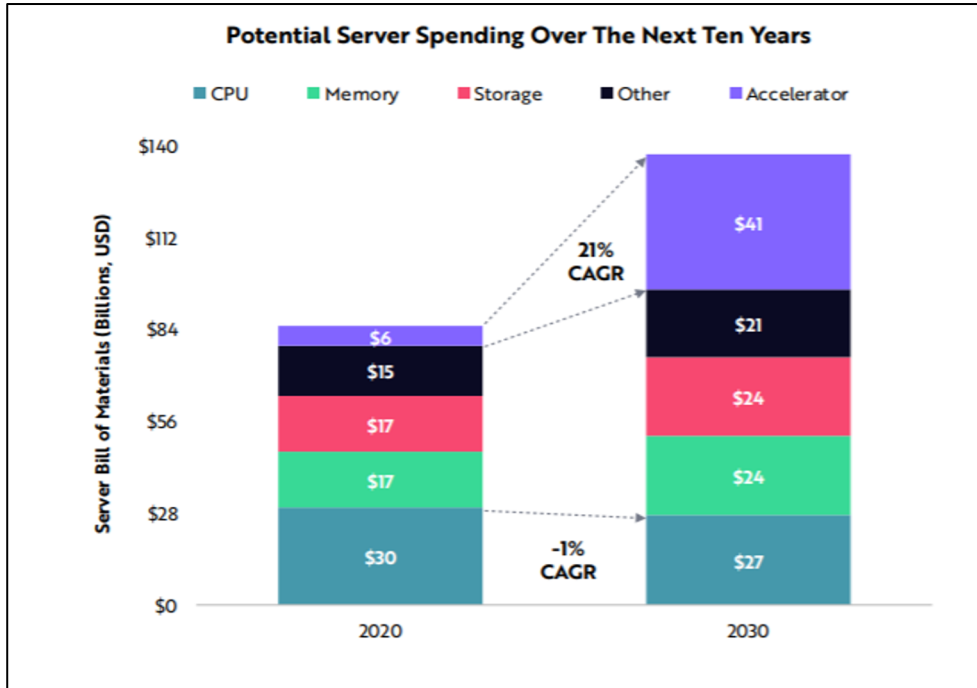
bigstream

Spark Use Cases Across Financial Services Sectors

Financial Services Sector	Customer Insights & Retention	Fraud Detection	Valuation & Market Models	Risk & Compliance	Telematics
Retail Banks	x	x		x	
Payments	x	x			
Investment banks			x	x	
Investment Management	x		x	x	
Insurance	x	x			x
Tech / FinTech	x	x			
Exchanges / Govt		x	x	x	

bigstream

Hardware Accelerators

- **Problem: CPU cluster scaling is running into limitations**

- **Programmable hardware is an attractive alternative: FPGA, GPU**

- **Designed for efficient execution of specialized code. A departure from general-purpose CPUs**

- **FPGA, GPU: provide flexibility, but don't natively connect to big data platforms**
  - **Middleware needed to automate this**

- **Both can provide power and cost advantages w.r.t CPU scale-up, scale-out**

- **Designed to physically attach to existing servers simply (i.e. PCIe slot)**
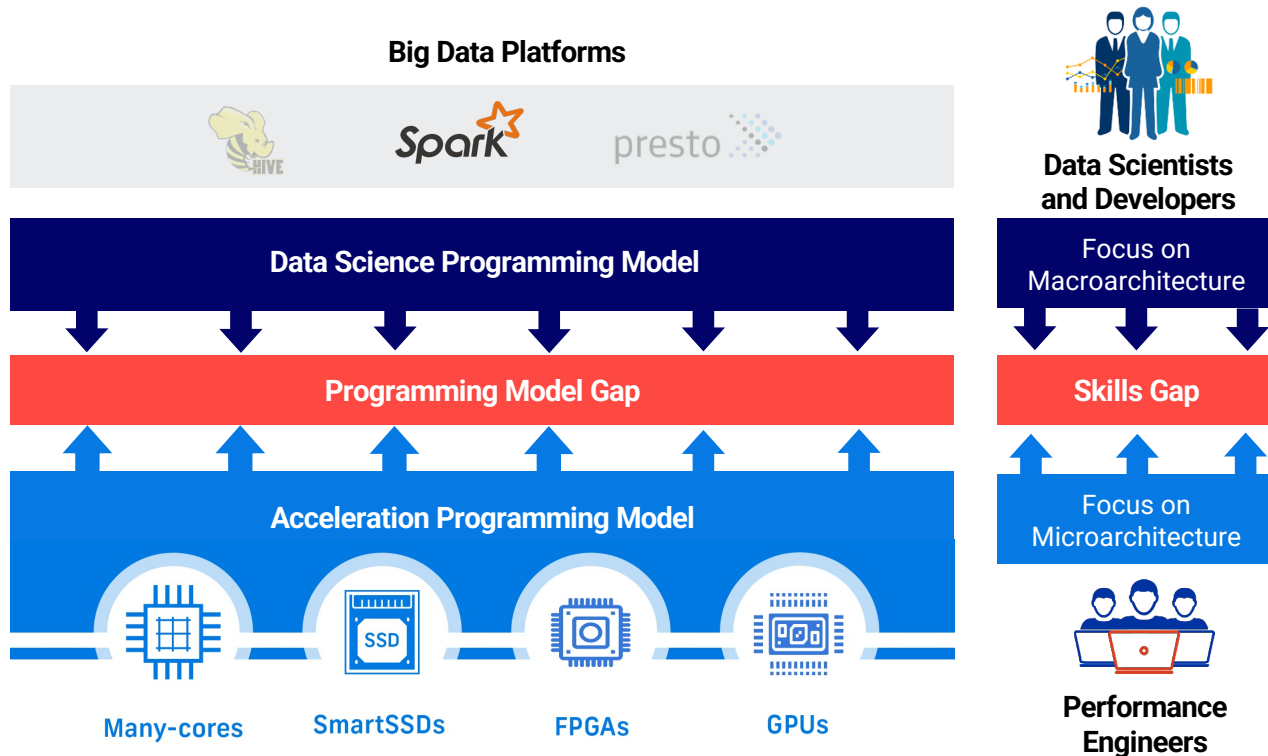
HW Accelerator Market is Trending

Potential Server Spending Over The Next Ten Years

Legend: CPU, Memory, Storage, Other, Accelerator

2020:
- Accelerator: $6
- Other: $15
- Storage: $17
- Memory: $17
- CPU: $30

2030:
- Accelerator: $41
- Other: $21
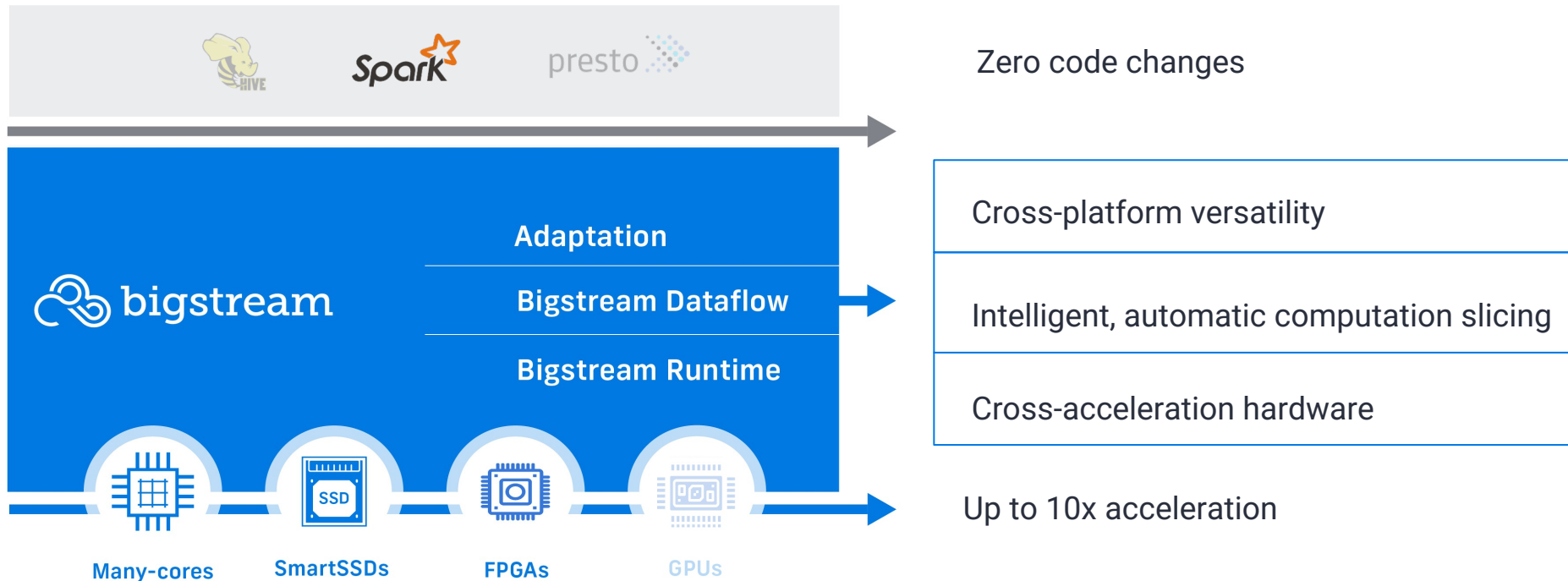- Storage: $24
- Memory: $24
- CPU: $27

21% CAGR

-1% CAGR

ARK Invest "Big Ideas 2021"

● Accelerators = GPUs, ASICs, and FPGAs

● $41 Billion industry in next 10 years, surpassing CPUs.

● Driven by big data analytics and AI

bigstream

Programming Model Gap

Big Data Platforms

Data Science Programming Model

Programming Model Gap

Acceleration Programming Model

Many-cores   SmartSSDs   FPGAs   GPUs

Data Scientists and Developers

Focus on Macroarchitecture

Skills Gap

Focus on Microarchitecture

Performance Engineers

- No automated way for big data platforms to leverage accelerators

- Requires highly specialized engineers to program, which are scarce and expensive.

- Technology needed to bridge the gap.

bigstream

Bigstream Hypercceleration Layer

Zero code changes

Adaptation

Bigstream Dataflow

Bigstream Runtime

Cross-platform versatility

Intelligent, automatic computation slicing

Cross-acceleration hardware

Up to 10x acceleration
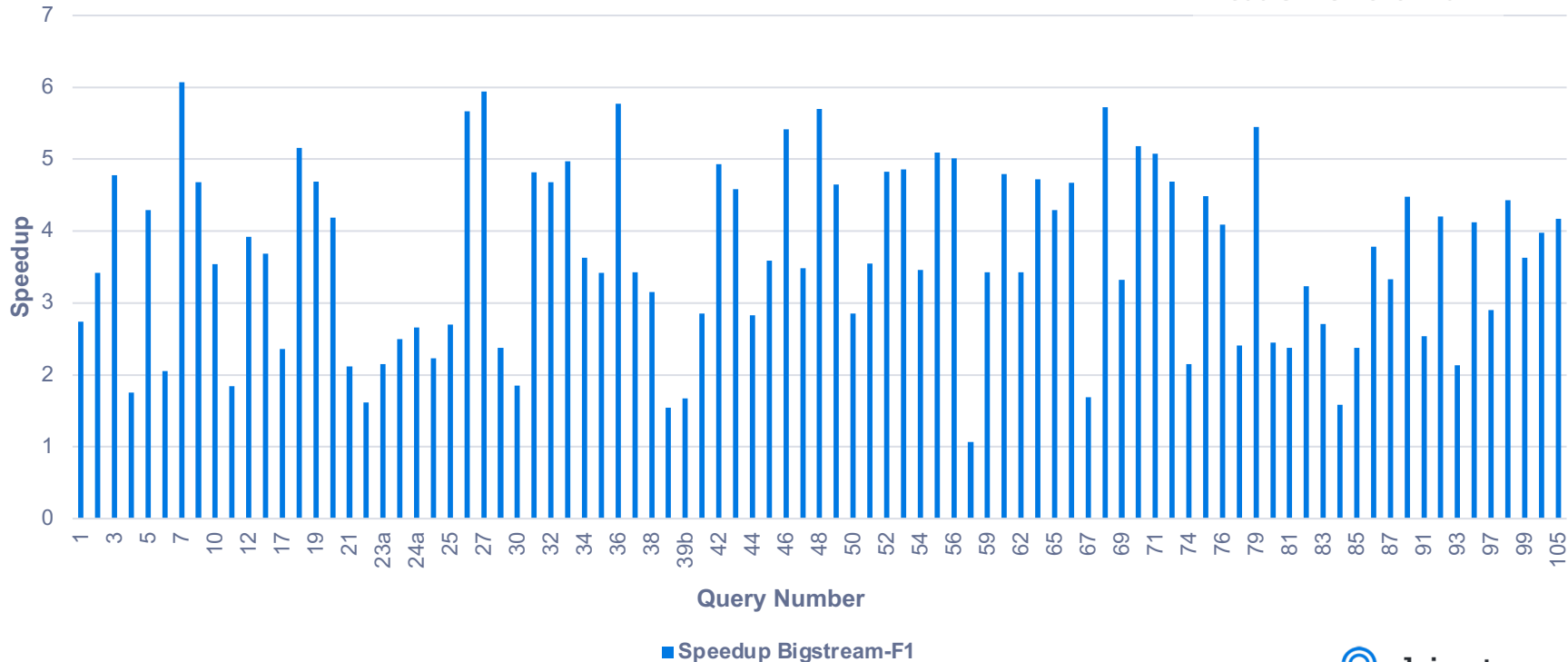
Many-cores   **SmartSSDs**   **FPGAs**   GPUs

Bigstream-F1 FPGA Speedup Results

- All runs on identical md5.2xl EMR clusters - Baseline: Spark 3.0
- 4 workers/cluster, S3 storage
- 250SF CSV gzip compressed standard data (approx. 72GB)

Bigstream Average Speedup: 3.6x

Not a STAC Benchmark

Speedup (y-axis) vs Query Number (x-axis)
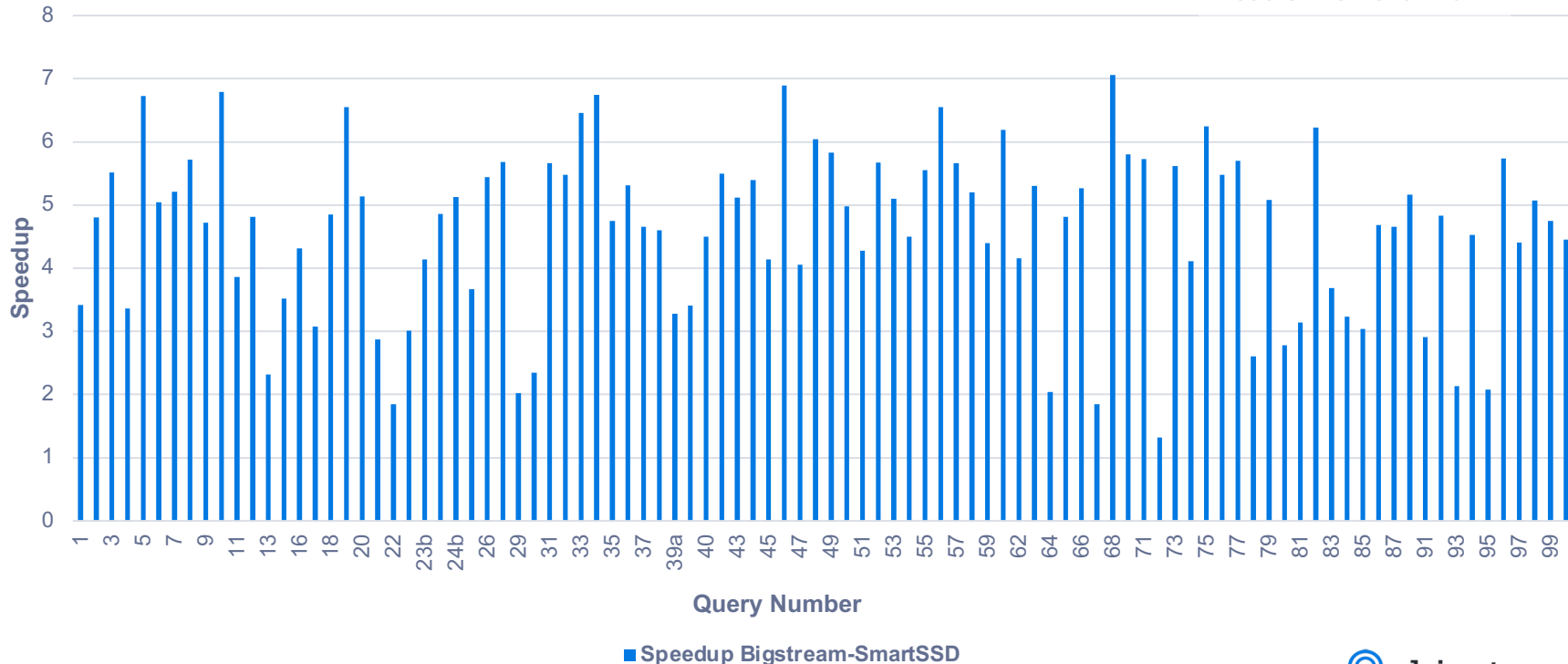
■ Speedup Bigstream-F1

bigstream

Bigstream-SmartSSD Speedup Results

- *All runs an identical Intel-based server - Baseline: Spark 2.4.6*
- *Single server, 3 SmartSSD*
- *250SF JSON gzip compressed standard data (approx. 110GB)*

Bigstream Average Speedup: 4.3x

Not a STAC Benchmark

■ Speedup Bigstream-SmartSSD

Bigstream Spark Acceleration: Options and Roadmap

Spark Deployment Option	Software-based Acceleration	Hardware-based Acceleration	
		FPGA	SmartSSD
Cloud (AWS)	✓ AWS EMR	✓ AWS EC2 + F1	
Cloud(Azure)	*1H '21*	*1H '21*	
On-Prem	✓	✓	✓

Primary Use Cases (current customers)
- Ingest / ETL / ELT / ML Prep
- Batch Analytics

Secondary Use Cases (road mapped)
- Ad Hoc Analytics
- ML Training

bigstream

Summary/Next Steps

- Bigstream: most mature and comprehensive Spark acceleration portfolio
  - From complex to easy
  - From fast to fastest
- Best customer fit - deployments
  - Managed cloud: Bigstream + EMR = plug & play, start today at Marketplace
  - FPGA on cloud – single script deploy
  - On prem – single script deploy
- Guaranteed TCO reduction vs existing infrastructure
- Fastest Spark with zero code change

bigstream

Thank You!

Please select **Bigstream** to start the conversation
roop@bigstream.co