

Adatmenedzsment Kihívások és válaszok

Andras Benczur

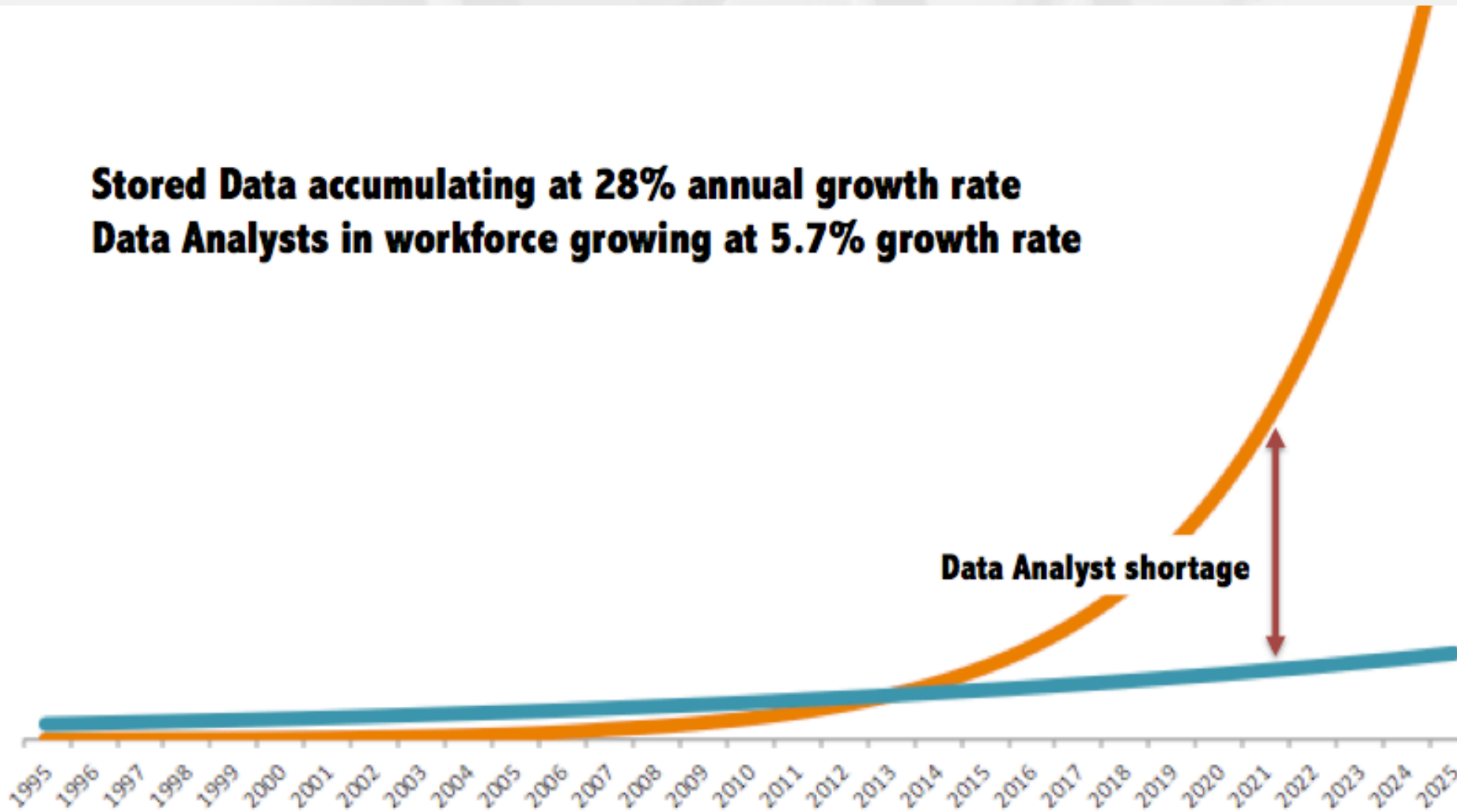
Institute for Computer Science and Control

Hungarian Academy of Sciences (MTA SZTAKI)

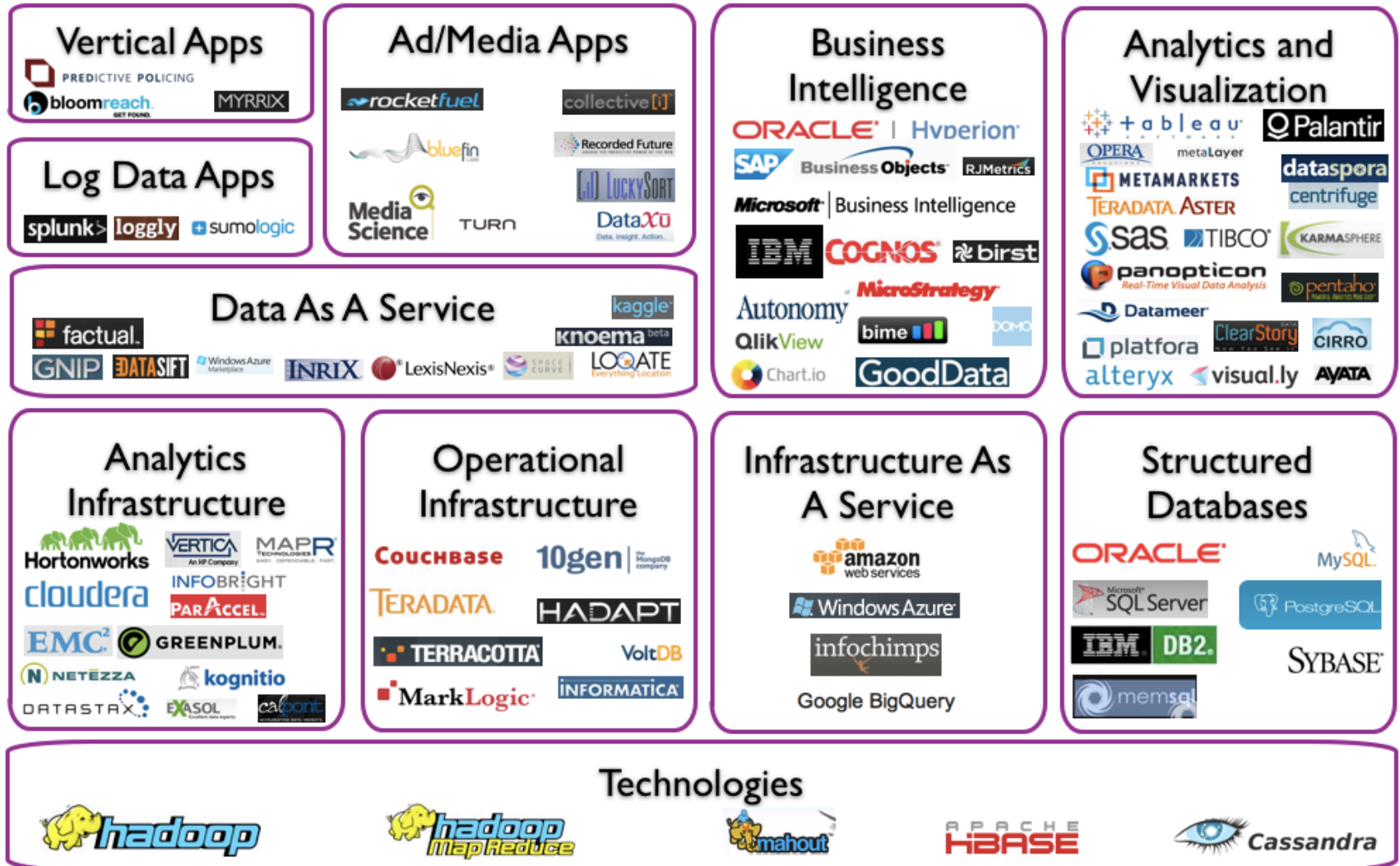
benczur@sztaki.mta.hu

<http://datamining.sztaki.hu>

Growth of Data vs. Growth of Data Analysts



Big Data Landscape



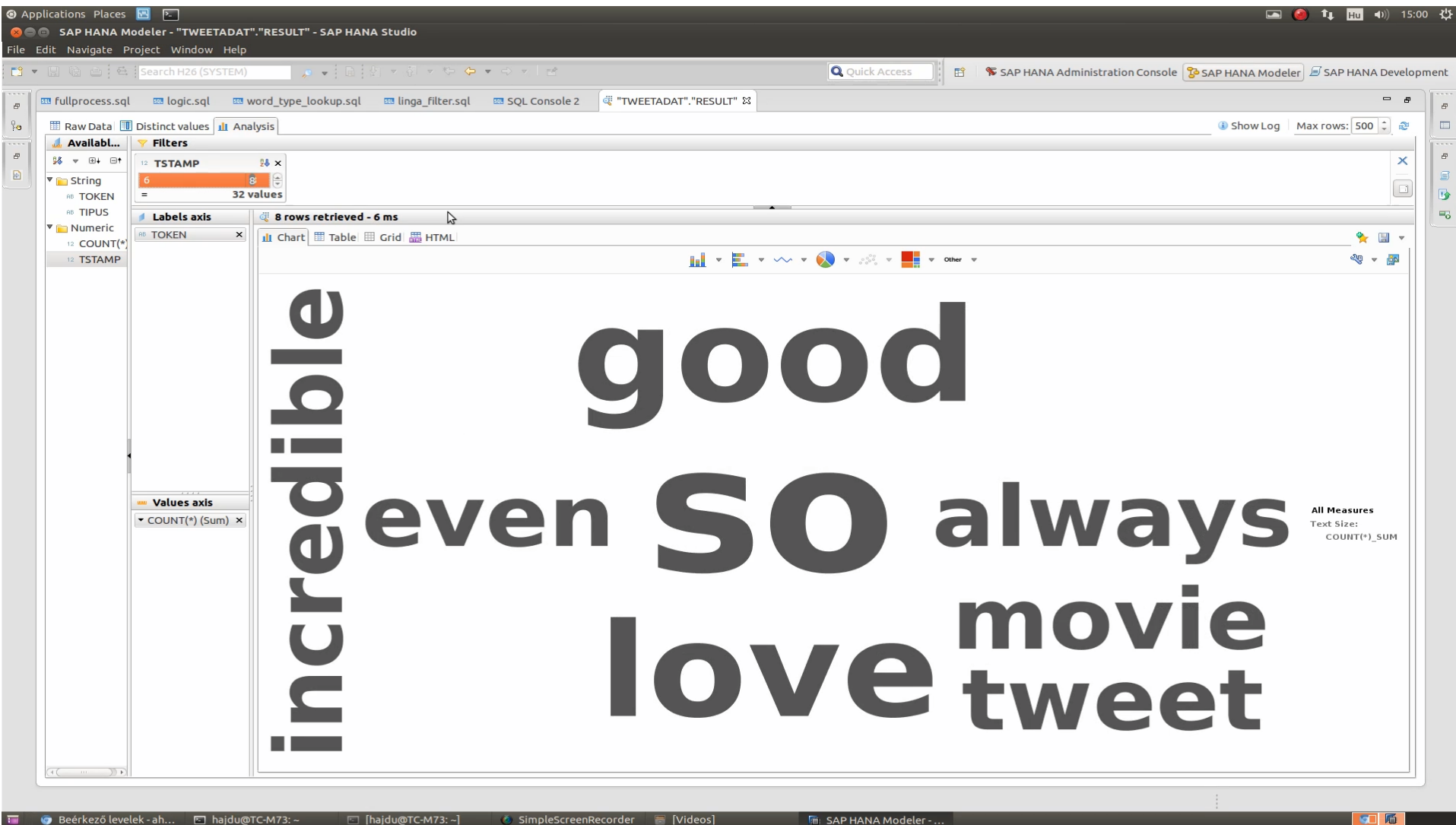
Big Data Landscape (Version 2.0)



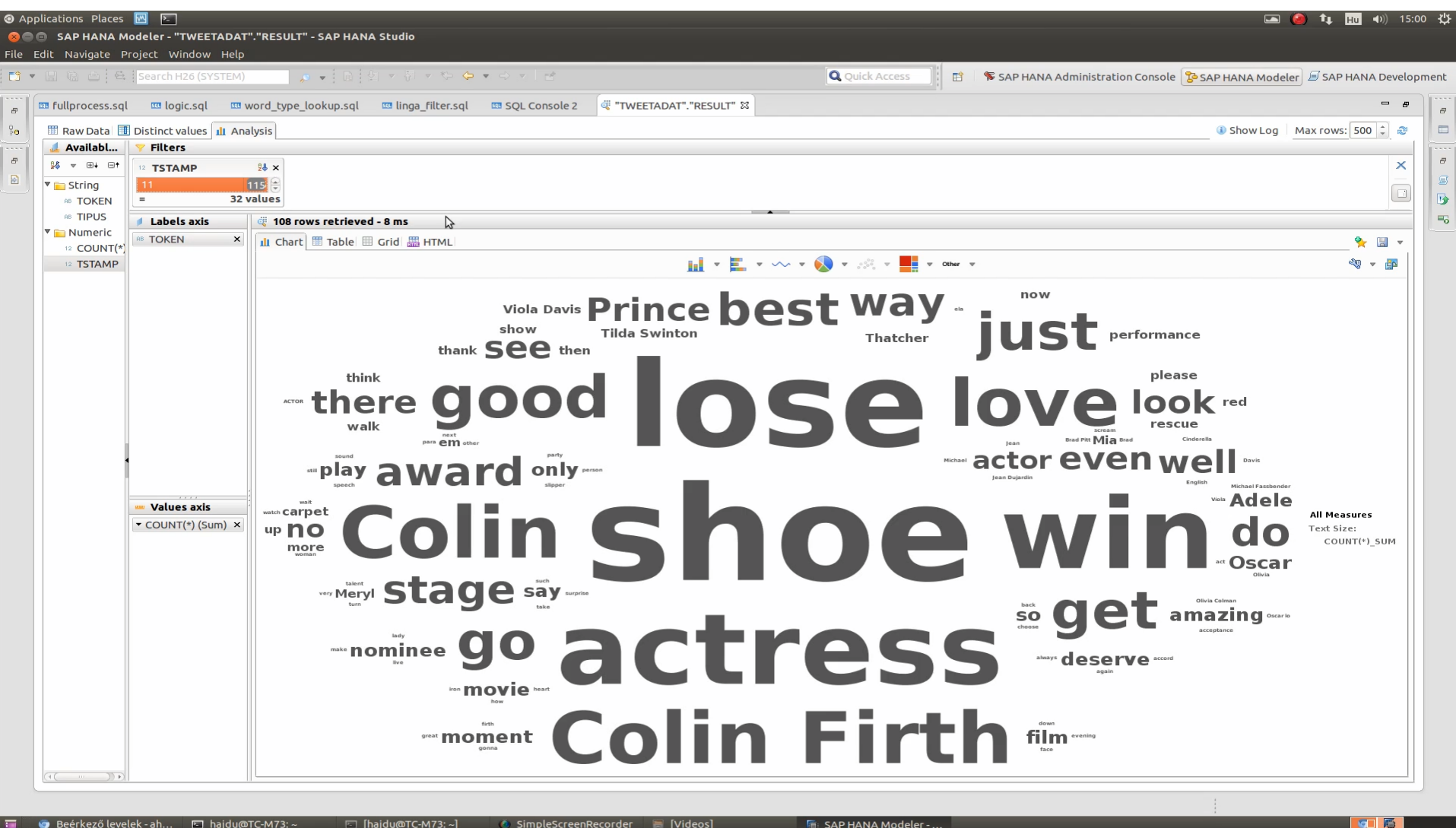
© Matt Turck (@mattturck) and ShivonZilis (@shivonz) Bloomberg Ventures

2013.02: <http://www.slideshare.net/mjft01/big-data-big-deal-a-big-data-101-presentation>

SAP HANA Demo: Meryl Streep, Oscar 2012



SAP HANA Demo: Meryl Streep, Oscar 2012



SAP HANA Demo: Meryl Streep, Oscar 2012

Applications Places SAP HANA Modeler - "TWEETADAT", "RESULT" - SAP HANA Studio

File Edit Navigate Project Window Help

Search H26 (SYSTEM)

fullprocess.sql logic.sql

Raw Data Distinct values

Availabl... Filters

String

TOKEN = 32

TIPUS

Numeric

COUNT(*)

TSTAMP

Labels axis

TOKEN

Values axis

COUNT(*) (Sum)

Meryl Streep loses shoe while collecting a BAFTA (Video)

Feb 13, 2012 10:07 By Caroline Corcoran

But no worries - Colin Firth is on hand to do a proper Prince Charming and put it back on her foot

Share f Share t Tweet g+ +1 Email



kép: <http://mirror.co.uk>

You've got to really hope it wasn't a bit sweaty

★ Recommended In TV

Aon Training Complex

RADAMEL FALCAO

New Manchester United hitman Radamel Falcao reveals worry over possible transfer collapse

MANCHESTER UNITED TRANSFER NEWS

Juventus eyeing January move for Manchester United midfielder Juan Mata

ISIS

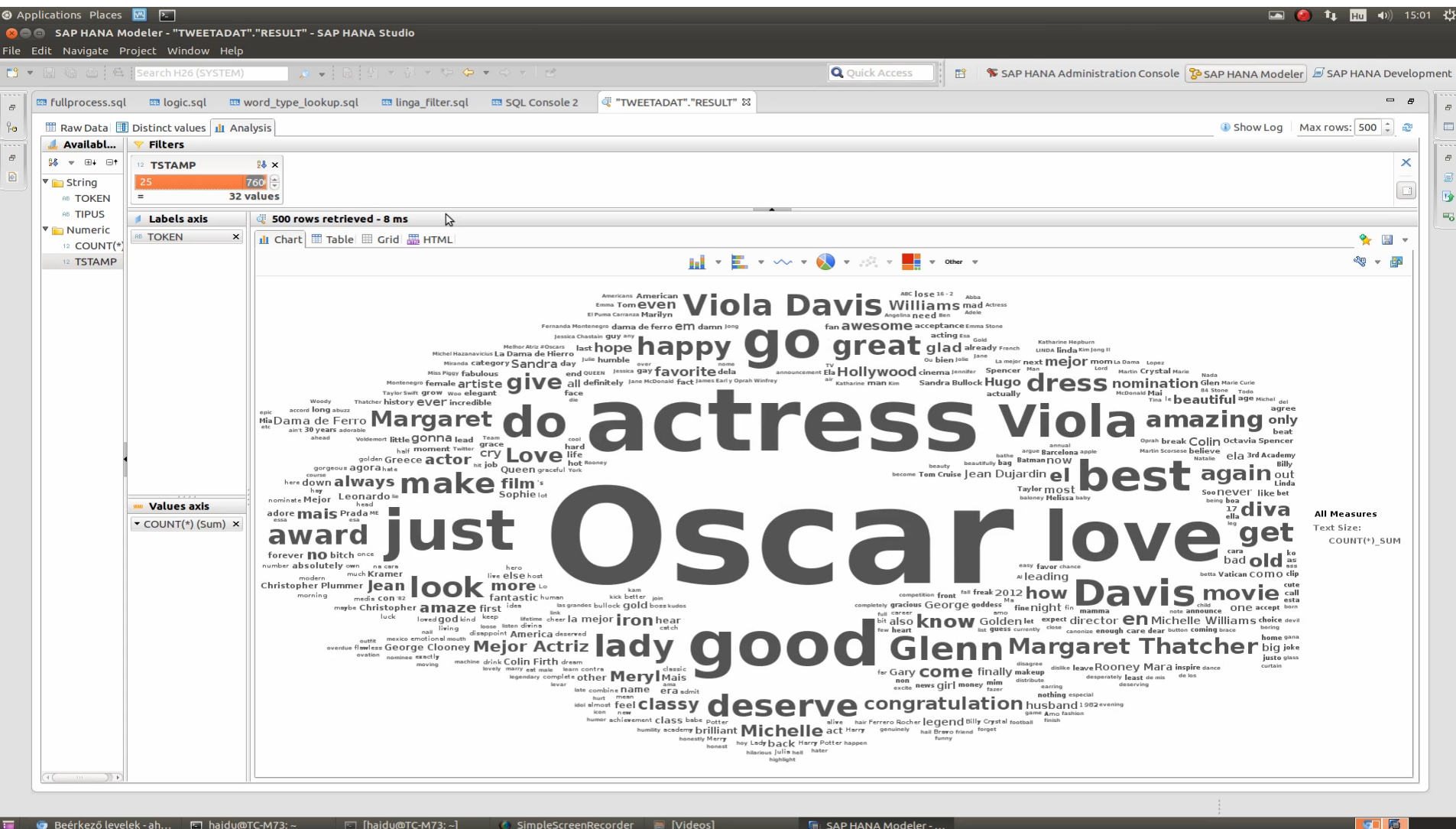
British jihadis flee ISIS maniacs and now want to return back home

JACK WILSHIRE

Jack Wilshere hits back at Jamie Redknapp criticism over his England and Arsenal performances

Beérkező levelek - ah... hajdu@TC-M73:~ [hajdu@TC-M73:~] SimpleScreenRecorder [Videos] SAP HANA Modeler - ...

SAP HANA Demo: Meryl Streep, Oscar 2012



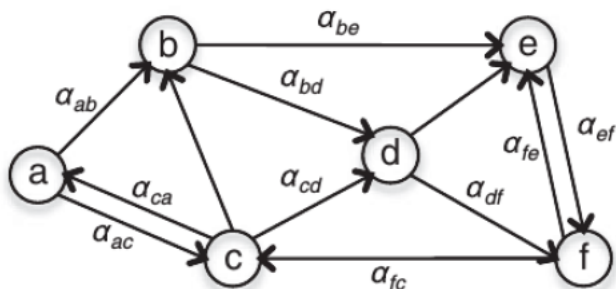
What was The Challenge in Analytics?

- One year 1 billion Tweet collection, 100GB
- Ad Hoc queries (Meryl Streep) may have 100,000+ hits
- Fast response needed to support the analyst
- Solutions
 - In Memory databases (such as SAP HANA)
 - Customized approximate data structures (Bloom filters, MinHash fingerprints)

Information spread prediction in time

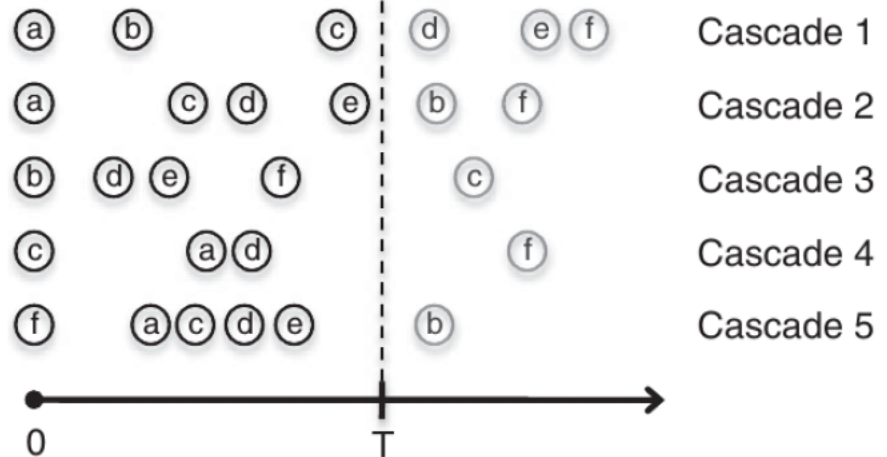


Diffusion Network



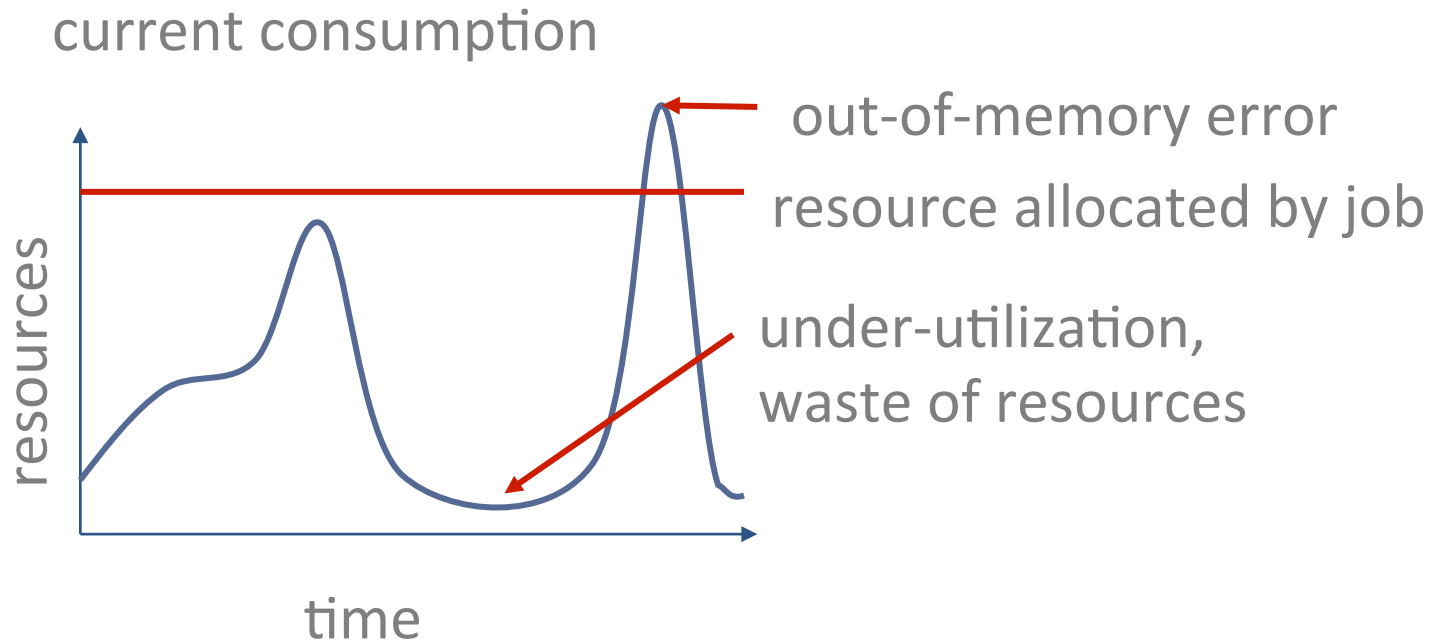
Observed

Unobserved



Collaboration w/ Ericsson on Spark Piggybacking

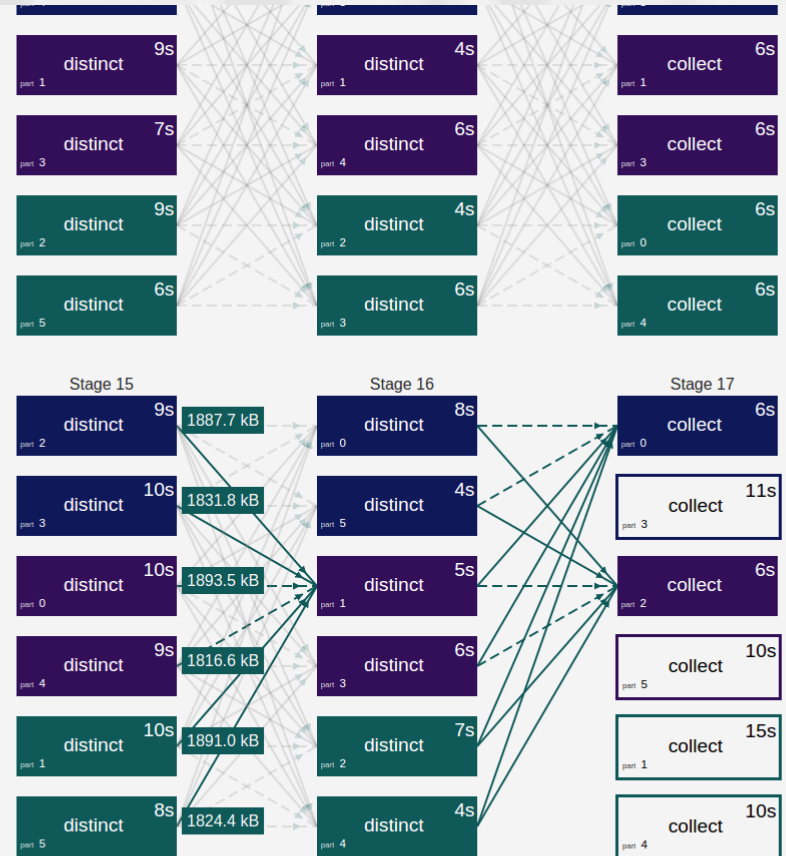
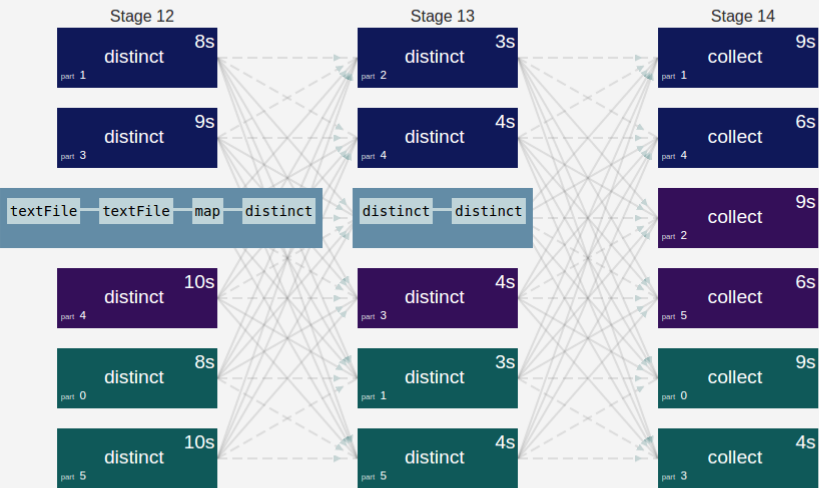
1. Hard to predict resource usage – overestimation hurts Spark
2. Peaks may cause out-of-memory errors – Spark will fail



Collaboration w/ Ericsson on Spark Piggybacking

ELAPSED TIME 4 min, 45 sec

1 nodes
3 executors
4 running tasks
104 completed tasks
0 failed tasks



CHANGE VIEW

SHOW NETWORK
TRAFFIC IN
KILOBYTES

ZOOM IN

ZOOM OUT

STOP

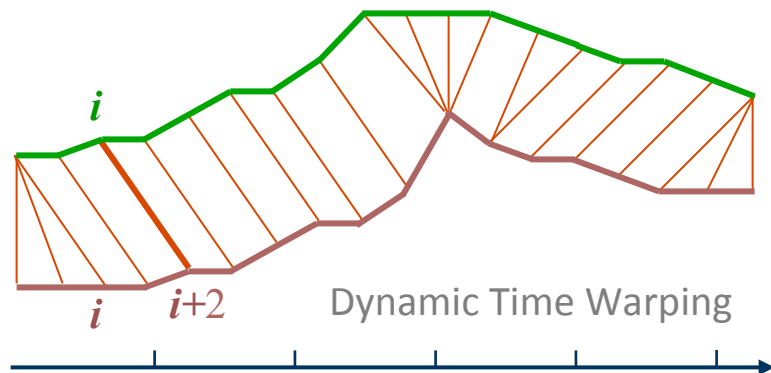
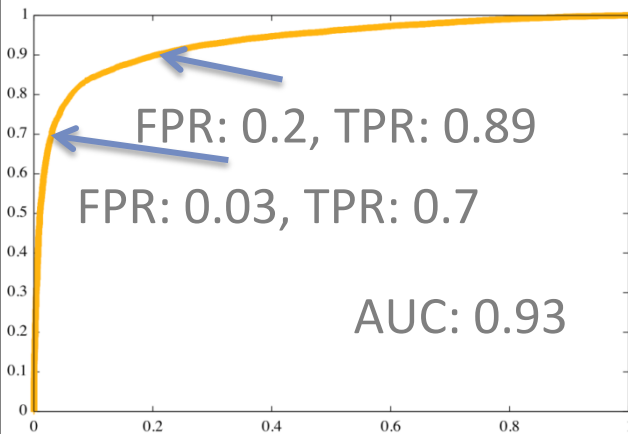
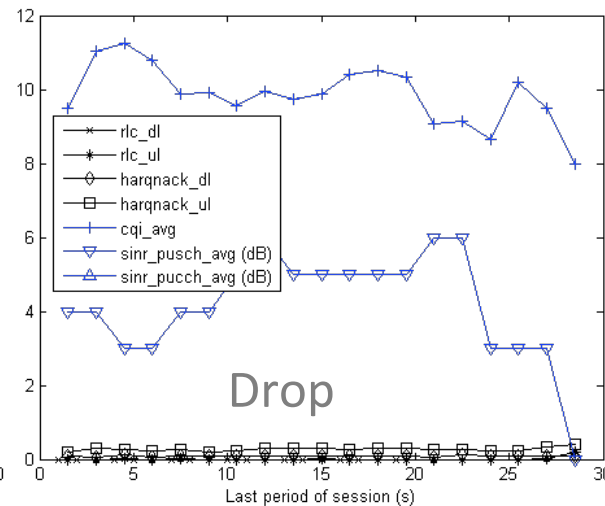
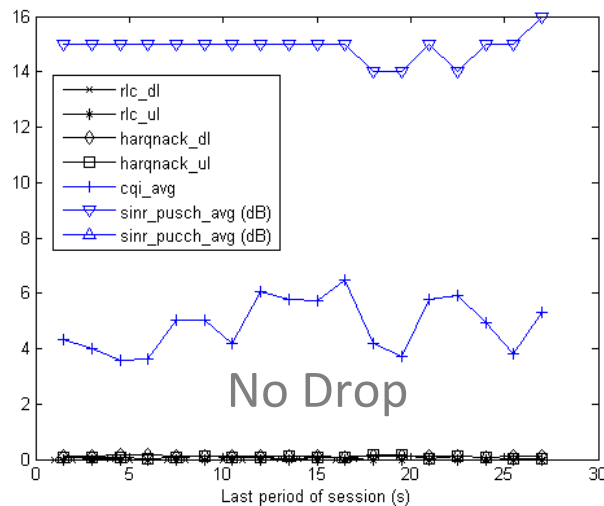
SHOW
ONLY
REMOTE
EDGES

AUTO
MODE

DROP FILES HERE

Data Stream Analytics: Mobile Session Drop

- Periodic radio channel physical parameter measurements
- Predict abnormal termination



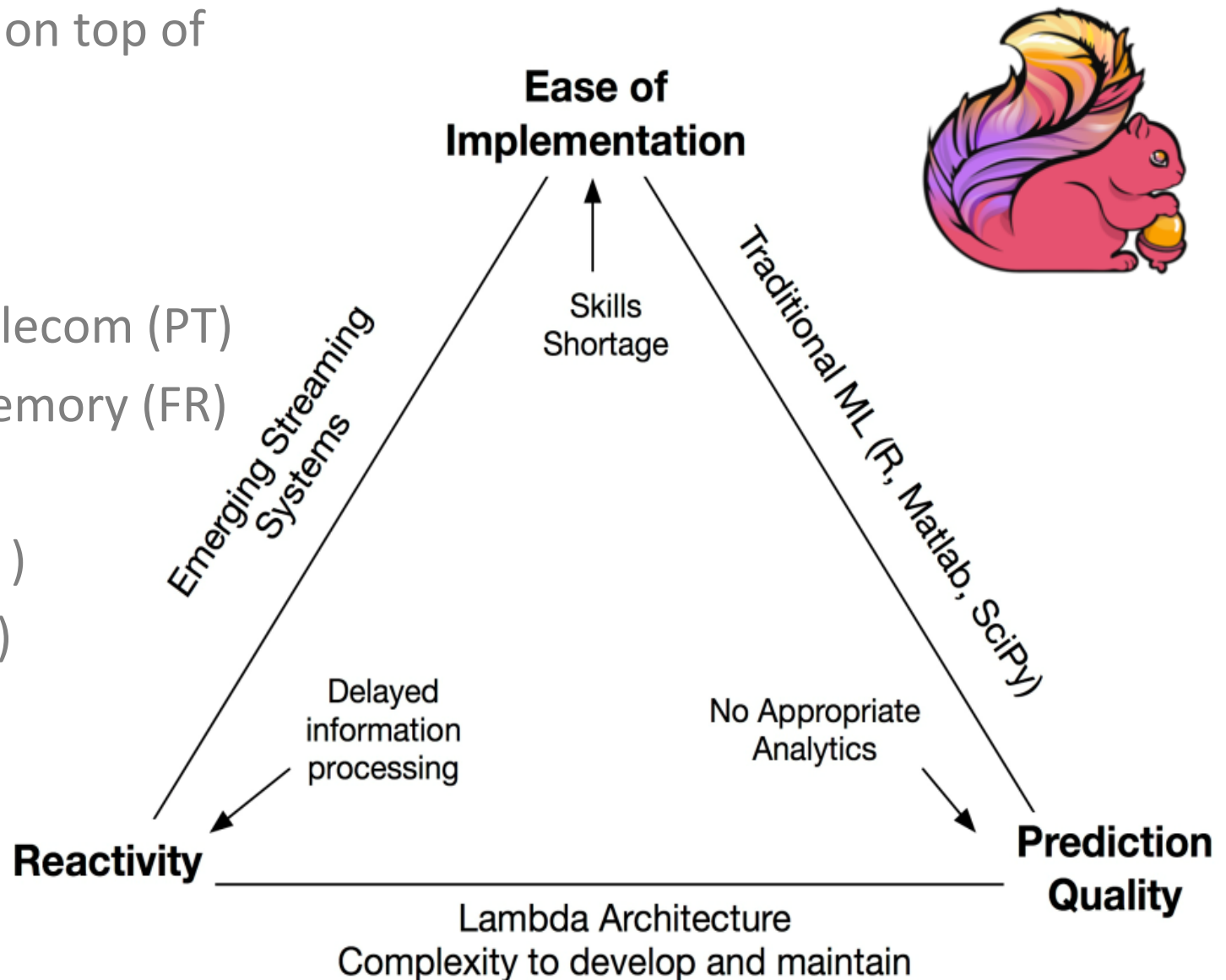
Best features selected by AdaBoost:

	Drop	Not drop
1. RLC NACK ratio UpL. Max	0.93447	0.12676
2. RLC NACK ratio UpL. Mean	0.11787	-0.11571
3. Harq NACK ratio Downl. Max	0.02061	0.00619
4. Delta RLC NACK ratio UpL. Mean	0.19277	0.18110
5. Signal-Interf+Noise Mean	1.92105	6.61538

STREAMLINE Magic Triangle

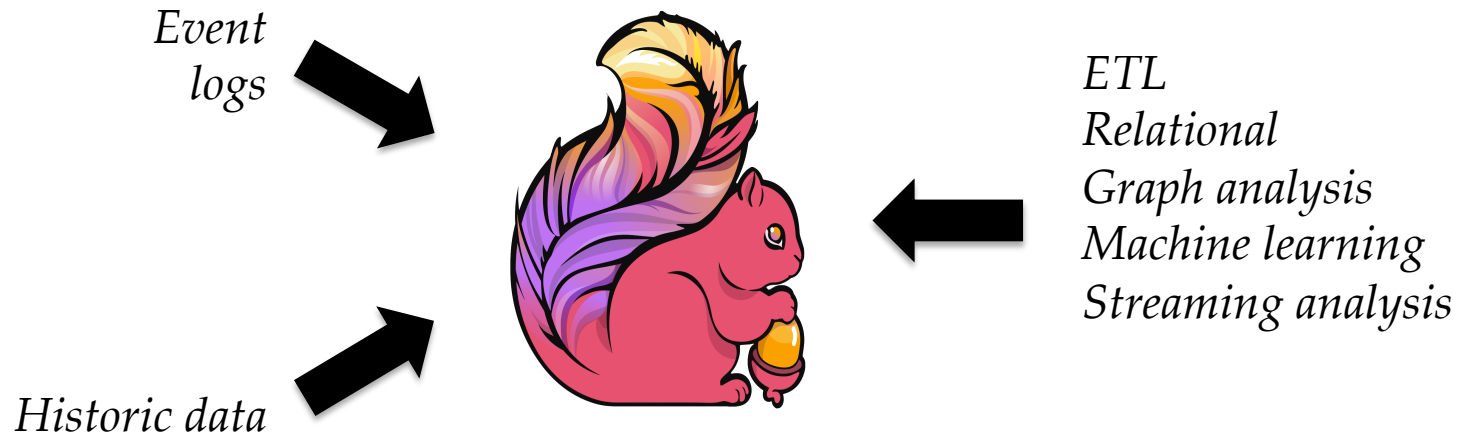
New initiative on top of
Apache Flink

- DFKI (DE)
- SICS (SE)
- Portugal Telecom (PT)
- Internet Memory (FR)
- Rovio (FI)
- NMusic (PT)
- SZTAKI (HU)



Batch and Streaming in Flink

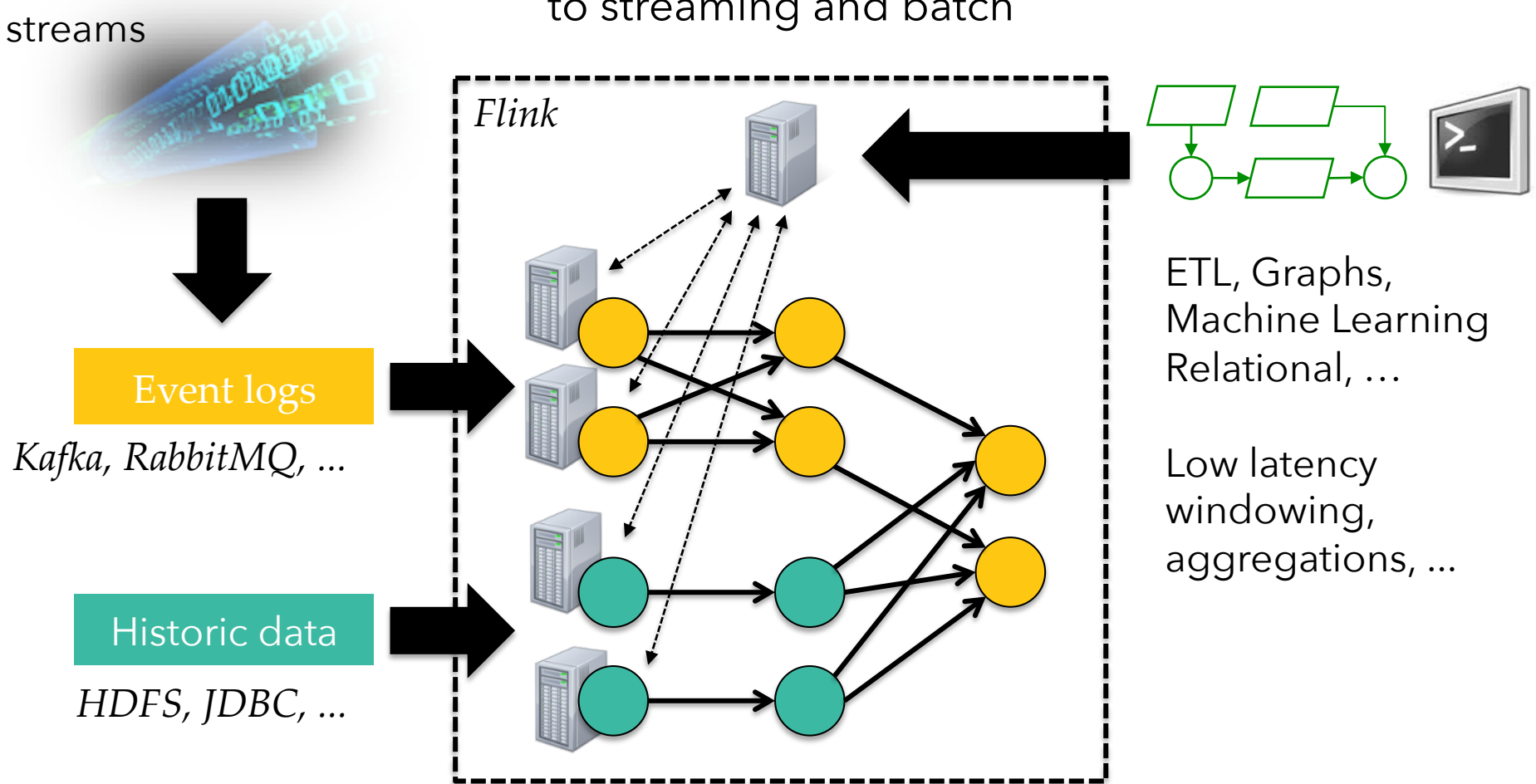
A "use-case complete" framework to unify batch and stream processing



Batch and stream: Same execution engine

Real-time data streams

An engine that puts equal emphasis to streaming and batch



Bike Share Challenge – one more month to go!

Welcome to the MOL Bubi Challenge 2015



MOL Bubi Challenge



Welcome



Dataset



Leaderboard BRP



Leaderboard DSDP



Username



Password

SIGN IN



Budapest Data Projects Meetup

Szeretettel meghívunk Benneteket a Budapest Data Projects Meetup következő, a MOL Bubi Challenge bemutatásával, az eddigi tapasztalatok megbeszélésével kapcsolatos rendezvényére 2015 október 28-án este 7-kor: [MOL Bubi challenge teams meetup](#)

Kérlek, a weboldalon jelezze mindenki a részvételi szándékát. Csapatonként elég egy regisztráció + további létszám jelzése.

New Data Scientists Wanted! Enter the MOL Bubi Challenge!

Participate in the [MOL Bubi public bike-sharing system](#) Analytics Challenge organized by the "Big Data - Momentum" research group of the [Hungarian Academy of Sciences \(MTA SZTAKI\)](#) and the Centre for Budapest Transport (BKK).

The Municipality of Budapest made a decision in 2008 to establish a public bike-sharing scheme. The MOL Bubi system went live in September 2014 with 1100 bicycles and 76 docking stations. You can use MOL Bubi day and night, with annual, semi-annual, and quarterly passes or one, three, and seven day tickets.



- New architecture for unified batch + stream needed
 - Apache Flink has the potential
- New machine learning is needed
 - We participate in turning research codes to open source software

Questions?