# Data In, Fact Out: Automated Monitoring of Facts by FactWatcher

Naeemul Hassan<sup>1</sup>, Afroza Sultana<sup>1</sup>, You Wu<sup>2</sup>, Gensheng Zhang<sup>1</sup>, Chengkai Li<sup>1</sup>, Jun Yang<sup>2</sup>, Cong Yu<sup>3</sup> <sup>1</sup>University of Texas at Arlington, <sup>2</sup>Duke University, <sup>3</sup>Google Research

# Motivation

- Factual statements backed by data are leads to news stories:
- Social Media: The social world's most viral photo ever generated 3.5 million likes, 170,000 comments and 460,000 shares by Wednesday.
- •Weather: This month the Chinese capital has experienced 10 days with a maximum temperature in around 35 degrees Celsius—the most for the month of July in a decade.
- Politics: Dick Durbin is one of the only three candidates who have raised more than 25% from PAC contributions and 25% from self-financing.

### **Concepts and Objectives**

- Append-only database  $R(\mathcal{D}, \mathcal{M})$ , latest tuple t
  - Dimension attributes  $\mathcal{D} = \{ \text{Player, Team, Season} \}$
  - Measure attributes  $\mathcal{M} = \{\text{Points}, \text{Assists}, \text{Blocks}\}$
  - o  $t = t_8$ : {Lamar Odom, Miami Heat, 2004, 30, 11, 11}
- C: a conjunctive expression on  $\mathcal{D}$ , M: a subspace of  $\mathcal{M}$ 
  - $\circ$  C: {Player = Lamar Odom, Team = \*, Season = 2004}
  - M: {Points, Blocks}
- Goal: Find 3 types of facts pertinent to the arrival of t
  - $\circ$  <u>Situational Facts<sup>[2]</sup></u>: (C, M) pairs where t is in the skyline
  - $\circ$  (C: {Player = \*, Team = \*, Season = 2004}, M: {Assists, Blocks})
  - season 2004 so far.
  - $\circ$  Prominent Streaks<sup>[3],[5]</sup>: (C, M) pairs where t generates prominent streaks
    - longest such streak in 2004.
  - $\circ$  <u>One-of-the-Few Objects<sup>[4]</sup></u>: (C, M) pairs where t is in the top-T skyband
  - (C: {Player = Lamar Odom, Team = \*, Season = \*}, M: {Points, Blocks})
  - this record. The same claim can be made for three performance records only.

### Challenges

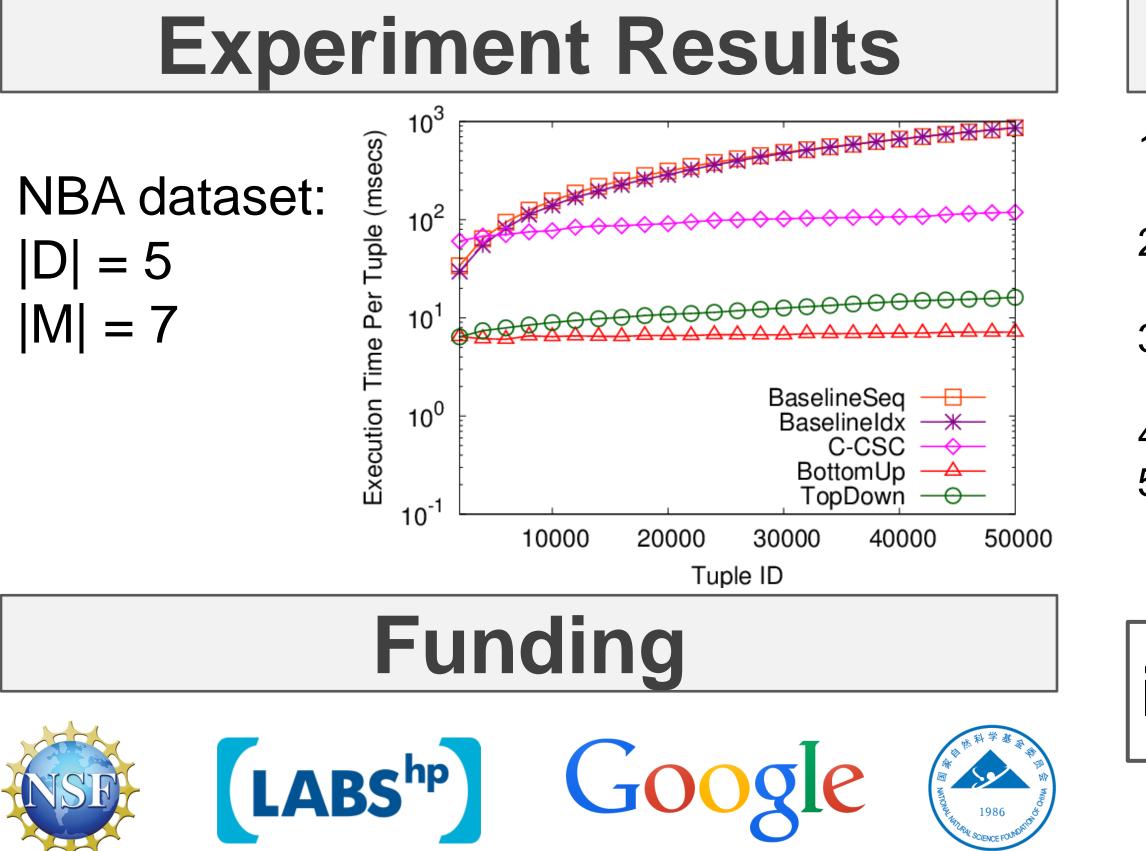
- •Huge space of possible facts
- $\circ$  216 (C,M) pairs in the above tiny table.
- •Fast, monitoring algorithm for reporting facts.
- Unified fact data model.
- •Unified fact ranking.



Fac

В
ß
_

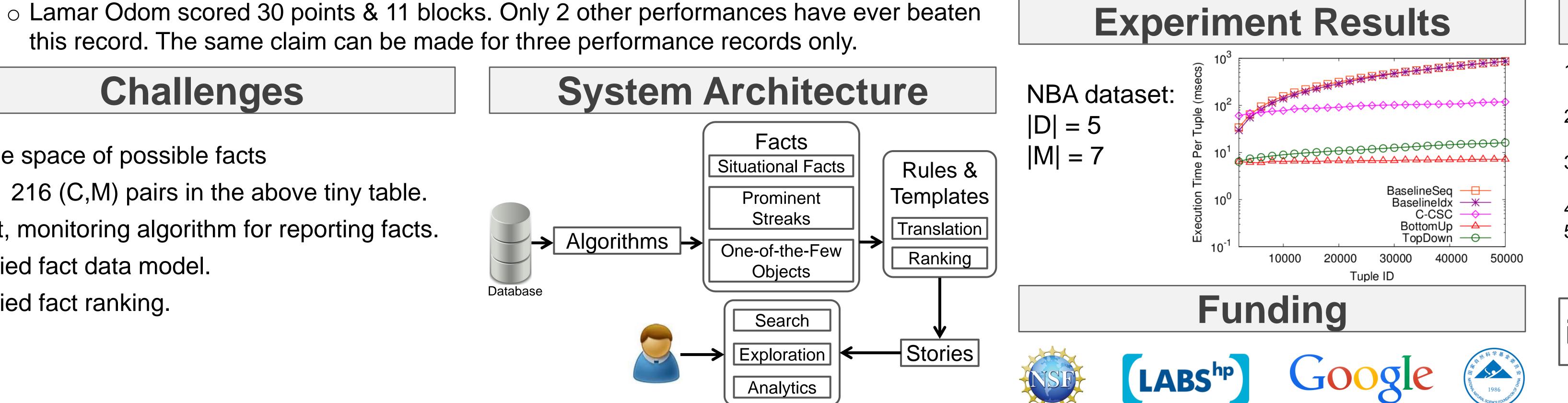


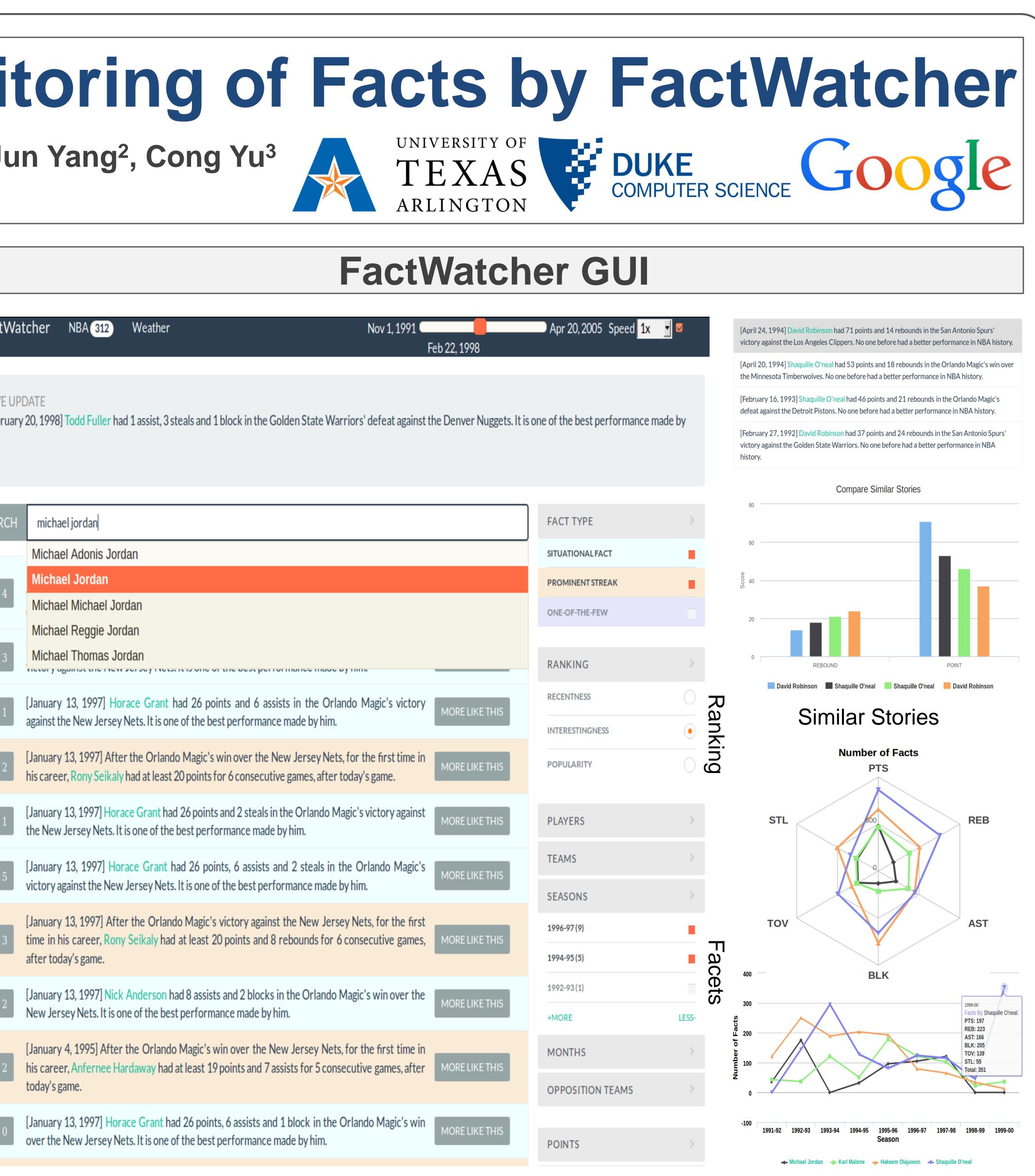


		<b>/</b>					
	Id	Player	Team	Season	Points	Assists	Blocks
}	$t_1$	Lamar Odom	LA Clippers	2003	30	11	12
	$t_2$	Eddie House	Miami Heat	2003	29	7	8
	$t_3$	Eddie House	Miami Heat	2003	28	6	9
	$t_4$	Lamar Odom	Miami Heat	2004	32	9	13
	$t_5$	Lamar Odom	Miami Heat	2004	28	11	6
	t <sub>6</sub>	Lamar Odom	Miami Heat	2004	29	9	7
	$t_7$	Eddie House	LA Clippers	2004	30	11	10
	$t_o$	Lamar Odom	Miami Heat	2004	30	11	11

Lamar Odom scored 11 assists and 11 blocks. No one made a better performance in

 $\circ$  (C: {Player = Lamar Odom, Team = 2004, Season = \*}, M: {Length, Points, Assists}) o Lamar Odom had at least 28 points and 9 or more assists for 4 consecutive games; the





ctWatcher	NBA 312	Weather	Nov 1, 1991	Apr 20, 2
			Feb 22, 1998	

#### »LIVE UPDATE

RCH	michaeljordan		FACT TY	
	Michael Adonis Jordan		SITUATION	
4	Michael Jordan			
	Michael Michael Jordan		ONE-OF-T	
	Michael Reggie Jordan			
3	Michael Thomas Jordan		RANKIN	
1	[January 13, 1997] Horace Grant had 26 points and 6 assists in the Orlando Magic's victory against the New Jersey Nets. It is one of the best performance made by him.	MORE LIKE THIS	RECENTNE	
2	[January 13, 1997] After the Orlando Magic's win over the New Jersey Nets, for the first time in his career, Rony Seikaly had at least 20 points for 6 consecutive games, after today's game.	MORE LIKE THIS	POPULAR	
1	[January 13, 1997] Horace Grant had 26 points and 2 steals in the Orlando Magic's victory against the New Jersey Nets. It is one of the best performance made by him.	MORE LIKE THIS	PLAYER	
5	[January 13, 1997] Horace Grant had 26 points, 6 assists and 2 steals in the Orlando Magic's victory against the New Jersey Nets. It is one of the best performance made by him.	MORE LIKE THIS	TEAMS SEASON	
	[January 13, 1997] After the Orlando Magic's victory against the New Jersey Nets, for the first time in his career, Rony Seikaly had at least 20 points and 8 rebounds for 6 consecutive games,	MORE LIKE THIS	1996-97 (9	
3	after today's game.		1994-95 (5	
	[January 13, 1997] Nick Anderson had 8 assists and 2 blocks in the Orlando Magic's win over the		1992-93(1	
2	New Jersey Nets. It is one of the best performance made by him.	MORE LIKE THIS	+MORE	
_	[January 4, 1995] After the Orlando Magic's win over the New Jersey Nets, for the first time in		MONTH	
	his career, Anfernee Hardaway had at least 19 points and 7 assists for 5 consecutive games, after today's game.	MORE LIKE THIS	OPPOSI	
	[January 13, 1997] Horace Grant had 26 points, 6 assists and 1 block in the Orlando Magic's win			
0	over the New Jersey Nets. It is one of the best performance made by him.	MORE LIKE THIS	POINTS	

### Main Interface

### Comparison of Players

# **Publications**

- 1. Data In, Fact Out. Automated Monitoring of Facts by FactWatcher. Hassan et al., VLDB 2014 demo.
- 2. Incremental discovery of prominent situational facts. Sultana et al., ICDE 2014.
- 3. Discovering general prominent streaks in sequence data. Zhang et al., TKDD 2014.
- 4. On one of the few objects. Wu et al., KDD 2012
- 5. Prominent streak discovery in sequence data. Jiang et al., KDD 2011

### idir.uta.edu/factwatcher

