

Azure Stream Analytics

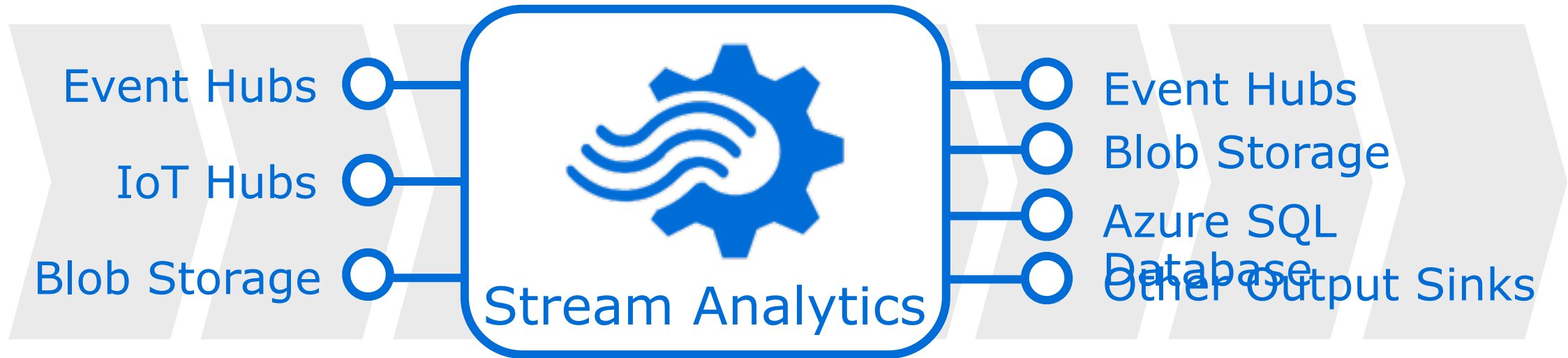
The Internet of Things (IoT)

- Currently 20 billion devices connected to the Internet
- By 2020, expect 50 billion or more
 - Health-monitoring devices
 - Thermostats, wind turbines, and solar farms
 - Cars, trucks, traffic lights, and drones
 - EVERYTHING will be connected
- How do you process all that data?
- How do you process it in real time?

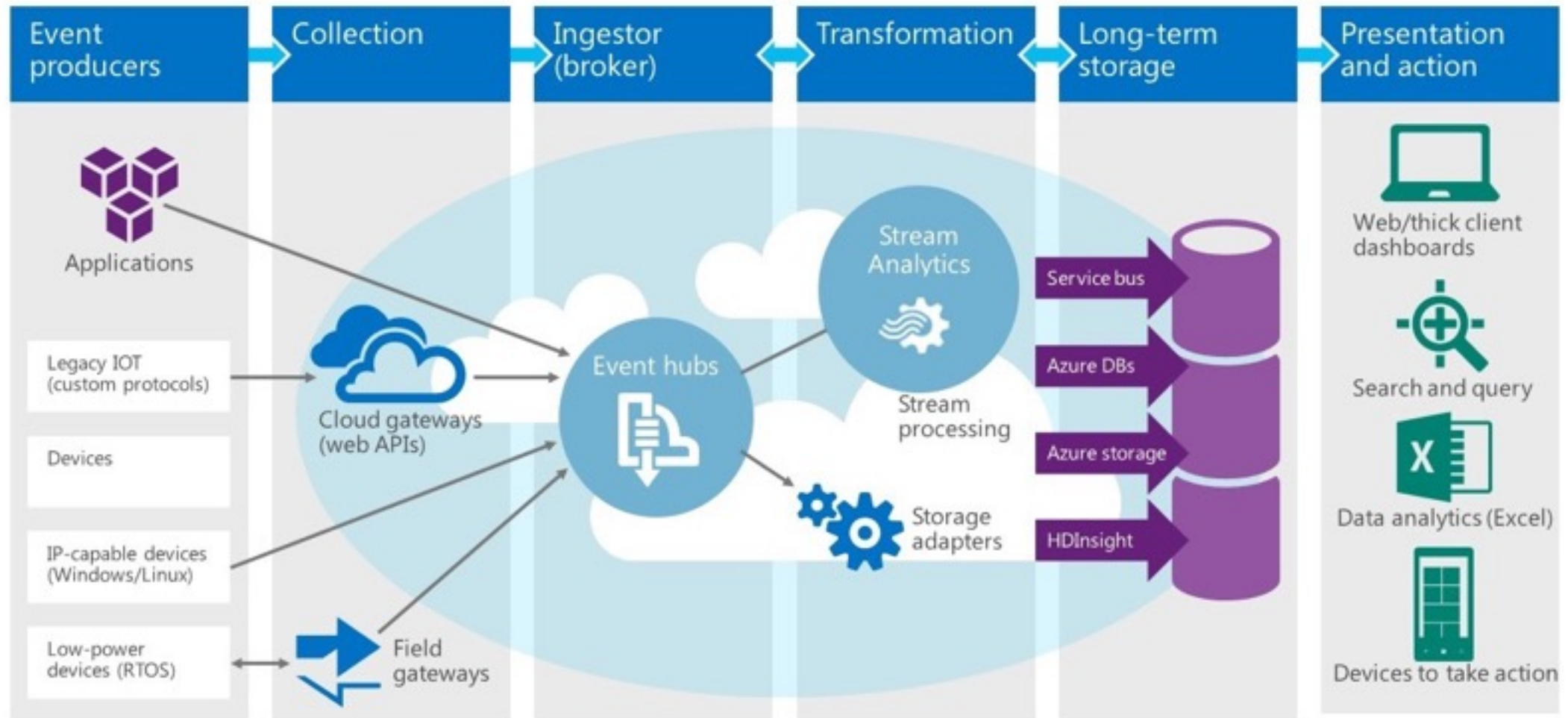


Azure Stream Analytics

- Highly scalable service for analyzing data in motion
- Supports SQL-like query language for data analysis
- Scales using Streaming Units (1 SU \sim 1 MB/sec)



Stream Analytics at Work



Stream Analytics Query Language

- SQL-like language for querying live data streams
 - Subset of T-SQL
 - Supports bigint, float, nvarchar(max), datetime, record, and array
 - Supports SELECT, FROM, WHERE, GROUP BY, and other common Data Manipulation Language (DML) statements
 - Supports COUNT, AVG, DATEDIFF, and other common functions
- Adds extensions such as TIMESTAMP BY and System.Timestamp
- Supports temporal grouping of events via "windowing"

Querying a Data Stream

- List all Connecticut cars that enter a toll booth, and include the entry time, toll booth ID, and license-plate number

```
SELECT EntryTime, TollId, LicensePlate  
FROM EntryData  
WHERE State = 'CT'
```

ENTRYTIME	TOLLID	LICENSEPLATE
2014-09-10T12:02:00+00:00	3	ABC 1004
2014-09-10T12:03:00+00:00	2	XYZ 1003
2014-09-10T12:11:00+00:00	1	NJB 1006

Designating a Field as the Event Time

- Designate the EntryTime field as the event time for calculations that involve event time

```
SELECT System.Timestamp AS [Entry Time],  
       TollId, LicensePlate  
FROM EntryData TIMESTAMP BY EntryTime  
WHERE State = 'CT'
```

ENTRYTIME	TOLLID	LICENSEPLATE
2014-09-10T12:02:00+00:00	3	ABC 1004
2014-09-10T12:03:00+00:00	2	XYZ 1003
2014-09-10T12:11:00+00:00	1	NJB 1006

JOINing Two Data Streams

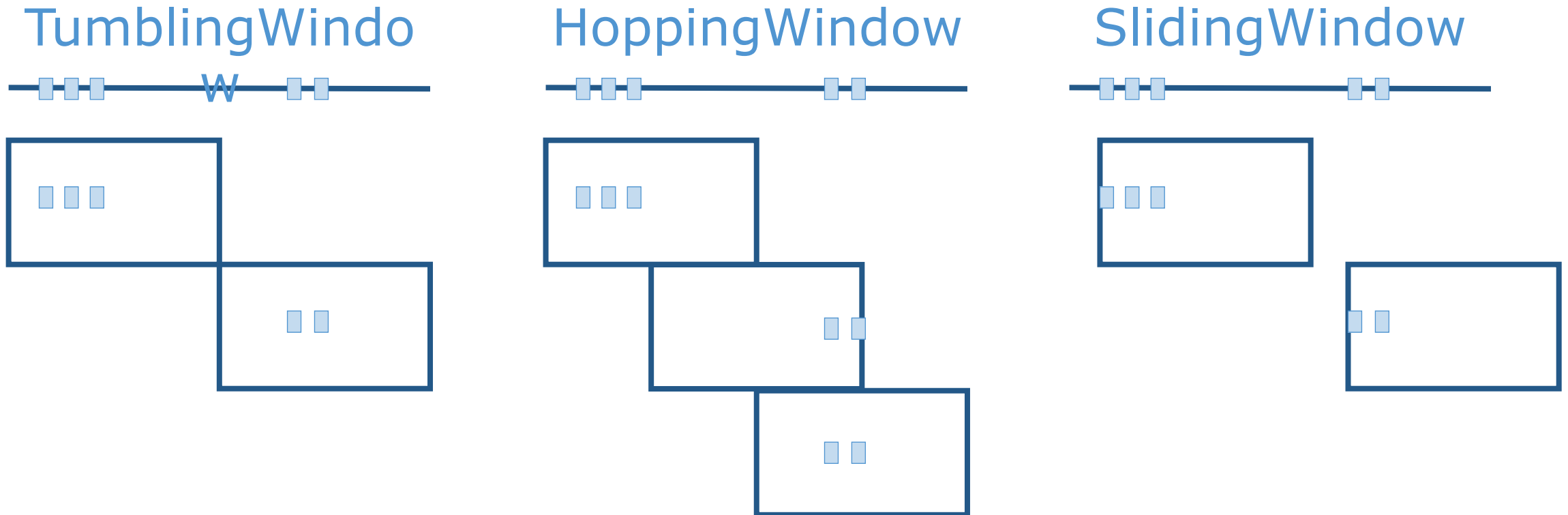
- How long does it take each car that enters a toll booth to pay the toll and exit the booth?

```
SELECT EN.TollId, EN.EntryTime, EN.LicensePlate,  
       DATEDIFF(minute, EN.EntryTime, EX.ExitTime) AS Minutes  
FROM EntryData EN TIMESTAMP BY EntryTime  
JOIN ExitData EX TIMESTAMP BY ExitTime  
  ON EN.TollId = EX.TollId  
  AND EN.LicensePlate = EX.LicensePlate  
  AND DATEDIFF(minute, EN, EX) BETWEEN 0 AND 60
```

TOLLID	ENTRYTIME	LICENSEPLATE	MINUTES
1	2014-09-10T12:01:00.000Z	JNB 7001	2
1	2014-09-10T12:02:00.000Z	YXZ 1001	1
3	2014-09-10T12:02:00.000Z	ABC 1004	2

Windowing

- Count or aggregate events over a specified time period



Using TumblingWindow

- How many New York cars enter a toll booth every 5 minutes?

```
SELECT DateAdd(minute, -5, System.TimeStamp)
       AS [Start Time], System.TimeStamp AS [End Time],
       COUNT(*)
FROM EntryData TIMESTAMP BY EntryTime
WHERE State = 'NY'
GROUP BY TumblingWindow(minute, 5)
```

START TIME	END TIME	COUNT
2014-09-10T12:00:00.000Z	2014-09-10T12:05:00.000Z	3
2014-09-10T12:05:00.000Z	2014-09-10T12:10:00.000Z	6
2014-09-10T12:15:00.000Z	2014-09-10T12:20:00.000Z	2

Using HoppingWindow

- What is the average wait time at all toll booths for the last 5 minutes, updated every 1 minute?

```
SELECT DateAdd(minute, -5, System.TimeStamp)
      AS [Start Time], System.TimeStamp AS [End Time],
      AVG(DATEDIFF(minute, EN.EntryTime, EX.ExitTime))
      AS [Average Wait Time]
FROM EntryData EN TIMESTAMP BY EntryTime
JOIN ExitData EX TIMESTAMP BY ExitTime
  ON EN.TollId = EX.TollId
  AND EN.LicensePlate = EX.LicensePlate
  AND DATEDIFF(minute, EN, EX) BETWEEN 0 AND 60
GROUP BY HoppingWindow(minute, 5, 1)
```

START TIME	END TIME	AVERAGE WAIT TIME
2014-09-10T11:58:00.000Z	2014-09-10T12:03:00.000Z	1.5
2014-09-10T11:59:00.000Z	2014-09-10T12:04:00.000Z	1.6666666666666667
2014-09-10T12:00:00.000Z	2014-09-10T12:05:00.000Z	1.6666666666666667

Using SlidingWindow

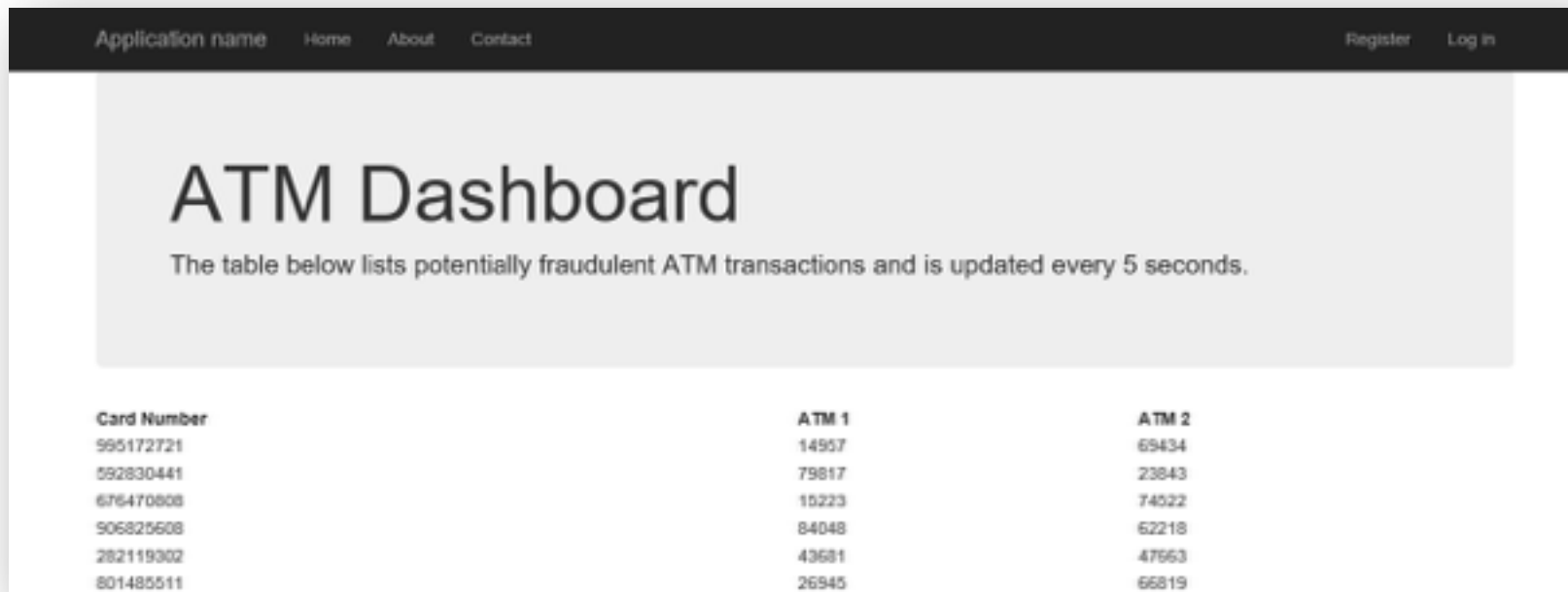
- In which 5-minute windows does at least one Connecticut car enter a toll booth?

```
SELECT DateAdd(minute, -5, System.TimeStamp)
       AS [Start Time], System.TimeStamp AS [End Time],
       TollId, COUNT(*)
FROM EntryData TIMESTAMP BY EntryTime
WHERE State = 'CT'
GROUP BY TollId, SlidingWindow(minute, 5)
HAVING COUNT(*) > 0
```

START TIME	END TIME	TOLLID	COUNT
2014-09-10T11:57:00.000Z	2014-09-10T12:02:00.000Z	3	1
2014-09-10T11:58:00.000Z	2014-09-10T12:03:00.000Z	2	1
2014-09-10T12:06:00.000Z	2014-09-10T12:11:00.000Z	1	1

Building Real-Time Dashboards

- Direct Stream Analytics output to an Azure event hub
- Write code that subscribes to events from the event hub



The screenshot shows a web application titled "ATM Dashboard". The header includes a navigation bar with links for "Application name", "Home", "About", "Contact", "Register", and "Log in". Below the header, the main content area features the title "ATM Dashboard" and a subtitle: "The table below lists potentially fraudulent ATM transactions and is updated every 5 seconds." Below this text is a table with three columns: "Card Number", "ATM 1", and "ATM 2". The table contains six rows of data.

Card Number	ATM 1	ATM 2
995172721	14957	69434
592830441	79817	23843
676470808	15223	74522
906825608	84048	62218
282119302	43681	47663
801485511	26945	66819