



Analitika v oblaku

Uroš Jemec

TEHNOLOGIJA



Agenda

- Uvod
- Predstavitev raziskave „BI Analytics and the Cloud“
- Arhitektura podatkovnega skladišča
 - Azure Data Factory
 - Azure SQL DataWarehouse
 - Azure Analysis Services
- Dva primera uporabe
- Zaključek



O meni

- Arhitekt podatkovnih skladišč
 - Studio Moderna
 - Pivovarna Laško Union
 - Bankart
 - Elektro Ljubljana
 - GEN-I
 - GEN Energija
- MCSE: Data Management and Analytics
- MCSE: Business Intelligence
- MS: Design and Implement Big Data Analytics Solutions
- MS-TSP

Reach us with #ntk17



Predstavitev raziskave TDWI

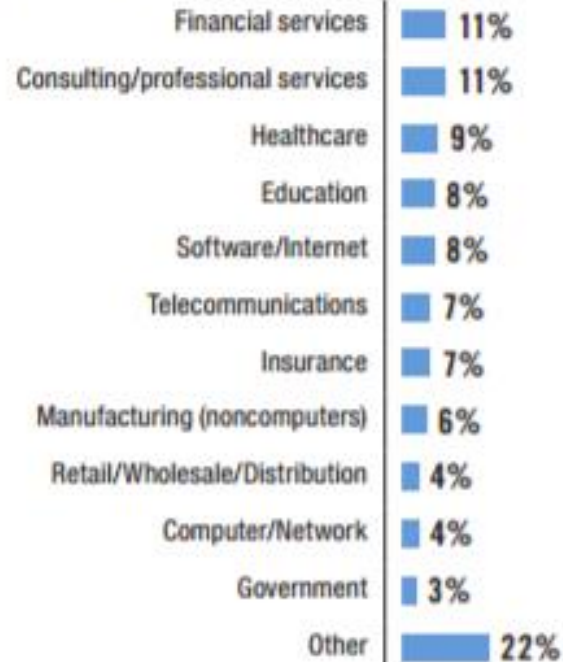
„BI Analytics and the Cloud“

Raziskovalna metodologija

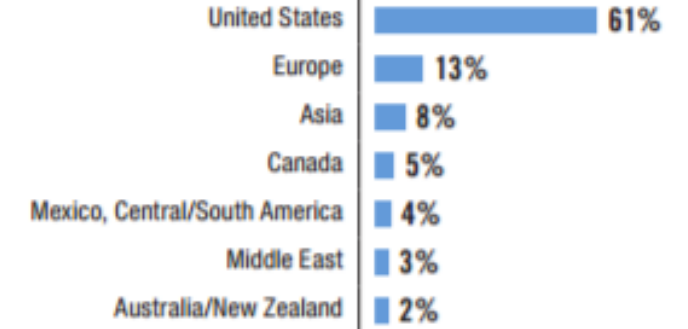
Position



Industry

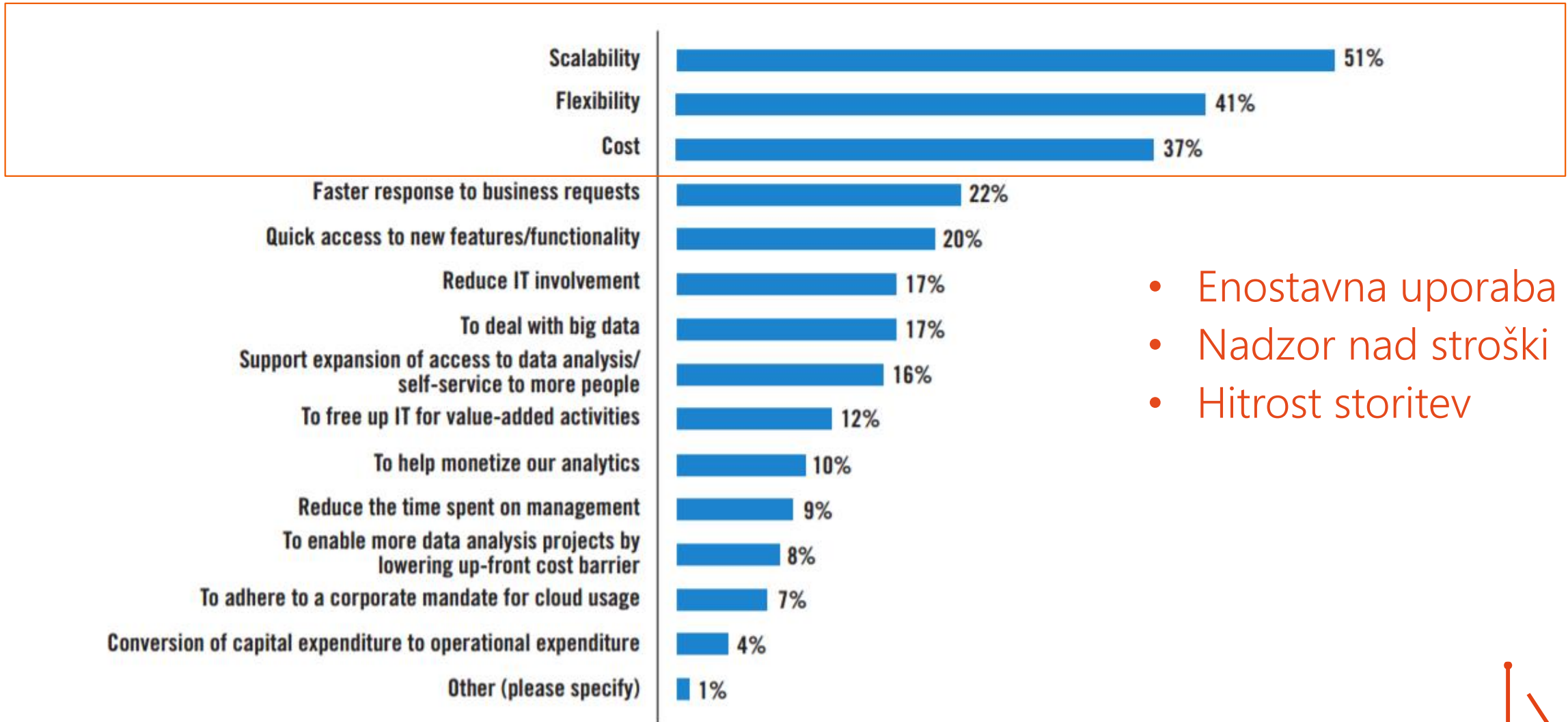


Geography



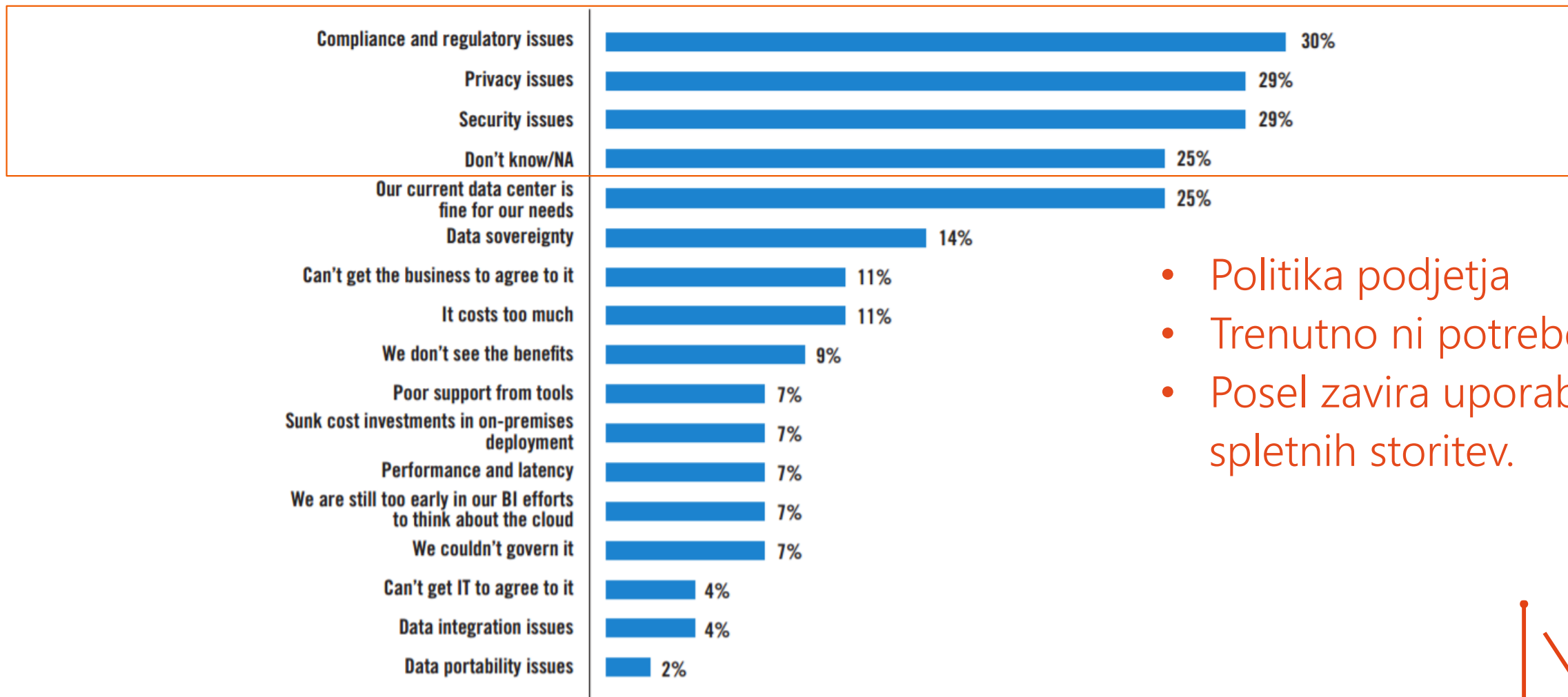
<https://tdwi.org/research/2016/10/best-practices-report-bi-analytics-and-the-cloud.aspx?tc=page0>

Kaj so ključni dejavniki, v prid analitiki v oblaku ?



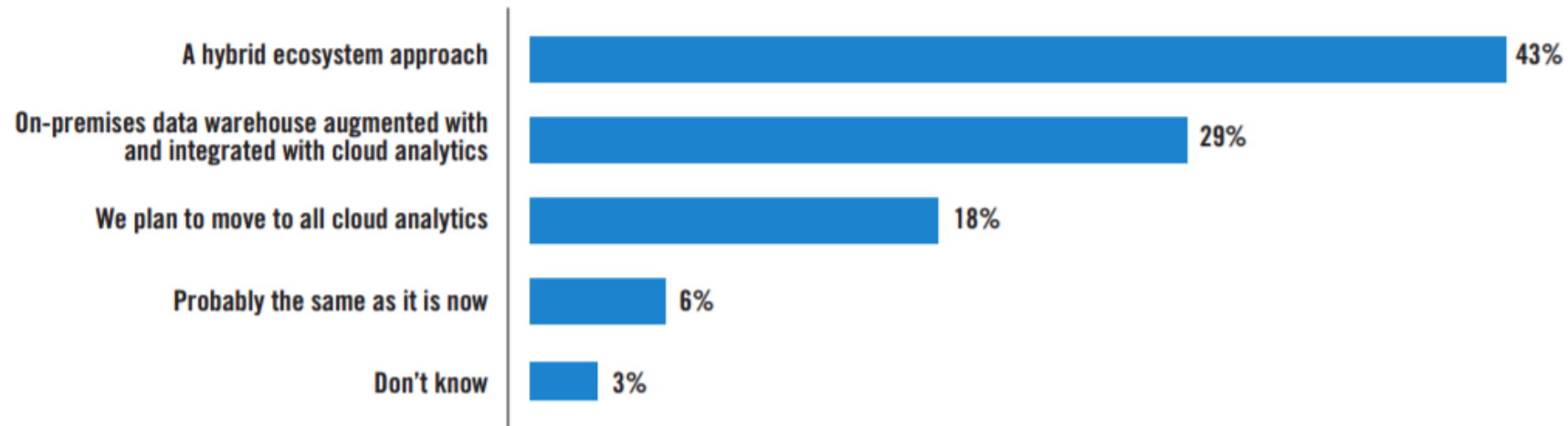
- Enostavna uporaba
- Nadzor nad stroški
- Hitrost storitev

Ključni dejavniki, da ne uporabljate analitiko v oblaku ?



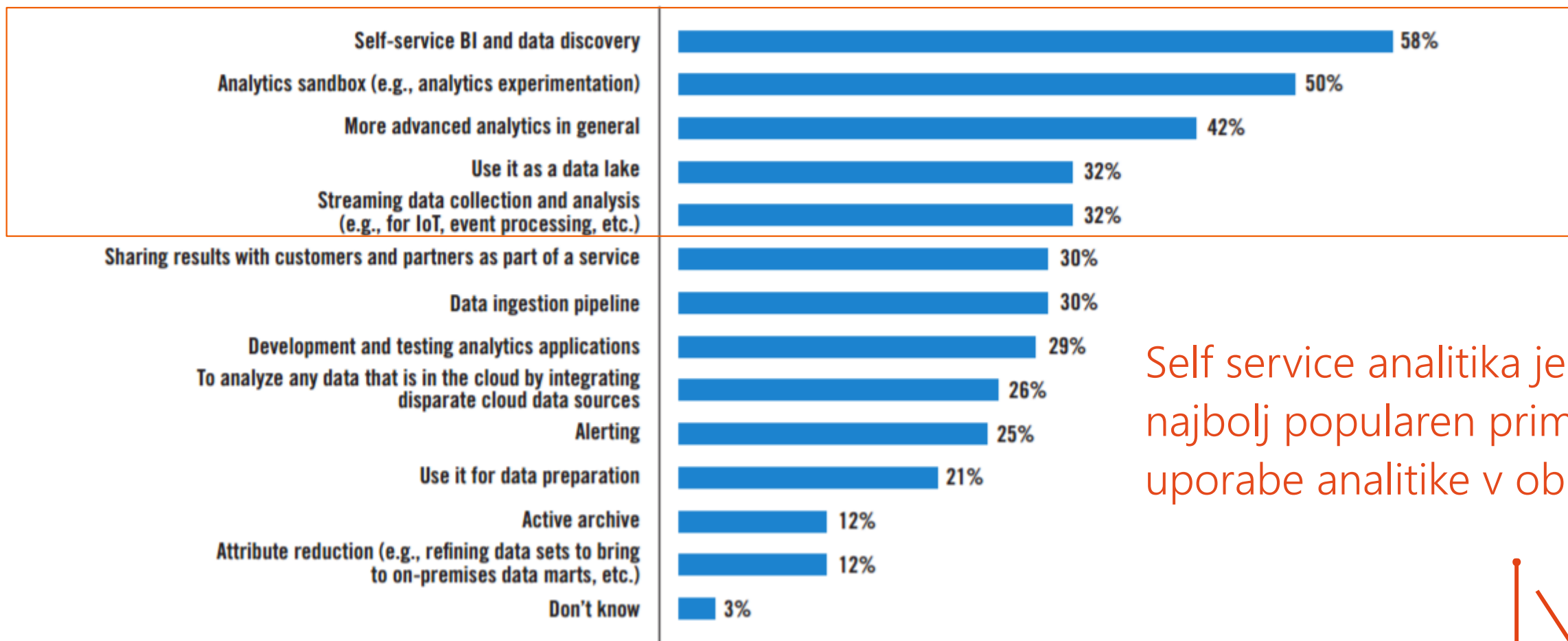
- Politika podjetja
- Trenutno ni potrebe
- Posel zavira uporabo spletnih storitev.

Kje vidite arhitekturo v naslednjih 3-5 let?



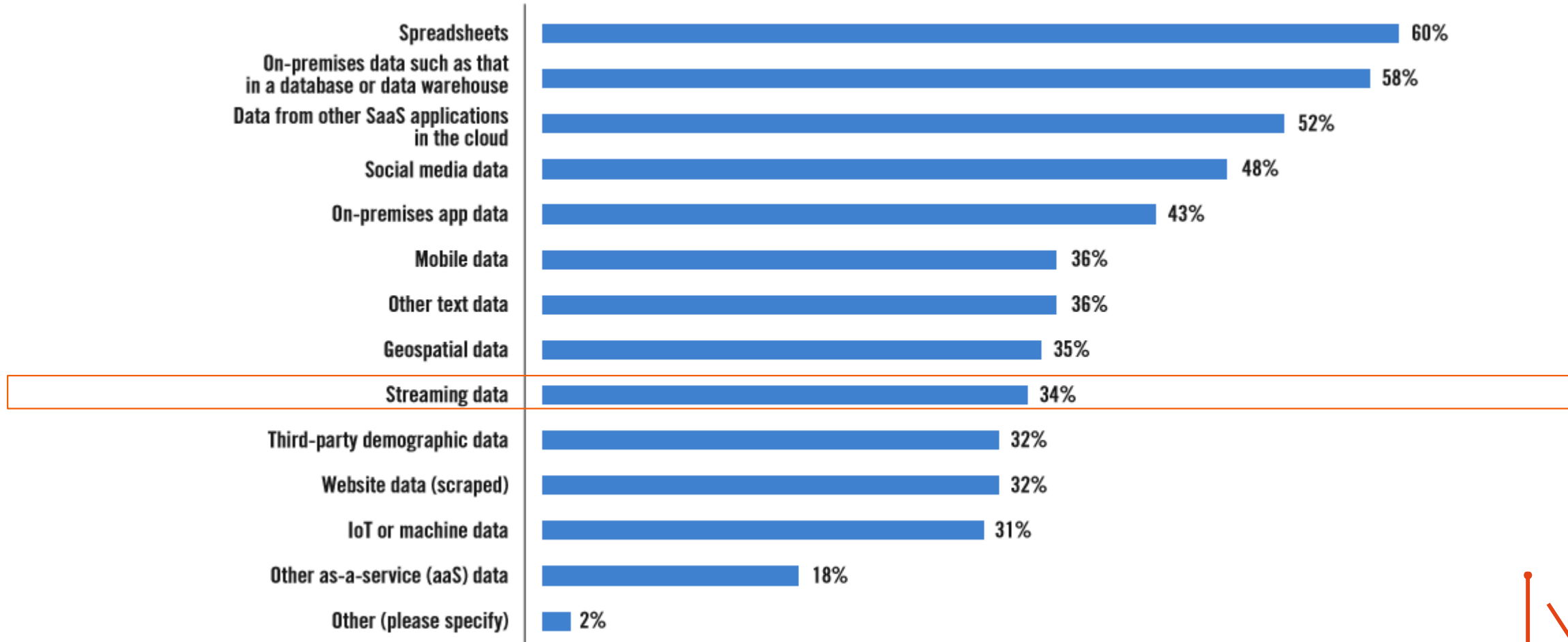
Več kot 40% uporabnikov meni
Do bo v naslednjih 3-5 let njihovo
Okolje hibridno

Kateri so potencialni scenariji uporabe za vas danes oziroma v bližnji prihodnosti?



Self service analitika je najbolj popularen primer uporabe analitike v oblaku

Katere vire podatkov rabite za povezovanje podatkov v oblak?



Arhitektura DWH okolja

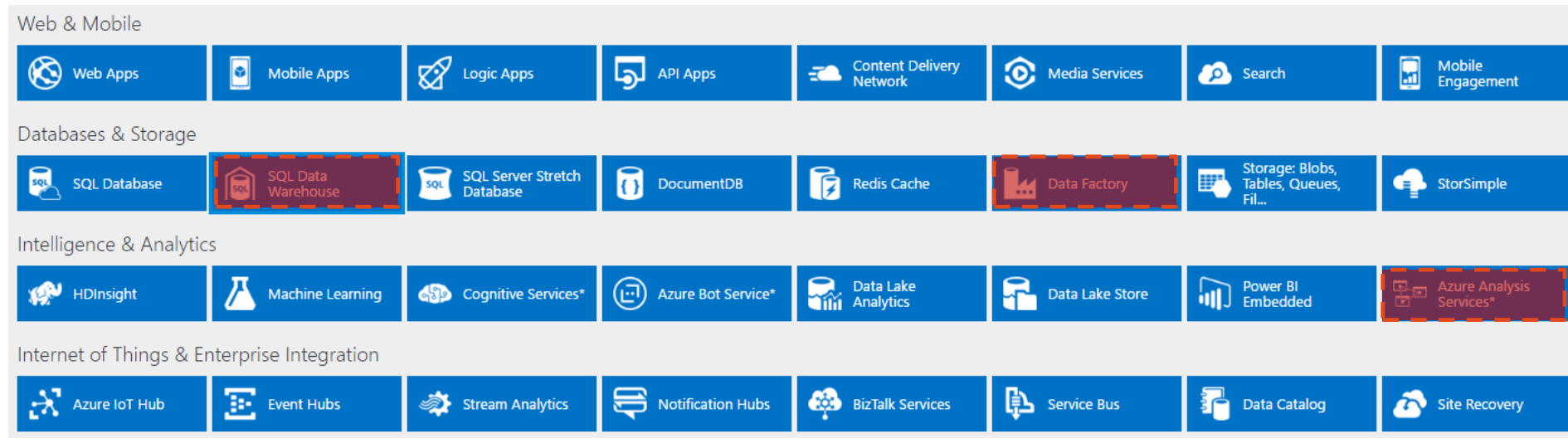
Iluzija ali realnost ?



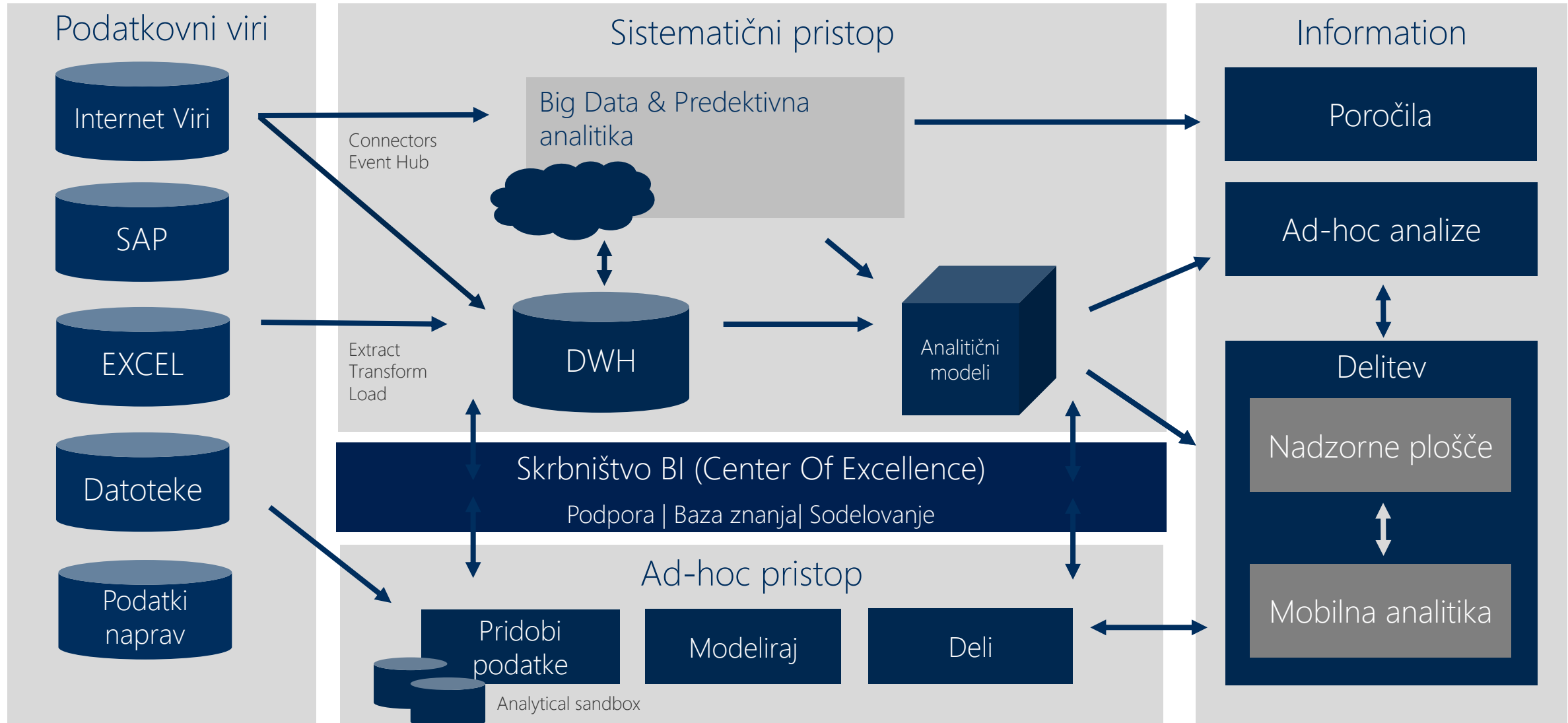
Entuziazem



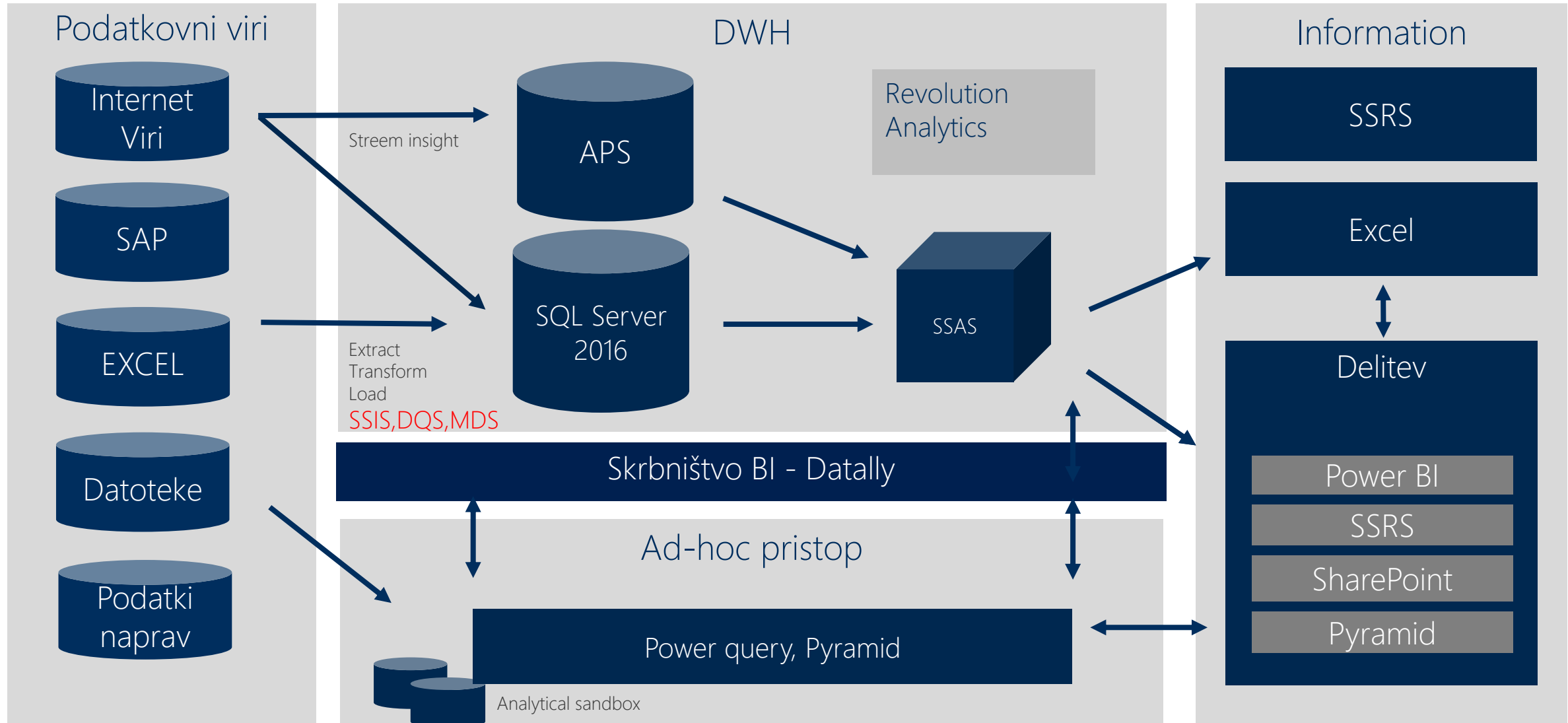
Realnost



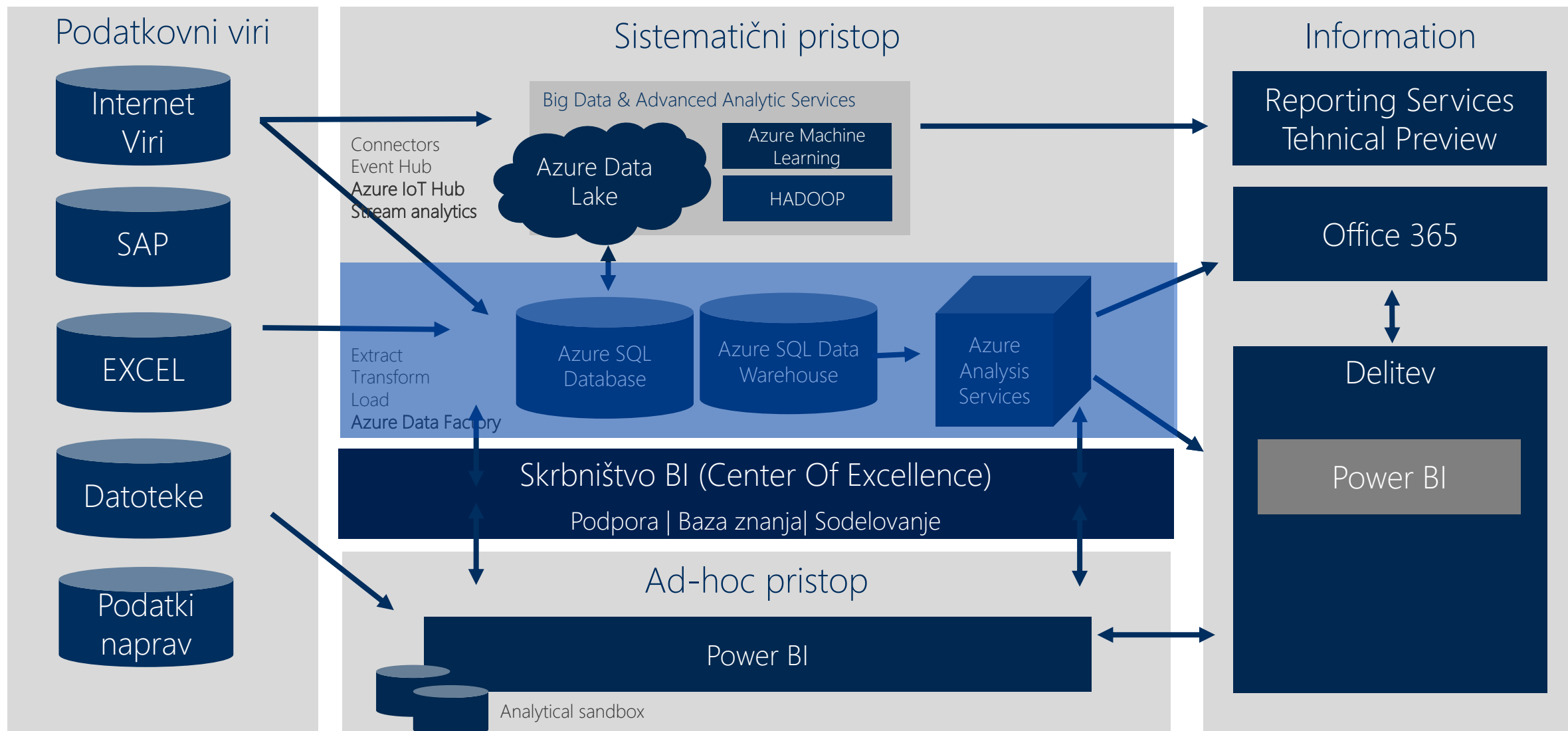
Sodobno DWH/BI okolje



Sodobno DWH/BI okolje – on Prem



Sodobno DWH/BI okolje – on Cloud



Azure Data Factory

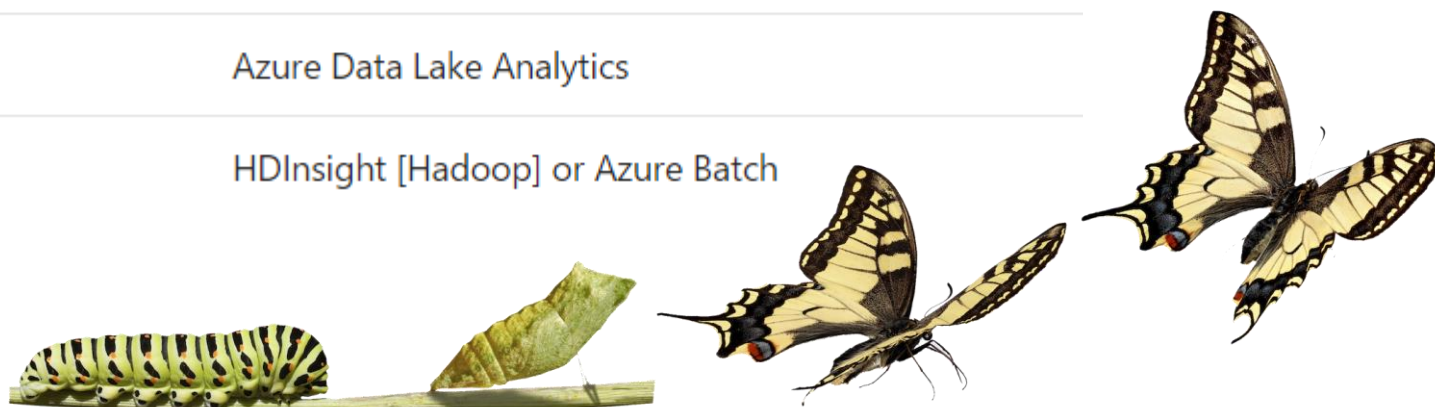
ETL orodje ki je v Oblaku?

Kaj lahko delam z Data Factory

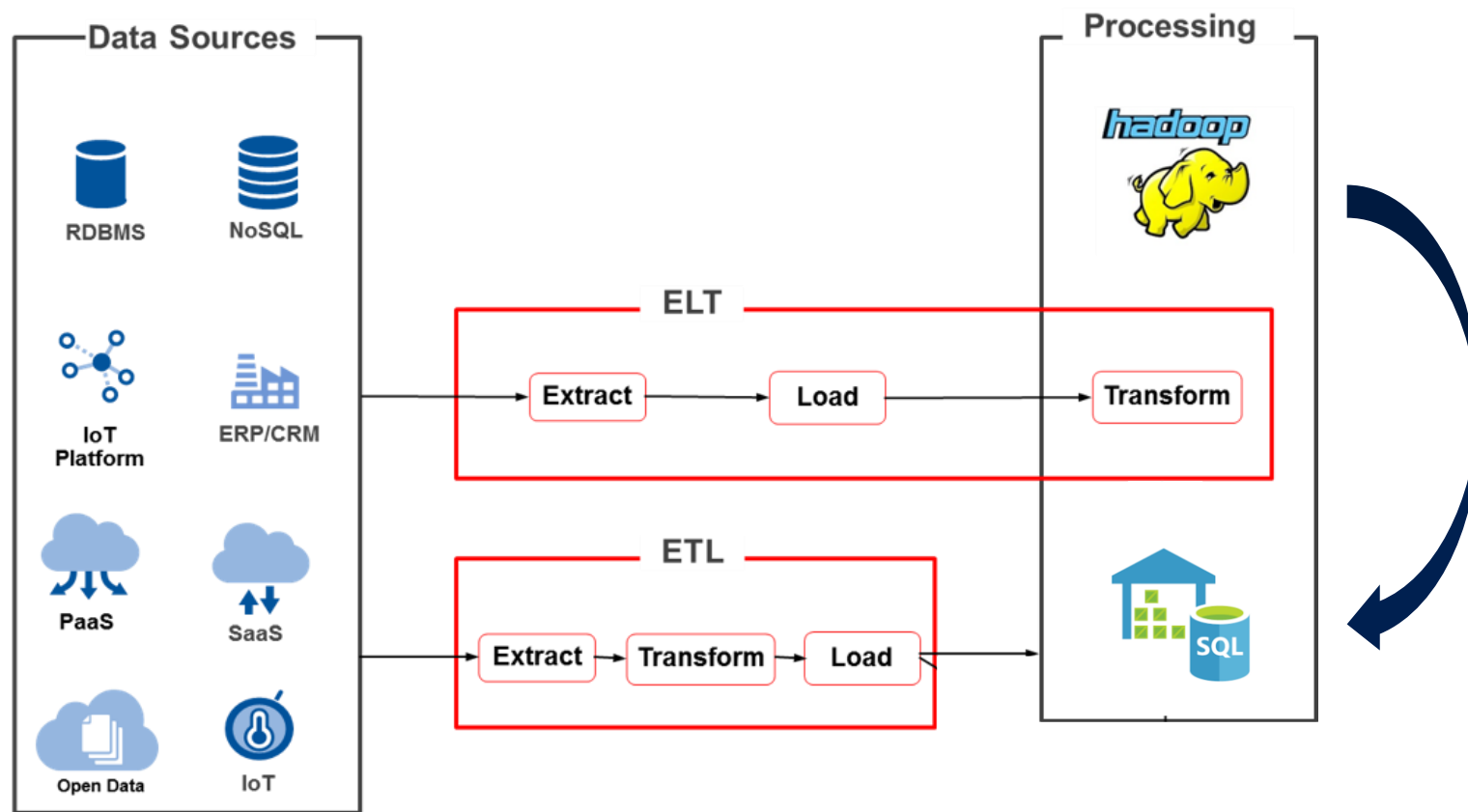


Transformacijski objekti

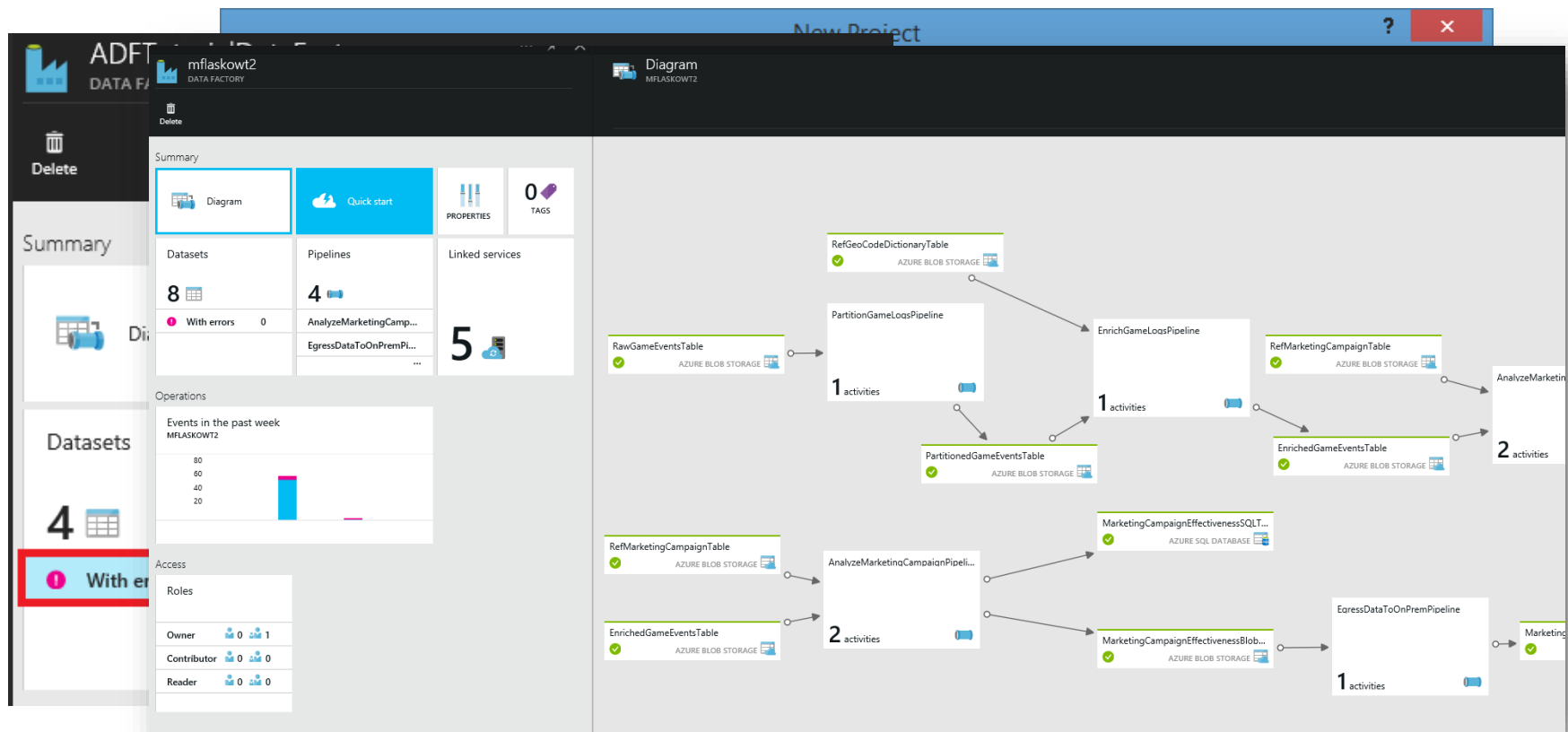
Hive	HDInsight [Hadoop]
Pig	HDInsight [Hadoop]
MapReduce	HDInsight [Hadoop]
Hadoop Streaming	HDInsight [Hadoop]
Spark	HDInsight [Hadoop]
Machine Learning activities: Batch Execution and Update Resource	Azure VM
Stored Procedure	Azure SQL, Azure SQL Data Warehouse, or SQL Server
Data Lake Analytics U-SQL	Azure Data Lake Analytics
DotNet	HDInsight [Hadoop] or Azure Batch



ETL vs ELT



DataFactory

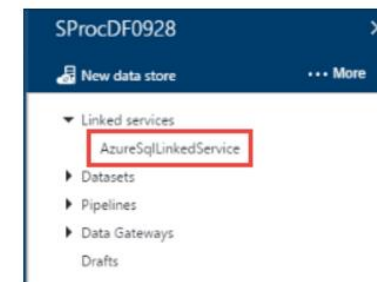


Primer Klica Procedure na Azure SQL Strežniku

- Create
- Store

```
{
  "name": "SprocActivitySamplePipeline",
  "properties": {
    "activities": [
      {
        "type": "SqlServerStoredProcedure",
        "typeProperties": {
          "storedProcedureName": "NTK_ImeStoredProcedure",
          "storedProcedureParameters": {
            "DateTime": "$Text.Format('{0:yyyy-MM-dd HH:mm:ss}', SliceStart)"
          }
        },
        "outputs": [
          {
            "name": "sprocsampleout"
          }
        ],
        "scheduler": {
          "frequency": "Hour",
          "interval": 1
        },
        "name": "NTK_SP_Activity"
      }
    ],
    "start": "2017-05-02T00:00:00Z",
    "end": "2017-05-02T05:00:00Z",
    "isPaused": false
  }
}
```

ota)



ETL Framework

Prednosti uporabe ETL Frameworka:

- **Standardizacija** razvoja podatkovnega skladišča
- V naprej pripravljeno podatkovno skladišče
- Generatorji nam omogočajo **hitrejši** razvoj

Dodana vrednost na projektu:

- Hiter razvoj
- **Konsistentna** koda
- Vsi razvijajo na poenoten način
- Enostavno vzdrževanje
- Enostavna možnost prilagajanja kode
- **Ponovljivost**

Veliko prihranjenega časa, pri izdelavi podatkovnega skladišča!

JSON Generator

EXEC [pr_DW_Source_ADM_GenerateJSON] 'FactProdaja'

75 %

Results Messages

	GeneratedJSON
1	{"name": "DatumNarocila", "type": "Date"}
2	{"name": "DatumDobave", "type": "Date"}
3	{"name": "ProduktSID", "type": "Int32"}
4	{"name": "StrankaSID", "type": "Int32"}
5	{"name": "prod_Kolicina", "type": "Int32"}
6	{"name": "prod_CenaNaEnoto", "type": "Decimal"}
7	{"name": "prod_PopustNaEnoto", "type": "Decimal"}
8	{"name": "prod_Skupaj", "type": "Decimal"}
9	{"name": "EtIInsertDate", "type": "Date"}
10	{"name": "EtIID", "type": "Int32"},

```
select
    DatumNarocila,
    ProduktSID,
    Prod_kolicina
from
    FactProdaja
FOR JSON PATH, ROOT('Prodaja')
```

75 %

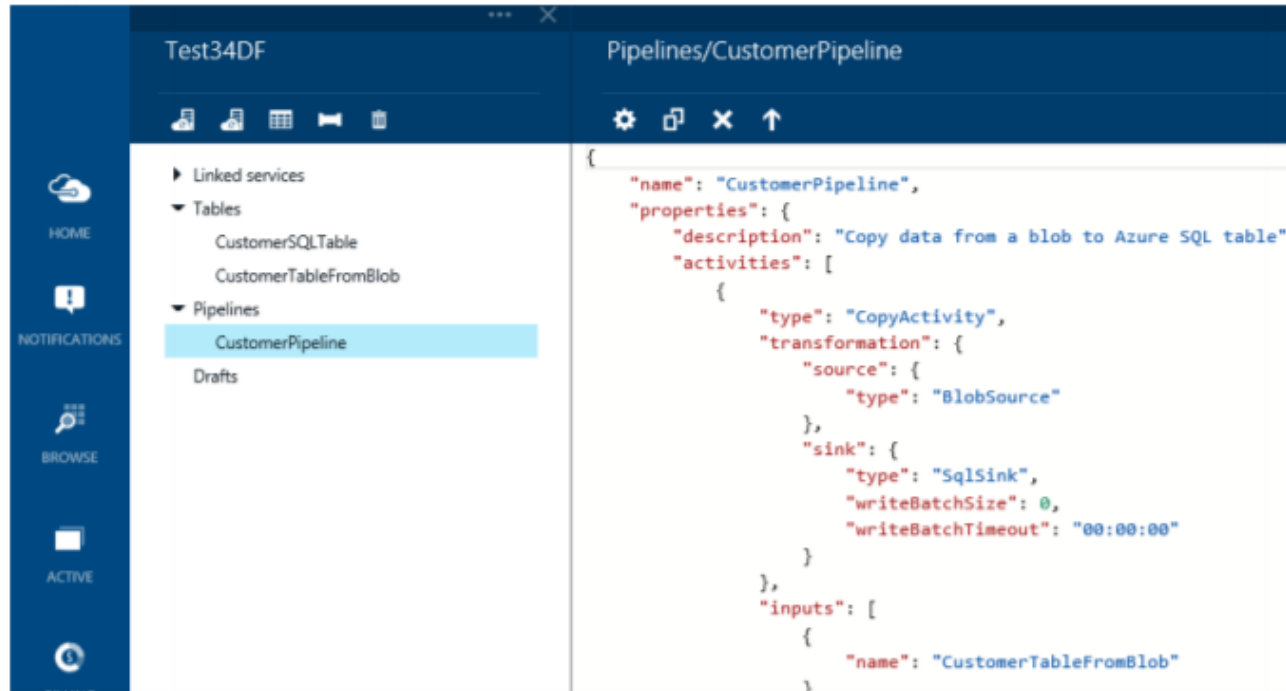
Results Messages

	JSON_F52E2B61-18A1-11d1-B105-00805F49916B
1	["Prodaja":{"DatumNarocila":"2008-01-11","Produ...

FOR JSON
OPENJSON
ISJSON
JSON_VALUE
JSON_QUERY
JSON_MODIFY

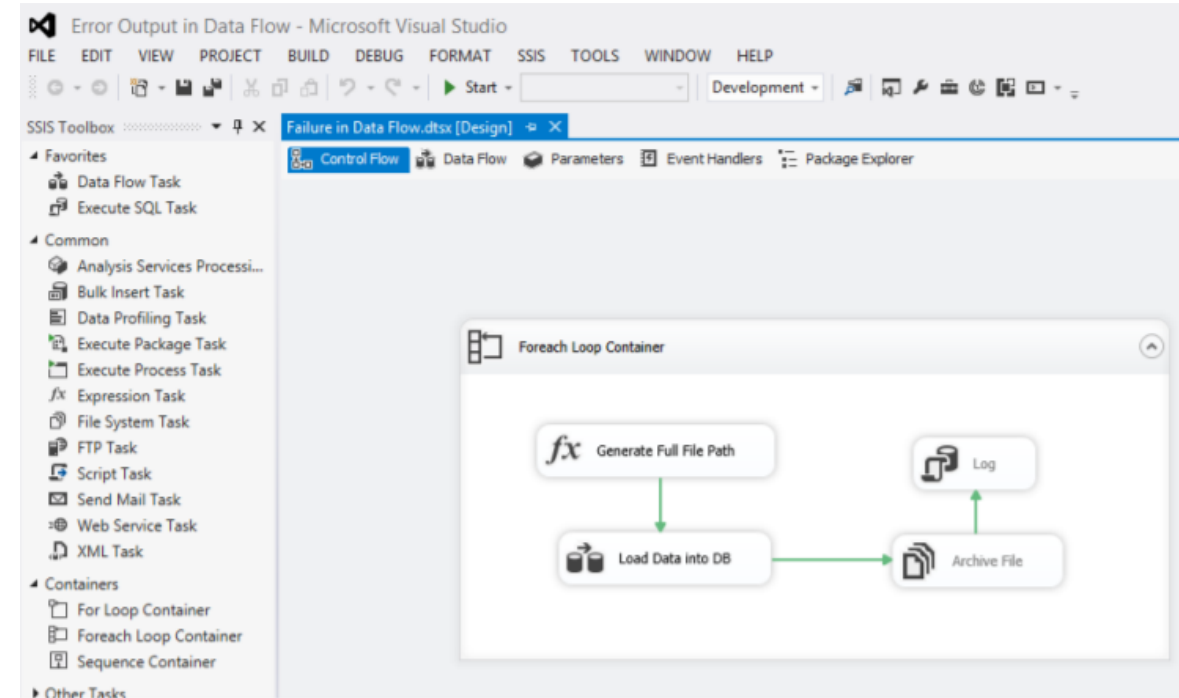
<https://docs.microsoft.com/en-us/azure/data-factory/data-factory-json-scripting-reference>

DataFactory vs SSIS



The screenshot shows the Azure Data Factory (ADF) interface. On the left, the 'Pipelines' section is expanded, showing 'CustomerPipeline' selected. The main editor displays the JSON definition for the 'CustomerPipeline'.

```
{
  "name": "CustomerPipeline",
  "properties": {
    "description": "Copy data from a blob to Azure SQL table",
    "activities": [
      {
        "type": "CopyActivity",
        "transformation": {
          "source": {
            "type": "BlobSource"
          },
          "sink": {
            "type": "SqlSink",
            "writeBatchSize": 0,
            "writeBatchTimeout": "00:00:00"
          }
        },
        "inputs": [
          {
            "name": "CustomerTableFromBlob"
          }
        ]
      }
    ]
  }
}
```



The screenshot shows the Microsoft Visual Studio SSIS (SQL Server Integration Services) interface. The 'SSIS Toolbox' on the left lists various tasks, including 'Data Flow Task', 'Execute SQL Task', 'Bulk Insert Task', 'Data Profiling Task', 'Execute Package Task', 'Execute Process Task', 'Expression Task', 'File System Task', 'FTP Task', 'Script Task', 'Send Mail Task', 'Web Service Task', and 'XML Task'. The 'Containers' section lists 'For Loop Container', 'Foreach Loop Container', and 'Sequence Container'. The main design view shows a 'Foreach Loop Container' with the following tasks:

- Generate Full File Path (Expression Task)
- Load Data into DB (Data Flow Task)
- Archive File (Data Flow Task)
- Log (Log Task)

The flow is: Generate Full File Path → Load Data into DB → Archive File → Log.

SSIS vs Azure Data Factory

	SQL Server Integration Services	Azure Data Factory
Razvojna orodja	Da	Da (potrebno več kodiranja)
Administracijska orodja	Da	Da (potrebno več kodiranja)
Viri in ponori	Ogromno virov	Tako internet kot tudi onprem viri
Objekti transformacije	Ogromno	Kodiranje
Cenovna politika	Po funkcionalnosti	Po porabi
Obvladovanje napak	Event Handlers, Alerts	Alerts
Logiranje	Da	Da
Monitoriranje	Da	Da
Data lineage		Da
HD Insight kompatibilnost	Omejena	Velika

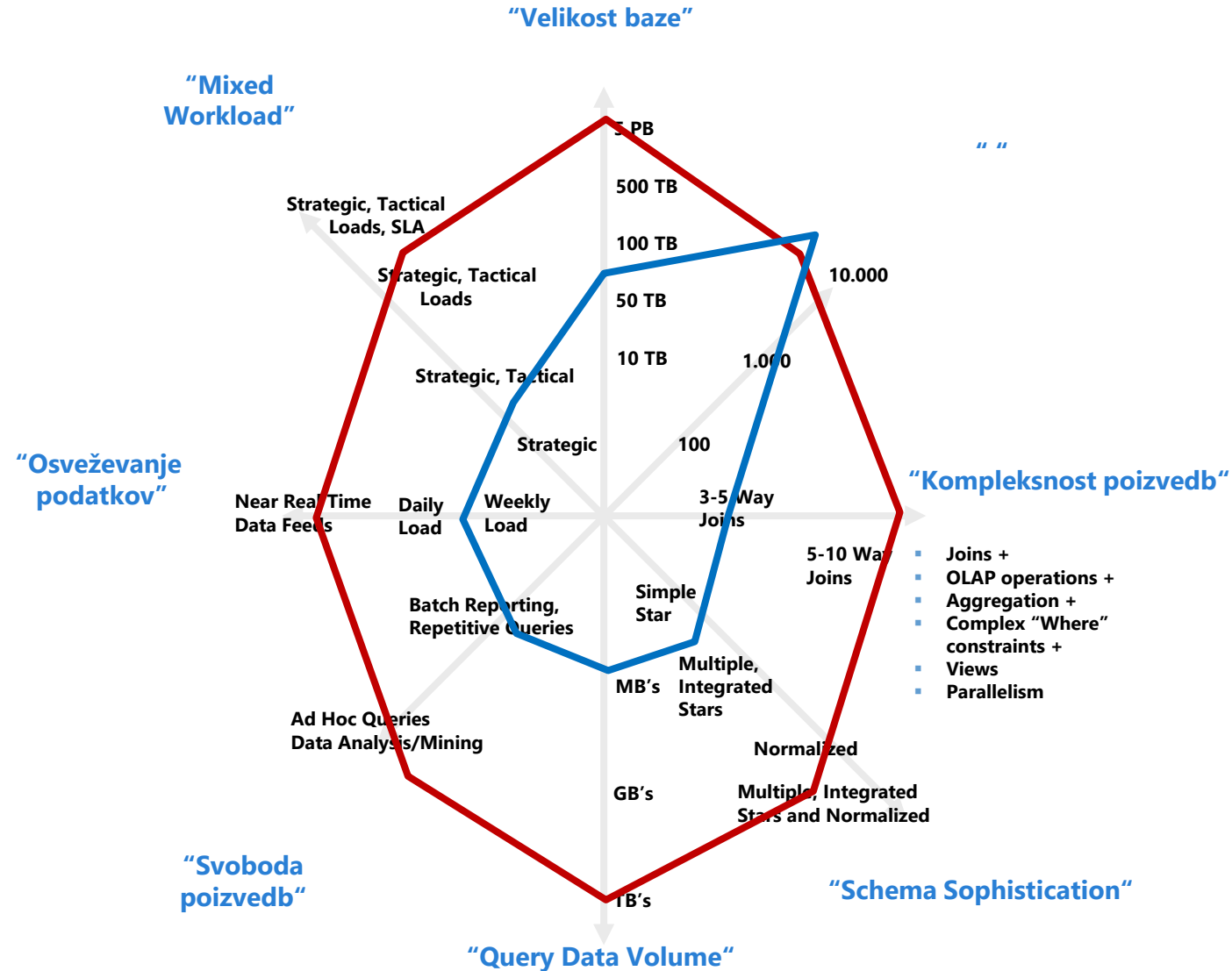
Azure Data Warehouse

MPP baza ki je v Oblaku

Osnovni princip MPP baz



DWH Skalabilnost



- MPP – Multidimensional Scalability
- SMP – Tunable in one dimension on cost of other dimensions

The spiderweb depicts important attributes to consider when evaluating Data Warehousing options.

Big Data support is newest dimension.

Azure SQL DataWarehouse



Kaj je SQL DW

- Podatkovno skladišče v oblaku kot service (**PaaS**)
- MPP v oblaku

Prednosti

- Ločeno diskovno polje od procesne moči. Na enostaven način lahko **povečamo zmogljivost** podatkovnega skladišča.
- Omogoča nam opcijo „**pause**“, v tem času ne plačujemo najemnine

Primeri uporabe

- Ogromna podatkovna skladišča
- Integracija z ogromno raznolikimi viri (**PolyBase**)
- Ni Primeren za **OLTP** aplikacije

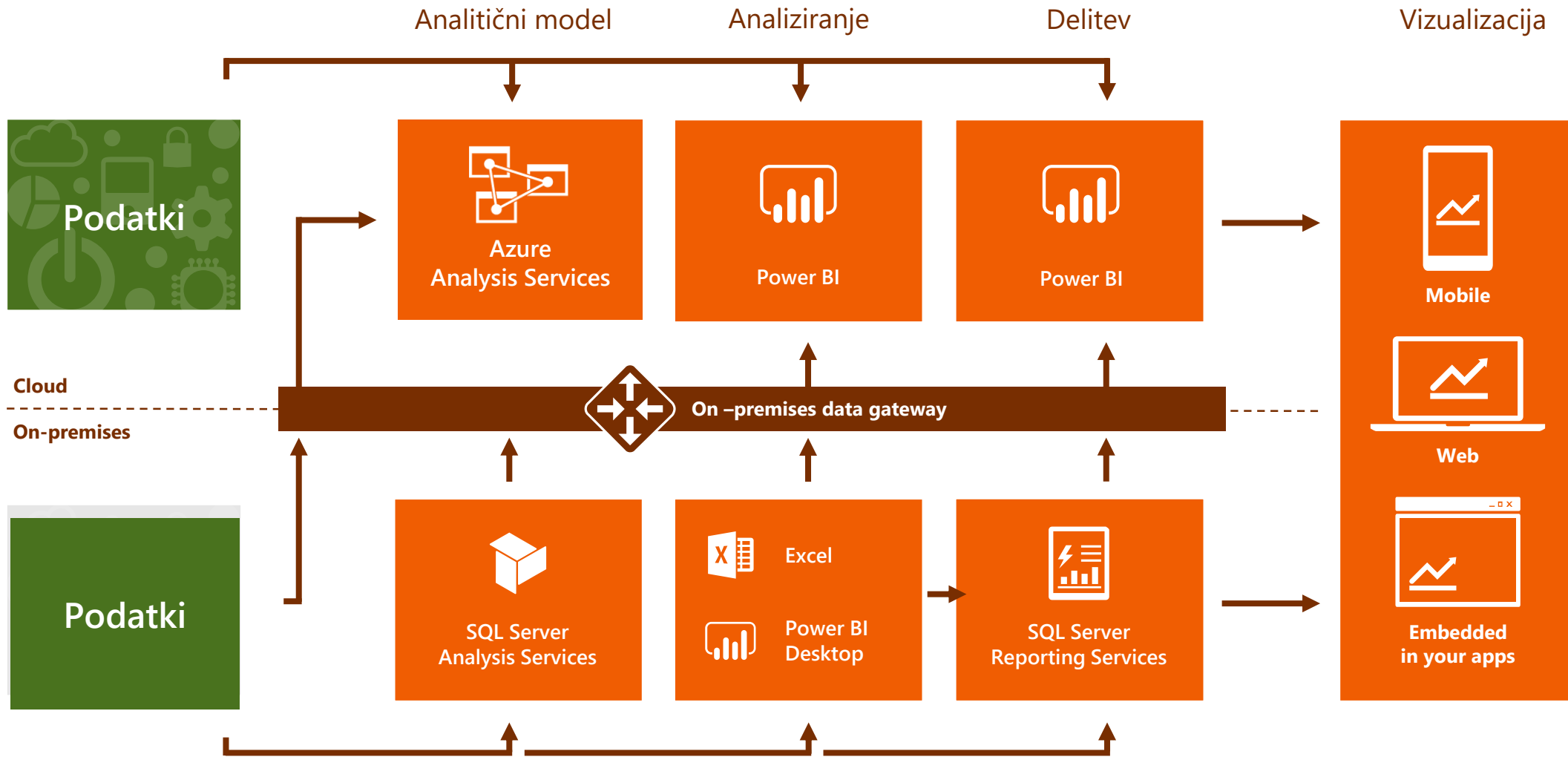
ADD & Azure SQL Data Warehouse

- Microsoft Country Partner of the Year 2016
- FLP 2015 Award for Big Data in EMEA region (Microsoft & HP)
- Izkušnje z Microsoft APS
- APS in Azure SQL DW Framework

Azure Analysis Services

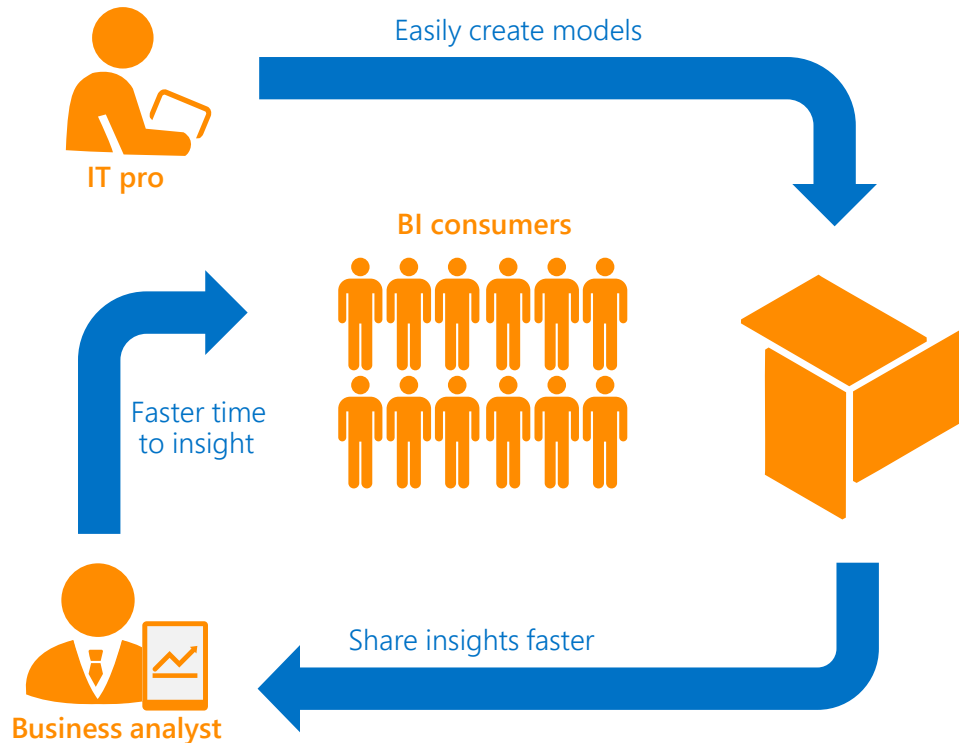
Analitiční model v oblaku

Analysis Services



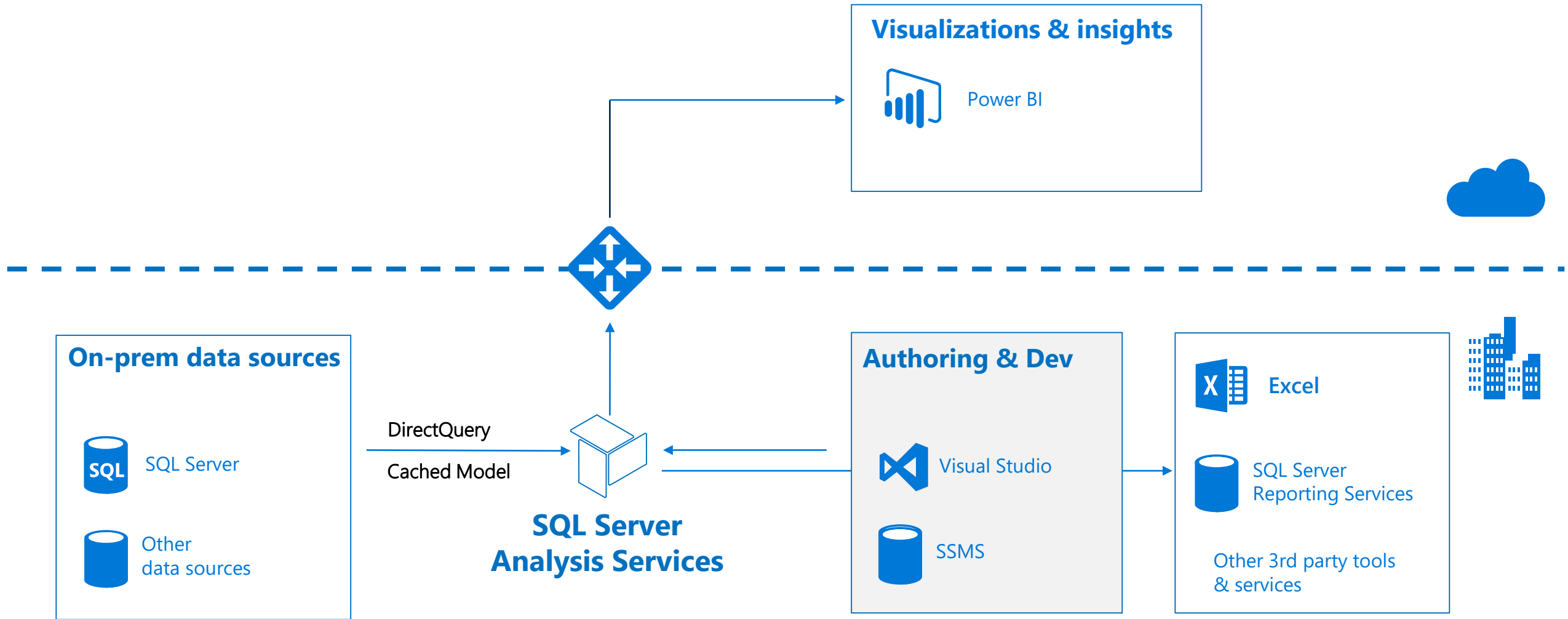
Analysis Services

Increase adoption

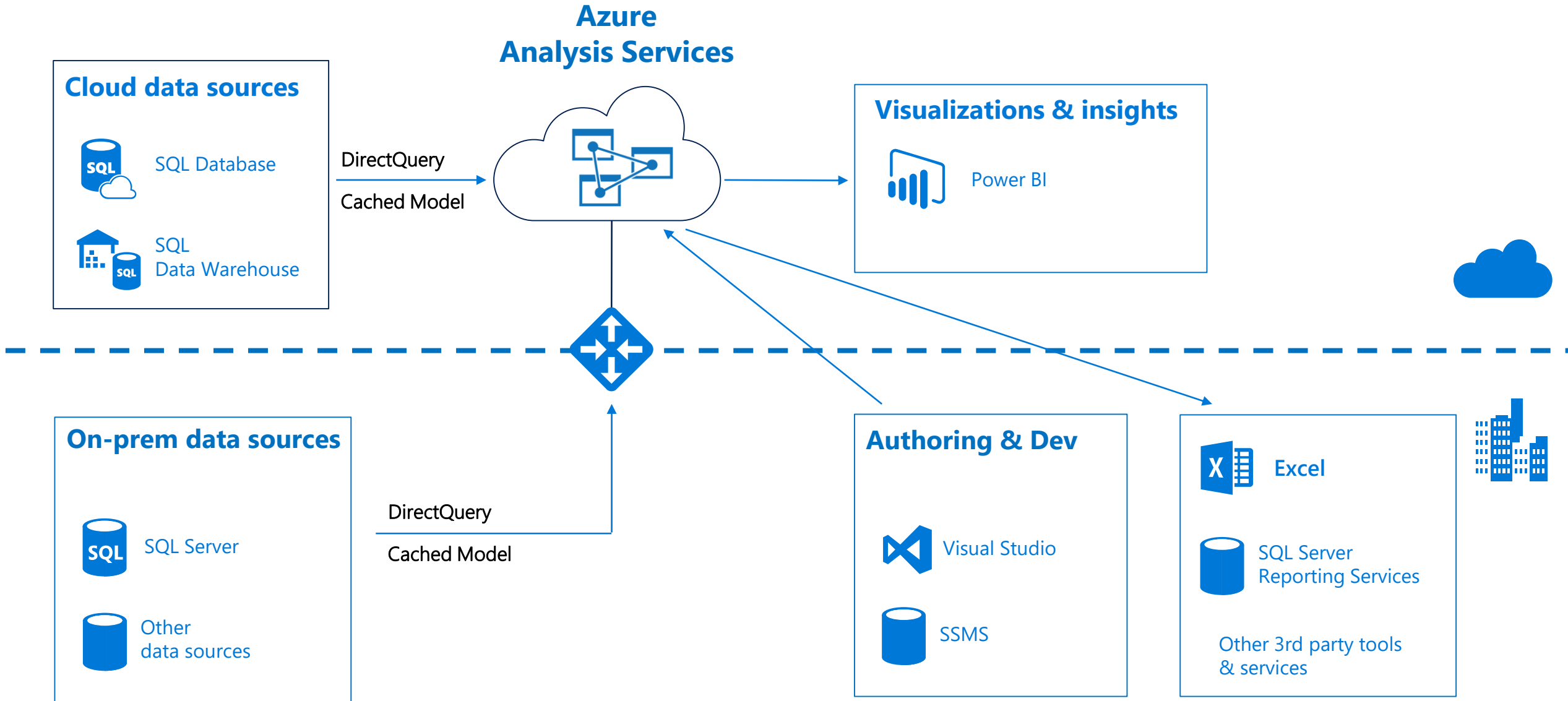


- Analysis Services kot semantični model
- Združevanje podatkov iz ogromno virov in postaviti poslovna pravila
- Hitro analiziranje podatkov v memoriji
- Integrirano z DevOps

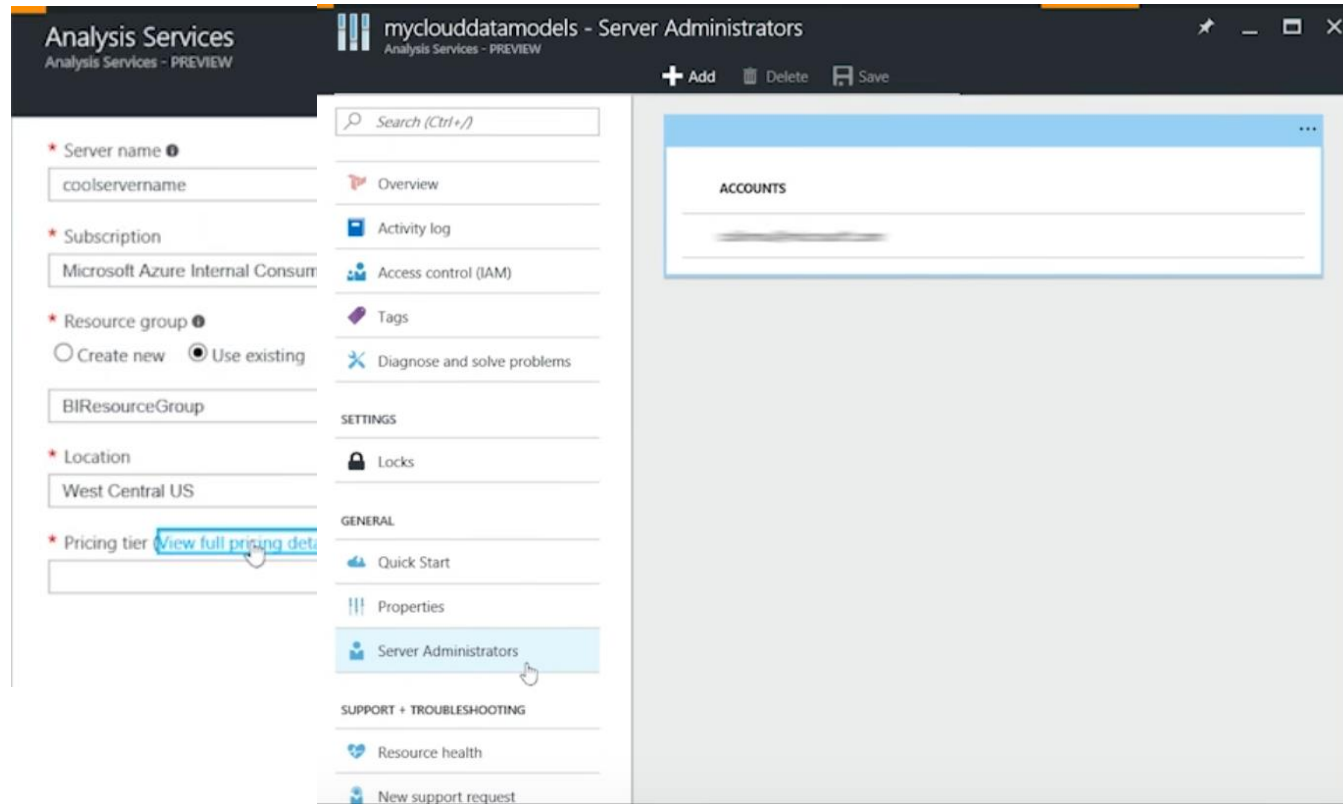
SQL Server Analysis Services



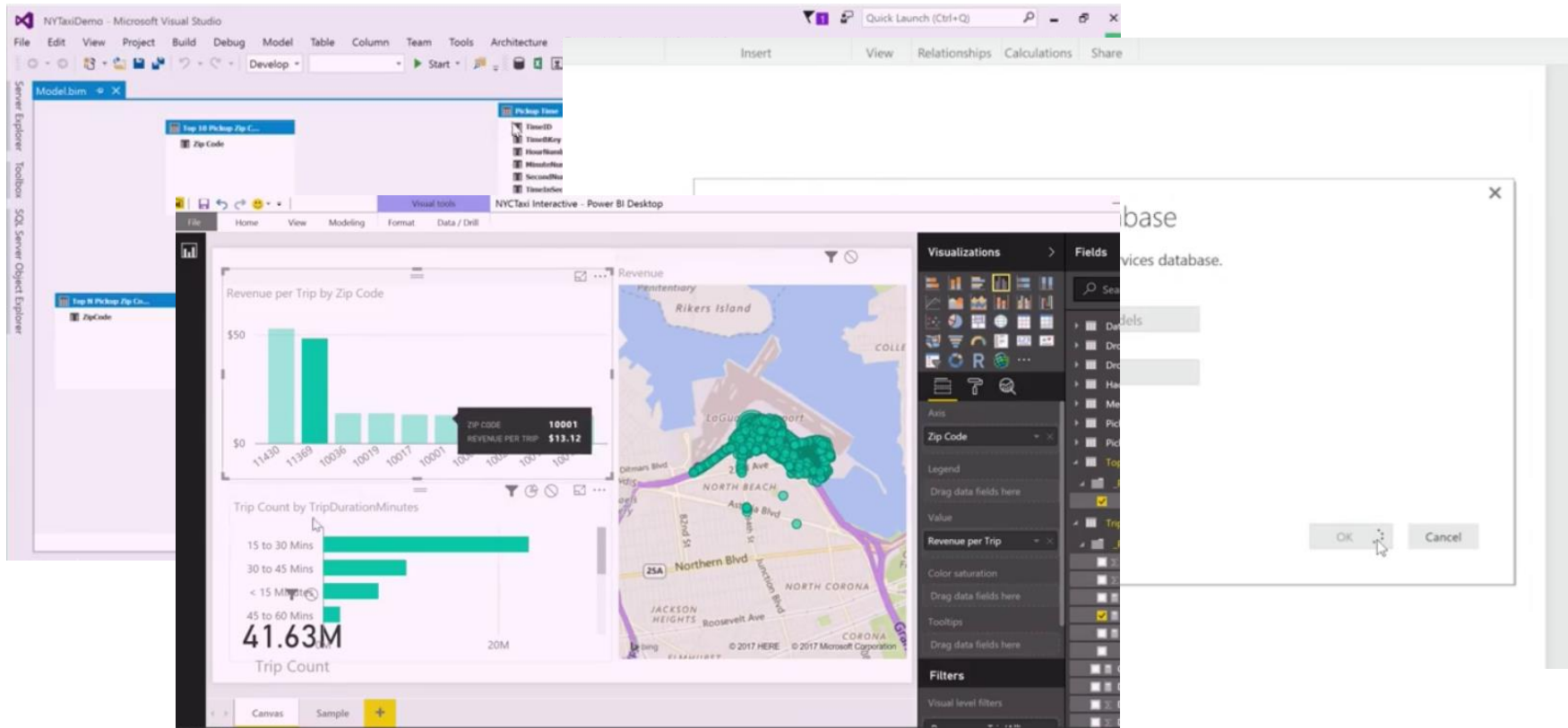
Azure Analysis Services



Azure Analysis Services



Azure Analysis Services



Primer uporabe „Internet Of Things“

Spremljanje karakteristik posamezne naprave



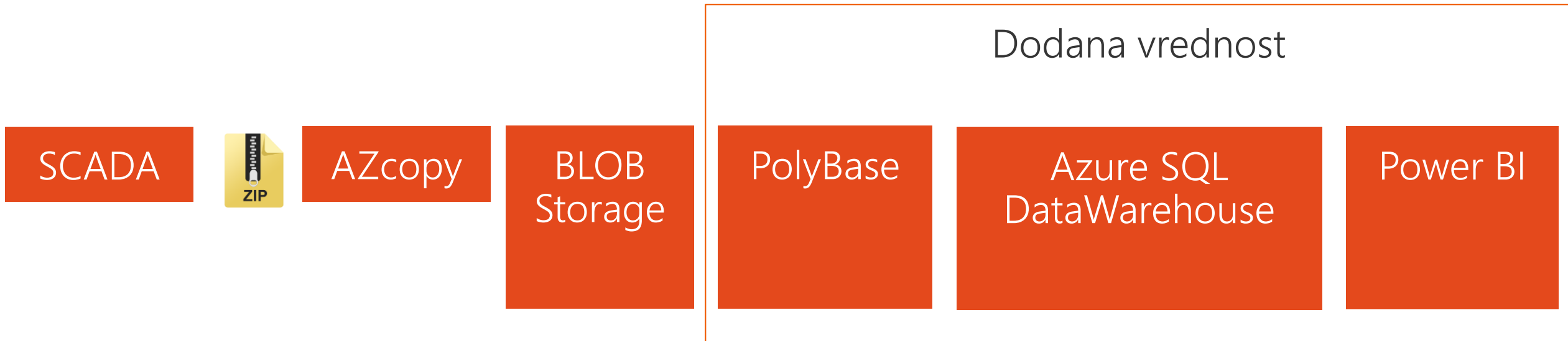
Opis

- Trenutno imajo okvirno 3.000.000.000 zapisov / napravo
- cc 1 TB podatkov
- Hranjenje podatkov za obdobje 10 let

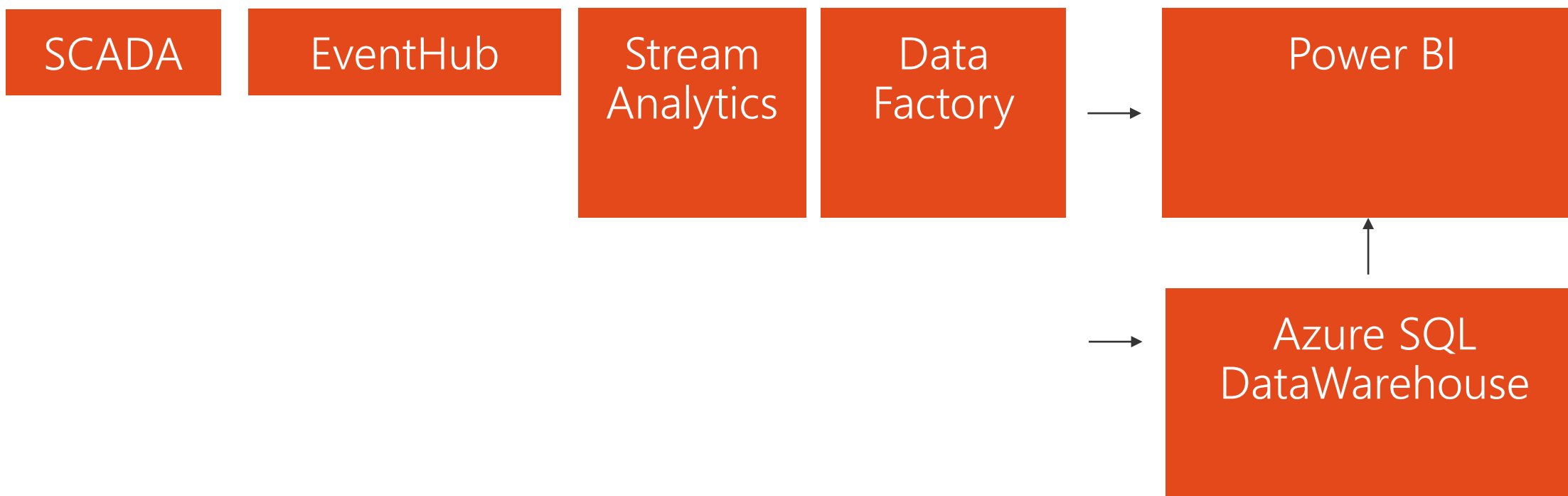


-
- Možnost analitike na zahtevo
 - Analitika v realnem času

Arhitektura Inicialno polenje



Arhitektura Analitika v realnem času



Primer uporabe „Internet Of Things“

Področje energetike

Trenutno imajo v upravljanju 10 elektrarn

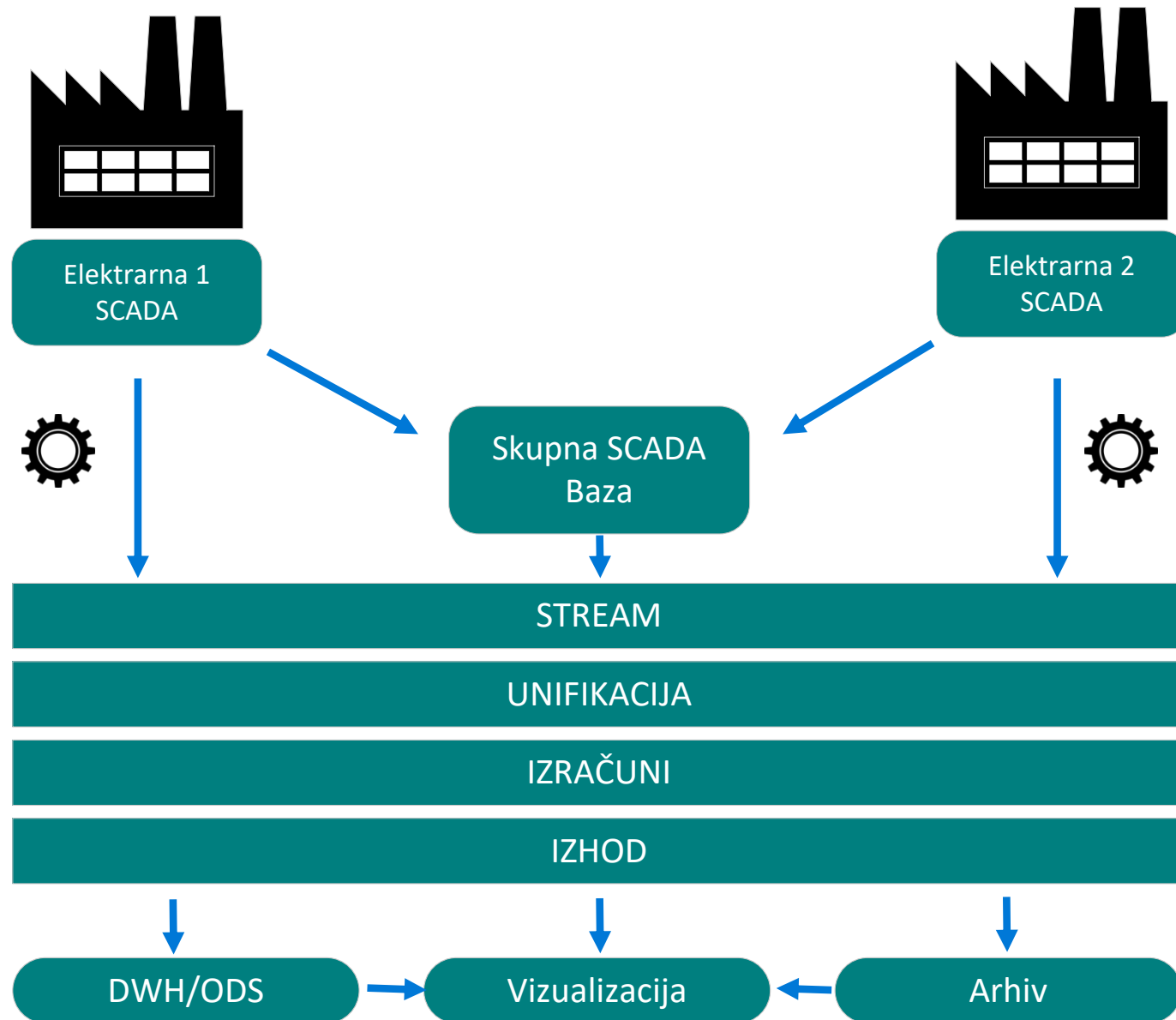
V več kot 90% proizvedene elektrike naredijo na podlagi skrilavca

Glavni cilj podjetja je izboljšati učinkovitost uporabe skrilavca pri proizvodnji energije

Delujejo na 70 km² in zaposlujejo 6500 ljudi



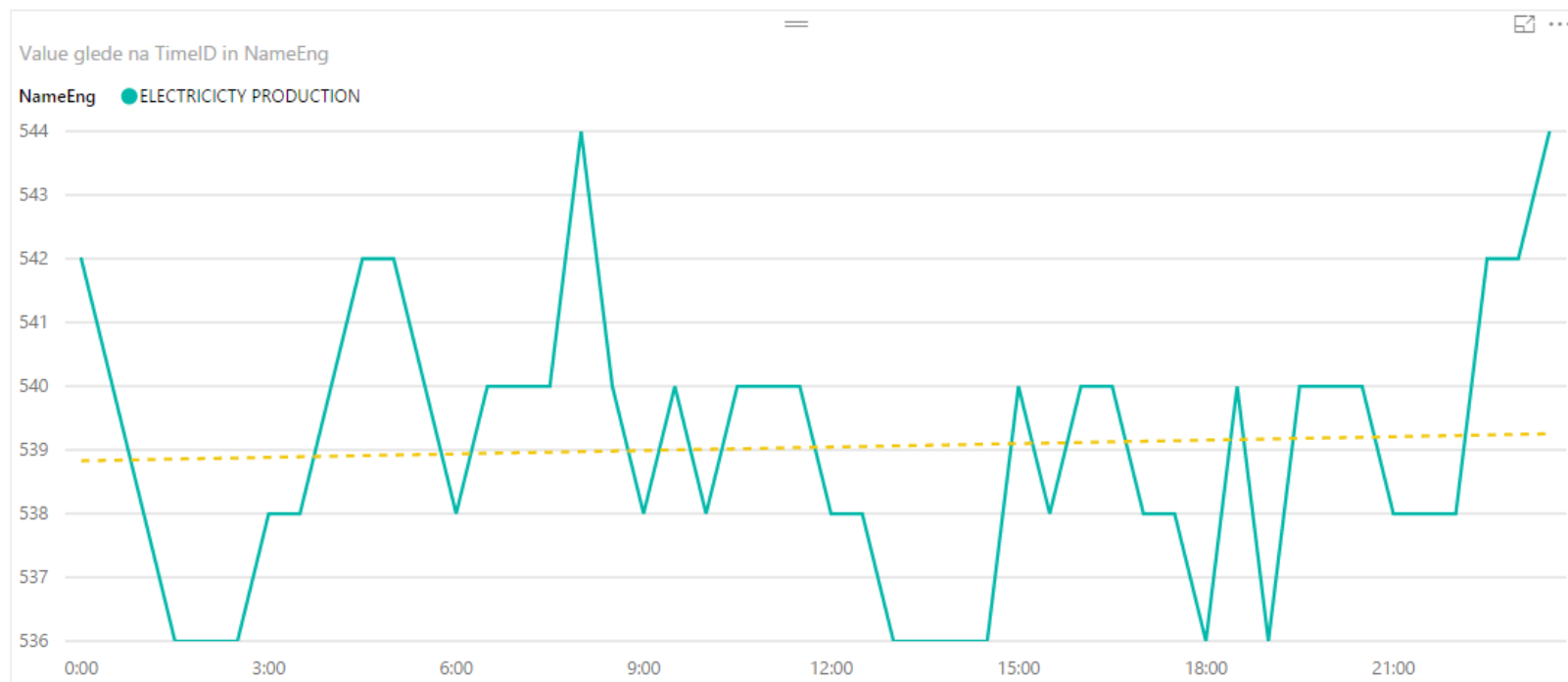
Primer uporabe



$$UNE = \frac{UNIT_NTWe_m}{Q_{(N)ZF}} \times 100$$

- Event Hubs
- Stream Analytics
- Datally
- Data Factory
- Machine Learning
- Power BI
- Data Lake

Power BI



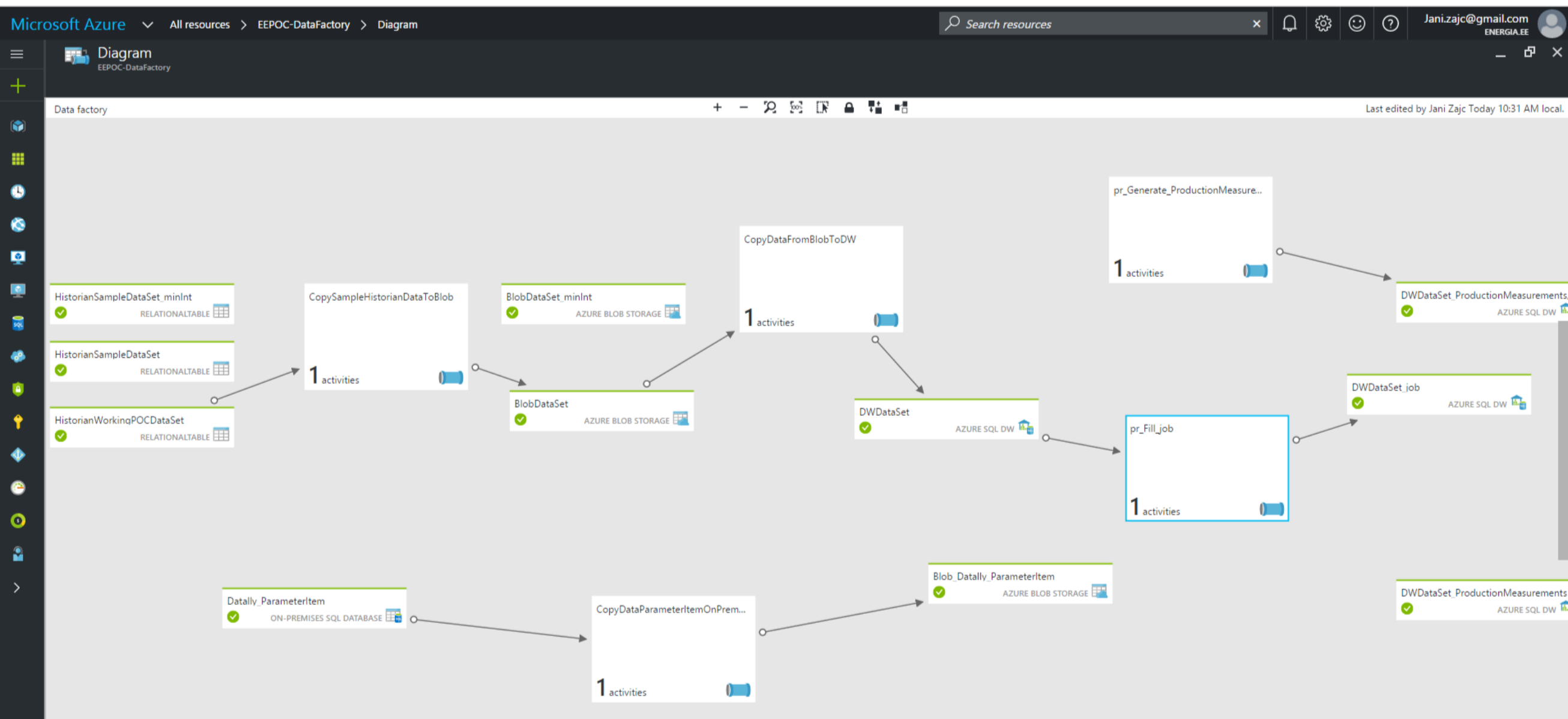
NameEng	Value	DateID	HalfHourName	Code
ELECTRICITY PRODUCTION	542,00	18. maj 2016	0:00 - 0:30	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	540,00	18. maj 2016	0:30 - 1:00	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	538,00	18. maj 2016	1:00 - 1:30	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	536,00	18. maj 2016	1:30 - 2:00	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	538,00	18. maj 2016	10:00 - 10:30	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	540,00	18. maj 2016	10:30 - 11:00	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	540,00	18. maj 2016	11:00 - 11:30	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	540,00	18. maj 2016	11:30 - 12:00	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	538,00	18. maj 2016	12:00 - 12:30	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	538,00	18. maj 2016	12:30 - 13:00	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	536,00	18. maj 2016	13:00 - 13:30	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	536,00	18. maj 2016	13:30 - 14:00	10BAA50CE030_XJ60
ELECTRICITY PRODUCTION	536,00	18. maj 2016	14:00 - 14:30	10BAA50CE030_XJ60

Skupaj 25.874,00

NameEng

- ☒ Izberi vse
- ☐ (Blank)
- ☐ ACTUAL HEAT TRANSFER RATE W/m2/°C
- ☐ AIR COMPRESSOR1 RUN TD (COMPRESS...
- ☐ AIR COMPRESSOR1 RUN TD (MOTOR BEA...
- ☐ AIR COMPRESSOR2 RUN TD (COMPRESS...
- ☐ AIR COMPRESSOR2 RUN TD (MOTOR BEA...
- ☐ AIR COMPRESSOR3 RUN TD (COMPRESS...
- ☐ AIR COMPRESSOR3 RUN TD (MOTOR BEA...
- ☐ AIR TEMPERATURE AFTER AIR PREHEATER
- ☐ AIR TEMPERATURE AFTER CALORIFER
- ☐ AIR TEMPERATURE BEFORE CALORIFER
- ☐ ANNUAL PEAK LOAD (net) AND TIME
- ☐ ASH CONTENT OF FUEL kg/kg
- ☐ AUXILIARY COOLING WATER MASS FLOW
- ☐ AUXILIARY POWER CONSUMPTION
- ☐ BAND SCREEN RUN TD
- ☐ BIOMASS / OIL SHALE RATIO SP U LIM
- ☐ BIOMASS CONSUMPTION FOR ELECTRICI...
- ☐ Biomass conveyor speed
- ☐ BIOMASS CONVEYORS RUN TD
- ☐ BIOMASS FLOW
- ☐ BIOMASS FLOW TOTAL
- ☐ Block Efficiency
- ☐ BOILER BRUTO PRODUCTION
- ☐ BOILER'S COOLING AIR TEMPERATURE
- ☐ CaCO3 CONTENT IN FUEL kg/kg
- ☐ CARBON CONTENT kg/kg
- ☐ CARBON DIOXIDE CONTENT
- ☐ CEP1 STAGE 1 RUN TD (MOTOR DE/NDE ...
- ☐ CEP1 STAGE 2 RUN TD (MOTOR BEARING)
- ☐ CEP2 STAGE 1 RUN TD (MOTOR DE/NDE ...
- ☐ CEP2 STAGE 2 RUN TD (MOTOR BEARING)
- ☐ CLOSED COOLING WTR PMP1 RUN TD (M...
- ☐ CO EMISSIONS

Data Factory



PoC učinki

Boljše razumevanje **efektivnosti** in **učinkovitosti** (OE analiza) elektrarn.

Možnost izdelave mesečnih finančnih poročil

Možnost odločanje na podlagi **živih podatkov**

Napoved v realnem času

Proizvodnje

Učinkovitosti

Porabe goriva.



???

