Playa del Rey, CA 90293 ♦ 323-309-0725 ♦ yuriym@msn.com

Yuriy Margulis

Synopsys

- ❖ Over 30 years of versatile software engineering experience including 18 years of Data Warehousing, Data Engineering, Big Data and Business Intelligence, 18 years of management primarily in the Director's role, hands-on experience with Data Warehouse architecture, development, operations and support, database administration, Big Data ecosystem deployments, SQL's and NoSQL's, ERP and CRM implementations, ETL engineering, application development.

- ❖ Established track record of successful project deliveries in the internet, e-commerce, media and entertainment fields for the last 18 years. Previous extensive IT experience in areas of advertising, marketing, real estate, mortgage, banking, accounting, finance, investments, annuities, health insurance, manufacturing, distribution, transportation, food processing, education and software development.

- ❖ Experienced in all aspects of IT management and technical leadership, including hiring, team building, mentoring, outsourcing management, customer and interdepartmental communications, projects planning, budgeting, tools evaluation and negotiations.

- ❖ Proficient in Data Warehousing Architecture, Big Data Architecture, OLAP Architecture, BI Architecture and Administration, ER Modeling, Dimensional Modeling, ETL/ELT Architecture, Meta Data Architecture, CRM and ERP Implementations.

- ❖ Hands-on experience in all areas of Software Engineering, Database Administration and programming for Oracle, Redshift, PostgreSQL, Sybase, MySQL, and MS SQL Server RDBMS, engineering in Spark/Hive/Presto/Hadoop/Pig environment, application programming in Python, C#, VB, Delphi and various scripting languages, Object-Oriented programming, API programming.

- ❖ Equally successful individually or as a part of a team. Excels both as a consultant and as a full-time employee. Self-starter, willing to learn new technologies and methods, successful in merging new knowledge with skills accumulated over the time.

2016-current

Big Data

Summary

Started experimenting with Big Data Hadoop/Pig/Hive platform first time in 2008, launched first Hadoop clusters in 2010, took Cloudera University class "Cloudera Developer Training for Apache Hadoop with Pig and Hive" in 2012. In 2015-2016, with introduction of Apache Spark and Presto as part of AWS EMR in combination with other AWS services such as RDS and S3 realized that the revolution in data processing started. In 2016 completely retooled from relational data warehousing to the big data cloud ecosystem, took BerkleyX series of online certification classes "Data Science and Engineering with Apache Spark", completed certification, joined Toptal platform and co-founded multiple startups. Over this period of time worked on multiple assignments lasting from couple month to a year, some contracts and some "permanent" positions – those also mostly looked like contracts. They were overlapping, sometimes two or even three simultaneously. Working in such mode allowed to better understand new philosophy, sharpen Python, Spark, AWS and Azure skills, shape up new architecture vision and adjust management techniques to the new realities. Although most importantly: by working on all those projects was able to filter out parts of previous experience still applicable while discarding those that became obsolete

Consulting Positions

04/2016 – current   Big Data Consultant   Playa del Rey, CA
Stack: AWS EMR - Spark, Hive, Presto, other AWS and HDP: Athena, Glue, RDS for PostgreSQL, S3, Oracle, MSSS, Redshift, Snowflake. Cloudera/Hortonworks. Azure Data Lake, Azure SQL Data Warehouse, Azure Databricks Streaming: Kafka, Kinesys. Fivetran, AWS DMS. Languages – SQL, Python/pySpark, pandas, boto3, Scala, APIs. Scheduling and workflow - Airflow, Luigi, AWS Lambda, AWS API, dbt. ETL -Fivetran, Stitch, pySpark. Pentaho (kettle) CE PDI, Reporting/Querying/Alerts - Redash on Athena and Postgres sources, Tableau, Looker. CI/CD – GitLab/GiHub/Bitbucket, Azure DevOps, Jira, Trello. Process - Agile

Performed as manager, data architect and data engineer, hands-on. Worked on multiple often overlapping projects directly at Crowd Consulting LLC (co-founder, CEO) and as subcontractor via multiple consulting companies – Toptal, Boston Consulting Group (via Toptal), Aviana Global Technologies, Elevon Data and more.

Joined Toptal in 2016. Toptal is a private network for the top 3% of freelance software engineers, designers, and finance experts ranked 12th fastest growing software company in North America by Deloitte. Worked on 10 Toptal projects.

Co-founded Crowd Consulting in 2016, developed all the business and managed the company, became Snowflake and Fivetran partner.

- *Maisonette* (via Toptal, **current**): data lake and data analytical plaatform, AWS

- *Teespring* (via Toptal): DW ETL re-platforming from Airflow/Redshift to Fivetran/Snowflake/Databricks Spark/pySpark, AWS

- *Carbon38*: real-time Big Data analytical ecosystem including AWS S3 Data Lake, AWS RDS Data Warehouse, BI Reporting System and comprehensive streaming ETL system, AWS.

- *Boston Consulting Group* (via Toptal): Three 3-month contracts. BCG's clients – two major pharmaceutical companies and one investment company in the fast food restaurant industry. Subcontracted by BCG GAMMA Advance Analytics and Data Science division to provide data engineering support for BCG's data scientists on multiple DMP and personalization projects. Main task on all projects – Feature Engineering, Spark ETL and data lake development. Additional tasks include: Python utilities, Airflow installations and administration, Spark to Excel Python conversions, various devops tasks. Designed and build dynamic S3-to-S3 RDS-driven (metadata in

Postgres) ETL system in Spark/Hive. All 3 ETL systems build based on modern Data Warehousing best practices. Documented systems and provided training. Proposed and enforced development standards, provided documentation, data validation procedures, operational support and maintenance guidelines. Designed and build multi-layered S3 Customer-360 Data Lake. All 3 projects utilized different stack – first project was on AWS, second – on Hortonworks Data Platform and third – on Azure Databricks. Though, common to all three projects was utilization Spark/Python as main development platform.

- *Mobivity*: full Big Data analytical ecosystem including Data Lake, Data Warehouse, Reporting Data Mart, Analytical Data Mart, BI Reporting System and comprehensive ETL system. Coded one of two subject areas completely, set up development methodology and standards. Personnel training on Big Data. Data validation, cleansing and governance. Technology selection and client's team augmentation. Managed Crowd Consulting's team, responsible for overall solution's architecture and project delivery

- *American Taekwondo Association* (via Toptal). ETL data transfer, data cleansing, validation and integration between new (MSSS) and legacy (Oracle) databases

- Smaller projects, POC's, support, team and account management
    - *CVS Health* (2020, 2 months, via Toptal) PoC, Rapids AI/dask/cuda/cudf, Azure
    - *Content.ad* (2019, 2 months, via Toptal) PoC, architecture, project support, AWS
    - *Oliver Wyman Labs* (2019, 2 weeks, via Toptal) PoC, architecture, Spark/EMR
    - *Ingrooves Music Group* (2017, 1 month, via Toptal) data pipeline support, MSSS, Druid
    - *AlgoLift* (2018, 2 weeks, Crowd Consulting) account and team management, AWS
    - *RCLCO* (2017, 1 month, Crowd Consulting) account and team management, AWS
    - *Enervee* (2017, 2 months, Crowd Consulting) team management, AWS

Permanent positions

10/2017 – 5/2018   Enervee   Venice, CA
Role: VP, Data
Stack: AWS EMR - Spark, Hive, Presto, other AWS: Redshift, RDS for PostgreSQL, MySQL, AuroraDB, PostgreSQL, S3, CloudWatch, Pipelines, Glue. Language - Python/Django, PyTest, Pandas. Scheduling and workflow - Airflow, AWS Lambda. ETL - pySpark, bash, Pentaho (kettle) CE PDI. Reporting/Querying/Alerts - Redash on Presto and Postgres sources. CI/CD –Bitbucket, Bamboo, Ansible, Jira. Process - Scrum Agile

Permanent position (contract-to-hire). Manage data team and built analytical system. Build AWS S3 Data Lake. Loaded Segment.com data from parsed Segment logs bypassing Segment Warehouse. Validated data against Segment Warehouse in AWS Redshift. Loaded PostgreSQL product catalog data to the Data Lake what allowed to develop site behavior marketing BI reporting and ML predictive analytics. Designed a framework for Enterprise Data Warehouse.

10/2016 - 10/2017   Investment Technology Group (a.k.a. ITG)   El Segundo, CA
Role: Big Data Architect
Stack: Apache Spark, Hive, Presto, Kafka. Databases - MemSQL, Sybase IQ, Sybase ASE, MySQL, kdb+. Languages – Q, Python. Docker, DataBricks, bash. Process – Scrum Agile

Permanent position with ITG's Transaction Cost Analysis (TCA) division to architect and design new TCA data analytical system and re-platform the data infrastructure from relational and columnar database environments to Big Data ecosystem. Installed 4-node Apache Hadoop cluster on ITG's private cloud. Designed and developed application integrating Apache Spark with Sybase ASE and Sybase IQ databases, legacy Java modules, API's and kdb+. Participated in company-wide platform kdb+ POC, converted Sybase IQ queries to Q language. Set up MemSQL cluster, MemSQL Kafka pipelines and Spark ETL on private cloud and on Docker

2001 - 2016

Data Warehousing

Summary

 Stating from 1/1/2001 was building, managing and running large Oracle data warehouses and BI environments for 16 years straight. Over this period of time worked for 5 companies on 5 data warehouses. All positions were very similar - manager or director "of Data Warehouse". The only exception was "Director DSS" but that was just the popular at that time name for the same data warehousing and BI. In addition to data warehousing carried variety of related responsibilities. Started in media and entertainment industry (MediaLive and UMG), from 2007 for 9 years switched to internet comparative shopping e-commerce at Edmunds, PriceGrabber and Connexity. All the time fully hands-on as the data engineer, devops engineer, DW and ETL architect in addition to people management. In 2006, co-founded and in 2015 closed Speed-Lead.com where developed and supported C# .Net/MSSS-based web application.

Positions

06/2015 – 04/ 2016   Connexity (a.k.a. Shopzilla)   Los Angeles, CA
Role: Director Data Warehouse
Stack: Oracle, Exadata, Cognos, Cognos Transformer, OBIEE, perl, bash, Hortonworks Hadoop, Pig, Hive, Mahout.
Process – Kanban Agile

> 9-month permanent position, post-acquisition, transfer from PriceGrabber. Connexity (Shopzilla, BizRate) acquired PriceGrabber.com in June 2015. Was transferred to Connexity at the same capacity. Continued with all responsibilities over PriceGrabber DW and BI systems. Assumed responsibilities over Shopzilla.com DW and BI function and team. Worked on PriceGrabber/Shopzilla integration and merger. Participated in PriceGrabber to Connexity data center move, was responsible for data warehouse and BI systems transition. Retired PriceGrabber Data Warehouse in April, 2016.

12/2008 – 06/2015   PriceGrabber   Los Angeles, CA
Role: Director Data Warehouse
Stack: Oracle, Exadata, Cognos, Cognos Transformer, perl, python, bash, Cloudera and Hortonworks Hadoop, Pig, Hive, Mahout, AWS Redshift, Pentaho (kettle) CE PDI. Process – waterfall.

> 7-years permanent position. Joined PriceGrabber on 12/17/08 as a Senior Manager, Data Warehouse. Assumed ETL lead's responsibilities in 2012, was promoted to a Director Data Warehouse in 2013. Was responsible for all aspects of management, architecture and operations of Enterprise Data Warehouse, ETL (ELT) system and BI analytical reporting environment.
> Managed of the development and support teams, concluded hiring both in US and India, resource allocation and contract management for all DW and BI related projects and activities. Rebuilt in-house team, hired outsourcing support team, established 24x7 operations and provided SLA compliance.
> Managed Data Warehouse iterative growth from 5 to 17 subject areas and from 100 million to 2 billion rows processed per day. Managed performance tuning of ETL system, database and storage network, what resulted in stabilization of the ETL batch processing window and 70 percent decrease in processing time.
> Implemented Master Data Management solution by sourcing legacy reporting data marts from the data warehouse, achieving data consolidation across the enterprise.
> Managed replacement of custom-build SQL-based aggregation systems with Cognos Transformer OLAP cubes, complete redesign and rebuilding of Cognos BI environment. Manage massive additions to existing star schemas and ETL system and development of the reporting data marts.
> Managed move of the Data Warehouse from commodity hardware and NetApp storage to Oracle Exadata X3 Database Machine and ZFS storage appliance.
> 100% hands-on as EDW architect, ETL engineer, data analyst, dev-ops lead and Oracle development DBA

07/2007 –08/ 2008   Edmunds   Santa Monica, CA
Role: Director Data Warehouse
Stack: Oracle, Informatica, perl, bash. Process – Scrum Agile

> 1.5-year permanent position. Managed Enterprise Data Warehouse, ETL and revenue-generating reporting Data Marts. Responsibilities included management of the development team, operational support, release management, strategic planning, upgrades and maintenance.
> Managed Data Warehousing development team through multiple simultaneous ETL and reporting projects, addition of new subject areas and a reporting data marts.
> Led database and data warehouse architecture effort, managed DW releases.
> Supported Data Warehouse operations during the nightly loads and daily activities

11/2003 – 08/2007   Universal Music Group   Universal City, CA
Role: Manager Data Warehouse
Stack: Oracle, Cognos, MSSS/MSAS, perl, bash, ERWin

> 4-years permanent continued into 3-years part-time consulting position. Joined UMG in November 2003 as Data Warehouse Architect, was promoted to a Manager in 2005, from 2007 continued as Temp ETL consultant until the end-of-life of the eCRM system in 2010.
> UMG eCRM system provided all recording labels with campaign management, e-marketing and analytics services by blasting over 250 million emails per year.
> Responsibilities included development and support for Data Warehouse, custom ETL system, Data Marts, Marketing Reporting System for Labels utilizing MSAS OLAP Data Marts, MSRS and Cognos reporting.
> Managed data warehouse and BI system through multiple iterations. Redesigned ETL system, rebuild reporting system, build new data marts and analytical application. Decreased batch processing window from 24+ to 10 hours
> Provided database and data warehouse architecture for the DW and all Data Marts, OLAP architecture, team augmentation and management, project management, database administration for new development (Oracle and SQL Server), ETL/ELT development, data modeling, Cognos administration and 24/7 operational support.

01/2001 – 09/2003   MediaLive Intl. (formerly Key3Media Events)   Los Angeles, CA
Role: Director Decision Support and Financial Systems
Stack: Oracle, Epiphany, EPI*CHANNEL, Oracle Discoverer, SPSS, MSSS, MSAS, MS DTS, Excel VBA, ERWin, eBusiness Suite, SalesForce, Sun Solaris

> 2.5-year permanent position. MediaLive (a.k.a. Key3Media, Ziff-Davis Events, Softbank Forums Japan) producer of the major technology events, including COMDEX, NetWorld+Interop, Seybold Seminars, JavaOne, and others
> Started data warehousing initiative and managed data warehouse through multiple iterations
> Was responsible for Data Warehousing including EDW, ODS, Data Marts, ETL system, KPI System, OLAP reporting systems, CRM including Epiphany Campaign Management, Epiphany Campaign Management for Emails, integration with Experian, IBM WebSphere and SPSS
> In 2002 also assumed responsibilities over ERP team and system including Oracle eBusiness Suite: Financials, Contracts, Order Management
> Defined overall architecture for the Data Warehouse, all data marts and reporting systems. Participated in both logical and physical design, data mapping, analysis, auditing and modeling. Designed Oracle Discoverer reporting system
> Hired DW team, managed both DW and Oracle Application teams
> Lead all data research activities for legal filings, financial audits and other special requests form executive management, finance, marketing and IT departments
> Managed ongoing support, maintenance and upgrade of Oracle Application
> Maintained relations with vendors and solution providers
> Hand-on as a development DBA, performed as both Oracle and MS SQL Server ETL engineer
> Developed Sales Pipeline Application and SaleForce.com-based reporting in Excel VBA

1999 - 2001

BI, OLAP and Data Warehousing Consulting

Summary

Stared working on first data warehousing project in 1997 as a consultant at PacifiCare. Switched to BI development in 1998 at permanent position of Principal Software Engineer with Platinum Software (a.k.a. Epicore). Became an early adopter and development partner for the first version of Microsoft SQL Server Analysis Services (MS DSS "Plato"). Designed and developed first commercial application utilizing Microsoft DSS/DTS as an add-on to Epicore Suite. Did OLAP development on Microsoft and Essbase platforms for a few years, but gradually became more involved in RDBMS Data Warehousing and ETL and finally preferred it over OLAP. Still was managing and administering BI systems for many years, but only as a secondary to Data Warehousing function.

Positions

1999 – 2001   Hyperion Solutions (now Oracle)   Sunnyvale/Fullerton, CA
Role: Senior Principal Consultant
Stack: Oracle, MSSS, Hyperion Essbase, Essbase Add-in and API for Excel, Excel VBA, Visual Basic, MSSS DTS, Data Junction, GEAK, SAP, Hyperion Pillar, Hyperion Enterprise, Erwin

> 2-year permanent position. Joined Data Into Action consulting company (DIA) in 1999 as a Consultant, was promoted to Practice Lead, Data Warehouse in 3 month. After DIA was acquired by Hyperion in 2000, became Senior Principal Consultant with Hyperion Professional Services, Essbase Practice. Later Hyperion was acquired by Oracle.
>
> Assignments:
> - Balance Scorecard Methodology-based KPI System. Client: leading semiconductor manufacturer.
> - Sales (RSM) reporting application project. Client: leading international electronics and computer equipment manufacturer
> - Lead a team of Hyperion's consultants through a direct face-to-face proof-of-concept competition against Cognos professional services team over the OLAP reporting project. Client: largest annuities company
> - Participated in 3 sequential data warehousing and data integration initiatives. Client: major real estate development and investments company.
> - Supported all Oracle-related projects in the pipeline of DIA consulting.

1998 – 1999   Platinum Software Corporation (now Epicor) Irvine, CA
Role: Principal Software Engineer
Stack: MSSS, MS DTS, MS DSS, Visual Basic, Erwin, Platinum SQL

> Application development, research, data modeling

before 1998

Early Career

Summary

Held multiple positions in software engineering, application development, education and research including two entrepreneurial endeavors in Odessa, Ukraine and two Oracle Application integration projects in Greater Los Angeles, CA. Coded in Clipper, VB, VBA, Delphi, C++, MS Access, Oracle PL/SQL, Informix. Industries included Mortgage, Retail Banking, International Banking, Entertainment, Education, Recreational Medicine, Agriculture and Manufacturing. Details upon request.

Certification

2016   edX/UC BerkleyX (2016): Certification in Machine Learning with Apache Spark

Education

Odessa State University, Ukraine: Applied Mathematics Diploma (MS Equivalent)

UCI Extension - Oracle DBA class: Oracle Administration and Implementation

Professional Training (15 classes, 1998 - 2016)
- 2016   edX/UC BerkleyX Series "Data Science and Engineering with Apache Spark" (3 classes)
- 2012   Cloudera University "Cloudera Developer Training for Apache Hadoop with Pig and Hive
- 2008   Agile for Data Warehousing
- 2007   Informatica PowerCener Fundamentals
- 2003   Oracle Financials Functional Foundation
- 2002   E.piphany course (3 classes)
- 2000   Hyperion course (4 classes)
- 1998   Platinum SQL: AP, AR
- 1997   Erwin/ERX, BPwin and ModelMart Data and Process Modeling